

FIGURE 14.2 T1 committee structure.

those of the ITU. This coordination with other standards setting bodies is also reflected in T1's involvement with Latin-American standards, through the formation of an ad hoc group with CITEL's permanent technical committee 1 (PTC 1/T1). Further coordination with ETSI and other standards setting bodies is accomplished through the global standards collaboration process.

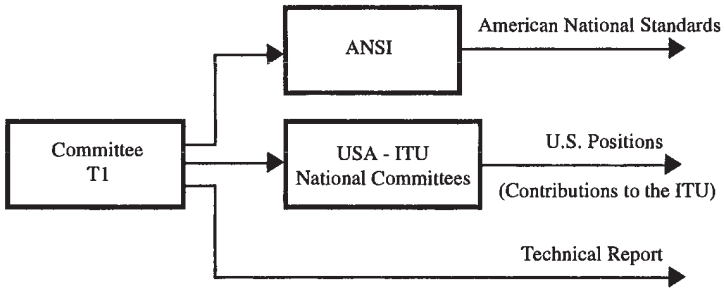


FIGURE 14.3 Committee T1 output.

14.4.2 TIA

The TIA is a full-service trade organization that provides its members with numerous services including government relations, market support activities, educational programs, and standards setting activities.

TIA is a member-driven organization. Policy is formulated by 25 board members selected from member companies, and is carried out by a permanent professional staff located in Washington D.C. TIA comprises six issue-oriented standing committees, each of which is chaired by a board member. The six committees are membership scope and development, international, marketing and trade shows, public policy and government relations, and technical. It is this last committee that in 1992 was accredited by ANSI in the United States to standardize telecommunications products. Technology standardization activities are reflected by TIA's four product-oriented divisions, namely, user premises equipment, network equipment, mobile and personal communications equipment, and fiber optics.

In these divisions the legislative and regulatory concerns of product manufacturers and the preparation of standards dealing with performance testing and compatibility are addressed. For example, modem and telematic standards, as well as much of the cellular standards technology, has been standardized in the United States under the mandate of TIA.

14.4.3 TTC

The third national committee to be addressed is the TTC in Japan. TTC was established in October 1985 to develop and disseminate Japanese domestic standards for deregulated technical items and protocols. It is a nongovernmental, nonprofit standards setting organization established to ensure fair and transparent standardization procedures.

TTC's primary emphasis is to develop, conduct studies and research, and disseminate protocols and standards for the connection of telecommunications networks. TTC is organized along six technical subcommittees that report to a board of directors through a technical assembly (see Fig. 14.4).

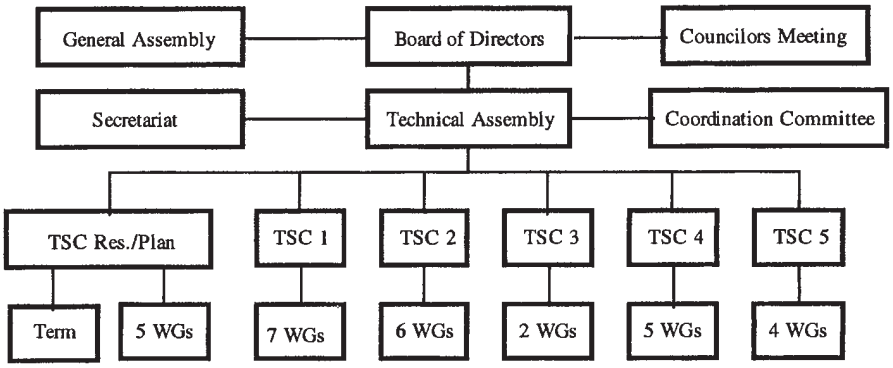


FIGURE 14.4 Organization of TTC.

The TTC organization comprises a general assembly, which is in charge of matters such as business plans and budgets. The councilors meeting examines standards development procedures in order to assure impartiality and clarity. The secretariat provides overall support to the organization; the technical assembly develops standards and handles technical matters including surveys and research. Each technical subcommittee is partitioned into two or more working groups (WG). The coordination committee handles all issues in or between the TSCs and WGs, and it assures the smooth running of all technical committee meetings.

Under the coordination committee, a subcommittee examines users' requests and studies their applicability to the five-year standardization-project plan. This subcommittee also conducts user-request surveys. The areas of involvement of each of the five subcommittees are shown in Table 14.6.

TTC membership is divided into four categories. Type I telecommunications carriers, that is, those carriers that own telecommunications circuits and facilities; type II telecom-

TABLE 14.6 TTC SubCommittees

Strategic Research and Planning Committee: Technical Survey and International Collaboration	
TSC 1	Network-to-network interfaces, mobile communications
TSC 2	User-network interfaces
TSC 3	PBX, LAN
TSC 4	Higher level protocols
TSC 5	Voice and video signal coding scheme and systems

munications carriers, that is, those with telecommunications circuits leased from type I carriers; related equipment manufacturers; and others, including users.

Underlying objectives that guide TTC's approach to standards development are 1) to conform to international recommendations or standards; 2) standardize items, where either international recommendations or standards are not clear, or where national standards need to be set, and where a consensus is achieved; and 3) to conduct further studies into any of the items just mentioned whenever the technical assembly is unable to arrive at a consensus.

These objectives, which give highest priority in developing standards that are compatible with international recommendations or standards, have often driven TTC to adapt international standards for national use through the use of supplements that:

- Give guidelines for users of TTC standards on how to apply them
- Help clarify the contents of standards
- Help with the implementation of standards in terminal equipment and adaptors
- Assure interconnection between terminal equipment and adaptors
- Provide background information regarding the content of standards
- Assure interconnection.

These supplements also include questions and answers that help in implementing the standards, including encoding examples of various parameters and explanation of the practical meaning of a standard.

14.5 Intellectual Property

In the deregulating telecommunication arena patents have become increasingly more important. New ideas that are incorporated in standards often have global market potential and patent holders are seeking to obtain an income from their intellectual property as well as from products. In addition, the general effort to develop standards quickly places them closer to the leading edge of technology. There are some cases, for example speech encoding algorithms, where terms of reference for performance are typically set as objectives that no one can meet when the objectives are defined. The state of the art is being pushed by goals of the standards development organization. In this environment, incorporation of some intellectual property in standards is practically unavoidable.

With regard to intellectual property rights in the ITU, the TSB has developed a "code of practice" which may be summarized as follows.

The TSB requests members putting forth standards to draw the attention of the TSB to any known patent or patent pending application relevant to the developing standard. Where such information has been declared to the TSB, a log of registered patent holders for each affected recommendation is maintained for the convenience of users of ITU standards. If a recommendation, which is a nonbinding international standard, is developed and contains patented intellectual property there are three situations that may arise.

- The patent holder waives the rights and the recommendation is freely accessible to everybody.
- The patent holder will not waive the rights but is willing to negotiate licenses with other parties on a nondiscriminatory basis and on reasonable terms and conditions. What is reasonable is not defined, and the ITU-T will not participate in such negotiations.

- The patent holder is not willing to comply with either of the above two situations, in which case the ITU-T will not approve a recommendation containing such intellectual property.

The patent policy of the American National Standards Institute (ANSI), which governs all standards development organizations accredited by ANSI, is defined in ANSI procedures 1.2.11. It is similar to that of the ITU in that it requires a statement from patent holders or identified parties to indicate granting of a royalty-free license, willingness to license on reasonable and nondiscriminatory terms and conditions, or a disclaimer of no patent. Unlike the ITU, ANSI advises that is prepared to get involved in resolving disputes of what is considered "nondiscriminatory" and "reasonable." Additional information on ANSI patent guidelines can be found at `http://web.ansi.org/public/library/guides/ppguide.html`.

As mentioned earlier ETSI produces a combination of mandatory and voluntary standards. This can create additional complications when intellectual property issues are encapsulated within the standards. To formally address these issues an intellectual property rights committee defines ETSI's policy in the area of intellectual property.

Given the different patent policies adopted by various standards organizations, it is recommended that companies developing products based on standards investigate and understand the patent policy of the associated standards body and the patent statements filed regarding the standard being implemented.

14.6 Standards Coordination

The pace of technological advancements coupled with deregulation has given rise to increased global telecommunications standards activities. At the same time a growth of regional standards bodies has occurred which has increased the potential for duplication of work, wasting resources, and creating conflicting standards. This potentially adverse situation was addressed by a number of interregional telecommunications standardization conferences (ITSCs) that were held in the early 1990s. A global standards collaboration (GSC) group was established to oversee collaborative activities including electronic document handling (EDH) and five high-interest standards subjects:

- Broadband integrated services digital network (B-ISDN)
- Intelligent Networks (IN)
- Transmission management network (TMN)
- Universal personal telecommunications (UPT)
- Synchronous digital hierarchy/synchronous optical network (SDH/SONET)

This early activity was successful in avoiding duplication of effort and coordinating activities on these major standardization efforts. Today the level of cooperative activities, again driven by the pressure to avoid wasting valuable resources and reaching agreed standards more rapidly, are being driven to lower levels through the use of liaison statements between regional standards groups and permitting "documents of information" to flow between standards development organizations. The processes for this information flow are evolving and the electronic addresses provided at the end of this chapter should be consulted for the current interstandards organization communication mechanisms.

14.7 Scientific

Another global, scientifically based organization that has been particularly active in standards development (more recently emphasizing information processing) is the IEEE. Responsibility for standards adoption within the IEEE lies with the IEEE standards board. The board is supported by nine standing committees (see Fig. 14.5).

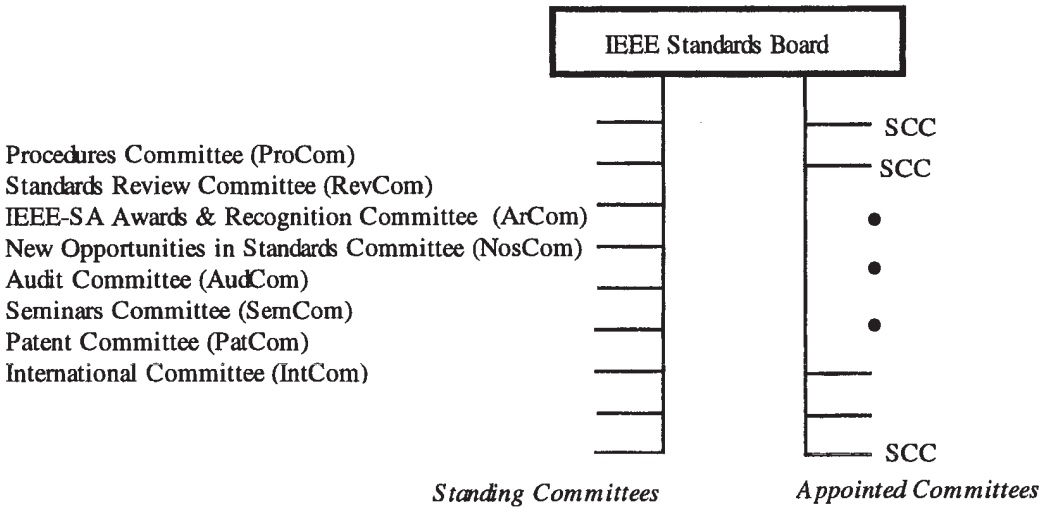


FIGURE 14.5 IEEE standards board organization.

Proposed standards are normally developed in the technical committees of the IEEE societies. There are occasions, however, when the scope of activity is too broad to be encompassed by a single society or where the societies are not able to do so for other reasons. In this case the standards board establishes its own standards developing committees, namely, the standards coordinating committees (**SCC**), to perform this function.

The adoption of IEEE standards is based on projects that have been approved by the IEEE standards board, while each project is the responsibility of a sponsor. Sponsors need not be an SCC, but can also include technical committees of IEEE societies; a standards, or standards coordinating committee of an IEEE Society; an accredited standards committee; or another organization approved by the IEEE standards board.

14.8 Standards Development Cycle

Although the manner in which standards are developed and approved somewhat varies between standards organizations, there are common characteristics to be found.

For most standards, first a set of requirements is defined. This may be done either by the standards committee actually developing the standard or by another entity in collaboration with such a committee. Subsequently, the technical details of a standard are developed. The actual entity developing a standard may be a member of the standards committee, or the actual standards committee itself. Outsiders may also contribute to standards development but, typically, only if sponsored by a committee member. Membership in the standards committee and the right to contribute technical information towards the development of

the standard differs among the various standards' organizations, as indicated. This process is illustrated in Fig. 14.6.

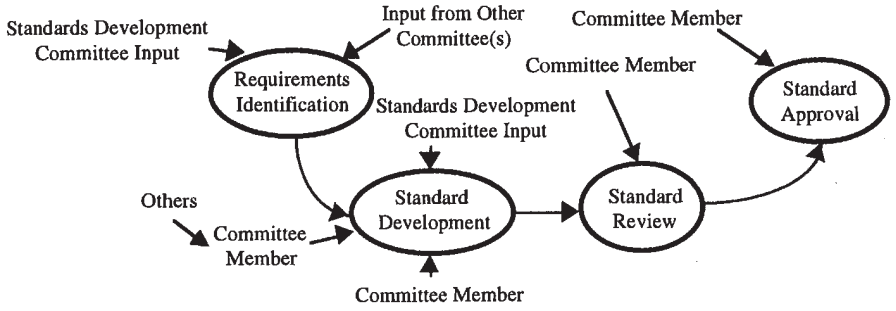


FIGURE 14.6 Typical standards development and approval process.

Finally, once the standard has been fully developed, it is placed under an approval cycle. Each standards setting body typically has precisely defined and often complex procedures for reviewing and then approving proposed standards, which although different in detail, are typically consensus driven.

Defining Terms

ANSI: The American National Standards Institute.

CCIR: The International Radio Consultative Committee, the predecessor of the ITU-R.

CCITT: The International Telephone and Telegraph Consultative Committee, the predecessor of the ITU-T.

CEPT: The European Conference for Post and Telecommunications, a predecessor of ETSI.

CITEL: Inter-American Telecommunications Commission, a standards setting body for the Americas.

ETS: A European (ETSI) technical standard.

ETSI: The European Telecommunications Standards Institute.

GSC: The Global Standards Collaboration group.

ICG: Intersector Coordination Group, a group which coordinates activities between the ITU-T and ITU-R.

IEC: The International Electrotechnical Commission.

IEEE: The Institute of Electrical and Electronics Engineers.

ISO: The International Standards Organization.

ITU: The International Telecommunications Union, an international treaty organization, which is part of the United Nations.

ITU-R: The radio communications sector of the ITU, the successor of the CCIR.

ITU-T: The standardization sector of the ITU, the successor of the CCITT.

JCG: The Joint Coordination Group, which oversees the coordination of common work between ITU-T study groups.

Recommendation: An ITU technical standard.

SCC: A standard's coordinating committee within the IEEE organization.

Standard: A publicly approved technical specification.

T1: An ANSI-approved standards body, which develops telecommunications standards in the United States.

T1AG: The primary advisory group within ANSI Committee T1 on Telecommunications.

TIA: The Telecommunications Industry Association, which is an ANSI-approved standards body that develops terminal equipment standards.

TTC: The Telecommunications Technology Committee, a Japanese standards setting body.

Further Information

[1] Irmer, T., Shaping future telecommunications: the challenge of global standardization, *IEEE Comm. Mag.*, 32(1), 20–28, 1994.

[2] Matute, M.A., CITEL: formulating telecommunications in the Americas. *IEEE Comm. Mag.*, 32(1), 38–39, 1994.

[3] Robin, G., The European perspective for telecommunications standards. *IEEE Comm. Mag.*, 32(1), 40–50, 1994.

[4] Reilly, A.K., A U.S. perspective on standards development. *IEEE Comm. Mag.*, 32(1), 30–36, 1994.

[5] Iida, T., Domestic standards in a changing world. *IEEE Comm. Mag.*, 32(1), 46–50, 1994.

[6] Habara, K., Cooperation in standardization. *IEEE Comm. Mag.*, 32(1), 78–84, 1994.

[7] *IEEE Standards Board Bylaws.* Institute of Electrical and Electronics Engineers. Dec. 1993.

[8] Chiarottino, W. and Pirani, G., International telecommunications standards organizations, *CSELT Tech. Repts.*, XXI(2), 207–236, 1993.

[9] ITU, Book No. 1. Resolutions; Recommendations on the organization of the work of ITU-T (series A); study groups and other groups; list of study questions (1993-1996). World Standardization Conf. Helsinki, 1–12, Mar. 1993.

[10] Standards Committee T1., *Telecommunications.* Procedures Manual. 7th Iss. Jun. 1992.

The standards' organizations often undergo structural and substantive changes. It is recommended that the following web sites be visited for the most updated information.

ANSI	`http://www.ansi.org/`
CITEL	`http://www.oas.org`
ETSI	`http://www.etsi.org`
IEC	`http://www.iec.ch`
IEEE	`http://www.ieee.org`
ISO	`http://www.iso.ch`
ITU	`http://www.itu.ch`
T1	`http://www.t1.org`
TIA	`http://www.tia.org`
TTC	`http://www.ttc.or.jp`

II

Wireless

15 Wireless Personal Communications: A Perspective *Donald C. Cox* 15-1
Introduction • Background and Issues • Evolution of Technologies, Systems, and Services • Evolution Toward the Future and to Low-Tier Personal Communications Services • Comparisons with Other Technologies • Quality, Capacity, and Economic Issues • Other Issues • Infrastructure Networks • Conclusion

16 Modulation Methods *Gordon L. Stüber* 16-1
Introduction • Basic Description of Modulated Signals • Analog Frequency Modulation • Phase Shift Keying (PSK) and $\pi/4$-QPSK • Continuous Phase Modulation (CPM) and MSK • Gaussian Minimum Shift Keying • Orthogonal Frequency Division Multiplexing (OFDM) • Conclusions

17 Access Methods *Bernd-Peter Paris* 17-1
Introduction • Relevant Wireless Communication System Characteristics • Frequency Division Multiple Access • Time Division Multiple Access • Code Division Multiple Access • Comparison and Outlook

18 Rayleigh Fading Channels *Bernard Sklar* 18-1
Introduction • The Challenge of a Fading Channel • Mobile-Radio Propagation: Large-Scale Fading and Small-Scale Fading • Signal Time-Spreading Viewed in the Time-Delay Domain: Figure 18.1, Block 7—The Multipath Intensity Profile • Signal Time-Spreading Viewed in the Frequency Domain: Figure 18.1, Block 10—The Spaced-Frequency Correlation Function • Typical Examples of Flat Fading and Frequency-Selective Fading Manifestations • Time Variance Viewed in the Time Domain: Figure 18.1, Block 13—The Spaced-Time Correlation Function • Time Variance Viewed in the Doppler-Shift Domain: Figure 18.1, Block 16—The Doppler Power Spectrum • Analogy Between Spectral Broadening in Fading Channels and Spectral Broadening in Digital Signal Keying • Degradation Categories due to Time Variance, Viewed in the Doppler-Shift Domain • Mitigation Methods • Summary of the Key Parameters Characterizing Fading Channels • The Viterbi Equalizer as Applied to GSM • The Rake Receiver Applied to Direct-Sequence Spread-Spectrum (DS/SS) Systems • Conclusion

19 Space-Time Processing *Arogyaswami J. Paulraj* 19-1
Introduction • The Space-Time Wireless Channel • Signal Models • ST Receive Processing (Base) • ST Transmit Processing (Base) • Summary

20 Location Strategies for Personal Communications Services *Ravi Jain, Yi-Bing Lin, and Seshadri Mohan* 20-1
Introduction • An Overview of PCS • IS-41 Preliminaries • Global System for Mobile Communications • Analysis of Database Traffic Rate for IS-41 and GSM • Reducing Signalling During Call Delivery • Per-User Location Caching • Caching Threshold Analysis • Techniques for Estimating Users' LCMR • Discussion • Conclusions

21 Cell Design Principles *Michel Daoud Yacoub* 21-1
Introduction • Cellular Principles • Performance Measures and System Requirements • System Expansion Techniques • Basic Design Steps • Traffic Engineering • Cell Coverage • Interference • Conclusions

22 Microcellular Radio Communications *Raymond Steele* 22-1
Introducing Microcells • Highway Microcells • City Street Microcells • Indoor Microcells • Microcellular Infrastructure • Multiple Access Issues • Discussion

23 Fixed and Dynamic Channel Assignment *Bijan Jabbari* 23-1
Introduction • The Resource Assignment Problem • Fixed Channel Assignment • Enhanced Fixed Channel Assignment • Dynamic Channel Assignment • CDMA Systems • Conclusion

24 Radiolocation Techniques *Gordon L. Stüber and James J. Caffery, Jr.* 24-1
Introduction • Description of Radiolocation Methods • Location Algorithms • Measures of Location Accuracy • Location in Cellular Systems • Sources of Location Error • Summary and Conclusions

25 Power Control *Roman Pichna and Qiang Wang* 25-1
Introduction • Cellular Systems and Power Control • Power Control Examples • Summary

26 Enhancements in Second Generation Systems *Marc Delprat and Vinod Kumar* 26-1
Introduction • Overview of Second Generation Systems • Capacity Enhancement • Quality Enhancement • High Bit Rate Data Transmission • Conclusion

27 The Pan-European Cellular System *Lajos Hanzo* 27-1
Introduction • Overview • Logical and Physical Channels • Speech and Data Transmission • Transmission of Control Signals • Synchronization Issues • Gaussian Minimum Shift Keying Modulation • Wideband Channel Models • Adaptive Link Control • Discontinuous Transmission • Summary

28 Speech and Channel Coding for North American TDMA Cellular Systems *Paul Mermelstein* 28-1
Introduction • Modulation of Digital Voice and Data Signals • Speech Coding Fundamentals • Channel Coding Considerations • VSELP Encoder • Linear Prediction Analysis and Quantization • Bandwidth Expansion • Quantizing and Encoding the Reflection Coefficients • VSELP Codebook Search • Long-Term Filter Search • Orthogonalization of the Codebooks • Quantizing the Excitation and Signal Gains • Channel Coding and Interleaving • Bad Frame Masking • ACELP Encoder • Algebraic Codebook Structure and Search • Quantization of the Gains for ACELP Encoding • Channel Coding for ACELP Encoding • Conclusions

29 The British Cordless Telephone Standard: CT-2 *Lajos Hanzo* 29-1
History and Background • The CT-2 Standard • The Radio Interface • Burst Formats • Signalling Layer Two (L2) • CPP-Initiated Link Setup Procedures • CFP-Initiated Link Setup Procedures • Handshaking • Main Features of the CT-2 System

30 Half-Rate Standards *Wai-Yip Chan, Ira Gerson, and Toshio Miki* 30-1

Introduction • Speech Coding for Cellular Mobile Radio Communications • Codec Selection and Performance Requirements • Speech Coding Techniques in the Half-Rate Standards • Channel Coding Techniques in the Half-Rate Standards • The Japanese Half-Rate Standard • The European GSM Half-Rate Standard • Conclusions

31 Wireless Video Communications *Madhukar Budagavi and Raj Talluri* 31-1
Introduction • Wireless Video Communications • Error Resilient Video Coding • MPEG-4 Error Resilience Tools • H.263 Error Resilience Tools • Discussion

32 Wireless LANs *Suresh Singh* 32-1
Introduction • Physical Layer Design • MAC Layer Protocols • Network Layer Issues • Transport Layer Design • Conclusions

33 Wireless Data *Allen H. Levesque and Kaveh Pahlavan* 33-1
Introduction • Characteristics of Wireless Data Networks • Market Issues • Modem Services Over Cellular Networks • Private Data Networks • Cellular Data Networks and Services • Other Developing Standards • Conclusions

34 Wireless ATM: Interworking Aspects *Melbourne Barton, Matthew Cheng, and Li Fung Chang* 34-1
Introduction • Background and Issues • Wireless Interworking With Transit ATM Networks • The PCS-to-ATM Interworking Scenario • QoS Support • Conclusions

35 Wireless ATM: QoS and Mobility Management *Bala Rajagopalan and Daniel Reininger* 35-1
Introduction • QoS in Wireless ATM • Mobility Management in Wireless ATM • Summary and Conclusions

36 An Overview of cdma2000, WCDMA, and EDGE *Tero Ojanperä and Steven D. Gray* 36-1
Introduction • CDMA-Based Schemes • CDMA System Design Issues • WCDMA • cdma2000 • TDMA-Based Schemes • Time Division Duplex (TDD) • Conclusions

15

Wireless Personal Communications: A Perspective

15.1 Introduction 15-1
15.2 Background and Issues 15-3
Mobility and Freedom from Tethers
15.3 Evolution of Technologies, Systems, and Services 15-5
Cordless Telephones • Cellular Mobile Radio Systems • Wide-Area Wireless Data Systems • High-Speed Wireless Local-Area Networks (WLANs) • Paging/Messaging Systems • Satellite-Based Mobile Systems • {Fixed Point-to-Multipoint Wireless Loops • Reality Check
15.4 Evolution Toward the Future and to Low-Tier Personal Communications Services 15-21
15.5 Comparisons with Other Technologies 15-25
Complexity/Coverage Area Comparisons • {Coverage, Range, Speed, and Environments
15.6 Quality, Capacity, and Economic Issues 15-29
Capacity, Quality, and Complexity • Economics, System Capacity, and Coverage Area Size • {Loop Evolution and Economics
15.7 Other Issues 15-34
Improvement of Batteries • People Only Want One Handset • Other Environments • Speech Quality Issues • New Technology • High-Tier to Low-Tier or Low-Tier to High-Tier Dual Mode
15.8 Infrastructure Networks 15-44
15.9 Conclusion 15-45
References 15-46

Donald C. Cox
Stanford University

{ This chapter has been updated using { } as indicators of inserts into the text of the original chapter of the same title that appeared in the first edition of this Handbook in 1996. }

15.1 Introduction

Wireless personal communications has captured the attention of the media and with it, the imagination of the public. Hardly a week goes by without one seeing an article on the subject appearing in a popular U.S. newspaper or magazine. Articles ranging from a short

0-8493-8597-0/99/$0.00+$.50

© 1999 by CRC Press LLC

paragraph to many pages regularly appear in local newspapers, as well as in nationwide print media, e.g., *The Wall Street Journal, The New York Times, Business Week,* and *U.S. News and World Report.* Countless marketing surveys continue to project enormous demand, often projecting that at least half of the households, or half of the people, want wireless personal communications. Trade magazines, newsletters, conferences, and seminars on the subject by many different names have become too numerous to keep track of, and technical journals, magazines, conferences, and symposia continue to proliferate and to have ever increasing attendance and numbers of papers presented. It is clear that wireless personal communications is, by any measure, the fastest growing segment of telecommunications. { The explosive growth of wireless personal communications has continued unabated worldwide. Cellular and high-tier PCS pocketphones, pagers, and cordless telephones have become so common in many countries that few people even notice them anymore. These items have become an expected part of everyday life in most developed countries and in many developing countries around the world. }

If you look carefully at the seemingly endless discussions of the topic, however, you cannot help but note that they are often describing different things, i.e., different versions of wireless personal communications [29, 50]. Some discuss pagers, or messaging, or data systems, or access to the national information infrastructure, whereas others emphasize cellular radio, or cordless telephones, or dense systems of satellites. Many make reference to popular fiction entities such as Dick Tracy, Maxwell Smart, or *Star Trek.* { In addition to the things noted above, the topic of wireless loops [24], [30], [32] has also become popular in the widespread discussions of wireless communications. As discussed in [30], this topic includes several fixed wireless applications as well as the low-tier PCS application that was discussed originally under the wireless loop designation [24, 32]. The fixed wireless applications are aimed at reducing the cost of wireline loop-ends, i.e., the so-called "last mile" or "last km" of wireline telecommunications. }

Thus, it appears that almost everyone wants wireless personal communications, but *What is it?* There are many different ways to segment the complex topic into different communications applications, modes, functions, extent of coverage, or mobility [29, 30, 50]. The complexity of the issues has resulted in considerable confusion in the industry, as evidenced by the many different wireless systems, technologies, and services being offered, planned, or proposed. Many different industry groups and regulatory entities are becoming involved. The confusion is a natural consequence of the massive dislocations that are occurring, and will continue to occur, as we progress along this large change in the paradigm of the way we communicate. Among the different changes that are occurring in our communications paradigm, perhaps the major constituent is the change from wired fixed place-to-place communications to wireless mobile person-to-person communications. Within this major change are also many other changes, e.g., an increase in the significance of data and message communications, a perception of possible changes in video applications, and changes in the regulatory and political climates. { The fixed wireless loop applications noted earlier do not fit the new mobile communications paradigm. After many years of decline of fixed wireless communications applications, e.g., intercontinental HF radio and later satellites, point-to-point terrestrial microwave radio, and tropospheric scatter, it is interesting to see this rebirth of interest in fixed wireless applications. This rebirth is riding on the gigantic "wireless wave" resulting from the rapid public acceptance of mobile wireless communications. It will be interesting to observe this rebirth to see if communications history repeats; certainly mobility is wireless, but there is also considerable historical evidence that wireless is also mobility. }

This chapter attempts to identify different issues and to put many of the activities in wireless into a framework that can provide perspective on what is driving them, and perhaps

even to yield some indication of where they appear to be going in the future. Like any attempt to categorize many complex interrelated issues, however, there are some that do not quite fit into neat categories, and so there will remain some dangling loose ends. Like any major paradigm shift, there will continue to be considerable confusion as many entities attempt to interpret the different needs and expectations associated with the new paradigm.

15.2 Background and Issues

15.2.1 Mobility and Freedom from Tethers

Perhaps the clearest constituents in all of the wireless personal communications activity are the desire for mobility in communications and the companion desire to be free from tethers, i.e., from physical connections to communications networks. These desires are clear from the very rapid growth of mobile technologies that provide primarily two-way voice services, even though economical wireline voice services are readily available. For example, cellular mobile radio has experienced rapid growth. Growth rates have been between 35 and 60% per year in the United States for a decade, with the total number of subscribers reaching 20 million by year-end 1994. The often neglected wireless companions to cellular radio, i.e., cordless telephones, have experienced even more rapid, but harder to quantify, growth with sales rates often exceeding 10 million sets a year in the United States, and with an estimated usage significantly exceeding 50 million in 1994. Telephones in airlines have also become commonplace. Similar or even greater growth in these wireless technologies has been experienced throughout the world. { The explosive growth in cellular and its identical companion, high-tier PCS, has continued to about 55 million subscribers in the U.S. at year-end 1997 and a similar number worldwide. In Sweden the penetration of cellular subscribers by 1997 was over one-third of the total population, i.e., the total including every man, woman, and child! And the growth has continued since. Similar penetrations of mobile wireless services are seen in some other developed nations, e.g., Japan. The growth in users of cordless telephones also has continued to the point that they have become the dominant subscriber terminal on wireline telephone loops in the U.S. It would appear that, taking into account cordless telephones and cellular and high-tier PCS phones, half of all telephone calls in the U.S. terminate with at least one end on a wireless device. }

{ Perhaps the most significant event in wireless personal communications since the writing of this original chapter was the widespread deployment and start of commercial service of personal handphone (PHS) in Japan in July of 1995 and its very rapid early acceptance by the consumer market [53]. By year-end 1996 there were 5 million PHS subscribers in Japan with the growth rate exceeding one-half million/month for some months. The PHS "phenomena" was one of the fastest adoptions of a new technology ever experienced. However, the PHS success story [41] peaked at a little over 7 million subscribers in 1997 and has declined slightly to a little under 7 million in mid-1998. This was the first mass deployment of a low-tier-like PCS technology (see later sections of this chapter), but PHS has some significant limitations. Perhaps the most significant limitation is the inability to successfully handoff at vehicular speeds. This handoff limitation is a result of the cumbersome radio link structure and control algorithms used to implement dynamic channel allocation (DCA) in PHS. DCA significantly increases channel occupancy (base station capacity) but incurs considerable complexity in implementing handoff. Another significant limitation of the PHS standard has been insufficient receiver sensitivity to permit "adequate" coverage from a "reasonably" dense deployment of base stations. These technology deficiencies coupled with heavy price cutting by the cellular service providers to compete with the rapid advancing of the PHS market were significant contributors to the leveling out of PHS growth. It is

again evident, as with CT-2 phone point discussed in a later section, that low-tier PCS has very attractive features that can attract many subscribers, but it must also provide vehicle speed handoff and widespread coverage of highways as well as populated regions.

Others might point out the deployment and start of service of CDMA systems as a significant event since the first edition. However, the major significance of this CDMA activity is that it confirmed that CDMA performance was no better than other less-complex technologies and that those, including this author, who had been branded as "unbelieving skeptics" were correct in their assessments of the shortcomings of this technology. The overwhelming failure of CDMA technology to live up to the early claims for it can hardly be seen as a significant positive event in the evolution of wireless communication. It was, of course, a significant negative event. After years of struggling with the problems of this technology, service providers still have significantly fewer subscribers on CDMA worldwide than there are PHS subscribers in Japan alone! CDMA issues are discussed more in later sections dealing with technology issues. }

Paging and associated messaging, although not providing two-way voice, do provide a form of tetherless mobile communications to many subscribers worldwide. These services have also experienced significant growth { and have continued to grow since 1996. } There is even a glimmer of a market in the many different specialized wireless data applications evident in the many wireless local area network (WLAN) products on the market, the several wide area data services being offered, and the specialized satellite-based message services being provided to trucks on highways. { Wireless data technologies still have many supporters, but they still have fallen far short of the rapid deployment and growth of the more voice oriented wireless technologies. However, hope appears to be eternal in the wireless data arena. }

The topics discussed in the preceding two paragraphs indicate a dominant issue separating the different evolutions of wireless personal communications. That issue is the voice versus data communications issue that permeates all of communications today; this division also is very evident in fixed networks. The packet-oriented computer communications community and the circuit-oriented voice telecommunications (telephone) community hardly talk to each other and often speak different languages in addressing similar issues. Although they often converge to similar overall solutions at large scales (e.g., hierarchical routing with exceptions for embedded high-usage routes), the small-scale initial solutions are frequently quite different. Asynchronous transfer mode (ATM-) based networks are an attempt to integrate, at least partially, the needs of both the packet-data and circuit-oriented communities.

Superimposed on the voice-data issue is an issue of competing modes of communications that exist in both fixed and mobile forms. These different modes include the following.

Messaging is where the communication is not real time but is by way of message transmission, storage, and retrieval. This mode is represented by voice mail, electronic facsimile (fax), and electronic mail (e-mail), the latter of which appears to be a modern automated version of an evolution that includes telegraph and telex. Radio paging systems often provide limited one-way messaging, ranging from transmitting only the number of a calling party to longer alpha-numeric text messages.

Real-time two-way communications are represented by the telephone, cellular mobile radio telephone, and interactive text (and graphics) exchange over data networks. Two-way video phone always captures significant attention and fits into this mode; however, its benefit/cost ratio has yet to exceed a value that customers are willing to pay.

Paging, i.e., broadcast with no return channel, alerts a paged party that someone wants to communicate with him/her. Paging is like the ringer on a telephone without having the capability for completing the communications.

Agents are new high-level software applications or entities being incorporated into some computer networks. When launched into a data network, an agent is aimed at finding information by some title or characteristic and returning the information to the point from which the agent was launched. { The rapid growth of the worldwide web is based on this mode of communications. }

There are still other ways in which wireless communications have been segmented in attempts to optimize a technology to satisfy the needs of some particular group. Examples include 1) user location, which can be differentiated by indoors or outdoors, or on an airplane or a train and 2) degree of mobility, which can be differentiated either by speed, e.g., vehicular, pedestrian, or stationary, or by size of area throughout which communications are provided. { As noted earlier, wireless local loop with stationary terminals has become a major segment in the pursuit of wireless technology. }

At this point one should again ask; wireless personal communications—*What is it?* The evidence suggests that what is being sought by users, and produced by providers, can be categorized according to the following two main characteristics.

Communications portability and mobility on many different scales:

- Within a house or building [cordless telephone, (WLANs)]
- Within a campus, a town, or a city (cellular radio, WLANs, wide area wireless data, radio paging, extended cordless telephone)
- Throughout a state or region (cellular radio, wide area wireless data, radio paging, satellite-based wireless)
- Throughout a large country or continent (cellular radio, paging, satellite-based wireless)
- Throughout the world?

Communications by many different modes for many different applications:

- Two-way voice
- Data
- Messaging
- Video?

Thus, it is clear why wireless personal communications today is not one technology, not one system, and not one service but encompasses many technologies, systems, and services optimized for different applications.

15.3 Evolution of Technologies, Systems, and Services

Technologies and systems [27, 29, 30, 39, 50, 59, 67, 87], that are currently providing, or are proposed to provide, wireless communications services can be grouped into about seven relatively distinct groups, { the seven previous groups are still evident in the technology but with the addition of the fixed point-to-multipoint wireless loops there are now eight, } although there may be some disagreement on the group definitions, and in what group some particular technology or system belongs. All of the technologies and systems are evolving as technology advances and perceived needs change. Some trends are becoming evident in the evolutions. In this section, different groups and evolutionary trends are explored along with factors that influence the characteristics of members of the groups. The grouping is generally with respect to scale of mobility and communications applications or modes.

15.3.1 Cordless Telephones

Cordless telephones [29, 39, 50] generally can be categorized as providing low-mobility, low-power, two-way tetherless voice communications, with low mobility applying both to the range and the user's speed. Cordless telephones using analog radio technologies appeared in the late 1970s, and have experienced spectacular growth. They have evolved to digital radio technologies in the forms of second-generation cordless telephone (CT-2), and digital European cordless telephone (DECT) standards in Europe, and several different industrial scientific medical (ISM) band technologies in the United States.[1]

{ Personal handyphone (PHS) noted earlier and discussed in later sections and inserts can be considered either as a quite advanced digital cordless telephone similar to DECT or as a somewhat limited low-tier PCS technology. It has most of the attributes of similarity of the digital cordless telephones listed later in this section except that PHS uses $\pi/4$ QPSK modulation. }

Cordless telephones were originally aimed at providing economical, tetherless voice communications inside residences, i.e., at using a short wireless link to replace the cord between a telephone base unit and its handset. The most significant considerations in design compromises made for these technologies are to minimize total cost, while maximizing the talk time away from the battery charger. For digital cordless phones intended to be carried away from home in a pocket, e.g., CT-2 or DECT, handset weight and size are also major factors. These considerations drive designs toward minimizing complexity and minimizing the power used for signal processing and for transmitting.

Cordless telephones compete with wireline telephones. Therefore, high circuit quality has become a requirement. Early cordless sets had marginal quality. They were purchased by the millions, and discarded by the millions, until manufacturers produced higher-quality sets. Cordless telephones sales then exploded. Their usage has become commonplace, approaching, and perhaps exceeding, usage of corded telephones.

The compromises accepted in cordless telephone design in order to meet the cost, weight, and talk-time objectives are the following.

- Few users per megahertz
- Few users per base unit (many link together a particular handset and base unit)
- Large number of base units per unit area; one or more base units per wireline access line (in high-rise apartment buildings the density of base units is very large)
- Short transmission range

There is no added network complexity since a base unit looks to a telephone network like a wireline telephone. These issues are also discussed in [29, 50].

Digital cordless telephones in Europe have been evolving for a few years to extend their domain of use beyond the limits of inside residences. Cordless telephone, second generation, (CT-2) has evolved to provide telepoint or phone-point services. Base units are located in places where people congregate. e.g., along city streets and in shopping malls, train stations, etc. Handsets registered with the phone-point provider can place calls when within range of a telepoint. CT-2 does not provide capability for transferring (handing off) active wireless calls from one phone point to another if a user moves out of range of the one to which the

[1] These ISM technologies either use spread spectrum techniques (direct sequence or frequency hopping) or very low-transmitter power ($<\sim$ 1 mW) as required by the ISM band regulations.

call was initiated. A CT-2+ technology, evolved from CT-2 and providing limited handoff capability, is being deployed in Canada. { CT-2+ deployment was never completed. } Phone-point service was introduced in the United Kingdom twice, but failed to attract enough customers to become a viable service. In Singapore and Hong Kong, however, CT-2 phone point has grown rapidly, reaching over 150,000 subscribers in Hong Kong [75] in mid-1994. The reasons for success in some places and failure in others are still being debated, but it is clear that the compactness of the Hong Kong and Singapore populations make the service more widely available, using fewer base stations than in more spreadout cities. Complaints of CT-2 phone-point users in trials have been that the radio coverage was not complete enough, and/or they could not tell whether there was coverage at a particular place, and the lack of handoff was inconvenient. In order to provide the alerting or ringing function for phone-point service, conventional radio pagers have been built into some CT-2 handsets. (The telephone network to which a CT-2 phone point is attached has no way of knowing from which base units to send a ringing message, even though the CT-2 handsets can be rung from a home base unit). { CT-2 phone points in Hong Kong peaked at about 170,000 subscribers. There was then a precipitous decline as these subscribers abandoned the limited service CT-2 in favor of high-tier PCS and cellular services which had reduced their prices to compete with phone points. Phone points have now been removed from Hong Kong. CT-2 phone points were also deployed in Korea, again with initial success followed by decline. The message is clear from the CT-2 and PHS experiences. The attributes of cordless phone-like low-tier PCS are very attractive, but need widespread coverage and vehicle speed handoff in order to be long term attractive and viable in the market. }

Another European evolution of cordless telephones is DECT, which was optimized for use inside buildings. Base units are attached through a controller to private branch exchanges (PBXs), key telephone systems, or phone company CENTREX telephone lines. DECT controllers can hand off active calls from one base unit to another as users move, and can page or ring handsets as a user walks through areas covered by different base units. { DECT has increased in deployment in some countries, notably Italy, for several applications including cordless telephone and wireless loop. There are reported to be perhaps 7 million or more DECT users in Europe. }

These cordless telephone evolutions to more widespread usage outside and inside with telepoints and to usage inside large buildings are illustrated in Fig. 15.1, along with the integration of paging into handsets to provide alerting for phone-point services. They represent the first attempts to increase the service area of mobility for low-power cordless telephones.

Some of the characteristics of the digital cordless telephone technologies, CT-2 and DECT, are listed in Table 15.1. Additional information can be found in [32, 50]. Even though there are significant differences between these technologies, e.g., multiple access technology [frequency division multiple access (FDMA) or time division multiple access (TDMA)/FDMA], and channel bit rate, there are many similarities that are fundamental to the design objectives discussed earlier and to a user's perception of them. These similarities and their implications are as follows.

32-kb/s Adaptive Differential Pulse Code Modulation (ADPCM) Digital Speech Encoding: This is a low-complexity (low-signal processing power) speech encoding process that provides wireline speech quality and is an international standard.

Average Transmitter Power $\leq$ 10 mW: This permits many hours of talk time with small, low-cost, lightweight batteries, but provides limited radio range.

Low-Complexity Radio Signal Processing: There is no forward error correction and no complex multipath mitigation (i.e., no equalization or spread spectrum).

TABLE 15.1 Wireless PCS Technologies

	High-Power Systems				Low-Power Systems			
	Digital Cellular (High-Tier PCS)				Low-Tier PCS		Digital Cordless	
System	IS-54	IS-95 (DS)	GSM	DCS-1800	WACS/PACS	Handi-Phone	DECT	CT-2
Multiple Access	TDMA/ FDMA	CDMA/ FDMA	TDMA/ FDMA	TDMA/ FDMA	TDMA/ FDMA	TDMA/ FDMA	TDMA/ FDMA	FDMA
Freq. band, MHz						1895–1907	1880–1990	864–868
Uplink, MHz	869–894	869–894	935–960	1710–1785	Emerg.			
Downlink, MHz	824–849 (USA)	824–849 (USA)	890–915 (Eur.)	1805–1880 (UK)	Tech.* (USA)	(Japan)	(Eur.)	(Eur. and Asia)
RF ch. spacing						300	1728	100
Downlink, KHz	30	1250	200	200	300			
Uplink, KHz	30	1250	200	200	300			
Modulation	$\pi/4$ DQPSK	BPSK/QPSK	GMSK	GMSK	$\pi/4$ QPSK	$\pi/4$ DQPSK	GFSK	GFSK
Portable txmit Power, max./avg.	600 mW/ 200 mW	600 mW	1 W/ 125 mW	1 W/ 125 mW	200 mW/ 25 mW	80 mW/ 10 mW	250 mW/ 10 mW	10 mW/ 5 mW
Speech coding	VSELP	QCELP	RPE-LTP	RPE-LTP	ADPCM	ADPCM	ADPCM	ADPCM
Speech rate, kb/s	7.95	8 (var.)	13	13	32/16/8	32	32	32
Speech ch./RF ch.	3	—	8	8	8/16/32	4	12	1
Ch. Bit rate, kb/s						384	1152	72
Uplink, kb/s	48.6		270.833	270.833	384			
Downlink, kb/s	48.6		270.833	270.833	384			
Ch. coding	1/2 rate conv.	1/2 rate fwd. 1/3 rate rev.	1/2 rate conv.	1/2 rate conv.	CRC	CRC	CRC (control)	None
Frame, ms	40	20	4.615	4.615	2.5	5	10	2

* Spectrum is 1.85–2.2 GHz allocated by the FCC for emerging technologies; DS is direct sequence.

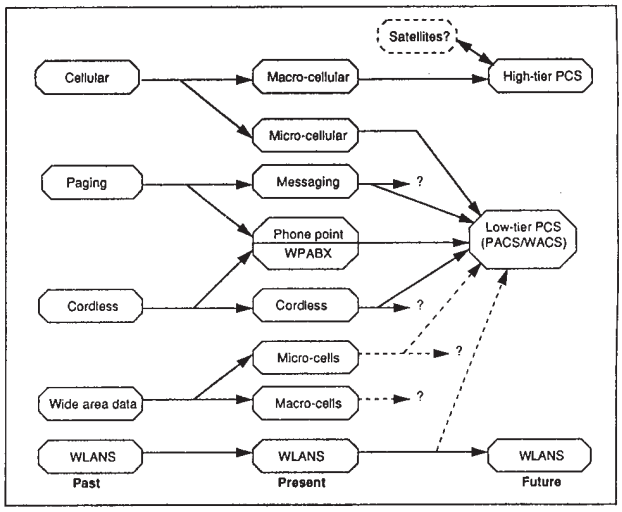


FIGURE 15.1

Low Transmission Delay, e.g., < 50 ms, and for CT-2 < 10-ms Round Trip: This is a speech-quality and network-complexity issue. A maximum of 10 ms should be allowed, taking into account additional inevitable delay in long-distance networks. Echo cancellation is generally required for delays > 10 ms.

Simple Frequency-Shift Modulation and Noncoherent Detection: Although still being low in complexity, the slightly more complex 4QPSK modulation with coherent detection provides significantly more spectrum efficiency, range, and interference immunity.

Dynamic Channel Allocation: Although this technique has potential for improved system capacity, the cordless-telephone implementations do not take full advantage of this feature for handoff and, thus, cannot reap the full benefit for moving users [15, 19].

Time Division Duplex (TDD): This technique permits the use of a single contiguous frequency band and implementation of diversity from one end of a radio link. Unless all base station transmissions are synchronized in time, however, it can incur severe cochannel interference penalties in outside environments [15, 16]. Of course, for cordless telephones used inside with base stations not having a propagation advantage, this is not a problem. Also, for small indoor PBX networks, synchronization of base station transmission is easier than is synchronization throughout a widespread outdoor network, which can have many adjacent base stations connected to different geographic locations for central control and switching.

15.3.2 Cellular Mobile Radio Systems

Cellular mobile radio systems are becoming known in the United States as high-tier personal communications service (PCS), particularly when implemented in the new 1.9-GHz PCS bands [20]. These systems generally can be categorized as providing high-mobility, wide-ranging, two-way tetherless voice communications. In these systems, high mobility refers to vehicular speeds, and also to widespread regional to nationwide coverage [27, 29, 50]. Mobile radio has been evolving for over 50 years. Cellular radio integrates wireless access with large-scale networks having sophisticated intelligence to manage mobility of users.

Cellular radio was designed to provide voice service to wide-ranging vehicles on streets and highways [29, 39, 50, 82], and generally uses transmitter power on the order of 100

times that of cordless telephones ($\approx$ 2 W for cellular). Thus, cellular systems can only provide reduced service to handheld sets that are disadvantaged by using somewhat lower transmitter power ($<$ 0.5 W) and less efficient antennas than vehicular sets. Handheld sets used inside buildings have the further disadvantage of attenuation through walls that is not taken into account in system design.

Cellular radio or high-tier PCS has experienced large growth as noted earlier. In spite of the limitations on usage of handheld sets already noted, handheld cellular sets have become very popular, with their sales becoming comparable to the sales of vehicular sets. Frequent complaints from handheld cellular users are that batteries are too large and heavy, and both talk time and standby time are inadequate. { Cellular and high-tier PCS pocket handsets have continued to decrease in size and weight and more efficient lithium batteries have been incorporated. This has increased their attractiveness (more on this in the later section "Reality Check"). For several years there have been many more pocket handsets sold than vehicular mounted sets every year. However, despite the improvements in these handsets and batteries, the complaints of weight and limited talk time still persist. The electronics have become essentially weightless compared to the batteries required for these high-tier PCS and cellular handsets. }

Cellular radio at 800 MHz has evolved to digital radio technologies [29, 39, 50] in the forms of the deployed systems standards

- Global Standard for Mobile (GSM) in Europe
- Japanese or personal digital cellular (JDC or PDC) in Japan
- U.S. TDMA digital cellular known as USDC or IS-54.

and in the form of the code division multiple access (CDMA) standard, IS-95, which is under development but not yet deployed. { Since the first edition was published, CDMA systems have been deployed in the U.S., Korea, Hong Kong, and other countries after many months (years) of redesign, reprogramming, and adjustment. These CDMA issues are discussed later in the section "New Technology." }

The most significant consideration in the design compromises made for the U.S. digital cellular or high-tier PCS systems was the high cost of cell sites (base stations). A figure often quoted is U.S. $1 million for a cell site. This consideration drove digital system designs to maximize users per megahertz and to maximize the users per cell site.

Because of the need to cover highways running through low-population-density regions between cities, the relatively high transmitter power requirement was retained to provide maximum range from high antenna locations.

Compromises that were accepted while maximizing the two just cited parameters are as follows.

- High transmitter power consumption.
- High user-set complexity, and thus high signal-processing power consumption.
- Low circuit quality.
- High network complexity, e.g., the new IS-95 technology will require complex new switching and control equipment in the network, as well as high-complexity wireless-access technology.

Cellular radio or high-tier PCS has also been evolving for a few years in a different direction, toward very small coverage areas or microcells. This evolution provides increased capacity in areas having high user density, as well as improved coverage of shadowed areas. Some microcell base stations are being installed inside, in conference center lobbies and

similar places of high user concentrations. Of course, microcells, also permit lower transmitter power that conserves battery power when power control is implemented, and base stations inside buildings circumvent the outside wall attenuation. Low-complexity microcell base stations also are considerably less expensive than conventional cell sites, perhaps two orders of magnitude less expensive. Thus, the use of microcell base stations provides large increases in overall system capacity, while also reducing the cost per available radio channel and the battery drain on portable subscriber equipment. This microcell evolution, illustrated in Fig. 15.1, moves handheld cellular sets in a direction similar to that of the expanded-coverage evolution of cordless telephones to phone points and wireless PBX.

Some of the characteristics of digital-cellular or high-tier PCS technologies are listed in Table 15.1 for IS-54, IS-95, and GSM at 900 MHz, and DCS-1800, which is GSM at 1800 MHz. { The technology listed here as IS-54 has also become known as IS-136 having more sophisticated digital control channels. These technologies, IS-54/IS-136 are also sometimes known as DAMPS (i.e., Digital AMPS), as U.S. TDMA or North American TDMA, or sometimes just as "TDMA." } Additional information can be found in [29, 39, 50]. The JDC or PDC technology, not listed, is similar to IS-54. As with the digital cordless technologies, there are significant differences among these cellular technologies, e.g., modulation type, multiple access technology, and channel bit rate. There are also many similarities, however, that are fundamental to the design objectives discussed earlier. These similarities and their implications are as follows.

Low Bit-Rate Speech Coding $\leq$13 kb/s with Some $\leq$8 kb/s: Low bit-rate speech coding obviously increases the number of users per megahertz and per cell site. However, it also significantly reduces speech quality [29], and does not permit speech encodings in tandem while traversing a network; see also the section on Other Issues later in this chapter.

Some Implementations Make Use of Speech Inactivity: This further increases the number of users per cell site, i.e., the cell-site capacity. It also further reduces speech quality [29], however, because of the difficulty of detecting the onset of speech. This problem is even worse in an acoustically noisy environment like an automobile.

High Transmission Delay; $\approx$200-ms Round Trip: This is another important circuit-quality issue. Such large delay is about the same as one-way transmission through a synchronous-orbit communications satellite. A voice circuit with digital cellular technology on both ends will experience the delay of a full satellite circuit. It should be recalled that one reason long-distance circuits have been removed from satellites and put onto fiber-optic cable is because customers find the delay to be objectionable. This delay in digital cellular technology results from both computation for speech bit-rate reduction and from complex signal processing, e.g., bit interleaving, error correction decoding, and multipath mitigation [equalization or spread spectrum code division multiple access (CDMA)].

High-Complexity Signal Processing, Both for Speech Encoding and for Demodulation: Signal processing has been allowed to grow without bound and is about a factor of 10 greater than that used in the low-complexity digital cordless telephones [29]. Since several watts are required from a battery to produce the high transmitter power in a cellular or high-tier PCS set, signal-processing power is not as significant as it is in the low-power cordless telephones; see also the section on Complexity/Coverage Area Comparisons later in this chapter.

Fixed Channel Allocation: The difficulties associated with implementing capacity-increasing dynamic channel allocation to work with handoff [15, 19] have impeded its adoption in systems requiring reliable and frequent handoff.

Frequency Division Duplex (FDD): Cellular systems have already been allocated paired-frequency bands suitable for FDD. Thus, the network or system complexity required

for providing synchronized transmissions [15, 16] from all cell sites for TDD has not been embraced in these digital cellular systems. Note that TDD has not been employed in IS-95 even though such synchronization is required for other reasons.

Mobile/Portable Set Power Control: The benefits of increased capacity from lower overall cochannel interference and reduced battery drain have been sought by incorporating power control in the digital cellular technologies.

15.3.3 Wide-Area Wireless Data Systems

Existing wide area data systems generally can be categorized as providing high mobility, wide-ranging, low-data-rate digital data communications to both vehicles and pedestrians [29, 50]. These systems have not experienced the rapid growth that the two-way voice technologies have, even though they have been deployed in many cities for a few years and have established a base of customers in several countries. Examples of these packet data systems are shown in Table 15.2.

TABLE 15.2 Wide-Area Wireless Packet Data Systems

	CDPD[1]	RAM Mobile (Mobitex)	ARDIS[2] (KDT)	Metricom (MDN)[3]
Data rate, kb/s	19.2	8 (19.2)	4.8 (19.2)	76
Modulation	GMSK BT = 0.5	GMSK	GMSK	GMSK
Frequency, MHz	800	900	800	915
Chan. spacing, kHz	30	12.5	25	160
Status	1994 service	Full service	Full service	In service
Access means	Unused AMPS channels	Slotted Aloha CSMA		FH SS (ISM)
Transmit power, W			40	1

Note: Data in parentheses ( ) indicates proposed.
[1]Cellular Digital Packet Data
[2]Advanced Radio Data Information Service
[3]Microcellular Data Network

The earliest and best known of these systems in the United States are the ARDIS network developed and run by Motorola, and the RAM mobile data network based on Ericsson Mobitex Technology. These technologies were designed to make use of standard, two-way voice, land mobile-radio channels, with 12.5- or 25-kHz channel spacing. In the United States these are specialized mobile radio services (SMRS) allocations around 450 MHz and 900 MHz. Initially, the data rates were low: 4.8 kb/s for ARDIS and 8 kb/s for RAM. The systems use high transmitter power (several tens of watts) to cover large regions from a few base stations having high antennas. The relatively low data capacity of a relatively expensive base station has resulted in economics that have not favored rapid growth.

The wide-area mobile data systems also are evolving in several different directions in an attempt to improve base station capacity, economics, and the attractiveness of the service. The technologies used in both the ARDIS and RAM networks are evolving to higher channel bit rates of 19.2 kb/s.

The cellular carriers and several manufacturers in the United States are developing and deploying a new wide area packet data network as an overlay to the cellular radio networks. This cellular digital packet data (CDPD) technology shares the 30-kHz spaced 800-MHz

voice channels used by the analog FM advanced mobile phone service (AMPS) systems. Data rate is 19.2 kb/s. The CDPD base station equipment also shares cell sites with the voice cellular radio system. The aim is to reduce the cost of providing packet data service by sharing the costs of base stations with the better established and higher cell-site capacity cellular systems. This is a strategy similar to that used by nationwide fixed wireline packet data networks that could not provide an economically viable data service if they did not share costs by leasing a small amount of the capacity of the interexchange networks that are paid for largely by voice traffic. { CDPD has been deployed in many U.S. cities for several years. However, it has not lived up to early expectations and has become "just another" wireless data service with some subscribers, but not with the large growth envisioned earlier. }

Another evolutionary path in wide-area wireless packet data networks is toward smaller coverage areas or microcells. This evolutionary path also is indicated on Fig. 15.1. The microcell data networks are aimed at stationary or low-speed users. The design compromises are aimed at reducing service costs by making very small and inexpensive base stations that can be attached to utility poles, the sides of buildings and inside buildings and can be widely distributed throughout a region. Base-station-to-base-station wireless links are used to reduce the cost of the interconnecting data network. In one network this decreases the overall capacity to serve users, since it uses the same radio channels that are used to provide service. Capacity is expected to be made up by increasing the number of base stations that have connections to a fixed-distribution network as service demand increases. Another such network uses other dedicated radio channels to interconnect base stations. In the high-capacity limit, these networks will look more like a conventional cellular network architecture, with closely spaced, small, inexpensive base stations, i.e., microcells, connected to a fixed infrastructure. Specialized wireless data networks have been built to provide metering and control of electric power distributions, e.g., Celldata and Metricom in California.

A large microcell network of small inexpensive base stations has been installed in the lower San Francisco Bay Area by Metricom, and public packet-data service was offered during early 1994. Most of the small (shoe-box size) base stations are mounted on street light poles. Reliable data rates are about 75 kb/s. The technology is based on slow frequency-hopped spread spectrum in the 902–928 MHz U.S. ISM band. Transmitter power is 1 W maximum, and power control is used to minimize interference and maximize battery life time. { The metricom network has been improved and significantly expanded in the San Francisco Bay Area and has been deployed in Washington, D.C. and a few other places in the U.S. However, like all wireless data services so far, it has failed to grow as rapidly or to attract as many subscribers as was originally expected. Wireless data overall has had only very limited success compared to that of the more voice-oriented technologies, systems, and services. }

15.3.4 High-Speed Wireless Local-Area Networks (WLANs)

Wireless local-area data networks can be categorized as providing low-mobility high-data-rate data communications within a confined region, e.g., a campus or a large building. Coverage range from a wireless data terminal is short, tens to hundreds of feet, like cordless telephones. Coverage is limited to within a room or to several rooms in a building. WLANs have been evolving for a few years, but overall the situation is chaotic, with many different products being offered by many different vendors [29, 59]. There is no stable definition of the needs or design objectives for WLANs, with data rates ranging from hundreds of kb/s to more than 10 Mb/s, and with several products providing one or two Mb/s wireless

link rates. The best description of the WLAN evolutionary process is: having severe birth pains. An IEEE standards committee, 802.11, has been attempting to put some order into this topic, but their success has been somewhat limited. A partial list of some advertised products is given in Table 15.3. Users of WLANs are not nearly as numerous as the users of more voice-oriented wireless systems. Part of the difficulty stems from these systems being driven by the computer industry that views the wireless system as just another plug-in interface card, without giving sufficient consideration to the vagaries and needs of a reliable radio system. { This section still describes the WLAN situation in spite of some attempts at standards in the U.S. and Europe, and continuing industry efforts. Some of the products in Table 15.3 have been discontinued because of lack of market and some new products have been offered, but the manufacturers still continue to struggle to find enough customers to support their efforts. Optimism remains high in the WLAN community that "eventually" they will find the "right" technology, service, or application to make WLANs "take off" — but the world still waits. Success is still quite limited. }

There are two overall network architectures pursued by WLAN designers. One is a centrally coordinated and controlled network that resembles other wireless systems. There are base stations in these networks that exercise overall control over channel access [44].

The other type of network architecture is the self-organizing and distributed controlled network where every terminal has the same function as every other terminal, and networks are formed ad hoc by communications exchanges among terminals. Such ad hoc networks are more like citizen band (CB) radio networks, with similar expected limitations if they were ever to become very widespread.

Nearly all WLANs in the United States have attempted to use one of the ISM frequency bands for unlicensed operation under part 15 of the FCC rules. These bands are 902–928 MHz, 2400–2483.5 MHz, and 5725–5850 MHz, and they require users to accept interference from any interfering source that may also be using the frequency. The use of ISM bands has further handicapped WLAN development because of the requirement for use of either frequency hopping or direct sequence spread spectrum as an access technology, if transmitter power is to be adequate to cover more than a few feet. One exception to the ISM band implementations is the Motorola ALTAIR, which operates in a licensed band at 18 GHz. { It appears that ALTAIR has been discontinued because of the limited market. } The technical and economic challenges of operation at 18 GHz have hampered the adoption of this 10–15 Mb/s technology. The frequency-spectrum constraints have been improved in the United States with the recent FCC allocation of spectrum from 1910–1930 MHz for unlicensed data PCS applications. Use of this new spectrum requires implementation of an access etiquette incorporating listen before transmit in an attempt to provide some coordination of an otherwise potentially chaotic, uncontrolled environment [68]. Also, since spread spectrum is not a requirement, access technologies and multipath mitigation techniques more compatible with the needs of packet-data transmission [59], e.g., multipath equalization or multicarrier transmission can be incorporated into new WLAN designs. { The FCC is allocating spectrum at 5 GHz for wideband wireless data for internet and next generation data network access, BUT it remains to be seen whether this initiative is any more successful than past wireless data attempts. Optimism is again high, BUT... }

Three other widely different WLAN activities also need mentioning. One is a large European Telecommunications Standards Institute (ETSI) activity to produce a standard for high performance radio local area network (HIPERLAN), a 20-Mb/s WLAN technology to operate near 5 GHz. Other activities are large U.S. Advance Research Projects Agency-(ARPA-) sponsored, WLAN research projects at the Universities of California at Berkeley (UCB), and at Los Angeles (UCLA). The UCB Infopad project is based on a coordinated network architecture with fixed coordinating nodes and direct-sequence spread spectrum

TABLE 15.3 Partial List of WLAN Products

Product Company Location	Freq., MHz	Link Rate, Mb/s	User Rate	Protocol(s)	Access	No. of chan. or Spread Factor	Mod./ Coding	Power, mW	Network Topology
Altair Plus Motorola Arlington Hts, IL	18–19 GHz	15	5.7 Mb/s	Ethernet			4-level FSK	25 peak	Eight devices/ radio; radio to base to ethernet
WaveLAN NCR/AT&T Dayton, OH	902–928	2	1.6 Mb/s	Ethernet-like	DS SS		DQPSK	250	Peer-to-peer
AirLan Solectek San Diego, CA	902–928		2 Mb/s	Ethernet	DS SS		DQPSK	250	PCMCIA w/ant.; radio to hub
Freeport Windata Inc. Northboro, MA	902–928	16	5.7 Mb/s	Ethernet	DS SS	32 chips/bit	16 PSK trellis coding	650	Hub
Intersect Persoft Inc. Madison, WI	902–928		2 Mb/s	Ethernet token ring	DS SS		DQPSK	250	Hub
LAWN O'Neill Comm. Horsham, PA	902–928		38.4 kb/s	AX.25	SS	20 users/chan.; max. 4 chan.		20	Peer-to-peer
WILAN Wi-LAN Inc. Calgary, Alberta	902–928	20	1.5 Mb/s/ chan.	Ethernet, token ring	CDMA/ TDMA	3 chan. 10–15 links each	unconven-tional	30	Peer-to-peer
RadioPort ALPS Electric USA	902–928		242 kb/s	Ethernet	SS	7/3 channels		100	Peer-to-peer
ArLAN 600 Telesys. SLW Don Mills, Ont.	902–928; 2.4 GHz		1.35 Mb/s	Ethernet	SS			1 W max	PCs with ant.; radio to hub
Radio Link Cal. Microwave Sunnyvale, CA	902–928; 2.4 GHz	250 kb/s	64 kb/s		FH SS	250 ms/hop 500 kHz space			Hub
Range LAN Proxim, Inc. Mountain View, CA	902–928		242 kb/s	Ethernet, token ring	DS SS	3 chan.		100	
RangeLAN 2 Proxim, Inc. Mountain View, CA	2.4 GHz	1.6	50 kb/s max.	Ethernet, token ring	FH SS	10 chan. at 5 kb/s; 15 sub-ch. each		100	Peer-to-peer bridge
Netwave Xircom Calabasas, CA	2.4 GHz	1/adaptor		Ethernet, token ring	FH SS	82 l-MHz chn. or "hops"			Hub
Freelink Cabletron Sys. Rochester, NH	2.4 and 5.8 GHz		5.7 Mb/s	Ethernet	DS SS	32 chips/bit	16 PSK trellis coding	100	Hub

(CDMA), whereas, the UCLA project is aimed at peer-to-peer networks and uses frequency hopping. Both ARPA sponsored projects are concentrated on the 900-MHz ISM band.

As computers shrink in size from desktop to laptop to palmtop, mobility in data network access is becoming more important to the user. This fact, coupled with the availability of more usable frequency spectrum, and perhaps some progress on standards, may speed the evolution and adoption of wireless mobile access to WLANs. From the large number of companies making products, it is obvious that many believe in the future of this market. { It should be noted that the objective for 10 MB/s data service with widespread coverage from a sparse distribution of widely separated base stations equivalent to cellular is unrealistic and unrealizable. This can be readily seen by considering a simple example. Consider a cellular coverage area that requires full cellular power of 0.5 watt to cover from a handset. Consider the handset to use a typical digital cellular bit rate of about 10 kb/s (perhaps 8 kb/s speech coding + overhead). With all else in the system the same, e.g., antennas, antenna height, receiver noise figure, detection sensitivity, etc., the 10 MB/s data would require 10 MB/s $\div$ 10 kb/s = 1000 times as much power as the 10 kb/s cellular. Thus, it would require 0.5 $\times$ 1000 = 500 watts for the wireless data transmitter. This is a totally unrealistic situation. If the data system operates at a higher frequency (e.g., 5 GHz) than the cellular system (e.g., 1 or 2 GHz) then there will be even more power required to overcome the additional loss at a higher frequency. The sometimes expressed desire by the wireless data community for a system to provide network access to users in and around buildings and to provide 10 MB/s over 10 miles with 10 milliwatts of transmitter power and costing $10.00 is totally impossible. It requires violation of the "laws of physics." }

15.3.5 Paging/Messaging Systems

Radio paging began many years ago as a one-bit messaging system. The one bit was: some one wants to communicate with you. More generally, paging can be categorized as one-way messaging over wide areas. The one-way radio link is optimized to take advantage of the asymmetry. High transmitter power (hundreds of watts to kilowatts), and high antennas at the fixed base stations permit low-complexity, very low-power-consumption, pocket paging receivers that provide long usage time from small batteries. This combination provides the large radio-link margins needed to penetrate walls of buildings without burdening the user set battery. Paging has experienced steady rapid growth for many years and serves about 15 million subscribers in the United States.

Paging also has evolved in several different directions. It has changed from analog tone coding for user identification to digitally encoded messages. It has evolved from the 1-b message, someone wants you, to multibit messages from, first, the calling party's telephone number to, now, short e-mail text messages. This evolution is noted in Fig. 15.1.

The region over which a page is transmitted has also increased from 1) local, around one transmitting antenna; to 2) regional, from multiple widely-dispersed antennas; to 3) nationwide, from large networks of interconnected paging transmitters. The integration of paging with CT-2 user sets for phone-point call alerting was noted previously.

Another evolutionary paging route sometimes proposed is two-way paging. This is an ambiguous and unrealizable concept, however, since the requirement for two-way communications destroys the asymmetrical link advantage so well exploited by paging. Two-way paging puts a transmitter in the user's set and brings along with it all of the design compromises that must be faced in such a two-way radio system. Thus, the word paging is not appropriate to describe a system that provides two-way communications. { The two-way paging situation is as unrealistic as that noted earlier for wide-area, high-speed, low-power wireless data. This can be seen by looking at the asymmetry situation in paging. In or-

der to achieve comparable coverage uplink and downlink, a 500-watt paging transmitter downlink advantage must be overcome in the uplink. Even considering the relatively high cellular handset transmit power levels on the order of 0.5 watt results in a factor of 1000 disadvantage, and 0.5 watt is completely incompatible with the low power paging receiver power requirements. If the same uplink and downlink coverage is required for an equivalent set of system parameters, then the only variable left to work with is bandwidth. If the paging link bit rate is taken to be 10 kb/sec (much higher than many paging systems), then the usable uplink rate is 10 kb/s/1000 = 10 B/s, an unusably low rate. Of course, some uplink benefit can be gained because of better base station receiver noise figure and by using forward error correction and perhaps ARQ. However, this is unlikely to raise the allowable rate to greater than 100 B/s which even though likely overoptimistic is still unrealistically low and we have assumed an unrealistically high transmit power in the two-way "pager!" }

15.3.6 Satellite-Based Mobile Systems

Satellite-based mobile systems are the epitome of wide-area coverage, expensive base station systems. They generally can be categorized as providing two-way (or one-way) limited quality voice and/or very limited data or messaging to very wide-ranging vehicles (or fixed locations). These systems can provide very widespread, often global, coverage, e.g., to ships at sea by INMARSAT. There are a few messaging systems in operation, e.g., to trucks on highways in the United States by Qualcomm's Omnitracs system.

A few large-scale mobile satellite systems have been proposed and are being pursued: perhaps the best known is Motorola's Iridium; others include Odyssey, Globalstar, and Teledesic. The strength of satellite systems is their ability to provide large regional or global coverage to users outside buildings. However, it is very difficult to provide adequate link margin to cover inside buildings, or even to cover locations shadowed by buildings, trees, or mountains. A satellite system's weakness is also its large coverage area. It is very difficult to provide from Earth orbit the small coverage cells that are necessary for providing high overall systems capacity from frequency reuse. This fact, coupled with the high cost of the orbital base stations, results in low capacity along with the wide overall coverage but also in expensive service. Thus, satellite systems are not likely to compete favorably with terrestrial systems in populated areas or even along well-traveled highways. They can complement terrestrial cellular or PCS systems in low-population-density areas. It remains to be seen whether there will be enough users with enough money in low-population-density regions of the world to make satellite mobile systems economically viable. { Some of the mobile satellite systems have been withdrawn, e.g., Odyssey. Some satellites in the Iridium and Globalstar systems have been launched. The industry will soon find out whether these systems are economically viable. }

Proposed satellite systems range from 1) low-Earth-orbit systems (LEOS) having tens to hundreds of satellites through 2) intermediate- or medium-height systems (MEOS) to 3) geostationary or geosynchronous orbit systems (GEOS) having fewer than ten satellites. LEOS require more, but less expensive, satellites to cover the Earth, but they can more easily produce smaller coverage areas and, thus, provide higher capacity within a given spectrum allocation. Also, their transmission delay is significantly less (perhaps two orders of magnitude!), providing higher quality voice links, as discussed previously. On the other hand, GEOS require only a few, somewhat more expensive, satellites (perhaps only three) and are likely to provide lower capacity within a given spectrum allocation and suffer severe transmission-delay impairment on the order of 0.5 s. Of course, MEOS fall in between these extremes. The possible evolution of satellite systems to complement high-tier PCS is indicated in Fig. 15.1.

15.3.7 {Fixed Point-to-Multipoint Wireless Loops

Wideband point-to-multipoint wireless loop technologies sometimes have been referred to earlier as "wireless cable" when they were proposed as an approach for providing interactive video services to homes [30]. However, as the video application started to appear less attractive, the application emphasis shifted to providing wideband data access for the internet, the worldwide web, and future wideband data networks. Potentially lower costs are the motivation for this wireless application. As such, these technologies will have to compete with existing coaxial cable and fiber/coax distribution by CATV companies, with satellites, and with fiber and fiber/coax systems being installed or proposed by telephone companies and other entities [30]. Another competitor is asymmetric digital subscriber line technology, which uses advanced digital signal processing to provide high-bandwidth digital distribution over twisted copper wire pairs.

In the U.S. two widely different frequency bands are being pursued for fixed point-to-multipoint wireless loops. These bands are at 28 GHz for local multipoint distribution systems or services (LMDS) [52] and 2.5 to 2.7 GHz for microwave or metropolitan distribution systems (MMDS) [74]. The goal of low-cost fixed wireless loops is based on the low cost of point-to-multipoint line-of-sight wireless technology. However, significant challenges are presented by the inevitable blockage by trees, terrain, and houses, and by buildings in heavily built-up residential areas. Attenuation in rainstorms presents an additional problem at 28 GHz in some localities. Even at the 2.5-GHz MMDS frequencies, the large bandwidth required for distribution of many video channels presents a challenge to provide adequate radio-link margin over obstructed paths. From mobile satellite investigations it is known that trees can often produce over 15 dB additional path attenuation [38]. Studies of blockage by buildings in cities have shown that it is difficult to have line-of-sight access to more than 60% of the buildings from a single base station [55]. Measurements in a region in Brooklyn, NY [60], suggest that access from a single base station can range from 25% to 85% for subscriber antenna heights of 10 to 35 ft and a base station height of about 290 ft. While less blockage by houses could be expected in residential areas, such numbers would suggest that greater than 90% access to houses could be difficult, even from multiple elevated locations, when mixes of one- and two-story houses, trees, and hills are present. In regions where tree cover is heavy, e.g., the eastern half of the U.S., tree cover in many places will present a significant obstacle. Heavy rainfall is an additional problem at 28 GHz in some regions. In spite of these challenges, the lure of low-cost wireless loops is attracting many participants, both service providers and equipment manufacturers. }

15.3.8 Reality Check

Before we go on to consider other applications and compromises, perhaps it would be helpful to see if there is any indication that the previous discussion is valid. For this check, we could look at cordless telephones for telepoint use (i.e., pocketphones) and at pocket cellular telephones that existed in the 1993 time frame.

Two products from one United States manufacturer are good for this comparison. One is a third-generation hand-portable analog FM cellular phone from this manufacturer that represents their second generation of pocketphones. The other is a first-generation digital cordless phone built to the United Kingdom CT-2 common air interface (CAI) standard. Both units are of flip phone type with the earpiece on the main handset body and the mouthpiece formed by or on the flip-down part. Both operate near 900 MHz and have 1/4 wavelength pull-out antennas. Both are fully functional within their class of operation (i.e., full number of U.S. cellular channels, full number of CT-2 channels, automatic channel

setup, etc.) Table 15.4 compares characteristics of these two wireless access pocketphones from the same manufacturer.

TABLE 15.4 Comparison of CT-2 and Cellular Pocket Size Flip-Phones from the Same Manufacturer

Characteristics/Parameter	CT-2	Cellular
Weight, oz		
Flip phone only	5.2	4.2
Battery[1] only	1.9	3.6
Total unit	7.1	7.8
Size (max.dimensions), in		
Flip phone only	5.9 × 2.2 × 0.95 8.5 in^3	5.5 × 2.4 × 0.9 —
Battery[1] only	1.9 × 1.3 × 0.5 internal	4.7 × 2.3 × 0.4 external
Total unit	5.9 × 2.2 × 0.95 8.5 in^3	5.9 × 2.4 × 1.1 11.6 in^3
Talk-time, min (h)		
Rechargeable battery[2]	180 (3)	45
Nonrechargeable battery	600 (10)	N/A
Standby time, h		
Rechargeable battery	30	8
Nonrechargeable battery	100	N/A
Speech quality	32 kb/s telephone quality	30 kHz FM depends on channel quality
Transmit power avg., W	0.005	0.5

[1] Rechargeable battery.
[2] Ni-cad battery.

The following are the most important items to note in the Table 15.4 comparison.

1. The talk time of the low-power pocketphone is four times that of the high-power pocket-phone.
2. The battery inside the low-power pocketphone is about one-half the weight and size of the battery attached to the high-power pocketphone.
3. The battery-usage ratio, talk time/weight of battery, is eight times greater, almost an order of magnitude, for the low-power pocketphone compared to the high-power pocketphone!
4. Additionally, the lower power (5 mW) digital cordless pocketphone is slightly smaller and lighter than the high-power (500 mW) analog FM cellular mobile pocketphone.

{ Similar comparisons can be made between PHS advanced cordless/low-tier PCS phones and advanced cellular/high-tier PCS pocketphones. New lithium batteries have permitted increased talk time in pocketphones. Digital control/paging channels facilitate significantly extended standby time. Advances in solid-state circuits have reduced the size and weight of cellular pocketphone electronics so that they are almost insignificant compared to the battery required for the high power transmitter and complex digital signal processing. However, even with all these changes, there is still a very significant weight and talk time benefit in the low complexity PHS handsets compared to the most advanced cellular/high-tier PCS handsets. Picking typical minimum size and weight handsets for both technologies results in the following comparisons.

	PHS	Cellular
weight, oz		
total unit	3	4.5
size		
total unit	4.2 in^3	—
talk-time, h	8	3
standby time, h	600	48

From the table, the battery usage ratio has been reduced to a factor of about 4 from a factor of 8, but this is based on total weight, not battery weight alone as used for the earlier CT-2 and cellular comparison.

Thus, there is still a large talk time and weight benefit for low-power low-complexity low-tier PCS compared to higher power high-complexity, high tier PCS and cellular. }

The following should also be noted.

1. The room for technology improvement of the CT-2 cordless phone is greater since it is first generation and the cellular phone is second/third generation.
2. A digital cellular phone built to the IS-54, GSM, or JDC standard, or in the proposed United States CDMA technology, would either have less talk time or be heavier and larger than the analog FM phone, because: a) the low-bit-rate digital speech coder is more complex and will consume more power than the analog speech processing circuits; b) the digital units have complex digital signal-processing circuits for forward error correction—either for delay dispersion equalizing or for spread-spectrum processing—that will consume significant amounts of power and that have no equivalents in the analog FM unit; and c) power amplifiers for the shaped-pulse nonconstant-envelope digital signals will be less efficient than the amplifiers for constant-envelope analog FM. Although it may be suggested that transmitter power control will reduce the weight and size of a CDMA handset and battery, if that handset is to be capable of operating at full power in fringe areas, it will have to have capabilities similar to other cellular sets. Similar power control applied to a CT-2-like low-maximum-power set would also reduce its power consumption and thus also its weight and size.

The major difference in size, weight, and talk time between the two pocketphones is directly attributable to the two orders of magnitude difference in average transmitter power. The generation of transmitter power dominates power consumption in the analog cellular phone. Power consumption in the digital CT-2 phone is more evenly divided between transmitter-power generation and digital signal processing. Therefore, power consumption in complex digital signal processing would have more impact on talk time in small low-power personal communicators than in cellular handsets where the transmitter-power generation is so large. Other than reducing power consumption for both functions, the only alternative for increasing talk time and reducing battery weight is to invent new battery technology having greater density; see section on Other Issues later in this chapter.

In contrast, lowering the transmitter power requirement, modestly applying digital signal processing, and shifting some of the radio coverage burden to a higher density of small, low-power, low-complexity, low-cost fixed radio ports has the effect of shifting some of the talk time, weight, and cost constraints from battery technology to solid state electronics technology, which continues to experience orders-of-magnitude improvements in the span of several years. Digital signal-processing complexity, however, cannot be permitted to overwhelm power consumption in low-power handsets; whereas small differences in complexity

will not matter much, orders-of-magnitude differences in complexity will continue to be significant.

Thus, it can be seen from Table 15.4 that the size, weight, and quality arguments in the preceding sections generally hold for these examples. It also is evident from the preceding paragraphs that they will be even more notable when comparing digital cordless pocketphones with digital cellular pocketphones of the same development generations.

15.4 Evolution Toward the Future and to Low-Tier Personal Communications Services

After looking at the evolution of several wireless technologies and systems in the preceding sections it appears appropriate to ask again: wireless personal communications, What is it? All of the technologies in the preceding sections claim to provide wireless personal communications, and all do to some extent. All have significant limitations, however, and all are evolving in attempts to overcome the limitations. It seems appropriate to ask, what are the likely endpoints? Perhaps some hint of the endpoints can be found by exploring what users see as limitations of existing technologies and systems and by looking at the evolutionary trends.

In order to do so, we summarize some important clues from the preceding sections and project them, along with some U.S. standards activity, toward the future.

Digital Cordless Telephones

- Strengths: good circuit quality; long talk time; small lightweight battery; low-cost sets and service.
- Limitations: limited range; limited usage regions.
- Evolutionary trends: phone points in public places; wireless PBX in business.
- Remaining limitations and issues: limited usage regions and coverage holes; limited or no handoff; limited range.

{ Experience with PHS and CT-2 phone point have provided more emphasis on the need for vehicle speed handoff and continuous widespread coverage of populated areas and of highways in between. }

Digital Cellular Pocket Handsets

- Strength: widespread service availability.
- Limitations: limited talk time; large heavy batteries; high-cost sets and service; marginal circuit quality; holes in coverage and poor in-building coverage; limited data capabilities; complex technologies.
- Evolutionary trends: microcells to increase capacity and in-building coverage and to reduce battery drain; satellite systems to extend coverage.
- Remaining limitations and issues: limited talk time and large battery; marginal circuit quality; complex technologies.

Wide Area Data

- Strength: digital messages.
- Limitations: no voice, limited data rate; high cost.
- Evolutionary trends: microcells to increase capacity and reduce cost; share facilities with voice systems to reduce cost.

- Remaining limitations and issues: no voice; limited capacity.

Wireless Local Area Networks (WLANs)

- Strength: high data rate.
- Limitations: insufficient capacity for voice, limited coverage; no standards; chaos.
- Evolutionary trends: hard to discern from all of the churning.

Paging/messaging

- Strengths: widespread coverage; long battery life; small lightweight sets and batteries; economical.
- Limitations: one-way message only; limited capacity.
- Evolutionary desire: two-way messaging and/or voice; capacity.
- Limitations and issues: two-way link cannot exploit the advantages of one-way link asymmetry.

{**Fixed Wireless Loops**

- Strength: High data rates.
- Limitations: no mobility. }

There is a strong trajectory evident in these systems and technologies aimed at providing the following features.

High Quality Voice and Data

- To small, lightweight, pocket carried communicators.
- Having small lightweight batteries.
- Having long talk time and long standby battery life.
- Providing service over large coverage regions.
- For pedestrians in populated areas (but not requiring high population density).
- Including low to moderate speed mobility with handoff. { It has become evident from the experience with PHS and CT-2 phone point that vehicle speed handoff is essential so that handsets can be used in vehicles also. }

Economical Service

- Low subscriber-set cost.
- Low network-service cost.

Privacy and Security of Communications

- Encrypted radio links.

This trajectory is evident in all of the evolving technologies but can only be partially satisfied by any of the existing and evolving systems and technologies! Trajectories from all of the evolving technologies and systems are illustrated in Fig. 15.1 as being aimed at low-tier personal communications systems or services, i.e., low-tier PCS. Taking characteristics from cordless, cellular, wide-area data and, at least moderate-rate, WLANs, suggests the following attributes for this low-tier PCS.

1. 32 kb/s ADPCM speech encoding in the near future to take advantage of the low complexity and low power consumption, and to provide low-delay high-quality speech.
2. Flexible radio link architecture that will support multiple data rates from several kilobits per second. This is needed to permit evolution in the future to lower bit rate speech as technology improvements permit high quality without excessive power consumption or transmission delay and to provide multiple data rates for data transmission and messaging.
3. Low transmitter power ($\leq$ 25 mW average) with adaptive power control to maximize talk time and data transmission time. This incurs short radio range that requires many base stations to cover a large region. Thus, base stations must be small and inexpensive, like cordless telephone phone points or the Metricom wireless data base stations. { The lower power will require somewhat closer spacing of base stations in cluttered environments with many buildings, etc. This issue is dealt with in more detail in Section 15.5. The issues associated with greater base station spacing along highways are also considered in Section 15.5. }
4. Low-complexity signal processing to minimize power consumption. Complexity one-tenth that of digital cellular or high-tier PCS technologies is required [29]. With only several tens of milliwatts (or less under power control) required for transmitter power, signal processing power becomes significant.
5. Low cochannel interference and high coverage area design criteria. In order to provide high-quality service over a large region, at least 99% of any covered area must receive good or better coverage and be below acceptable cochannel interference limits. This implies less than 1% of a region will receive marginal service. This is an order-of-magnitude higher service requirement than the 10% of a region permitted to receive marginal service in vehicular cellular system (high-tier PCS) design criteria.
6. Four-level phase modulation with coherent detection to maximize radio link performance and capacity with low complexity.
7. Frequency division duplexing to relax the requirement for synchronizing base station transmissions over a large region. { PHS uses time division duplexing and requires base station synchronization. In first deployments, one provider did not implement this synchronization. The expected serious performance degradation prompted system upgrades to provide the needed synchronization. While this is not a big issue, it does add complexity to the system and decreases the overall robustness. }
8. { As noted previously, experience with PHS and CT-2 phone point have emphasized the need for vehicular speed handoff in these low-tier PCS systems. Such handoff is readily implemented in PACS and has been demonstrated in the field [51]. This issue is discussed in more detail later in this section. }

Such technologies and systems have been designed, prototyped, and laboratory and field tested and evaluated for several years [7, 23, 24, 25, 26, 27, 28, 29, 31, 32, 50]. The viewpoint expressed here is consistent with the progress in the Joint Technical Committee (JTC) of the U.S. standards bodies, Telecommunications Industry Association (TIA) and Committee T1 of the Alliance for Telecommunications Industry Solutions (ATIS). Many technologies and systems were submitted to the JTC for consideration for wireless PCS in the new 1.9-GHz frequency bands for use in the United States [20]. Essentially all of the technologies and systems listed in Table 15.1, and some others, were submitted in late 1993. It was evident

that there were at least two and perhaps three distinctly different classes of submissions. No systems optimized for packet data were submitted, but some of the technologies are optimized for voice.

One class of submissions was the group labeled high-power systems, digital cellular (high-tier PCS) in Table 15.1. These are the technologies discussed previously in this chapter. They are highly optimized for low-bit-rate voice and, therefore, have somewhat limited capability for serving packet-data applications. Since it is clear that wireless services to wide ranging high-speed mobiles will continue to be needed, and that the technology already described for low-tier PCS may not be optimum for such services, Fig. 15.1 shows a continuing evolution and need in the future for high-tier PCS systems that are the equivalent of today's cellular radio. There are more than 100 million vehicles in the United States alone. In the future, most, if not all, of these will be equipped with high-tier cellular mobile phones. Therefore, there will be a continuing and rapidly expanding market for high-tier systems.

Another class of submissions to the JTC [20] included the Japanese personal handyphone system (PHS) and a technology and system originally developed at Bellcore but carried forward to prototypes and submitted to the JTC by Motorola and Hughes Network Systems. This system was known as wireless access communications systems (WACS).[2] These two submissions were so similar in their design objectives and system characteristics that, with the agreement of the delegations from Japan and the United States, the PHS and WACS submissions were combined under a new name, personal access communication systems (PACS), that was to incorporate the best features of both. This advanced, low-power wireless access system, PACS, was to be known as low-tier PCS. Both WACS/PACS and Handyphone (PHS) are shown in Table 15.1 as low-tier PCS and represent the evolution to low-tier PCS in Fig. 15.1. The WACS/PACS/ UDPC system and technology are discussed in [7, 23, 24, 25, 26, 28, 29, 31, 32, 50].

In the JTC, submissions for PCS of DECT and CT-2 and their variations were also lumped under the class of low-tier PCS, even though these advanced digital cordless telephone technologies were somewhat more limited in their ability to serve all of the low-tier PCS needs. They are included under digital cordless technologies in Table 15.1. Other technologies and systems were also submitted to the JTC for high-tier and low-tier applications, but they have not received widespread industry support.

One wireless access application discussed earlier that is not addressed by either high-tier or low-tier PCS is the high-speed WLAN application. Specialized high-speed WLANs also are likely to find a place in the future. Therefore, their evolution is also continued in Fig. 15.1. The figure also recognizes that widespread low-tier PCS can support data at several hundred kilobits per second and, thus, can satisfy many of the needs of WLAN users.

It is not clear what the future roles are for paging/messaging, cordless telephone appliances, or wide-area packet-data networks in an environment with widespread contiguous coverage by low-tier and high-tier PCS. Thus, their extensions into the future are indicated with a question mark in Fig. 15.1.

Those who may object to the separation of wireless PCS into high-tier and low-tier should review this section again, and note that we have two tiers of PCS now. On the voice side there is cellular radio, i.e., high-tier PCS, and cordless telephone, i.e., an early form of low-tier PCS. On the data side there is wide-area data, i.e., high-tier data PCS, and WLANs,

[2]WACS was known previously as Universal Digital Portable Communications (UDPC).

i.e., perhaps a form of low-tier data PCS. In their evolutions, these all have the trajectories discussed and shown in Fig. 15.1 that point surely toward low-tier PCS. It is this low-tier PCS that marketing studies continue to project is wanted by more than half of the U.S. households or by half of the people, a potential market of over 100 million subscribers in the United States alone. Similar projections have been made worldwide.

{ PACS technology [6] has been prototyped by several manufacturers. In 1995 field demonstrations were run in Boulder, CO at a U.S. West test site using radio ports (base stations) and radio port control units made by NEC. "Handset" prototypes made by Motorola and Panasonic were trialed. The handsets and ports were brought together for the first time in Boulder. The highly successful trial demonstrated the ease of integrating the subsystems of the low-complexity PACS technology and the overall advantages of PACS from a user's perspective as noted throughout this chapter. Effective vehicular speed operation was demonstrated in these tests. Also, Hughes Network Systems (HNS) has developed and tested many sets of PACS infrastructure technology with different handsets in several settings and has many times demonstrated highly reliable high vehicular speed (in excess of 70 mi/hr) operation and handoff among several radio ports. Motorola also has demonstrated PACS equipment in several settings at vehicular speeds as well as for wireless loop applications. Highly successful demonstrations of PACS prototypes have been conducted from Alaska to Florida, from New York to California, and in China and elsewhere.

A PACS deployment in China using NEC equipment started to provide service in 1998. The U.S. Service Provider, 21st Century Telesis, is poised to begin a PACS deployment in several states in the U.S. using infrastructure equipment from HNS and handsets and switching equipment from different suppliers. Perhaps, with a little more support of these deployments, the public will finally be able to obtain the benefits of low-tier PCS. }

15.5 Comparisons with Other Technologies

15.5.1 Complexity/Coverage Area Comparisons

Experimental research prototypes of radio ports and subscriber sets [64, 66] have been constructed to demonstrate the technical feasibility of the radio link requirements in [7]. These WACS prototypes generally have the characteristics and parameters previously noted, with the exceptions that 1) the portable transmitter power is lower (10 mW average, 100 mW peak), 2) dynamic power control and automatic time slot transfer are not implemented, and 3) a rudimentary automatic link-transfer implementation is based only on received power. The experimental base stations transmit near 2.17 GHz; the experimental subscriber sets transmit near 2.12 GHz. Both operated under a Bellcore experimental license. The experimental prototypes incorporate application-specific, very large-scale integrated circuits[3] fabricated to demonstrate the feasibility of the low-complexity high-performance digital signal-processing techniques [63, 64] for symbol timing and coherent bit detection. These techniques permit the efficient short TDMA bursts having only 100 b that are necessary for low-delay TDMA implementations. Other digital signal-processing functions in the prototypes are implemented in programmable logic devices. All of the digital signal-processing functions combined require about 1/10 of the logic gates that are required for digital signal processing in vehicular digital cellular mobile implementations [42, 62, 63]; that is, this low-complexity PCS implementation having no delay-dispersion-compensating circuits and

[3] Applications specific integrated circuits (ASIC), very large-scale integration (VLSI).

no forward error-correction decoding and is about 1/10 as complex as the digital cellular implementations that include these functions.[4] The 32 kb/s ADPCM speech-encoding in the low-complexity PCS implementation is also about 1/10 as complex as the less than 10-kb/s speech encoding used in digital cellular implementations. This significantly lower complexity will continue to translate into lower power consumption and cost. It is particularly important for low-power pocket personal communicators with power control in which the DC power expended for radio frequency transmitting can be only tens of milliwatts for significant lengths of time.

The experimental radio links have been tested in the laboratory for detection sensitivity [bit error rate (BER) vs SNR] [18, 61, 66] and for performance against cochannel interference [1] and intersymbol interference caused by multipath delay spread [66]. These laboratory tests confirm the performance of the radio-link techniques. In addition to the laboratory tests, qualitative tests have been made in several PCS environments to compare these experimental prototypes with several United States CT-1 cordless telephones at 50 MHz, with CT-2 cordless telephones at 900 MHz, and with DCT-900 cordless telephones at 900 MHz. Some of these comparisons have been reported [8, 71, 84, 85]. In general, depending on the criteria, e.g., either no degradation or limited degradation of circuit quality, these WACS experimental prototypes covered areas inside buildings that ranged from 1.4 to 4 times the areas covered by the other technologies. The coverage areas for the experimental prototypes were always substantially limited in two or three directions by the outside walls of the buildings. These area factors could be expected to be even larger if the coverage were not limited by walls, i.e., once all of a building is covered in one direction, no more area can be covered no matter what the radio link margin. The earlier comparisons [8, 84, 85] were made with only two-branch uplink diversity before subscriber-set transmitting antenna switching was implemented and, with only one radio port before automatic radio-link transfer was implemented. The later tests [71] included these implementations. These reported comparisons agree with similar unreported comparisons made in a Bellcore Laboratory building. Similar coverage comparison results have been noted for a 900-MHz ISM-band cordless telephone compared to the 2-GHz experimental prototype. The area coverage factors (e.g., $\times 1.4$ to $\times 4$) could be expected to be even greater if the cordless technologies had also been operated at 2 GHz since attenuation inside buildings between similar small antennas is about 7 dB greater at 2 GHz than at 900 MHz [35, 36] and the 900 MHz handsets transmitted only 3 dB less average power than the 2-GHz experimental prototypes. The greater area coverage demonstrated for this technology is expected because of the different compromises noted earlier; the following, in particular.

1. Coherent detection of QAM provides more detection sensitivity than noncoherent detection of frequency-shift modulations [17].
2. Antenna diversity mitigates bursts of errors from multipath fading [66].
3. Error detection and blanking of TDMA bursts having errors significantly improves perceived speech quality [72]. (Undetected errors in the most significant bit cause sharp audio pops that seriously degrade perceived speech quality.)
4. Robust symbol timing and burst and frame synchronization reduce the number of frames in error due to imperfect timing and synchronization [66].

[4]Some indication of VLSI complexity can be seen by the number of people required to design the circuits. For the low-complexity TDMA ASIC set, only one person part time plus a student part time were required; the complex CDMA ASIC has six authors on the paper alone.

5. Transmitting more power from the radio port compared to the subscriber set offsets the less sensitive subscriber set receiver compared to the port receiver that results from power and complexity compromises made in a portable set.

Of course, as expected, the low-power (10-mW) radio links cover less area than high-power (0.5-W) cellular mobile pocketphone radio links because of the 17-dB transmitter power difference resulting from the compromises discussed previously. In the case of vehicular mounted sets, even more radio-link advantage accrues to the mobile set because of the higher gain of vehicle-mounted antennas and higher transmitter power (3 W).

{ The power difference between a low-tier PACS handset and a high-tier PCS or cellular pocket handset is not as significant in limiting range as is often portrayed. Other differences in deployment scenarios for low-tier and high-tier systems are as large or even larger factors, e.g., base station antenna height and antenna gain. This can be seen by considering using the same antennas and receiver noise figures at base stations and looking at the range of high-tier and low-tier handsets. High-tier handsets typically transmit a maximum average power of 0.5 watt. The PACS handset average transmit power is 25 milliwatts (peak power is higher for TDMA, but equal comparisons can be made considering average power and equivalent receiver sensitivities). This power ratio of $\times 20$ translates to approximately a range reduction of a factor of about 0.5 for an environment with a distance dependence of $1/(\mathrm{d})^4$ or a factor of about 0.4 for a $1/(\mathrm{d})^{3.5}$ environment. These represent typical values of distance dependence for PCS and cellular environments. Thus, if the high-tier handset would provide a range of 5 miles in some environment, the low-tier handset would provide a range of 2 to 2.5 miles in the same environment, if the base station antennas and receiver noise figures were the same. This difference in range is no greater than the difference in range between high-tier PCS handsets used in 1.9 GHz systems and cellular handsets with the same power used in 800 MHz systems. Range factors are discussed further in the next section. }

15.5.2 {Coverage, Range, Speed, and Environments

Interest has been expressed in having greater range for low-tier PCS technology for low-population-density areas. One should first note that the range of a wireless link is highly dependent on the amount of clutter or obstructions in the environment in which it is operated. For example, radio link calculations that result in a 1400-ft base station (radio-port) separation at 1.9 GHz contain over 50-dB margin for shadowing from obstructions and multipath effects [25, 37]. Thus, in an environment without obstructions, e.g., along a highway, the base station separation can be increased at least by a factor of 4 to over a mile, i.e., 25 dB for an attenuation characteristic of d^{-4}, while providing the same quality of service, without any changes to the base station or subscriber transceivers, and while still allowing over 25-dB margin for multipath and some shadowing. This remaining margin allows for operation of a handset inside an automobile. In such an unobstructed environment, multipath RMS delay spread [21, 33] will still be less than the 0.5 μs in which PACS was designed to operate [28].

Operation at still greater range along unobstructed highways or at a range of a mile along more obstructed streets can be obtained in several ways. Additional link gain of 6 dB can be obtained by going from omnidirectional antennas at base stations to 90° sectored antennas (four sectors). Another 6 dB can be obtained by raising base station antennas by a factor of 2 from 27 ft to 55 ft in height. This additional 12 dB will allow another factor of 2 increase in range to 2-mile base station separation along highways, or to about 3000-ft separation in residential areas. Even higher-gain and taller antennas could be used

to concentrate coverage along highways, particularly in rural areas. Of course, range could be further increased by increasing the power transmitted.

As the range of the low-tier PACS technology is extended in cluttered areas by increasing link gain, increased RMS delay spread is likely to be encountered. This will require increasing complexity in receivers. A factor of 2 in tolerance of delay spread can be obtained by interference-canceling signal combining [76, 77, 78, 79, 80] from two antennas instead of the simpler selection diversity combining originally used in PACS. This will provide adequate delay-spread tolerance for most suburban environments [21, 33].

The PACS downlink contains synchronization words that could be used to train a conventional delay-spread equalizer in subscriber set receivers. Constant-modulus (blind) equalization will provide greater tolerance to delay spread in base station receivers on the uplink [45, 46, 47, 48] than can be obtained by interference-cancellation combining from only two antennas. The use of more base-station antennas and receivers can also help mitigate uplink delay spread. Thus, with some added complexity, the low-tier PACS technology can work effectively in the RMS delay spreads expected in cluttered environments for base station separations of 2 miles or so.

The guard time in the PACS TDMA uplink is adequate for 1-mile range, i.e., 2-mile separation between base station and subscriber transceivers. A separation of up to 3 miles between transceivers could be allowed if some statistical outage were accepted for the few times when adjacent uplink timeslots are occupied by subscribers at the extremes of range (near–far). With some added complexity in assigning timeslots, the assignment of subscribers at very different ranges to adjacent timeslots could be avoided, and the base station separation could be increased to several miles without incurring adjacent slot interference. A simple alternative in low-density (rural) areas, where lower capacity could be acceptable and greater range could be desirable, would be to use every other timeslot to ensure adequate guard time for range differences of many tens of miles. Also, the capability of transmitter time advance has been added to PACS standard in order to increase the range of operation. Such time advance is applied in the cellular TDMA technologies.

The synchronization, carrier recovery, and detection in the low-complexity PACS transceivers will perform well at highway speeds. The two-receiver diversity used in uplink transceivers also will perform well at highway speeds. The performance of the single-receiver selection diversity used in the low-complexity PACS downlink transceivers begins to deteriorate at speeds above about 30 mi/h. However, at any speed, the performance is always at least as good as that of a single transceiver without the low-complexity diversity. Also, fading in the relatively uncluttered environment of a highway is likely to have a less severe Ricean distribution, so diversity will be less needed for mitigating the fading. Cellular handsets do not have diversity. Of course, more complex two-receiver diversity could be added to downlink transceivers to provide two-branch diversity performance at highway speeds. It should be noted that the very short 2.5-ms TDMA frames incorporated into PACS to provide low transmission delay (for high speech quality) also make the technology significantly less sensitive to high-speed fading than the longer-frame-period cellular technologies. The short frame also facilitates the rapid coordination needed to make reliable high-speed handoffs between base stations. Measurements on radio links to potential handoff base stations can be made rapidly, i.e., a measurement on at least one radio link every 2.5 ms. Once a handoff decision is made, signalling exchanges every 2.5 ms ensure that the radio link handoff is completed quickly. In contrast, the long frame periods in the high-tier (cellular) technologies prolong the time it takes to complete a handoff. As noted earlier, high speed handoff has been demonstrated many times with PACS technology and at speeds over 70 mi/hr. }

15.6 Quality, Capacity, and Economic Issues

Although the several trajectories toward low-tier PCS discussed in the preceding section are clear, it does not fit the existing wireless communications paradigms. Thus, low-tier PCS has attracted less attention than the systems and technologies that are compatible with the existing paradigms. Some examples are cited in the following paragraphs.

The need for intense interaction with an intelligent network infrastructure in order to manage mobility is not compatible with the cordless telephone appliance paradigm. In that paradigm, independence of network intelligence and base units that mimic wireline telephones are paramount.

Wireless data systems often do not admit to the dominance of wireless voice communications and, thus, do not take advantage of the economics of sharing network infrastructure and base station equipment. Also, wireless voice systems often do not recognize the importance of data and messaging and, thus, only add them in as bandaids to systems.

The need for a dense collection of many low-complexity, low-cost, low-tier PCS base stations interconnected with inexpensive fixed-network facilities (copper or fiber based) does not fit the cellular high-tier paradigm that expects sparsely distributed $1 million cell sites. Also, the need for high transmission quality to compete with wireline telephones is not compatible with the drive toward maximizing users-per-cell-site and per megahertz to minimize the number of expensive cell sites. These concerns, of course, ignore the hallmark of frequency-reusing cellular systems. That hallmark is the production of almost unlimited overall system capacity by reducing the separation between base stations. The cellular paradigm does not recognize the fact that almost all houses in the U.S. have inexpensive copper wires connecting telephones to the telephone network. The use of low-tier PCS base stations that concentrate individual user services before backhauling in the network will result in less fixed interconnecting facilities than exist now for wireline telephones. Thus, inexpensive techniques for interconnecting many low-tier base stations are already deployed to provide wireline telephones to almost all houses. { The cost of backhaul to many base stations (radio ports) in a low-tier system is often cited as an economic disadvantage that cannot be overcome. However, this perception is based on existing tariffs for T1 digital lines which are excessive considering current digital subscriber line technology. These tariffs were established many years ago when digital subscriber line electronics were very expensive. With modern low-cost high-rate digital subscriber line (HDSL) electronics, the cost of backhaul could be greatly reduced. If efforts were made to revise tariffs for digital line backhaul based on low cost electronics and copper loops like residential loops, the resulting backhaul costs would more nearly approach the cost of residential telephone lines. As it is now, backhaul costs are calculated based on antiquated high T1 line tariffs that were established for "antique" high cost electronics. }

This list could be extended, but the preceding examples are sufficient, along with the earlier sections of the paper, to indicate the many complex interactions among circuit quality, spectrum utilization, complexity (circuit and network), system capacity, and economics that are involved in the design compromises for a large, high-capacity wireless-access system. Unfortunately, the tendency has been to ignore many of the issues and focus on only one, e.g., the focus on cell site capacity that drove the development of digital-cellular high-tier systems in the United States. Interactions among circuit quality, complexity, capacity, and economics are considered in the following sections.

15.6.1 Capacity, Quality, and Complexity

Although capacity comparisons frequently are made without regard to circuit quality, complexity, or cost per base station, such comparisons are not meaningful. An example in Table 15.5 compares capacity factors for U.S. cellular or high-tier PCS technologies with the low-tier PCS technology, PACS/WACS. The mean opinion scores (MOS) (noted in Table 15.5) for speech coding are discussed later. Detection of speech activity and turning off the transmitter during times of no activity is implemented in IS-95. Its impact on MOS also is noted later. A similar technique has been proposed as E-TDMA for use with IS-54 and is discussed with respect to TDMA system in [29]. Note that the use of low-bit-rate speech coding combined with speech activity degrades the high-tier system's quality by nearly one full MOS point on the five-point MOS scale when compared to 32 kb/s ADPCM. Tandem encoding is discussed in a later section. These speech quality degrading factors alone provide a base station capacity increasing factor of $\times 4 \times 2.5 = \times 10$ over the high-speech-quality low-tier system! Speech coding, of course, directly affects base station capacity and, thus, overall system capacity by its effect on the number of speech channels that can fit into a given bandwidth.

TABLE 15.5 Comparison of Cellular (IS-54/IS-95) and Low-Tier PCS (WACS/PACS). Capacity Comparisons Made without Regard to Quality Factors, Complexity, and Cost per Base Station Are not Meaningful

Parameter	Cellular (High-Tier)	Low-Tier PCS	Capacity Factor
Speech Coding, kb/s	8 (MOS 3.4) No tandem coding	32 (MOS 4.1) 3 or 4 tandem	×4
Speech activity	Yes (MOS 3.2)	No (MOS 4.1)	×2.5
Percentage of good areas, %	90	99	×2
Propagation σ, dB	8	10	×1.5
Total: trading quality for capacity			×30

The allowance of extra system margin to provide coverage of 99% of an area for low-tier PCS versus 90% coverage for high-tier is discussed in the previous section and [29]. This additional quality factor costs a capacity factor of ×2. The last item in Table 15.5 does not change the actual system, but only changes the way that frequency reuse is calculated. The additional 2-dB margin in standard deviation σ, allowed for coverage into houses and small buildings for low-tier PCS, costs yet another factor of ×1.5 in calculation only. Frequency reuse factors affect the number of sets of frequencies required and, thus, the bandwidth available for use at each base station. Thus, these factors also affect the base station capacity and the overall system capacity.

For the example in Table 15.5, significant speech and coverage quality has been traded for a factor of ×30 in base station capacity! Whereas base station capacity affects overall system capacity directly, it should be remembered that overall system capacity can be increased arbitrarily by decreasing the spacing between base stations. Thus, if the PACS low-tier PCS technology were to start with a base station capacity of ×0.5 of AMPS cellular[5] (a

[5]Note that the ×0.5 factor is an arbitrary factor taken for illustrating this example. The so-called ×AMPS factors are only with regard to base station capacity, although they are often misused as system capacity.

much lower figure than the $\times 0.8$ sometimes quoted [20]), and then were degraded in quality as described above to yield the $\times 30$ capacity factor, it would have a resulting capacity of $\times 15$ of AMPS! Thus, it is obvious that making such a base station capacity comparison without including quality is not meaningful.

15.6.2 Economics, System Capacity, and Coverage Area Size

Claims are sometimes made that low-tier PCS cannot be provided economically, even though it is what the user wants. These claims are often made based on economic estimates from the cellular paradigm. These include the following.

- Very low estimates of market penetration, much less than cordless telephones, and often even less than cellular.
- High estimates of base station costs more appropriate to high-complexity, high-cost cellular technology than to low-complexity, low-cost, low-tier technology.
- Very low estimates of circuit usage time more appropriate to cellular than to cordless/wireline telephone usage, which is more likely for low-tier PCS.
- { Backhaul costs based on existing T1 line tariffs that are based on "antique" high cost digital loop electronics. (See discussion in fourth paragraph at start of Section 15.6.) }

Such economic estimates are often done by making absolute economic calculations based on very uncertain input data. The resulting estimates for low-tier and high-tier are often closer together than the large uncertainties in the input data. A perhaps more realistic approach for comparing such systems is to vary only one or two parameters while holding all others fixed and then looking at relative economics between high-tier and low-tier systems. This is the approach used in the following examples.

EXAMPLE 15.1:

In the first example (see Table 15.6), the number of channels per megahertz is held constant for cellular and for low-tier PCS. Only the spacing is varied between base stations, e.g., cell sites for cellular and radio ports for low-tier PCS, to account for the differences in transmitter power, antenna height, etc. In this example, overall system capacity varies directly as the square of base station spacing, but base station capacity is the same for both cellular and low-tier PCS. For the typical values in the example, the resulting low-tier system capacity is $\times 400$ greater, only because of the closer base station spacing. If the two systems were to cost the same, the equivalent low-tier PCS base stations would have to cost less than \$2,500.

This cost is well within the range of estimates for such base stations, including equivalent infrastructure. These low-tier PCS base stations are of comparable or lower complexity than cellular vehicular subscriber sets, and large-scale manufacture will be needed to produce the millions that will be required. Also, land, building, antenna tower and legal fees for zoning approval, or rental of expensive space on top of commercial buildings, represent large expenses for cellular cell sites. Low-tier PCS base stations that are mounted on utility poles and sides of buildings will not incur such large additional expenses. Therefore, costs of the order of magnitude indicated seem reasonable in large quantities. Note that, with these estimates, the per-wireless-circuit cost of the low-tier PCS circuits would be only \$14/circuit compared to \$5,555/circuit for the high-tier circuits. Even if there were a factor

TABLE 15.6 System Capacity/Coverage Area Size/Economics

Example 15.1

Assume channels/MHz are the same for cellular and PCS
Cell site: spacing = 20.000 ft cost $ = 1 M
PCS port: spacing = 1,000 ft
PCS system capacity is $(20000/1000)^2 = 400 \times$ cellular capacity

Then, for the system costs to be the same
Port cost = ($ 1 M/400) $2,500 a reasonable figure

If, cell site and port each have 180 channels
Cellular cost/circuit = $ 1 M/180 = $5,555/circuit
PCS cost/circuit = $2500/180 = $14/circuit

Example 15.2

Assume equal cellular and PCS system capacity
Cell site: spacing = 20,000 ft
PCS port: spacing = 1,000 ft

If, a cell site has 180 channels
then, for equal system capacity, a PCS port needs 180/400 < 1 channel/port

Example 15.3

Quality/cost trade
Cell site: Spacing = 20,000 ft cost = $1 M channels = 180
PCS port: Spacing = 1,000 ft cost = $2,500

Cellular to PCS, base station spacing capacity factor = × 400
PCS to cellular quality reduction factors:

32 to 8 kb/s speech	×4
Voice activity (buying)	×2
99–90% good areas	×2
Both in same environment (same σ)	×1
Capacity factor traded	×16

180 ch/16 = 11.25 channels/port then, $2500/11.25 = $222/circuit
and remaining is ×400/16 = ×25 system capacity of PCS over cellular

of 10 error in cost estimates, or a reduction of channels per radio port of a factor of 10, the per-circuit cost of low-tier PCS would still be only $140/circuit, which is still much less than the per-circuit cost of high-tier.

EXAMPLE 15.2:

In the second example (see Table 15.6), the overall system capacity is held constant, and the number of channels/port, i.e., channels/(base station) is varied. In this example, less than 1/2 channel/port is needed, again indicating the tremendous capacity that can be produced with close-spaced low-complexity base stations.

EXAMPLE 15.3:

Since the first two examples are somewhat extreme, the third example (see Table 15.6) uses a more moderate, intermediate approach. In this example, some of the cellular high-tier channels/(base station) are traded to yield higher quality low-tier PCS as in the previous subsection. This reduces the channels/port to 11+, with an accompanying increase in cost/circuit up to $222/circuit, which is still much less than the $5,555/circuit for the high-tier system. Note, also, that the low-tier system still has ×25 the capacity of the high-tier system!

Low-tier base station (Port) cost would have to exceed $62,500 for the low-tier per-circuit cost to exceed that of the high-tier cellular system. Such a high port cost far exceeds any existing realistic estimate of low-tier system costs.

It can be seen from these examples, and particularly Example 15.3, that the circuit economics of low-tier PCS are significantly better than for high-tier PCS, if the user demand and density is sufficient to make use of the large system capacity. Considering the high penetration of cordless telephones, the rapid growth of cellular handsets, and the enormous market projections for wireless PCS noted earlier in this chapter, filling such high capacity in the future would appear to be certain. The major problem is providing rapidly the widespread coverage (buildout) required by the FCC in the United States. If this unrealistic regulatory demand can be overcome, low-tier wireless PCS promises to provide the wireless personal communications that everyone wants.

15.6.3 {Loop Evolution and Economics

It is interesting to note that several wireless loop applications are aimed at reducing cost by replacing parts of wireline or CATV loops with wireless links between transceivers. The economics of these applications are driven by the replacing of labor-intensive wireline and cable technologies with mass-produced solid-state electronics in transceivers.

Consider first a cordless telephone base unit. The cordless base-unit transceiver usually serves one or, at most, two handsets at the end of one wireline loop. Now consider moving such a base unit back along the copper-wire-pair loop end a distance that can be reliably covered by a low-power wireless link [25, 31], i.e., several hundred to a thousand feet or so, and mounting it on a utility pole or a street light pole. This replaces the copper loop end with the wireless link. Many additional copper loop ends to other subscribers will be contained within a circle around the pole having a maximum usable radius of this wireless link. Replace all of the copper loop ends within the circle with cordless base units on the same pole. Note that this process replaces the most expensive parts of these many loops, i.e., the many individual loop ends, with the wireless links from cordless handsets to"equivalent" cordless base units on a pole. Of course, being mounted outside will require somewhat stronger enclosures and means of powering the base units, but these additional costs are considerably more than offset by eliminating the many copper wire drops.

It is instructive to consider how many subscribers could be collected at a pole containing base units. Consider, as an example, a coverage square of 1400 ft on a side (PACS will provide good coverage over this range, i.e., for base unit pole separations of about 1400 ft, at 1.9 GHz). Within this square will be 45 houses for a 1 house/acre density typical of low-housing-density areas, or 180 houses for 4 house/acre density more typical of high-density single-family housing areas. These represent significant concentration of traffic at a pole.

Because of the trunking advantage of the significant number of subscribers concentrated at a pole, they can share a smaller number of base unit, i.e., wireless base unit transceivers, than there are wireless subscriber sets. Therefore, the total cost compared with having a cordless base unit per subscriber also is reduced by the concentration of users.

A single PACS transceiver will support simultaneously eight TDMA channels or circuits at 32 kb/s (or 16 at 16 kb/s or 32 at 8 kb/s) [56]. Of these, one channel is reserved for system control. The cost of such moderate-rate transceivers is relatively insensitive to the number of channels supported; i.e., the cost of such an 8-channel (or 16 or 32) transceiver will be significantly less than twice the cost of a similar one-channel transceiver. Thus, another economic advantage accrues to this wireless loop approach from using time-multiplexed (TDMA) transceivers instead of single-channel-per-transceiver cordless telephone base units.

For an offered traffic of 0.06 Erlang, a typical busy-hour value for a wireline subscriber, a seven-channel transceiver could serve about 40 subscribers at 1% blocking, based on the Erlang B queuing discipline. From the earlier example, such a transceiver could serve most

of the 45 houses within a 1400-ft square. Considering partial penetration, the transceiver capacity is more than adequate for the low-density housing.[6]

Considering the high-density example of 4 houses/acre, a seven-channel transceiver could serve only about 20% of the subscribers within a 1400-ft square. If the penetration became greater than about 20%, either additional transceivers, perhaps those of other service providers, or closer transceiver spacing would be required.

Another advantageous economic factor for wireless loops results when considering time-multiplexed transmission in the fixed distribution facilities. For copper or fiber digital subscriber loop carrier (SLC), e.g., T1 or high-rate digital subscriber line (HDSL), a demultiplexing/multiplexing terminal and drop interface are required at the end of the time-multiplexed SLC line to provide the individual circuits for each subscriber loop-end circuit, i.e., for each drop. The most expensive part of such an SLC terminating unit is the subscriber line cards that provide per-line interfaces for each subscriber drop. Terminating a T1 or HDSL line on a wireless loop transceiver eliminates all per-line interfaces, i.e., all line cards, the most expensive part of a SLC line termination. Thus, the greatly simplified SLC termination can be incorporated within a TDMA wireless loop transceiver, resulting in another cost savings over the conventional copper-wire-pair telephone loop end.

The purpose of the previous discussions is not to give an exact system design or economic analysis, but to illustrate the inherent economic advantages of low-power wireless loops over copper loop ends and over copper loop ends with cordless telephone base units. Some economic analyses have found wireless loop ends to be more economical than copper loop ends when subscribers use low-power wireless handsets. Rizzo and Sollenberger [56] have also discussed the advantageous economics of PACS wireless loop technology in the context of low-tier PCS.

The discussions in this section can be briefly summarized as follows. Replacing copper wire telephone loop ends with low-complexity wireless loop technology like PACS can produce economic benefits in at least four ways. These are.

1. Replacing the most expensive part of a loop, the per-subscriber loop-end, with a wireless link.
2. Taking advantage of trunking in concentrating many wireless subscriber loops into a smaller number of wireless transceiver channels.
3. Reducing the cost of wireless transceivers by time multiplexing (TDMA) a few (7, 15, or 31) wireless loop circuits (channels) in each transceiver.
4. Eliminating per-line interface cards in digital subscriber line terminations by terminating time-multiplexed subscriber lines in the wireless loop transceivers. }

15.7 Other Issues

Several issues in addition to those addressed in the previous two sections continue to be raised with respect to low-tier PCS. These are treated in this section.

[6]The range could be extended by using higher base unit antennas, by using higher-gain directional (sectored) antennas, and/or by increasing the maximum power that can be transmitted.

15.7.1 Improvement of Batteries

Frequently, the suggestion is made that battery technology will improve so that high-power handsets will be able to provide the desired 5 or 6 hours of talk time in addition to 10 or 12 hours of standby time, and still weigh less than one-fourth of the weight of today's smallest cellular handset batteries. This hope does not take into account the maturity of battery technology, and the long history (many decades) of concerted attempts to improve it. Increases in battery capacity have come in small increments, a few percent, and very slowly over many years, and the shortfall is well over a factor of 10. In contrast, integrated electronics and radio frequency devices needed for low-power low-tier PCS continue to improve and to decrease in cost by factors of greater than 2 in time spans on the order of a year or so. It also should be noted that, as the energy density of a battery is increased, the energy release rate per volume must also increase in order to supply the same amount of power. If energy storage density and release rate are increased significantly, the difference between a battery and a bomb become indistinguishable! The likelihood of a $\times 10$ improvement in battery capacity appears to be essentially zero. If even a modest improvement in battery capacity were possible, many people would be driving electric vehicles.

{ As noted in the addition to the "Reality Check" section, new lithium batteries have become the batteries of choice for the smallest cellular/high-tier PCS handsets. While these lithium batteries have higher energy density than earlier nickel cadmium batteries, they still fall far short of the factor of 10 improvement that was needed to make long talk time, small size, and low weight possible. With the much larger advance in electronics, the battery is even more dominant in the size and weight of the newest cellular handsets. The introduction of these batteries incurred considerable startup pain because of the greater fire and explosive hazard associated with lithium materials, i.e., closer approach to a bomb. Further attempts in this direction will be even more hazardous. }

15.7.2 People Only Want One Handset

This issue is often raised in support of high-tier cellular handsets over low-tier handsets. Whereas the statement is likely true, the assumption that the handset must work with high-tier cellular is not. Such a statement follows from the current large usage of cellular handsets; but such usage results because that is the only form of widespread wireless service currently available, not because it is what people want. The statement assumes inadequate coverage of a region by low-tier PCS, and that low-tier handsets will not work in vehicles. The only way that high-tier handsets could serve the desires of people discussed earlier would be for an unlikely breakthrough in battery technology to occur. A low-tier system, however, can cover economically any large region having some people in it. (It will not cover rural or isolated areas but, by definition, there is essentially no one there to want communications anyway.)

Low-tier handsets will work in vehicles on village and city streets at speeds up to 30 or 40 mi/h, and the required handoffs make use of computer technology that is rapidly becoming inexpensive. { As noted earlier, vehicular speed handoff is readily accomplished with PACS. Reliable handoff has been demonstrated for PACS at speeds in excess of 70 mi/hr. } Highways between populated areas, and also streets within them, will need to be covered by high-tier cellular PCS, but users are likely to use vehicular sets in these cellular systems. Frequently the vehicular mobile user will want a different communications device anyway, e.g., a hands-free phone. The use of hands-free phones in vehicles is becoming a legal requirement in some places now and is likely to become a requirement in many more places in the future. Thus, handsets may not be legally usable in vehicles anyway.

With widespread deployment of low-tier PCS systems, the one handset of choice will be the low-power, low-tier PCS pocket handset or voice/data communicator.

{ As discussed in earlier sections, it is quite feasible economically to cover highways between cities with low-tier systems, if the low-tier base stations have antennas with the same height and gain as used for cellular and high-tier PCS systems. (The range penalty for the lower power was noted earlier to be only on the order of 1/2, or about the same as the range penalty in going from 800 MHz cellular to 1.9 GHz high-tier PCS.) }

There are approaches for integrating low-tier pocket phones or pocket communicators with high-tier vehicular cellular mobile telephones. The user's identity could be contained either in memory in the low-tier set or in a small smart card inserted into the set, as is a feature of the European GSM system. When entering an automobile, the small low-tier communicator or card could be inserted into a receptacle in a high-tier vehicular cellular set installed in the automobile.[7] The user's identity would then be transferred to the mobile set. { "Car adapters" that have a cradle for a small cellular handset providing battery charging and connection to an outside antenna are quite common — e.g., in Sweden use of such adapters is commonplace. Thus, this concept has already evolved significantly, even for the disadvantaged cellular handsets when they are used in vehicles. } The mobile set could then initiate a data exchange with the high-tier system, indicating that the user could now receive calls at that mobile set. This information about the user's location would then be exchanged between the network intelligence so that calls to the user could be correctly routed.[8] In this approach the radio sets are optimized for their specific environments, high-power, high-tier vehicular or low-power, low-tier pedestrian, as discussed earlier, and the network access and call routing is coordinated by the interworking of network intelligence. This approach does not compromise the design of either radio set or radio system. It places the burden on network intelligence technology that benefits from the large and rapid advances in computer technology.

The approach of using different communications devices for pedestrians than for vehicles is consistent with what has actually happened in other applications of technology in similarly different environments. For example, consider the case of audio cassette tape players. Pedestrians often carry and listen to small portable tape players with lightweight headsets (e.g., a Walkman).[9] When one of these people enters an automobile, he or she often removes the tape from the Walkman and inserts it into a tape player installed in the automobile. The automobile player has speakers that fill the car with sound. The Walkman is optimized for a pedestrian, whereas the vehicular-mounted player is optimized for an automobile. Both use the same tape, but they have separate tape heads, tape transports, audio preamps, etc. They do not attempt to share electronics. In this example, the tape cassette is the information-carrying entity similar to the user identification in the personal communications example discussed earlier. The main points are that the information is shared among different devices but that the devices are optimized for their environments and do not share electronics.

Similarly, a high-tier vehicular-cellular set does not need to share oscillators, synthesizers, signal processing, or even frequency bands or protocols with a low-tier pocket-size communicator. Only the information identifying the user and where he or she can be reached

[7] Inserting the small personal communicator in the vehicular set would also facilitate charging the personal communicator's battery.

[8] This is a feature proposed for FPLMTS in CCIR Rec. 687.

[9] Walkman is a registered trademark of Sony Corporation.

needs to be shared among the intelligence elements, e.g., routing logic, databases, and common channel signalling [26, 29] of the infrastructure networks. This information exchange between network intelligence functions can be standardized and coordinated among infrastructure subnetworks owned and operated by different business entities (e.g., vehicular cellular mobile radio networks and intelligent low-tier PCS networks). Such standardization and coordination are the same as are required today to pass intelligence among local exchange networks and interexchange carrier networks.

15.7.3 Other Environments

Low-tier personal communications can be provided to occupants of airplanes, trains, and buses by installing compatible low-tier radio access ports inside these vehicles. The ports can be connected to high-power, high-tier vehicular cellular mobile sets or to special air-ground or satellite-based mobile communications sets. Intelligence between the internal ports and mobile sets could interact with cellular mobile, air-ground, or satellite networks in one direction, using protocols and spectrum allocated for that purpose, and with low-tier personal communicators in the other direction to exchange user identification and route calls to and from users inside these large vehicles. Radio isolation between the low-power units inside the large metal vehicles and low-power systems outside the vehicles can be ensured by using windows that are opaque to the radio frequencies. Such an approach also has been considered for automobiles, i.e., a radio port for low-tier personal communications connected to a cellular mobile set in a vehicle so that the low-tier personal communicator can access a high-tier cellular network. (This could be done in the United States using unlicensed PCS frequencies within the vehicle.)

15.7.4 Speech Quality Issues

All of the PCS and cordless telephone technologies that use CCITT standardized 32-kb/s ADPCM speech encoding can provide similar error-free speech distortion quality. This quality often is rated on a five-point subjective mean opinion score (MOS) with 5 excellent, 4 good, 3 fair, 2 poor, and 1 very poor. The error-free MOS of 32-kb/s ADPCM is about 4.1 and degrades very slightly with tandem encodings. Tandem encodings could be expected in going from a digital-radio PCS access link, through a network using analog transmission or 64-kb/s PCM, and back to another digital-radio PCS access link on the other end of the circuit. In contrast, a low-bit-rate (<10-kb/s) vocoder proposed for a digital cellular system was recently reported [54] to yield an MOS of 3.4 on an error-free link without speech-activity detection. This score dropped to 3.2 when speech-activity detection was implemented to increase system capacity. This nearly one full point decrease on the five-point MOS score indicates significant degradation below accepted CCITT wireline speech distortion quality. Either almost half of the population must have rated it as poor or most of the population must have rated it as only fair. It should also be noted that these MOS scores may not reflect additional degradation that may occur in low-bit-rate speech encoding when the speech being encoded is combined with acoustical noise in a mobile environment, e.g., tire, wind, and engine noise in automobiles and street noise, background talking, etc., in handheld phone environments along streets and in buildings. Comments from actual users of low-bit-rate speech technology in acoustically noisy environments suggest that the MOS scores just quoted are significantly degraded in these real world environments. Waveform coders, e.g., ADPCM are not likely to experience degradation from such background noise. In addition, the low-bit-rate speech encoding is not at all tolerant of the tandem speech encodings that will inevitably occur for PCS for many years. That is, when low-bit-rate

speech is transcoded to a different encoding format, e.g., to 64 kb/s as is used in many networks or from an IS-54 phone on one end to a GSM or IS-95 phone on the other end, the speech quality deteriorates precipitously. Although this may not be a serious issue for a vehicular mobile user who has no choice other than not to communicate at all, it is likely to be a serious issue in an environment where a wireline telephone is available as an alternative. It is also less serious when there are few mobile-to-mobile calls through the network, but as wireless usage increases and digital mobile-to-mobile calls become commonplace, the marginal transcoded speech quality is likely to become a serious issue. These comments in this paragraph are generally applicable to speech encoding at rates of 13 kb/s or less. { The predictions above about the seriousness of the speech coding quality issues have been proven to be true in virtually all cellular technologies [2, 5]. The much touted speech quality of the original IS-95 CDMA 8 kb/s QCELP speech coder proved to be unacceptable to users.

The 8 kb/s speech coder was replaced with a 13 kb/s speech coder when IS-95 CDMA technology was deployed for high-tier PCS at 1.9 GHz. Many of the 800 MHz CDMA systems are converting or considering converting to the higher bit rate coder to improve their speech quality. This change was accompanied by a base station capacity reduction of 1/3, but the CDMA advocates seldom acknowledge this fact. The GSM effort to create a "half-rate" speech coder was redirected to develop an improved speech quality coder at 13 kb/s that would be less sensitive to acoustic background noise, bit errors, tandem encoding, etc. The quality of the new coder is better under some conditions. The 8 kb/s North American TDMA/IS-136 speech coder was redesigned to improve speech quality. Some improvement has been achieved. In all these revised coder cases, the issues discussed above in the original article still apply. }

In the arena of transmission delay, the short-frame (2-ms) FDMA/TDD and TDMA technologies (e.g., CT-2 and WACS noted earlier) can readily provide single-radio-link round-trip delays of <10 ms and, perhaps, even <5 ms. The longer frame (10 ms and greater) cordless-phone TDMA technologies, e.g., DCT-900/CT-3/DECT and some ISM-band implementations, inherently have a single-link round-trip delay of at least 20 ms and can range 30-40 ms or more in some implementations. As mentioned earlier, the digital vehicular-cellular technologies with low-bit-rate speech encoding, bit interleaving, forward error-correction decoding, and relatively long frame time (~16–20 ms) result in single-link round-trip delays on the order of 200 ms, well over an order of magnitude greater than the short-frame technologies, and on the same order of magnitude as a single-direction synchronous satellite link. { Even with the somewhat improved low bit rate speech coders noted previously, the transmission delay remains excessive for all of the cellular/high-tier PCS digital technologies. } It should be noted that almost all United States domestic long-distance telephone circuits have been removed from such satellite links, and many international satellite links also are being replaced by undersea fiber links. These long-distance-circuit technology changes are made partially to reduce the perceptual impairment of long transmission delay.

15.7.5 New Technology

New technology, e.g., spread spectrum or CDMA, is sometimes offered as a solution to both the higher-tier cell site capacity and transmitter power issues. As these new technologies are pursued vigorously, however, it becomes increasingly evident that the early projections were considerably overoptimistic, that the base station capacity will be about the same as other technologies [29], and that the high complexity will result in more, not less, power consumption.

With the continuing problems and delays in initial deployments, there is increasing concern throughout the industry as to whether CDMA is a viable technology for high-capacity cellular applications. With the passage of time, it is becoming more obvious that Viterbi was correct in his 1985 paper in which he questioned the use of spread spectrum for commercial communications [73].

The IS-95 proposal is considerably more technically sophisticated than earlier spread spectrum proposals. It includes fast feedback control of mobile transmitter power, heavy forward error correction, speech detection and speech-encoding rate adjustment to take advantage of speech inactivity, and multiple receiver correlators to latch onto and track resolvable multipath maxima [57]. The spreading sequence rate is 1.23 MHz.

The near–far problem is addressed directly and elegantly on the uplink by a combination of the fast-feedback power control and a technique called soft handoff that permits the instantaneous selection of the best paths between a mobile and two cell sites. Path selection is done on a frame-by-frame basis when paths between a mobile and the two cell sites are within a specified average level (perhaps 6–10 dB) for each other. This soft handoff provides a form of macroscopic diversity [10] between pairs of cell sites when it is advantageous. Increasing capacity by soft handoff requires precise time synchronization (on the order of a microsecond) among all cell sites in a system. An advantage of this proposal is that frequency coordination is not needed among cell sites since all sites can share a frequency channel. Coordination of the absolute time delays of spreading sequences among cell sites is required, however, since these sequence delays are used to distinguish different cell sites for initial access and for soft handoff. Also, handoff from one frequency to another is complicated.

Initially, the projected cell site capacity of this CDMA system, determined by mathematical analysis and computer simulation of simplified versions of the system, was $\times 20$ to $\times 40$ that of the analog AMPS, with a coverage criterion of 99% of the covered area [11]. However, some other early estimates [83] suggested that the factors were more likely to be $\times 6$ to $\times 8$ of AMPS.

A limited experiment was run in San Diego, California, during the fourth quarter of 1991 under the observation of cellular equipment vendors and service providers. This experiment had 42–62 mobile units in fewer than that many vehicles,[10] and four or five cell sites, one with three sectors. Well over half of the mobiles needed to provide the interference environment for system capacity tests were simulated by hardware noise simulation by a method not yet revealed for technical assessment. Estimates of the cell site capacity from this CDMA experiment center around $\times 10$ that of AMPS [12, 54][11] with coverage criteria $<99\%$, perhaps 90-95%, and with other capacity estimates ranging between $\times 8$ and $\times 15$.

This experiment did not exercise several potential capacity-reducing factors; the following, for example.

1. Only four cells participated in capacity tests. The test mobiles were all located in a relatively limited area and had limited choices of cell sites with which to communicate for soft handoffs. This excludes the effects of selecting a strong cell-site downlink for soft handoff that does not have the lowest uplink attenuation because of uncorrelated uplink and downlink multipath fading at slow vehicle speeds [3].

[10] Some vehicles contained more than one mobile unit.

[11] AT&T stated that the San Diego data supported a capacity improvement over analog cellular of at least $\times 10$.

2. The distribution of time-dispersed energy in hilly environments like San Diego usually is more concentrated around one or two delays than is the dispersed energy scattered about in heavily built-up urban areas like downtown Manhattan [13, 14], or Chicago. Energy concentrated at one or two delays is more fully captured by the limited number of receiver correlators than is energy more evenly dispersed in time.
3. Network delay in setting up soft handoff channels can result in stronger paths to other cell sites than to the one controlling uplink transmitter power. This effect can be more pronounced when coming out of shadows of tall buildings at intersections in heavily built-up areas. The effect will not occur as frequently in a system with four or five cell sites as it will in a large, many cell site system.

All of these effects and others [4] will increase the interference in a large system, similarly to the increase in interference that results from additional mobiles and, thus, will decrease cell site capacity significantly below that estimated in the San Diego trial. Factors like these have been shown to reduce the San Diego estimate of $\times 10$ to an expected CDMA capacity of $\times 5$ or $\times 6$ of analog AMPS [3, 4, 13]. This further reduction in going from a limited experiment to a large-scale system in a large metropolitan area is consistent with the reduction already experienced in going from simplified theoretical estimates to the limited experiment in a restricted environment. { There are many "small" effects that decrease the performance and, thus, the base station capacity of actual mobile CDMA systems compared to calculations of performance based on ideal assumptions — e.g., stationary channel and additive white Gaussian noise analysis. Some of these effects, in addition to those noted above, included inaccuracy in assigning correlators ("fingers") to time varying statistically nonstationary multiple "paths" at different delays when many "paths" exist, and lack of diversity because many environments have delay spreads less than the resolution of correlators. }

The San Diego trial also indicated a higher rate of soft(er) handoffs [54] between antenna sectors at a single cell site than expected for sectors well isolated by antenna patterns. This result suggests a lower realizable sectorization gain because of reflected energy than would be expected from more idealized antennas and locations. This could further reduce the estimated cell site capacity of a large-scale system.

Even considering the aforementioned factors, capacity increases of $\times 5$ or $\times 6$ are significant. These estimates, however, are consistent with the factor of $\times 3$ obtained from low-bit-rate (<10-kb/s) speech coding and the $\times 2$ to $\times 2.5$ obtained by taking advantage of speech pauses. These factors result in an expected increase of $\times 6$–7.5, with none of these speech-processing-related contributions being directly attributable to the spread-spectrum processing in CDMA. These results are consistent with the factor of $\times 6$–8 estimate made earlier [83] and are not far from the factor of $\times 8$ quoted recently [86].

Thus, it is clear that new high-complexity, high-tier technology will not be a substitute for low-complexity, low-power, low-tier PCS.

{ Since the writing for the first edition of this handbook, IS-95 based CDMA systems have been deployed and are providing service in several parts of the world, e.g., the U.S.A., Korea, and Hong Kong. The "high priests" and the "disciples" of CDMA continue to espouse their "belief" in the superiority of their CDMA technology. Good hard data on system performance is still scarce and of questionable reliability. However, there is sufficient evidence from some "usually reliable sources" and from "conversations in the halls" to confirm that, not only does CDMA not perform as well as was originally promised, but also in many places it is likely to be providing performance that is inferior to the other high-tier digital technologies. In spite of the continuing dismal failures of CDMA technology to

meet the early claims for it, a significant fraction of the wireless industry appears to be committed to pursuing variations of CDMA for future applications. At this time it is not obvious how many more redesign and reprogramming failures, or perhaps even failures of manufacturing companies and service providers will be needed to overcome the religiously fanatic attachment of a part of the industry to CDMA. It appears that some are heavily committed again to "sweep on to the grand fallacy." Viterbi may even have been over optimistic about CDMA in 1985 [73], but he was correct in the way it attracts advocates.

IS-95 CDMA has failed to satisfy any of the major claims made for it in the late 1980s and early 1990s. The listing in Table 15.7 and the following discussion cites specific claims that have NOT been met now that reality has set in.

TABLE 15.7 CDMA Scoresheet

Early Claim	Reality
Easy to install and expand (no frequency planning)	Many "parameters" to adjust; adjustment is very difficult; many software revisions required
"Capacity" of AMPS ×20 (with high expectation of AMPS ×40)	AMPS ×3 or 4; AMPS × <3? (at unspecified but likely high call dropping rate) (continuously revised downward over time)
No more dropped calls	More dropped calls than other technologies (some reports say up to 40% dropped calls when system is loaded!)
No problem with interference (CDMA "loves" interference)	Can't live with interference from AMPS or other sources
Greater range and coverage than other technologies	System coverage "breathes" as loading increases; range often is less than other technologies
"Toll quality" 8 kb/s QCELP speech	Original 8 kb/s QCELP speech quality worse than VSELP in TDMA; required change to 13 kb/s coder (change reduced "capacity")

Much of the original CDMA hype was focused on undefined cell site capacity (see text from original Handbook, particularly this section and the section on "Capacity, Quality, and Complexity.") Information from many sources (e.g., [9], Benjamin and Lusignan, 1997 — Benjamin is with Lucent Technologies; Lusignan with Stanford University; [40, 49, 69], etc.) suggest that with 13 kb/s speech coding, "fully loaded" systems may provide base station capacities on the order of AMPS ×3 or perhaps ×4. However, the meaning of fully loaded is not well defined. If it is taken to be at a very high call dropping rate (sometimes said to be as high as 40%!), perhaps even these low factors are again unrealistically inflated. A usable system loading must provide an acceptable call dropping rate, perhaps 1% or less. If loading is reduced to provide sufficient operating margin to prevent too many dropped calls, perhaps the usable base station capacity may be even significantly less than AMPS ×3. (Note: Publicly presented CDMA claims have been consistently inflated in the past. Why should we be any more inclined to believe them now?) The capacity situation is so bad that the CDMA crowd doesn't even want to discuss this topic anymore. They now attempt to shift attention to the fact that there are operating CDMA systems, and try to brush aside the fact CDMA performance is so very poor. Low capacity has a significant negative impact on the economics of systems; high call dropping rates have a significant negative impact on service quality.

Another big early claim for CDMA was that it was easy to install and expand, that is, that it needed little engineering and planning for installation because it didn't require frequency reuse and power coordination among base stations. In reality, many parameters

require adjustment and critical thresholds require setting for a system to provide even the low capacity numbers cited above. Many software revisions have been made in attempts to cope with fundamental problems that keep appearing and reappearing in new deployments and expansions. Deployments and service starts have been consistently late by many months and even years. [40, 69, and others]. A partial list of late CDMA deployments that required many months or even years of "parameter adjustment" and/or "software revisions" and/or "redesign" include those in Seattle, WA, U.S.A. (800 MHz); Los Angeles, CA, U.S.A. (800 MHz); Hong Kong, China (800 MHz); Trenton, NJ, U.S.A.; Seoul, Korea (800 MHz); Houston and Dallas, TX, U.S.A. (1.9 GHz); etc. As these systems were deployed a consistent pattern developed. Announcements would concentrate on a projected future service date, e.g., Seattle, 1Q 1994. When that date approached, news focused on yet another service date elsewhere, e.g., after Seattle was Los Angeles, to divert attention from the deployment problems causing large slips in earlier announcements. Sometimes when service date slips became excessive, sizes of deployed areas were revised downward, and "trials" were substituted for service starts, e.g., Los Angeles. This pattern continued through the early deployments and service eventually was started on systems, even though their performance was significantly below earlier claims. Excuses for slips were often vague, for example: software problems (e.g., Seattle, Hong Kong, and Korea); adjusting parameters (e.g., Seattle, Houston, and Dallas); too much interference from AMPS (e.g., Los Angeles); not enough multipath! (e.g., Seattle). These sound like descriptions of attempts to overcome some of the fundamental problems in mobile CDMA systems. In any event it is obvious that CDMA is not easy to install and is likely more difficult to adjust than other technologies.

The CDMA speech quality issue was discussed earlier in the "Speech Quality Issues" section. It is just another example of unrealistic initial claims.

Throughout the CDMA debate over almost 10 years, when unrealistic claims were questioned and even when sound technical analysis was cited by more conservative members of the wireless technical community, including this author, these members were immediately branded as heretics and their concerns were dismissed with rhetoric by the high priests and their disciples. (In the technical world, "never have so many been misled for so long by so few!") However, as deployments have proceeded and reality set in, even the more conservative initial assessments have, unfortunately, turned out to be somewhat optimistic. Thus, even the conservative heretics in the community now appear to be in the unfortunate position of having also been overoptimistic in their initial estimates of the performance of IS-95 CDMA. }

Statistical Multiplexing, Speech Activity, CDMA, and TDMA

Factors of $\times 2$–2.5 have been projected for the capacity increase possible by taking advantage of pauses in speech. It has been suggested that implementing statistical multiplexing is easier for CDMA systems because it is sometimes thought to be time consuming to negotiate channels for speech spurts for implementation in TDMA systems. The most negative quality-impacting factor in implementing statistical multiplexing for speech, however, is not in obtaining a channel when needed but is in the detection of the onset of speech, particularly in an acoustically noisy environment. The effect of clipping at the onset of speech is evident in the MOS scores noted for the speech-activity implementation in the United States cellular CDMA proposal discussed earlier (i.e., an MOS of 3.4 without statistical multiplexing and of 3.2 with it). The degradation in MOS can be expected to be even greater for encoding that starts with a higher MOS, e.g., 32-kb/s ADPCM.

It was noted earlier that the proposed cellular CDMA implementation was ×10 as complex as the proposed WACS wireless access for personal communications TDMA implementation. From earlier discussion, the CDMA round-trip delay approaches 200 ms, whereas the short 2-ms-frame TDMA delay is <10-ms round trip. It should be noted that the TDMA architecture could permit negotiation for time slots when speech activity is detected. Since the TDMA frames already have capability for exchange of signalling data, added complexity for statistical multiplexing of voice could readily be added within less than 200 ms of delay and less than ×10 in complexity. That TDMA implementation supports 8 circuits at 32 kb/s or 16 circuits at 16 kb/s for each frequency. These are enough circuits to gain benefit from statistical multiplexing. Even more gain could be obtained at radio ports that support two or three frequencies and, thus, have 16–48 circuits over which to multiplex.

A statistical multiplexing protocol for speech and data has been researched at Rutgers WINLAB [39]. The Rutgers packet reservation multiple access (PRMA) protocol has been used to demonstrate the feasibility of increasing capacity on TDMA radio links. These PRMA TDMA radio links are equivalent to slotted ALOHA packet-data networks. Transmission delays of less than 50 ms are realizable. The capacity increase achievable depends on the acceptable packet-dropping ratio. This increase is soft in that a small increase in users causes a small increase in packet-dropping ratio. This is analogous to the soft capacity claimed for CDMA.

Thus, for similar complexity and speech quality, there appears to be no inherent advantage of either CDMA or TDMA for the incorporation of statistical multiplexing. It is not included in the personal communications proposal but is included in cellular proposals because of the different speech-quality/complexity design compromises discussed throughout this chapter, not because of any inherent ease of incorporating it in any particular access technology.

15.7.6 High-Tier to Low-Tier or Low-Tier to High-Tier Dual Mode

Industry and the FCC in the United States appear willing to embrace multimode handsets for operating in very different high-tier cellular systems, e.g., analog FM AMPS, TDMA IS-54, and CDMA IS-95. Such sets incur significant penalties for dual mode operation with dissimilar air interface standards and, of course, incur the high-tier complexity penalties.

It has been suggested that multimode high-tier and low-tier handsets could be built around one air-interface standard, for example, TDMA IS-54 or GSM. When closely spaced low-power base stations were available, the handset could turn off unneeded power-consuming circuitry, e.g., the multipath equalizer. The problem with this approach is that the handset is still encumbered with power-consuming and quality-reducing signal processing inherent in the high-tier technology, e.g., error correction decoding and low-bit-rate speech encoding and decoding.

{ With widespread and successful deployment of GSM, perhaps an initial dual-mode strategy of GSM and PACS could be desirable. Of all the high-tier and low-tier technologies, GSM and PACS appear most compatible for dual-mode handset and base station implementation. They are considerably more compatible than the dual-mode combinations, e.g., AMPS/IS-54/IS-136 and AMPS/IS-95, already implemented. }

An alternative dual-mode low-tier, high-tier system based on a common air-interface standard can be configured around the low-tier PACS/WACS system, if such a dual-mode system is deemed desirable in spite of the discussion in this chapter. The range of PACS can readily be extended by increasing transmitter power and/or the height and gain of base station antennas. With increased range, the multipath delay spread will be more severe in some locations [21, 22, 33]. Two different solutions to the increased delay spread can be employed, one for the downlink and another for the uplink. The PACS radio-link

architecture has a specified bit sequence, i.e., a unique word, between each data word on the TDM downlink [7, 28]. This unique word can be used as a training sequence for setting the tap weights of a conventional equalizer added to subscriber sets for use in a high-tier PACS mode. Since received data can be stored digitally [62, 65], tap weights can be trimmed, if necessary, by additional passes through an adaptive equalizer algorithm, e.g., a decision feedback equalizer algorithm.

The PACS TDMA uplink has no unique word. The high-tier uplink, however, will terminate on a base station that can support greater complexity but still be no more complex than the high-tier cellular technologies. Research at Stanford University has indicated that blind equalization, using constant-modulus algorithms (CMA), [58, 70], can be effective for equalizing the PACS uplink. Techniques have been developed for converging the CMA equalizer on the short TDMA data burst.

{ See earlier added section on "Coverage, Range, Speed and Environments." }

The advantages of building a dual-mode high-tier, low-tier PCS system around the low-tier PACS air-interface standard follow.

1. The interface can still support small low-complexity, low-power, high-speech-quality low-tier handsets.
2. Both data and voice can be supported in a PACS personal communicator.
3. In high-tier low-tier dual mode PACS sets, circuits used for low-tier operation will also be used for high-tier operation, with additional circuits being activated only for high-tier operation.
4. The flexibility built into the PACS radio link to handle different data rates from 8 kb/s to several hundred kb/s will be available to both modes of operation.

15.8 Infrastructure Networks

It is beyond the scope of this chapter to consider the details of PCS network infrastructures. There are, however, perhaps as many network issues as there are wireless access issues discussed herein [26, 32, 43, 81]. With the possible exception of the self-organizing WLANS, wireless PCS technologies serve as access technologies to large integrated intelligent fixed communications infrastructure networks.

These infrastructure networks must incorporate intelligence, i.e., database storage, signalling, processing and protocols, to handle both small-scale mobility, i.e., handoff from base station to base station as users move, and large-scale mobility, i.e., providing service to users who roam over large distances, and perhaps from one network to another. The fixed infrastructure networks also must provide the interconnection among base stations and other network entities, e.g., switches, databases, and control processors. Of course, existing cellular mobile networks now contain or are incorporating these infrastructure network capabilities. Existing cellular networks, however, are small compared to the expected size of future high-tier and low-tier PCS networks, e.g., 20 million cellular users in the United States compared with perhaps 100 million users or more each in the future for high-tier and low-tier PCS.

Several other existing networks have some of the capabilities needed to serve as access networks for PCS. Existing networks that could provide fixed base station interconnection include:

- Local exchange networks that could provide interconnection using copper or glass-fiber distribution facilities

- Cable TV networks that could provide interconnection using new glass-fiber and coaxial-cable distribution facilities
- Metropolitan fiber digital networks that could provide interconnection in some cities in which they are being deployed

Networks that contain intelligence, e.g., databases, control processors, and signalling that is suitable or could be readily adapted to support PCS access include:

- Local exchange networks that are equipped with signalling system 7 common channel signalling (SS7 CCS), databases, and digital control processors
- Interexchange networks that are similarly equipped

Data networks, e.g., the internet, could perhaps be adapted to provide the needed intelligence for wireless data access, but they do not have the capacity needed to support large voice/data wireless low-tier PCS access.

Many entities and standards bodies worldwide are working on the access network aspects of wireless PCS. The signalling, control processing, and database interactions required for wireless access PCS are considerably greater than those required for fixed place-to-place networks, but that fact must be accepted when considering such networks.

Low-tier PCS, when viewed from a cellular high-tier paradigm, requires much greater fixed interconnection for the much closer spaced base stations. When viewed from a cordless telephone paradigm of a base unit for every handset and, perhaps, several base units per wireline, however, the requirement is much less fixed interconnection because of the concentration of users and trunking that occurs at the multiuser base stations. One should remember that there are economical fixed wireline connections to almost all houses and business offices in the United States now. If wireless access displaces some of the wireline connections, as expected, the overall need for fixed interconnection could decrease!

15.9 Conclusion

Wireless personal communications embraces about seven relatively distinct groups of tetherless voice and data applications or services having different degrees of mobility for operation in different environments. Many different technologies and systems are evolving to provide the different perceived needs of different groups. Different design compromises are evident in the different technologies and systems. The evidence suggests that the evolutionary trajectories are aimed toward at least three large groups of applications or services, namely, high-tier PCS (current cellular radio), high-speed wireless local-area networks (WLANS), and low-tier PCS (an evolution from several of the current groups). It is not clear to what extent several groups, e.g., cordless telephones, paging, and wide-area data, will remain after some merging with the three large groups. Major considerations that separate current cellular technologies from evolving low-tier low-power PCS technologies are speech quality, complexity, flexibility of radio-link architecture, economics for serving high-user-density or low-user-density areas, and power consumption in pocket carried handsets or communicators. High-tier technologies make use of large complex expensive cell sites and have attempted to increase capacity and reduce circuit costs by increasing the capacity of the expensive cell sites. Low-tier technologies increase capacity by reducing the spacing between base stations, and achieve low circuit cost by using low-complexity, low-cost base stations. The differences between these approaches result in significantly different compromises in circuit quality and power consumption in pocket-sized handsets or communicators. These

kinds of differences also can be seen in evolving wireless systems optimized for data. Advantages of the low-tier PACS/WACS technology are reviewed in the chapter, along with techniques for using that technology in high-tier PCS systems.

References

[1] Afrashteh, A., Sollenberg, N.R., and Chukurov, D.D., Signal to interference performance for a TDMA portable radio link. *IEEE VTC'91*, St. Louis, MO, 19–22, May, 1991.

[2] {Andrews, E.L., "When Digital Doesn't Always Mean Clear," *New York Times*, p. C1, 4, Jun. 26, 1995.}

[3] Ariyavisitakul, S., SIR-based power control in a CDMA system. *IEEE GLOBECOM'92*, Orlando, FL, Paper 26.3, Dec. 6–9, to be published in *IEEE Trans. Comm.*, 1992.

[4] Ariyavisitakul, S., et al., private communications.

[5] {Baugh, C.R., Wireless quality: a ticking time bomb, *Telephony*, 232(9), 50–58, Mar. 3, 1997.}

[6] {Baugh, C.R., Laborde, E., Pandey, V., and Varma, V., Personal access communications system: fixed wireless local loop and mobile configurations and services, *Proc. IEEE*, 1498–1506, July, 1998. }

[7] Bellcore Technical Advisories. Generic framework criteria for universal digital personal communications systems (PCS). FA-TSY-001013(1) March and FA-NWT-001013(2) Dec., and Tech. Ref. 1993. Generic criteria for version 0.1 wireless access communications systems (WACS). (1) Oct. rev. 1, Jun. 1994.

[8] BellSouth Services Inc., Quarterly progress report number 3 for experimental licenses KF2XFO and KF2XFN. To the Federal Communications Commission, Nov. 25, 1991.

[9] { Benjamin, B. and Lusignan, B., CDMA vs. TDMA, Stanford Univ. Center for Telecom. Symposium, The Acceleration of World Wide Wireless Communication, May 21, 1997.}

[10] Bernhardt, R.C., Macroscopic diversity in frequency reuse radio systems. *IEEE J. Sel. Areas in Comm.*, SAC-5, 862–870, Jun. 1987.

[11] Cellular Telecommunications Industries Association. CDMA digital cellular technology open forum. Jun. 6, 1989.

[12] Cellular Telecommunications Industries Association. CTIA Meeting, Washington, D.C., 5–6, Dec. 1991.

[13] Chang, L.F., Dispersive fading effects in CDMA radio systems. *Elect. Lett.*, 28(19), 1801–1802, 1992.

[14] Chang, L.F. and Ariyavisitakul, S., Performance of a CDMA radio communications system with feed-back power control and multipath dispersion. *IEEE GLOBECOM'91*, Phoenix, AZ, 1017–1021, Dec. 1991.

[15] Chuang, J.C.-I., Performance issues and algorithms for dynamic channel assignment. *IEEE JSAC*, Aug. 1993.

[16] Chuang, J.C.-I., Performance limitations of TDD wireless personal communications with asynchronous radio ports. *Electron. Lett.*, 28, 532–533, Mar. 1992.

[17] Chuang, J.C.-I., Comparison of coherent and differential detection of BPSK and QPSK in a quasistatic fading channel. *IEEE Trans. Comm.*, 565–576, May 1990.

[18] Chuang, J.C.-I. and Sollenberger, N.R., Burst coherent detection with robust frequency and timing estimation for portable radio communications. *IEEE GLOBECOM'88*, Hollywood, FL, 28–30, Nov. 1988.

[19] Chuang, J.C.-I., Sollenberger, N.R., and Cox, D.C., A pilot based dynamic channel assignment scheme for wireless access TDMA/FDMA systems. *Proc. IEEE ICUPC'93*, Ottawa, Canada, Oct. 12–15, 706–712, 1994. *Intl. J. Wireless Inf. Net.*, 1(1), 37–48, 1993.

[20] Cook, C.I., Development of air interface standards for PCS. *IEEE Personal Comm.*, 4th Quarter, 30–34, 1994.

[21] Cox, D.C., Delay-Doppler characteristics of multipath propagation at 910 MHz in a suburban mobile radio environment. *IEEE Trans. on Antennas and Propagation,* 625–635, Sep. 1972.

[22] Cox, D.C., Multipath delay spread and path loss correlation for 910 MHz urban mobile radio propagation. *IEEE Trans. on Veh. Tech.,* 340–344, Nov. 1977.

[23] Cox, D.C., Universal portable radio communications. *IEEE Trans. on Veh. Tech.,* 117–121, Aug. 1985.

[24] Cox, D.C., Research toward a wireless digital loop. *Bellcore Exchange,* 2, 2–7, Nov./Dec. 1986.

[25] Cox, D.C., Universal digital portable radio communications. *Proc. IEEE,* 75, 436–477, Apr. 1987.

[26] Cox, D.C., Portable digital radio communications—An approach to tetherless access. *IEEE Comm. Mag.,* 30–40, Jul. 1989.

[27] Cox, D.C., Personal communications—A viewpoint. *IEEE Comm. Mag.,* 8–20, Nov. 1990.

[28] Cox, D.C., A radio system proposal for widespread low-power tetherless communications. *IEEE Trans. on Comm.,* 324–335, Feb. 1991.

[29] Cox, D.C., Wireless network access for personal communications. *IEEE Comm. Mag.,* 96–115, Dec. 1992.

[30] {Cox, D.C., Wireless loops: what are they. *Intl. J. Wireless Inf. Net.,* 3(3), Plenum Press, 1996.}

[31] Cox, D.C., Arnold, H.W., and Porter, P.T., Universal digital portable communications—a system perspective. *IEEE JSAC,* JSAC-5, 764–773, Jun. 1987.

[32] Cox, D.C., Gifford, W.G., and Sherry, H., Low-power digital radio as a ubiquitous subscriber loop. *IEEE Comm. Mag.,* 92–95, Mar. 1991.

[33] {Devasirvatham, D.M.J., Radio propagation studies in a small city for universal portable communications. *IEEE VTC'88,* Conference Record. Philadelphia, PA, 100–104, Jun. 15–17, 1988.}

[34] {Devasirvatham, D.M.J., Radio propagation studies in a small city for universal portable communications, *IEEE VTC'88,* Philadelphia, PA, 100–104, Jun. 15–17, 1988.}

[35] Devasirvatham, D.M.J., et al., Radio propagation measurements at 850 MHz, 1.7 GHz and 4 GHz inside two dissimilar office buildings. *Elect. Lett.,* 26(7), 445–447, 1990a.

[36] Devasirvatham, D.M.J., et al., Multi-frequency radiowave propagation measurements in the portable radio environment. *IEEE ICC'90,* 1334–1340, Apr. 1990b.

[37] {Devasirvatham, D.M.J., Murray, R.R., Arnold, H.W., and Cox, D.C., Four-frequency CW measurements in residential environments for personal communications, symposium on personal indoor and mobile radio communications (PIMRC'93), Yokohama, Japan, 201–205, Sep. 9–11, 1993.}

[38] {Goldhirsh, J. and Vogel, W., Mobile satellite system fade and statistics for shadowing and multipath from roadside trees at UFH and L-band, *IEEE Transactions on Antennas and Propagation,* 37, Apr. 1989.}

[39] Goodman, D.J., Trends in cellular and cordless communications. *IEEE Comm. Mag.,* 31–40, Jun. 1991.

[40] {Hardy, Q., Jacob's Patter, The Wall Street Journal, Friday, Sep. 6, 1996.}

[41] {Hattori, T., PHS system deployment success story, Stanford Univ. Center for Telecom. Symposium, May 21, 1997.}

[42] Hinderling, J., et al., CDMA mobile station modem ASIC. *IEEE CICC 92,* Boston, MA, May. *Intl. J. Wireless Info. Net.,* 1(1), 37–48, Jan. 1994.

[43] Jabbari, B., et al., Network issues for wireless personal communications. *IEEE Comm. Mag.,* 88–98, Jan. 1995.

[44] Katz, R.H., Adaptation and mobility in wireless information systems. *IEEE Personal Comm.*, 1st Quarter, 6–17, 1994.

[45] {Kim, B.-J. and Cox, D.C., Blind equalization for short burst communications over frequency selective wireless channels, 1997 *IEEE VTC'97,* Phoenix, AZ, 2, 544–548, May 4–7, 1997a.}

[46] {Kim, B.-J. and Cox, D.C., Blind sequence estimation for short burst communications over time-varying wireless channels, *IEEE ICUPC'97,* San Diego, CA, 2, 713–717, Oct. 12–16, 1997b.}

[47] {Kim, B.-J. and Cox, D.C., Blind diversity combining equalization for short burst wireless communications, *IEEE Globecom'97,* Phoenix, AZ, 3, 1163–1167, Nov. 3–8, 1997c.}

[48] {Kim, B.-J., Blind equalization for short burst wireless communications, Stanford University Ph.D. Thesis, Jan. 1998.}

[49] {Korea Mobile Telecom, Article on CDMA, Mobile Communications International, Feb. 1997.}

[50] Padgett, J.E., Hattori, T., and Gunther, C., Overview of wireless personal communications. *IEEE Comm. Mag.*, 28–41, Jan. 1995.

[51] {Patel, V., Personal Access Communications System, Stanford Univ. Center for Telecom. Symposium, May 21, 1997.}

[52] {Phillips, B.W., Broadband in the Local Loop, Telecommunications, 37–42, Nov. 1994.}

[53] {PHS, World Congress, Singapore, Nov. 18–20, 1996.}

[54] Qualcomm Technology Forum, Open meeting on status of CDMA technology and review of San Diego experiment. San Diego, CA, Jan. 16–17, 1992.

[55] {Ranade, A., Local access radio interference due to building reflections, *IEEE Transaction on Communications,* 70–74, Jan. 1989.}

[56] {Rizzo, J.F. and Sollenberger, N.R., Multiuser Wireless Access, *IEEE Personal Communications Magazine,* 18–30, Jun. 1995.}

[57] Salmasi, A. and Gilhousen, K.S., On the system design aspects of code division multiple access (CDMA) applied to digital cellular and personal communications networks. *IEEE VTC'91,* St. Louis, MO, 57–62, May 1991.

[58] Sato, Y., A method of self-recovering equalization for multilevel amplitude modulation systems. *IEEE Trans. on Comm.*, 679–682, Jun. 1975.

[59] Schneideman, R., Spread spectrum gains wireless applications. *Microwaves and RF,* 31–42, May 1992.

[60] {Seidel, S.Y. and Arnold, H.W., Propagation measurements of 28 GHz to investigate the performance of local multipoint distribution services (LMDS), *IEEE Globecom'95,* 1995.}

[61] Sollenberger, N.R. and Chuang, J.C-I., Low overhead symbol timing and carrier recovery for TDMA portable radio systems. Third nordic sem. digital land mobile radio comm., Copenhagen, Denmark, 13–15, Sep. 1988.

[62] Sollenberger, N.R., An experimental VLSI implementation of low-overhead symbol timing and frequency offset estimation for TDMA portable radio applications. *IEEE Globecom'90,* San Diego, CA, 1701–1711, Dec. 1990.

[63] Sollenberger, N.R., An experimental TDMA modulation/demodulation CMOS VLSI chip-set. *IEEE CICC'91,* San Diego, CA, May 12–15, 1991.

[64] Sollenberger, N.R. and Afrashteh, A., An experimental low-delay TDMA portable radio link. *Wireless'91,* Calgary, Canada, Jul. 8–10, 1991.

[65] Sollenberger, N.R. and Chuang, J.C.-I., Low-overhead symbol timing and carrier recovery for TDMA portable radio systems. *IEEE Trans. on Comm.*, 1886–1892, Oct. 1990.

[66] Sollenberger, N.R., et al., Architecture and implementation of an efficient and robust TDMA frame structure for digital portable communications. *IEEE Trans. Veh. Tech.*, 40, 250–260, Feb. 1991.

[67] Steele, R., Deploying personal communications networks. *IEEE Comm. Mag.*, 12–15, Sep. 1990.

[68] Steer, D.G., Coexistence and access etiquette in the United States unlicensed PCS band. *IEEE Personal Comm. Mag.*, 4th Quarter, 36–43, 1994.

[69] {Titch, S., Blind Faith, Telephony, Sep. 8, 1997 and Global Telephony, Intertec Publishing Corp., Oct. 30, 1997.}

[70] Treichler, J.R. and Agee, B.G., A new approach to multipath correction of constant modulus signals, *IEEE Trans. on Acoustics, Speech and Signal Processing*, 459–472, Apr. 1983.

[71] Tuthill, J.P., Granger, B.S., and Wurtz, J.L., Request for a pioneer's preference. Before the Federal Communications Commission, Pacific Bell submission for FCC General Docket No. 90–314, RM-7140, and RM-7175, May 4, 1992.

[72] Varma, V.K., et al., Performance of sub-band and RPE coders in the portable communication environment. *Fourth Intl. Conf. on Land Mobile Radio*, Coventry, UK, 221–227, Dec. 14–17, 1987.

[73] Viterbi, A.J., When not to spread spectrum—A sequel. *IEEE Comm. Mag.*, 12–17, Apr. 1985.

[74] {Weseloh, C., Solutions for Wireless Communications, Stanford University Center for Telecommunications, Symposium, Stanford, CA, May 16, 1995.}

[75] Wong, A., Regulating public wireless networks. Workshop on lightwave, wireless and networking technologies, Chinese Univ. of Hong Kong, Hong Kong, Aug. 24, 1994.

[76] {Wong, P., Low-power, low-complexity diversity combining algorithm and VLSI circuit chip for hand-held PCS receivers, Stanford University Ph.D. Thesis, Apr. 1997.}

[77] {Wong, P. and Cox, D.C., Low-complexity cochannel interference cancellation and macroscopic diversity for high capacity PCS, *IEEE ICC'95 Digest*, Seattle, WA, 852–857, Jun. 19–22, 1995.}

[78] {Wong, P. and Cox, D.C., Low-complexity diversity combining algorithms and circuit architectures for cochannel interference cancellation and frequency selective fading mitigation, *IEEE ICC'96 Digest*, Dallas TX, Jun. 23–27, 1996a.}

[79] {Wong, P. and Cox, D.C., Low-complexity diversity combining algorithms and circuit architectures for cochannel interference cancellation and frequency-selective fading mitigation, *IEEE Trans. on Comm.*, 44(9), 1107–1116, Sep. 1996b.}

[80] {Wong, P. and Cox, D.C., Low-complexity interference cancellation and macroscopic diversity for high capacity PCS, *IEEE Trans. on Veh. Tech.*, 124–132, Feb. 1998.}

[81] Zaid, M., Personal mobility in PCS. *IEEE Personal Comm.*, 4th Quarter, 12–16, 1994.

[82] Special issue on advanced mobile phone service (AMPS). *Bell System Tech. J.*, (BSTS). 58, Jan. 1979.

[83] CDMA capacity seen as less than advertised. *Adv. Wireless Comm.*, (AWC). 1–2, Feb. 6, 1991.

[84] Bellcore's Wireless Prototype. *Microcell News*, (MN). 5–6, Jan. 25, 1992.

[85] Bellcore PCS phone shines in Bellsouth test; others fall short. *Adv. Wireless Comm.*, (AWC). 2–3, Feb. 19, 1992a.

[86] TDMA accelerates, but CDMA could still be second standard. *Adv. Wireless Comm.*, (AWC). 6, Mar. 4, 1992b.

[87] Special issue on wireless personal communications. *IEEE Comm. Mag.*, Jan. 1995.

16

Modulation Methods

16.1 Introduction 16-1
16.2 Basic Description of Modulated Signals 16-2
16.3 Analog Frequency Modulation 16-2
16.4 Phase Shift Keying (PSK) and $\pi/4$-QPSK 16-3
16.5 Continuous Phase Modulation (CPM) and MSK 16-7
16.6 Gaussian Minimum Shift Keying 16-9
16.7 Orthogonal Frequency Division Multiplexing (OFDM) 16-11
16.8 Conclusions 16-13
Defining Terms 16-13
References 16-14
Further Information 16-14

Gordon L. Stüber
Georgia Institute of Technology

16.1 Introduction

Modulation is the process where the message information is added to the radio carrier. Most first generation cellular systems such as the advanced mobile telephone system (AMPS) use analog **frequency modulation** (**FM**), because analog technology was very mature when these systems were first introduced. Digital modulation schemes, however, are the obvious choice for future wireless systems, especially if data services such as wireless multimedia are to be supported. Digital modulation can also improve spectral efficiency, because digital signals are more robust against channel impairments. Spectral efficiency is a key attribute of wireless systems that must operate in a crowded radio frequency spectrum.

To achieve high spectral efficiency, modulation schemes must be selected that have a high **bandwidth efficiency** as measured in units of bits per second per Hertz of bandwidth. Many wireless communication systems, such as cellular telephones, operate on the principle of frequency reuse, where the carrier frequencies are reused at geographically separated locations. The link quality in these systems is limited by cochannel interference. Hence, modulation schemes must be identified that are both bandwidth efficient and capable of tolerating high levels of cochannel interference. More specifically, digital modulation techniques are chosen for wireless systems that satisfy the following properties.

Compact Power Density Spectrum: To minimize the effect of adjacent channel interference, it is desirable that the power radiated into the adjacent channel be 60–80 dB below that in the desired channel. Hence, modulation techniques with a narrow main lobe and fast rolloff of sidelobes are desirable.

0-8493-8597-0/99/$0.00+$.50

© 1999 by CRC Press LLC

Good Bit-Error-Rate Performance: A low-bit-error probability should be achieved in the presence of cochannel interference, adjacent channel interference, thermal noise, and other channel impairments, such as fading and intersymbol interference.

Envelope Properties: Portable and mobile applications typically employ nonlinear (class C) power amplifiers to minimize battery drain. Nonlinear amplification may degrade the bit-error-rate performance of modulation schemes that transmit information in the amplitude of the carrier. Also, spectral shaping is usually performed prior to up-conversion and nonlinear amplification. To prevent the regrowth of spectral sidelobes during nonlinear amplification, the input signal must have a relatively constant envelope.

A variety of digital modulation techniques are currently being used in wireless communication systems. Two of the more widely used digital modulation techniques for cellular mobile radio are $\pi/4$ phase-shifted quadrature **phase shift keying** ($\pi/4$-QPSK) and **Gaussian minimum shift keying** (**GMSK**). The former is used in the North American IS-54 digital cellular system and Japanese Personal Digital Cellular (PDC), whereas the latter is used in the global system for mobile communications (GSM system). This chapter provides a discussion of these and other modulation techniques that are employed in wireless communication systems.

16.2 Basic Description of Modulated Signals

With any modulation technique, the bandpass signal can be expressed in the form

$$s(t) = \mathrm{Re}\left\{v(t)e^{j2\pi f_c t}\right\} \tag{16.1}$$

where $v(t)$ is the complex envelope, f_c is the carrier frequency, and $\mathrm{Re}\{z\}$ denotes the real part of z. For digital modulation schemes $v(t)$ has the general form

$$v(t) = A\sum_k b\left(t - kT, \boldsymbol{x}_k\right) \tag{16.2}$$

where A is the amplitude of the carrier $\boldsymbol{x}_k = (x_k, x_{k-1}, \ldots, x_{k-K})$ is the data sequence, T is the symbol or baud duration, and $b(t, \boldsymbol{x}_i)$ is an equivalent shaping function usually of duration T. The precise form of $b(t, \boldsymbol{x}_i)$ and the memory length K depends on the type of modulation that is employed. Several examples are provided in this chapter where information is transmitted in the amplitude, phase, or frequency of the bandpass signal.

The **power spectral density** of the bandpass signal $S_{ss}(f)$ is related to the power spectral density of the complex envelope $S_{vv}(f)$ by

$$S_{ss}(f) = \frac{1}{2}\left[S_{vv}\left(f - f_c\right) + S_{vv}\left(f + f_c\right)\right] \tag{16.3}$$

The power density spectrum of the complex envelope for a digital modulation scheme has the general form

$$S_{vv}(f) = \frac{A^2}{T}\sum_m S_{b,m}(f)e^{-j2\pi fmT} \tag{16.4}$$

where

$$S_{b,m}(f) = \frac{1}{2}E\left[B\left(f, \boldsymbol{x}_m\right)B^*\left(f, \boldsymbol{x}_0\right)\right] \tag{16.5}$$

$B(f, \boldsymbol{x}_m)$ is the Fourier transform of $b(t, \boldsymbol{x}_m)$, and $E[\cdot]$ denotes the expectation operator. Usually symmetric signal sets are chosen so that the complex envelope has zero mean, i.e.,

$E[b(t, \boldsymbol{x}_0)] = 0$. This implies that the power density spectrum has no discrete components. If, in addition, $\boldsymbol{x}_m$ and $\boldsymbol{x}_0$ are independent for $|m| > K$, then

$$S_{vv}(f) = \frac{A^2}{T} \sum_{|m|<K} S_{b,m}(f) e^{-j2\pi fmT} \tag{16.6}$$

16.3 Analog Frequency Modulation

With analog frequency modulation the complex envelope is

$$v(t) = A \exp\left[j2\pi k_f \int_0^t m(\tau)\,\mathrm{d}\tau\right] \tag{16.7}$$

where $m(t)$ is the modulating waveform and k_f in Hz/v is the frequency sensitivity of the FM modulator. The bandpass signal is

$$s(t) = A\cos\left[2\pi f_c t + 2\pi k_f \int_0^t m(t)\,\mathrm{d}t\right] . \tag{16.8}$$

The instantaneous frequency of the carrier $f_i(t) = f_c + k_f m(t)$ varies linearly with the waveform $m(t)$, hence, the name frequency modulation. Notice that FM has a constant envelope making it suitable for nonlinear amplification. However, the complex envelope is a nonlinear function of the modulating waveform $m(t)$ and, therefore, the spectral characteristics of $v(t)$ cannot be obtained directly from the spectral characteristics of $m(t)$.

With the sinusoidal modulating waveform $m(t) = A_m \cos(2\pi f_m t)$ the instantaneous carrier frequency is

$$f_i(t) = f_c + \Delta_f \cos(2\pi f_m t) \tag{16.9}$$

where $\Delta_f = k_f A_m$ is the peak frequency deviation. The complex envelope becomes

$$\begin{aligned} v(t) &= \exp\left[2\pi \int_0^t f_i(t)\,\mathrm{d}t\right] \\ &= \exp\left[2\pi f_c t + \beta \sin(2\pi f_m t)\right] \end{aligned} \tag{16.10}$$

where $\beta = \Delta_f / f_m$ is called the modulation index. The bandwidth of $v(t)$ depends on the value of β. If $\beta < 1$, then narrowband FM is generated, where the spectral widths of $v(t)$ and $m(t)$ are about the same, i.e., $2f_m$. If $\beta \gg 1$, then wideband FM is generated, where the spectral occupancy of $v(t)$ is slightly greater than $2\Delta_f$. In general, the approximate bandwidth of an FM signal is

$$W \approx 2\Delta_f + 2f_m = 2\Delta_f\left(1 + \frac{1}{\beta}\right) \tag{16.11}$$

which is a relation known as Carson's rule. Unfortunately, typical analog cellular radio systems use a modulation index in the range $1 \lesssim \beta \lesssim 3$ where Carson's rule is not accurate. Furthermore, the message waveform $m(t)$ is not a pure sinusoid so that Carson's rule does not directly apply.

In analog cellular systems the waveform $m(t)$ is obtained by first companding the speech waveform and then hard limiting the resulting signal. The purpose of the limiter is to control the peak frequency deviation Δ_f. The limiter introduces high-frequency components that

must be removed with a low-pass filter prior to modulation. To estimate the bandwidth occupancy, we first determine the ratio of the frequency deviation Δ_f corresponding to the maximum amplitude of $m(t)$, and the highest frequency component B that is present in $m(t)$. These two conditions are the most extreme cases, and the resulting ratio, $D = \Delta_f/B$, is called the *deviation ratio*. Then replace β by D and f_m by B in Carson's rule, giving

$$W \approx 2\Delta_f + 2B = 2\Delta_f \left(1 + \frac{1}{D}\right) \tag{16.12}$$

This approximation will overestimate the bandwidth requirements. A more accurate estimate of the bandwidth requirements must be obtained from simulation or measurements.

16.4 Phase Shift Keying (PSK) and $\pi/4$-QPSK

With **phase shift keying** (PSK), the equivalent shaping function in Eq. (16.2) has the form

$$b(t, \boldsymbol{x}_k) = \psi_T(t)\exp\left[j\frac{\pi}{M}x_k h_s(t)\right], \qquad \boldsymbol{x}_k = x_k \tag{16.13}$$

where $h_s(t)$ is a phase shaping pulse, $\psi_T(t)$ an amplitude shaping pulse, and M the size of the modulation alphabet. Notice that the phase varies linearly with the symbol sequence $\{x_k\}$, hence the name phase shift keying. For a modulation alphabet size of M, $x_k \in \{\pm 1, \pm 3, \ldots, \pm(M-1)\}$. Each symbol x_k is mapped onto $\log_2 M$ source bits. A QPSK signal is obtained by using $M = 4$, resulting in a transmission rate of 2 b/symbol.

Usually, the phase shaping pulse is chosen to be the rectangular pulse $h_s(t) = u_T(t) \triangleq u(t) - u(t-T)$, where $u(t)$ is the unit step function. The amplitude shaping pulse is very often chosen to be a square root raised cosine pulse, where the Fourier transform of $\psi_T(t)$ is

$$\Psi_T(f) = \begin{cases} \sqrt{T} & 0 \le |f| \le (1-\beta)/2T \\ \sqrt{\dfrac{T}{2}\left[1 - \sin\dfrac{\pi T}{\beta}\left(f - \dfrac{1}{2T}\right)\right]} & (1-\beta)/2T \le |f| \le (1+\beta)/2T \end{cases} \tag{16.14}$$

The receiver implements the same filter $\Psi_R(f) = \Psi_T(f)$ so that the overall pulse has the raised cosine spectrum $\Psi(f) = \Psi_R(f)\Psi_T(f) = |\Psi_T(f)|^2$. If the channel is affected by flat fading and additive white Gaussian noise, then this partitioning of the filtering operations between the transmitter and receiver will optimize the signal to noise ratio at the output of the receiver filter at the sampling instants. The rolloff factor β usually lies between 0 and 1 and defines the **excess bandwidth** $100\beta\%$. Using a smaller β results in a more compact power density spectrum, but the link performance becomes more sensitive to errors in the symbol timing. The IS-54 system uses $\beta = 0.35$, while PDC uses $\beta = 0.5$.

The time domain pulse corresponding to Eq. (16.14) can be obtained by taking the inverse Fourier transform, resulting in

$$\psi_T(t) = 4\beta\frac{\cos\left[(1+\beta)\,\pi t/T\right] + \sin\left[(1-\beta)\,\pi t/T\right](4\beta t/T)^{-1}}{\pi\sqrt{T}\left[1 - 16\beta^2 t^2/T^2\right]} \tag{16.15}$$

A typical square root raised cosine pulse with a rolloff factor of $\beta = 0.5$ is shown in Fig. 16.1. Strictly speaking the pulse $\psi_T(t)$ is noncausal, but in practice a truncated time domain pulse is used. For example, in Fig. 16.1 the pulse is truncated to $6T$ and time shifted by $3T$ to yield a causal pulse.

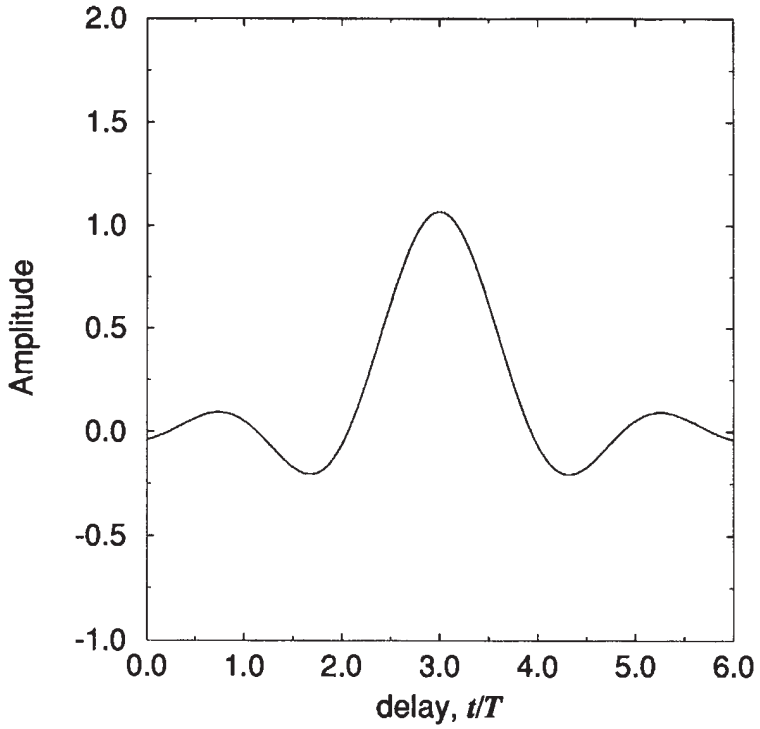


FIGURE 16.1 Square root raised cosine pulse with rolloff factor $\beta = 0.5$.

Unlike conventional QPSK that has four possible transmitted phases, $\pi/4$-QPSK has eight possible transmitted phases. Let $\theta(n)$ be the transmitted carrier phase for the nth epoch, and let $\Delta\theta(n) = \theta(n) - \theta(n-1)$ be the differential carrier phase between epochs n and $n-1$. With $\pi/4$-QPSK, the transmission rate is 2 b/symbol and the differential phase is related to the symbol sequence $\{x_n\}$ through the mapping

$$\Delta\theta(n) = \begin{cases} -3\pi/4, & x_n = -3 \\ -\pi/4, & x_n = -1 \\ \pi/4, & x_n = +1 \\ 3\pi/4, & x_n = +3 \end{cases} \tag{16.16}$$

Since the symbol sequence $\{x_n\}$ is random, the mapping in Eq. (16.16) is arbitrary, except that the phase differences must be $\pm\pi/4$ and $\pm 3\pi/4$. The phase difference with the given mapping can be written in the convenient algebraic form

$$\Delta\theta(n) = x_n \frac{\pi}{4} \tag{16.17}$$

which allows us to write the equivalent shaping function of the $\pi/4$-QPSK signal as

$$\begin{aligned} b\,(t, \underline{\boldsymbol{x}}_k) &= \psi(t)\exp\left\{j\left[\theta(k-1) + x_k\frac{\pi}{4}\right]\right\} \\ &= \psi_T(t)\exp\left[j\frac{\pi}{4}\left(\sum_{n=-\infty}^{k-1} x_n + x_k\right)\right] \end{aligned} \tag{16.18}$$

The summation in the exponent represents the accumulated carrier phase, whereas the last term is the phase change due to the kth symbol. Observe that the phase shaping function

is the rectangular pulse $u_T(t)$. The amplitude shaping function $\psi_T(t)$ is usually the square root raised cosine pulse in Eq. (16.15).

The phase states of QPSK and $\pi/4$-QPSK signals can be summarized by the signal space diagram in Fig. 16.2 that shows the phase states and allowable transitions between the phase states. However, it does not describe the actual phase trajectories. A typical diagram showing phase trajectories with square root raised cosine pulse shaping is shown in Fig. 16.3. Note that the phase trajectories do not pass through the origin. This reduces the envelope fluctuations of the signal making it less susceptible to amplifier nonlinearities and reduces the dynamic range required of the power amplifier.

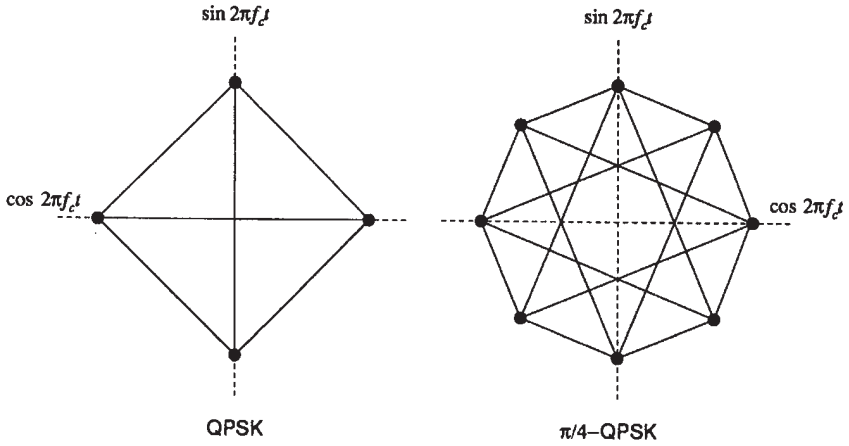


FIGURE 16.2 Signal-space constellations for QPSK and $\pi/4$-DQPSK.

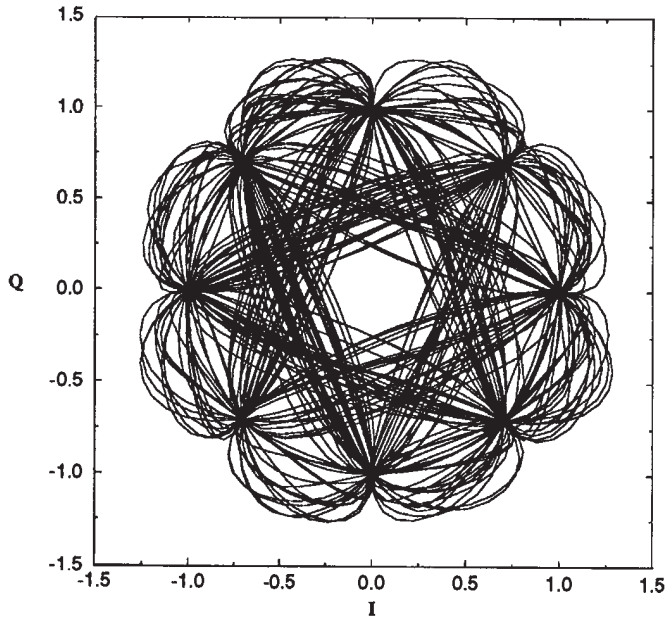


FIGURE 16.3 Phase diagram of $\pi/4$-QPSK with square root raised cosine pulse; $\beta = 0.5$.

The power density spectrum of QPSK and $\pi/4$-QPSK depends on both the amplitude and phase shaping pulses. For the rectangular phase shaping pulse $h_s(t) = u_T(t)$, the power density spectrum of the complex envelope is

$$S_{vv}(f) = \frac{A^2}{T} |\Psi_T(f)|^2 \tag{16.19}$$

With square root raised cosine pulse shaping, $\Psi_T(f)$ has the form defined in Eq. (16.14). The power density spectrum of a pulse $\tilde{\psi}_T(t)$ that is obtained by truncating $\psi_T(t)$ to length τ can be obtained by writing $\tilde{\psi}_T(t) = \psi_T(t)\text{rect}(t/\tau)$. Then $\tilde{\Psi}_T(f) = \Psi_T(f) * \tau\text{sinc}(f\tau)$, where $*$ denotes the operation of convolution, and the power density spectrum is again obtained by applying Eq. (16.19). Truncation of the pulse will regenerate some side lobes, thus causing adjacent channel interference. Figure 16.4 illustrates the power density spectrum of a truncated square root raised cosine pulse for various truncation lengths τ.

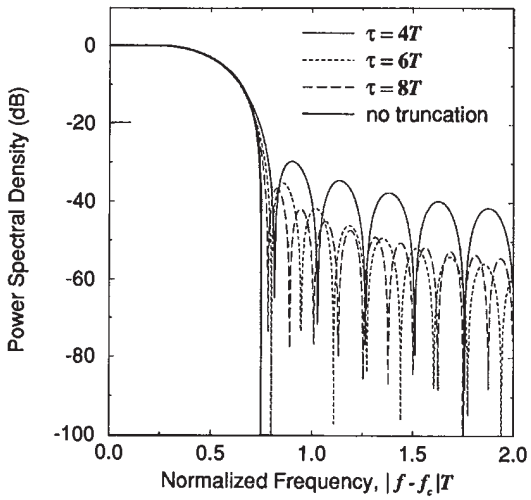


FIGURE 16.4 Power density spectrum of truncated square root raised cosine pulse with various truncation lengths; $\beta = 0.5$.

16.5 Continuous Phase Modulation (CPM) and MSK

Continuous phase modulation (CPM) refers to a broad class of frequency modulation techniques where the carrier phase varies in a continuous manner. A comprehensive treatment of CPM is provided in [1]. CPM schemes are attractive because they have constant envelope and excellent spectral characteristics. The complex envelope of any CPM signal is

$$v(t) = A \exp\left[j2\pi k_f \int_{-\infty}^{t} \sum_n x_n h_s(\tau - nT)\, \mathrm{d}\tau \right] \tag{16.20}$$

The instantaneous frequency deviation from the carrier is

$$f_{\text{dev}}(t) = k_f \sum_n x_n h_s(t - nT) \tag{16.21}$$

where k_f is the peak frequency deviation. If the frequency shaping pulse $h_s(t)$ has duration T, then the equivalent shaping function in Eq. (16.2) has the form

$$b(t, \boldsymbol{x}_k) = \exp\left\{ j \left[\beta(T) \sum_{n=-\infty}^{k-1} x_n + x_k \beta(t) \right] \right\} u_T(t) \tag{16.22}$$

where

$$\beta(t) = \begin{cases} 0, & t < 0 \\ \dfrac{\pi h}{\int_0^T h_s(\tau)\,\mathrm{d}\tau} \displaystyle\int_0^t h_s(\tau)\,\mathrm{d}\tau, & 0 \le t \le T \\ \pi h, & t \ge T \end{cases} \tag{16.23}$$

is the phase shaping pulse, and $h = \beta(T)/\pi$ is called the modulation index.

Minimum shift keying (MSK) is a special form of binary CPM ($x_k \in \{-1, +1\}$) that is defined by a rectangular frequency shaping pulse $h_s(t) = u_T(t)$, and a modulation index $h = 1/2$ so that

$$\beta(t) = \begin{cases} 0, & t < 0 \\ \pi t/2T, & 0 \le t \le T \\ \pi/2, & t \ge T \end{cases} \tag{16.24}$$

Therefore, the complex envelope is

$$v(t) = A \exp\left(j \frac{\pi}{2} \sum_{n=-\infty}^{k-1} x_n + \frac{\pi}{2} x_k \frac{t - kT}{T} \right) \tag{16.25}$$

A MSK signal can be described by the phase trellis diagram shown in Fig. 16.5 which plots the time behavior of the phase

$$\theta(t) = \frac{\pi}{2} \sum_{n=-\infty}^{k-1} x_n + \frac{\pi}{2} x_k \frac{t - kT}{T} \tag{16.26}$$

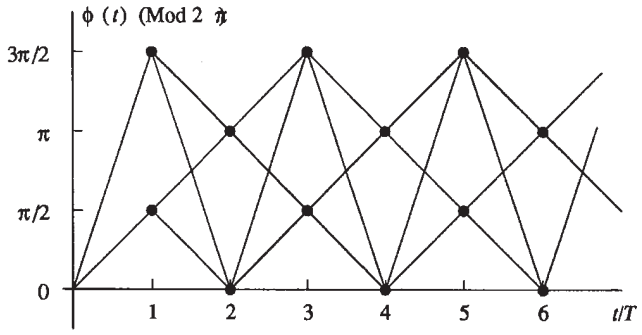


FIGURE 16.5 Phase-trellis diagram for MSK.

The MSK bandpass signal is

$$\begin{aligned} s(t) &= A \cos\left(2\pi f_c t + \frac{\pi}{2} \sum_{n=-\infty}^{k-1} x_n + \frac{\pi}{2} x_k \frac{t - kT}{T} \right) \\ &= A \cos\left[2\pi \left(f_c + \frac{x_k}{4T} \right) t - \frac{k\pi}{2} x_k + \frac{\pi}{2} \sum_{n=-\infty}^{k-1} x_n \right] \quad kT \le t \le (k+1)T \end{aligned} \tag{16.27}$$

From Eq. (16.27) we observe that the MSK signal has one of two possible frequencies $f_L = f_c - 1/4T$ or $f_U = f_c + 1/4T$ during each symbol interval. The difference between these frequencies is $f_U - f_L = 1/2T$. This is the minimum frequency difference between two sinusoids of duration T that will ensure orthogonality with coherent demodulation [7], hence, the name minimum shift keying. By applying various trigonometric identities to Eq. (16.27) we can write

$$s(t) = A\left[x_k^I \psi(t - k2T)\cos(2\pi f_c t) - x_k^Q \psi(t - k2T - T)\sin(2\pi f_c t)\right], \quad kT \le t \le (k+1)T \tag{16.28}$$

where

$$\begin{aligned} x_k^I &= -x_{k-1}^Q x_{2k-1} \\ x_k^Q &= x_k^I x_{2k} \\ \psi(t) &= \cos\left(\frac{\pi t}{2T}\right), \qquad -T \le t \le T \end{aligned}$$

Note that the x_k^I and x_k^Q are independent binary symbols that take on elements from the set $\{-1, +1\}$, and the half-sinusoid amplitude shaping pulse $\psi(t)$ has duration $2T$ and $\psi(t - T) = \sin(\pi t/2T), 0 \le t \le 2T$. Therefore, MSK is equivalent to offset quadrature amplitude shift keying (OQASK) with a half-sinusoid amplitude shaping pulse.

To obtain the power density spectrum of MSK, we observe from Eq. (16.28) that the equivalent shaping function of MSK has the form

$$b(t, \boldsymbol{x}_k) = x_k^I \psi(t) + j x_k^Q \psi(t - T) \tag{16.29}$$

The Fourier transform of Eq. (16.29) is

$$B(f, \boldsymbol{x}_k) = \left(x_k^I + j x_k^Q e^{-j2\pi fT}\right)\Psi(f) \tag{16.30}$$

Since the symbols x_k^I and x_k^Q are independent and zero mean, it follows from Eqs. (16.5) and (16.6) that

$$S_{vv}(f) = \frac{A^2|\Psi(f)|^2}{2T} \tag{16.31}$$

Therefore, the power density spectrum of MSK is determined solely by the Fourier transform of the half-sinusoid amplitude shaping pulse $\psi(t)$, resulting in

$$S_{vv}(f) = \frac{16A^2T}{\pi^2}\left[\frac{\cos 2\pi fT}{1 - 16f^2T^2}\right]^2 \tag{16.32}$$

The power spectral density of MSK is plotted in Fig. 16.8. Observe that an MSK signal has fairly large sidelobes compared to $\pi/4$-QPSK with a truncated square root raised cosine pulse (c.f., Fig. 16.4).

16.6 Gaussian Minimum Shift Keying

MSK signals have all of the desirable attributes for mobile radio, except for a compact power density spectrum. This can be alleviated by filtering the modulating signal $x(t) = \Sigma_n x_n u_T(t - nT)$ with a low-pass filter prior to frequency modulation, as shown in Fig. 16.6.

Such filtering removes the higher frequency components in $x(t)$ and, therefore, yields a more compact spectrum. The low-pass filter is chosen to have 1) narrow bandwidth and a sharp transition band, 2) low-overshoot impulse response, and 3) preservation of the output pulse area to ensure a phase shift of $\pi/2$.

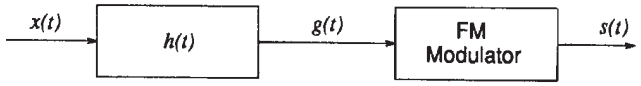


FIGURE 16.6 Premodulation filtered MSK.

GMSK uses a low-pass filter with the following transfer function:

$$H(f) = A \exp\left\{ -\left(\frac{f}{B}\right)^2 \frac{\ln 2}{2} \right\} \tag{16.33}$$

where B is the 3-dB bandwidth of the filter and A a constant. It is apparent that $H(f)$ is bell shaped about $f = 0$, hence the name Gaussian MSK. A rectangular pulse $\text{rect}(t/T) = u_T(t + T/2)$ transmitted through this filter yields the frequency shaping pulse

$$h_s(t) = A\sqrt{\frac{2\pi}{\ln 2}}(BT) \int_{t/T-1/2}^{t/T+1/2} \exp\left\{ -\frac{2\pi^2 (BT)^2 x^2}{\ln 2} \right\} \mathrm{d}x \tag{16.34}$$

The phase change over the time interval from $-T/2 \leq t \leq T/2$ is

$$\theta\left(\frac{T}{2}\right) - \theta\left(\frac{-T}{2}\right) = x_0 \beta_0(T) + \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} x_n \beta_n(T) \tag{16.35}$$

where

$$\beta_n(T) = \frac{\pi h}{\int_{-\infty}^{\infty} h_s(\nu)\, \mathrm{d}\nu} \int_{-T/2-nT}^{T/2-nT} h_s(\nu)\, \mathrm{d}\nu \tag{16.36}$$

The first term in Eq. (16.35) is the desired term, and the second term is the intersymbol interference (ISI) introduced by the premodulation filter. Once again, with GMSK $h = 1/2$ so that a total phase shift of $\pi/2$ is maintained.

Notice that the pulse $h_s(t)$ is noncausal so that a truncated pulse must be used in practice. Figure 16.7 plots a GMSK frequency shaping pulse that is truncated to $\tau = 5T$ and time shifted by $2.5T$, for various normalized filter bandwidths BT. Notice that the frequency shaping pulse has a duration greater than T so that ISI is introduced. As BT decreases, the induced ISI is increased. Thus, whereas a smaller value of BT results in a more compact power density spectrum, the induced ISI will degrade the bit-error-rate performance. Hence, there is a tradeoff in the choice of BT. Some studies have indicated that $BT = 0.25$ is a good choice for cellular radio systems [6].

The power density spectrum of GMSK is quite difficult to obtain, but can be computed by using published methods [3]. Figure 16.8 plots the power density spectrum for $BT = 0.2$, 0.25, and 0.3, obtained from Wesolowski, [8]. Observe that the spectral sidelobes are greatly reduced by the Gaussian low-pass filter.

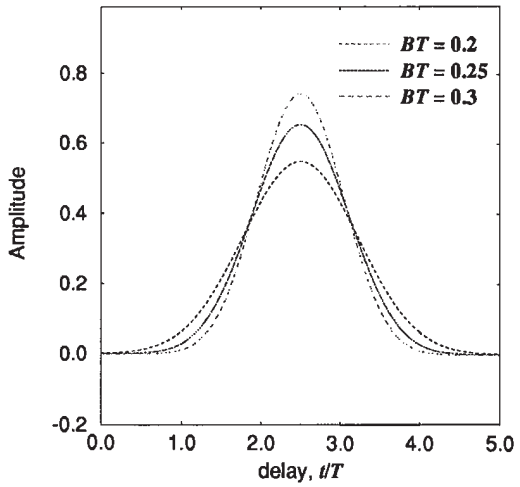


FIGURE 16.7 GMSK frequency shaping pulse for various normalized filter bandwidths BT.

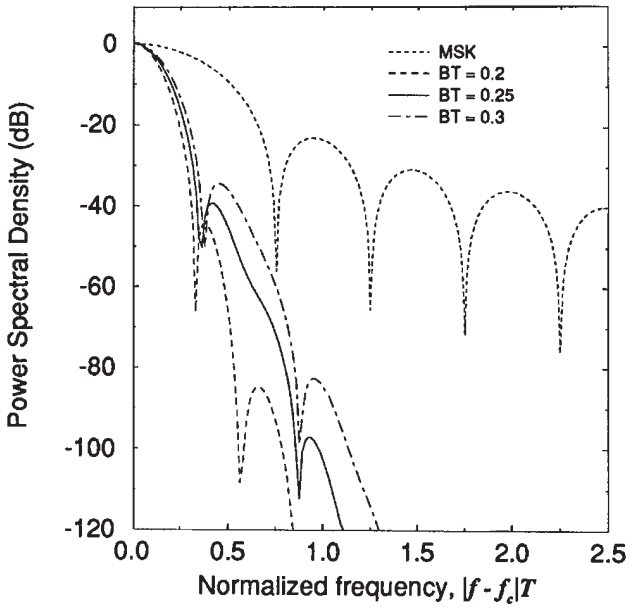


FIGURE 16.8 Power density spectrum of MSK and GMSK.

16.7 Orthogonal Frequency Division Multiplexing (OFDM)

Orthogonal frequency division multiplexing (OFDM) is a modulation technique that has been recently suggested for use in cellular radio [2], digital audio broadcasting [4], and digital video broadcasting. The basic idea of OFDM is to transmit blocks of symbols in parallel by employing a (large) number of orthogonal subcarriers. With block transmission, N serial source symbols each with period T_s are converted into a block of N parallel modulated symbols each with period $T = NT_s$. The block length N is chosen so that $NT_s \gg \sigma_\tau$, where σ_τ is the rms delay spread of the channel. Since the symbol rate on each subcarrier is much less than the serial source rate, the effects of delay spread are greatly reduced. This has practical advantages because it may reduce or even eliminate the need for equalization.

Although the block length N is chosen so that $NT_s \gg \sigma_\tau$, the channel dispersion will still cause consecutive blocks to overlap. This results in some residual ISI that will degrade the performance. This residual ISI can be eliminated at the expense of channel capacity by using guard intervals between the blocks that are at least as long as the effective channel impulse response.

The complex envelope of an OFDM signal is described by

$$v(t) = A \sum_k \sum_{n=0}^{N-1} x_{k,n} \phi_n(t - kT) \tag{16.37}$$

where

$$\phi_n(t) = \exp\left\{ j \frac{2\pi \left(n - \dfrac{N-1}{2} \right) t}{T} \right\} U_T(t), \qquad n = 0, 1, \ldots, N-1 \tag{16.38}$$

are orthogonal waveforms and $U_T(t)$ is a rectangular shaping function. The frequency separation of the subcarriers, $1/T$, ensures that the subcarriers are orthogonal and phase continuity is maintained from one symbol to the next, but is twice the minimum required for orthogonality with coherent detection. At epoch k, N-data symbols are transmitted by using the N distinct pulses. The data symbols $x_{k,n}$ are often chosen from an M-ary **quadrature amplitude modulation** (M-QAM) constellation, where $x_{k,n} = x_{k,n}^I + j x_{k,n}^Q$ with $x_{k,n}^I, x_{k,n}^Q \in \{\pm 1, \pm 3, \ldots, \pm(N-1)\}$ and $N = \sqrt{M}$.

A key advantage of using OFDM is that the modulation can be achieved in the discrete domain by using either an inverse discrete Fourier transform (IDFT) or the more computationally efficient inverse fast Fourier transform (IFFT). Considering the data block at epoch $k = 0$ and ignoring the frequency offset $\exp\{-j[2\pi(N-1)t/2T]\}$, the complex low-pass OFDM signal has the form

$$v(t) = \sum_{n=0}^{N-1} x_{0,n} \exp\left\{ \frac{j2\pi nt}{NT_s} \right\}, \qquad 0 \le t \le T \tag{16.39}$$

If this signal is sampled at epochs $t = kT_s$, then

$$v^k = v(kT_s) = \sum_{n=0}^{N-1} x_{0,n} \exp\left\{ \frac{j2\pi nk}{N} \right\}, \qquad k = 0, 1, \ldots, N-1 \tag{16.40}$$

Observe that the sampled OFDM signal has duration N and the samples $v^0, v^1, \ldots, v^{N-1}$ are just the IDFT of the data block $x_{0,0}, x_{0,1}, \ldots, x_{0,N-1}$. A block diagram of an OFDM transmitter is shown in Fig. 16.9.

The power spectral density of an OFDM signal can be obtained by treating OFDM as independent modulation on subcarriers that are separated in frequency by $1/T$. Because the subcarriers are only separated by $1/T$, significant spectral overlap results. Because the subcarriers are orthogonal, however, the overlap improves the spectral efficiency of the scheme. For a signal constellation with zero mean and the waveforms in Eq. (16.38), the power density spectrum of the complex envelope is

$$S_{vv}(f) = \frac{A^2}{T} \sigma_x^2 \sum_{n=0}^{N-1} \left| \operatorname{sinc}\left[fT - \left(n - \frac{N-1}{2} \right) \right] \right|^2 \tag{16.41}$$

where $\sigma_x^2 = \frac{1}{2} E[|x_{k,n}|^2]$ is the variance of the signal constellation. For example, the complex envelope power spectrum of OFDM with $N = 32$ subcarriers is shown in Fig. 16.10.

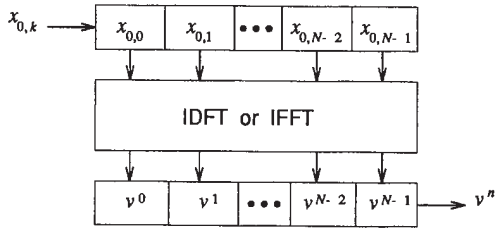


FIGURE 16.9 Block diagram of OFDM transmitter using IDFT or IFFT.

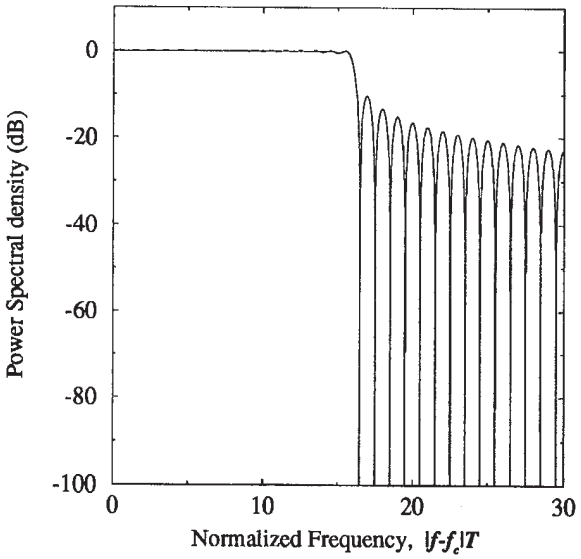


FIGURE 16.10 Power density spectrum of OFDM with $N = 32$.

16.8 Conclusions

A variety of modulation schemes are employed in wireless communication systems. Wireless modulation schemes must have a compact power density spectrum, while at the same time providing a good bit-error-rate performance in the presence of channel impairments such as cochannel interference and fading. The most popular digital modulation techniques employed in wireless systems are GMSK in the European GSM system, $\pi/4$-QPSK in the North American IS-54 and Japanese PDC systems, and OFDM in digital audio broadcasting systems.

Defining Terms

Bandwidth efficiency: Transmission efficiency of a digital modulation scheme measured in units of bits per second per Hertz of bandwidth.

Continuous phase modulation: Frequency modulation where the phase varies in a continuous manner.

Excess bandwidth: Percentage of bandwidth that is in excess of the minimum of $1/2T$ (T is the baud or symbol duration) required for data communication.

Frequency modulation: Modulation where the instantaneous frequency of the carrier varies linearly with the data signal.

Gaussian minimum shift keying: MSK where the data signal is prefiltered with a Gaussian filter prior to frequency modulation.

Minimum shift keying: A special form of continuous phase modulation having linear phase trajectories and a modulation index of 1/2.

Orthogonal frequency division multiplexing: Modulation by using a collection of low-bit-rate orthogonal subcarriers.

Phase shift keying: Modulation where the instantaneous phase of the carrier varies linearly with the data signal.

Power spectral density: Relative power in a modulated signal as a function of frequency.

Quadrature amplitude modulation: Modulation where information is transmitted in the amplitude of the cosine and sine components of the carrier.

References

[1] Anderson, J.B., Aulin, T., and Sundberg, C.-E., *Digital Phase Modulation,* Plenum Press, New York, 1986.

[2] Birchler, M.A. and Jasper, S.C., A 64 kbps digital land mobile radio system employing M-16QAM. *Proc. 5th Nordic Sem. Dig. Mobile Radio Commun.,* 237–241, Dec. 1992.

[3] Garrison, G.J., A power spectral density analysis for digital FM, *IEEE Trans. Commun.,* COM-23, 1228–1243, Nov. 1975.

[4] Le Floch, B., Halbert-Lassalle, R., and Castelain, D., Digital sound broadcasting to mobile receivers, *IEEE Trans. Consum. Elec.,* 35, Aug. 1989.

[5] Murota, K. and Hirade, K., GMSK modulation for digital mobile radio telephony, *IEEE Trans. Commun.,* COM-29, 1044–1050, Jul. 1981.

[6] Murota, K., Kinoshita, K., and Hirade, K., Spectral efficiency of GMSK land mobile radio. *Proc. ICC'81,* 23.8.1, Jun. 1981.

[7] Proakis, J.G., *Digital Communications,* 2nd ed., McGraw-Hill, New York, 1989.

[8] Wesolowski, K., Private Communication, 1994.

Further Information

A good discussion of digital modem techniques is presented in *Advanced Digital Communications,* edited by K. Feher, Prentice-Hall, 1987.

Proceedings of various IEEE conferences such as the Vehicular Technology Conference, International Conference on Communications, and Global Telecommunications Conference, document the lastest development in the field of wireless communications each year.

Journals such as the *IEEE Transactions on Communications* and *IEEE Transactions on Vehicular Technology* report advances in wireless modulation.

17

Access Methods

17.1 Introduction 17-1
17.2 Relevant Wireless Communication System Characteristics 17-2
17.3 Frequency Division Multiple Access 17-4
Channel Considerations • Influence of Antenna Height • Example 17.1: CT2 • Further Remarks
17.4 Time Division Multiple Access 17-7
Propagation Considerations • Initial Channel Assignment • Example 17.2: GSM • Further Remarks
17.5 Code Division Multiple Access 17-9
Propagation Considerations • Multiple-Access Interference • Further Remarks
17.6 Comparison and Outlook 17-11
Defining Terms 17-12
References 17-13
Further Information 17-13

Bernd-Peter Paris
George Mason University

17.1 Introduction

The radio channel is fundamentally a broadcast communication medium. Therefore, signals transmitted by one user can potentially be received by all other users within range of the transmitter. Although this high connectivity is very useful in some applications, like broadcast radio or television, it requires stringent access control in wireless communication systems to avoid, or at least to limit, interference between transmissions. Throughout, the term wireless communication systems is taken to mean communication systems that facilitate two-way communication between a portable radio communication terminal and the fixed network infrastructure. Such systems range from mobile cellular systems through personal communication systems (PCS) to cordless telephones.

The objective of wireless communication systems is to provide communication channels on demand between a portable radio station and a radio port or base station that connects the user to the fixed network infrastructure. Design criteria for such systems include **capacity**, cost of implementation, and quality of service. All of these measures are influenced by the method used for providing multiple-access capabilities. However, the opposite is also true: the access method should be chosen carefully in light of the relative importance of design criteria as well as the system characteristics.

Multiple access in wireless radio systems is based on insulating signals used in different connections from each other. The support of parallel transmissions on the uplink and

0-8493-8597-0/99/$0.00+$.50
© 1999 by CRC Press LLC

downlink, respectively, is called multiple access, whereas the exchange of information in both directions of a connection is referred to as **duplexing**. Hence, multiple access and duplexing are methods that facilitate the sharing of the broadcast communication medium. The necessary insulation is achieved by assigning to each transmission different components of the domains that contain the signals. The signal domains commonly used to provide multiple access capabilities include the following.

Spatial domain: All wireless communication systems exploit the fact that radio signals experience rapid attenuation during propagation. The propagation exponent ρ on typical radio channels lies between $\rho = 2$ and $\rho = 6$ with $\rho = 4$ a typical value. As signal strength decays inversely proportional to the ρth power of the distance, far away transmitters introduce interference that is negligible compared to the strength of the desired signal. The cellular design principle is based on the ability to reuse signals safely if a minimum reuse distance is maintained. Directional antennas can be used to enhance the insulation between signals. We will not focus further on the spatial domain in this treatment of access methods.

Frequency domain: Signals which occupy nonoverlapping frequency bands can be easily separated using appropriate bandpass filters. Hence, signals can be transmitted simultaneously without interfering with each other. This method of providing multiple access capabilities is called **frequency-division multiple access (FDMA)**.

Time domain: Signals can be transmitted in nonoverlapping time slots in a round-robin fashion. Thus, signals occupy the same frequency band but are easily separated based on their time of arrival. This multiple access method is called **time-division multiple access (TDMA)**.

Code domain: In **code-division multiple access (CDMA)** different users employ signals that have very small cross-correlation. Thus, correlators can be used to extract individual signals from a mixture of signals even though they are transmitted simultaneously and in the same frequency band. The term code-division multiple-access is used to denote this form of channel sharing. Two forms of CDMA are most widely employed and will be described in detail subsequently, frequency hopping (FH) and direct sequence (DS).

System designers have to decide in favor of one, or a combination, of the latter three domains to facilitate multiple access. The three access methods are illustrated in Fig. 17.1. The principal idea in all three of these access methods is to employ signals that are orthogonal or nearly orthogonal. Then, correlators that project the received signal into the subspace of the desired signal can be employed to extract a signal without interference from other transmissions.

Preference for one access method over another depends largely on overall system characteristics, as we will see in the sequel. No single access method is universally preferable, and system considerations should be carefully weighed before the design decision is made. Before going into the detailed description of the different access methods, we will discuss briefly the salient features of some wireless communication systems. This will allow us later to assess the relative merits of the access methods in different scenarios.

17.2 Relevant Wireless Communication System Characteristics

Modern wireless radio systems range from relatively simple cordless telephones to mobile cellular systems and the emerging personal communication systems (PCS). It is useful to consider such diverse systems as cordless telephone and mobile cellular radio to illustrate some of the fundamental characteristics of wireless communication systems [2].

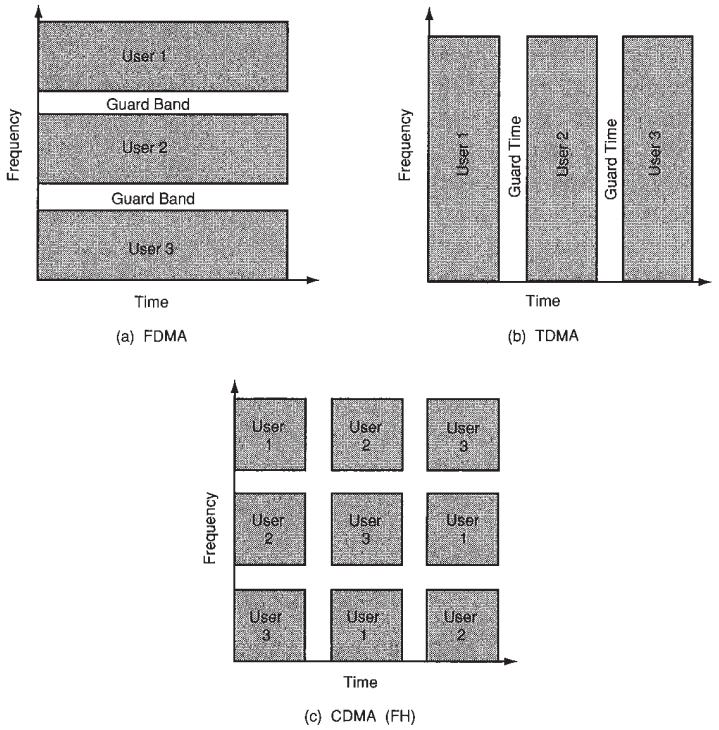


FIGURE 17.1 Multiple-access methods for wireless communication systems.

A summary of the relevant parameters and characteristics for cordless telephone and cellular radio is given in Table 17.1 As evident from that table, the fundamental differences between the two systems are speech quality and the area covered by a base station. The high speech quality requirement in the cordless application is the consequence of the availability of tethered access in the home and office and the resulting direct competition with wire-line telephone services. In the mobile cellular application, the user has no alternative to the wireless access and may be satisfied with lower, but still acceptable, quality of service.

TABLE 17.1 Summary of Relevant Characteristics of Cordless Telephone and Cellular Mobile Radio

Characteristic or Parameter	Cordless Telephone	Cellular Radio
Speech quality	Toll quality	Varying with channelquality; possibly decreased by speech pause exploitation
Transmission range	<100 m	100 m–30 km
Transmit power	Milliwatts	Approx. 1 W
Base station antenna height	Approx. 1 m	Tens of meters
Delay spread	Approx. 1 μs	Approx. 10 μs
Complexity of base station	Low	High
Complexity of user set	Low	High

In cordless telephone applications the transmission range is short because the base station can simply be moved to a conveniently located wire-line access point (wall jack) to provide wireless network access where desired. In contrast, the mobile cellular base station must provide access for users throughout a large geographical area of up to approximately 30 km

(20 mi) around the base station. This large coverage area is necessary to economically meet the promise of uninterrupted service to roaming users.

The different range requirements directly affect the transmit power and antenna height for the two systems. High-power transmitters used in mobile cellular user sets consume far more power than even complex signal processing hardware. Hence, sophisticated signal processing, including speech compression, voice activity detection, error correction and detection, and adaptive equalization, can be employed without substantial impact on the battery life in portable hand sets. Furthermore, such techniques are consistent with the goals of increased range and support of large numbers of users with a single, expensive base station. On the other hand, the high mobile cellular base station antennas introduce delay spreads that are one or two orders of magnitude larger than those commonly observed in cordless telephone applications.

Clearly, the two systems just considered are at extreme ends of the spectrum of wireless communications systems. Most notably, the emerging PCS systems fall somewhere between the two. However, the comparison above highlights some of the system characteristics that should be considered when discussing access methods for wireless communication systems.

17.3 Frequency Division Multiple Access

As mentioned in Section 17.1, in FDMA nonoverlapping frequency bands are allocated to different users on a continuous time basis. Hence, signals assigned to different users are clearly orthogonal, at least ideally. In practice, out-of-band spectral components can not be completely suppressed leaving signals not quite orthogonal. This necessitates the introduction of guard bands between frequency bands to reduce adjacent channel interference, i.e., inference from signals transmitted in adjacent frequency bands; see also Fig. 17.1(a).

It is advantageous to combine FDMA with time-division duplexing (TDD) to avoid simultaneous reception and transmission that would require insulation between receive and transmit antennas. In this scenario, the base station and portable take turns using the same frequency band for transmission. Nevertheless, combining FDMA and frequency division duplex is possible in principle, as is evident from the analog FM-based systems deployed throughout the world since the early 1980s.

17.3.1 Channel Considerations

In principle there exists the well-known duality between TDMA and FDMA; see [1, p. 113]. In the wireless environment, however, propagation related factors have a strong influence on the comparison between FDMA and TDMA. Specifically, the duration of a transmitted symbol is much longer in FDMA than in TDMA. As an immediate consequence, an equalizer is typically not required in an FDMA-based system because the delay spread is small compared to the symbol duration.

To illustrate this point, consider a hypothetical system that transmits information at a constant rate of 50 kb/s. This rate would be sufficient to support 32-kb/s adaptive differential pulse code modulation (ADPCM) speech encoding, some coding for error protection, and control overhead. If we assume further that some form of QPSK modulation is employed, the resulting symbol duration is 40 μs. In relation to delay spreads of approximately 1 μs in the cordless application and 10 μs in cellular systems, this duration is large enough that only little intersymbol interference is introduced. In other words, the channel is frequency nonselective, i.e., all spectral components of the signal are affected equally by the channel. In the cordless application an equalizer is certainly not required; cellular receivers

may require equalizers capable of removing intersymbol interference between adjacent bits. Furthermore, it is well known that intersymbol interference between adjacent bits can be removed without loss in SNR by using maximum-likelihood sequence estimation; e.g., [8, p. 622].

Hence, rather simple receivers can be employed in FDMA systems at these data rates. However, there is a flip side to the argument. Recall that the Doppler spread, which characterizes the rate at which the channel impulse response changes, is given approximately by $B_d = v/cf_c$, where v denotes the speed of the mobile user, c is the propagation speed of the electromagnetic waves carrying the signal, and f_c is the carrier frequency. Thus, for systems operating in the vicinity of 1 GHz, B_d will be less than 1 Hz in the cordless application and typically about 100 Hz for a mobile traveling on a highway. In either case, the signal bandwidth is much larger than the Doppler spread B_d, and the channel can be characterized as slowly fading. Whereas this allows tracking of the carrier phase and the use of coherent receivers, it also means that fade durations are long in comparison to the symbol duration and can cause long sequences of bits to be subject to poor channel conditions. The problem is compounded by the fact that the channel is frequency nonselective because it implies that the entire signal is affected by a fade.

To overcome these problems either time diversity, frequency diversity, or spatial diversity could be employed. Time diversity can be accomplished by a combination of coding and interleaving if the fading rate is sufficiently large. For very slowly fading channels, such as the cordless application, the necessary interleaving depth would introduce too much delay to be practical. Frequency diversity can be introduced simply by slow frequency hopping, a technique that prescribes users to change the carrier frequency periodically. Frequency hopping is a form of spectrum spreading because the bandwidth occupied by the resulting signal is much larger than the symbol rate. In contrast to direct sequence spread spectrum discussed subsequently, however, the instantaneous bandwidth is not increased. The jumps between different frequency bands effectively emulate the movement of the portable and, thus, should be combined with the just described time-diversity methods. Spatial diversity is provided by the use of several receive or transmit antennas. At carrier frequencies exceeding 1 GHz, antennas are small and two or more antennas can be accommodated even in the hand set. Furthermore, if FDMA is combined with time-division duplexing, multiple antennas at the base station can provide diversity on both uplink and downlink. This is possible because the channels for the two links are virtually identical, and the base station, using channel information gained from observing the portable's signal, can transmit signals at each antenna such that they combine coherently at the portable's antenna. Thus, signal processing complexity is moved to the base station extending the portable's battery life.

17.3.2 Influence of Antenna Height

In the cellular mobile environment base station antennas are raised considerably to increase the coverage area. Antennas mounted on towers and rooftops are a common sight, and antenna heights of 50 m above ground are no exceptions. Besides increasing the coverage area, this has the additional effect that frequently there exists a better propagation path between two base station antennas than between a mobile and the base station; see Fig. 17.2.

Assuming that FDMA is used in conjunction with TDD as specified at the beginning of this section, then base stations and mobiles transmit on the same frequency. Now, unless there is tight synchronization between all base stations, signals from other base stations will interfere with the reception of signals from portables at the base station. To keep the interference at acceptable levels, it is necessary to increase the reuse distance, i.e., the distance between cells using the same frequencies. In other words, sufficient insulation in

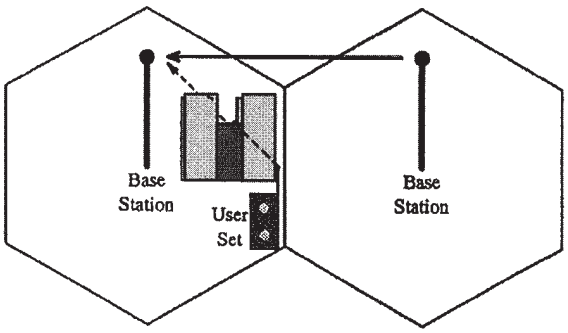


FIGURE 17.2 High base station antennas lead to stronger propagation paths between base stations than between a user set and its base stations.

the spatial domain must be provided to facilitate the separation of signals. Note that these comments apply equally to cochannel and adjacent channel interference.

This problem does not arise in cordless applications. Base station antennas are generally of the same height as user sets. Hence, interference created by base stations is subject to the same propagation conditions as signals from user sets. Furthermore, in cordless telephone applications there are frequently attenuating obstacles, such as walls, between base stations that reduce intracell interference further. Note that this reduction is vital for the proper functioning of cordless telephones since there is typically no network planning associated with installing a cordless telephone. As a safety feature, to overcome intercell interference, adaptive channel management strategies based on sensing interference levels can be employed.

17.3.3 Example 17.1: CT2

The CT2 standard was originally adopted in 1987 in Great Britain and improved with a common air interface (CAI) in 1989. The CAI facilitates interoperability between equipment from different vendors whereas the original standard only guarantees noninterference. The CT2 standard is used in home and office cordless telephone equipment and has been used for telepoint applications [5].

CT2 operates in the frequency band 864–868 MHz and uses carriers spaced at 100 kHz. FDMA with time division duplexing is employed. The combined gross bit rate is 72 kb/s, transmitted in frames of 2-ms duration of which the first-half carries downlink and the second-half carries uplink information. This setup supports a net bit rate of 32 kb/s of user data (32-kb/s ADPCM encoded speech) and 2-kb/s control information in each direction. The CT2 modulation technique is binary frequency shift keying.

17.3.4 Further Remarks

From the preceding discussion it is obvious that FDMA is a good candidate for applications like cordless telephone. In particular, the simple signal processing makes it a good choice for inexpensive implementation in the benign cordless environment. The possibility of concentration of signal processing functions in the base station strengthens this aspect.

In the cellular application, on the other hand, FDMA is inappropriate because of the lack of built-in diversity and the potential for severe intercell interference between base stations. A further complication arises from the difficulty of performing handovers if base-stations are not tightly synchronized.

For PCS the decision is not as obvious. Depending on whether the envisioned PCS application resembles more a cordless private branch exchange (PBX) than a cellular system, FDMA may be an appropriate choice. We will see later that it is probably better to opt for a combined TDMA/FDMA or a CDMA-based system to avoid the pitfalls of pure FDMA systems and still achieve moderate equipment complexities.

Finally, there is the problem of channel assignment. Clearly, it is not reasonable to assign a unique frequency to each user as there are not sufficient frequencies and the spectral resource would be unused whenever the user is idle. Instead, methods that allocate channels on demand can make much more efficient use of the spectrum. Such methods will be discussed further during the description of TDMA systems.

17.4 Time Division Multiple Access

In TDMA systems users share the same frequency band by accessing the channel in non-overlapping time intervals in a round-robin fashion [3]. Since the signals do not overlap, they are clearly orthogonal, and the signal of interest is easily extracted by switching the receiver on only during the transmission of the desired signal. Hence, the receiver filters are simply windows instead of the bandpass filters required in FDMA. As a consequence, the guard time between transmissions can be made as small as the synchronization of the network permits. Guard times of 30–50 μs between time slots are commonly used in TDMA-based systems. As a consequence, all users must be synchronized with the base station to within a fraction of the guard time. This is achievable by distributing a master clock signal on one of the base station's broadcast channels.

TDMA can be combined with TDD or frequency-division duplexing (FDD). The former duplexing scheme is used, for example, in the Digital European Cordless Telephone (DECT) standard and is well suited for systems in which base-to-base and mobile-to-base propagation paths are similar, i.e., systems without extremely high base station antennas. Since both the portable and the base station transmit on the same frequency, some signal processing functions for the downlink can be implemented in the base station, as discussed earlier for FDMA/TDD systems.

In the cellular application, the high base station antennas make FDD the more appropriate choice. In these systems, separate frequency bands are provided for uplink and downlink communication. Note that it is still possible and advisable to stagger the uplink and downlink transmission intervals such that they do not overlap, to avoid the situation that the portable must transmit and receive at the same time. With FDD the uplink and downlink channel are not identical and, hence, signal processing functions can not be implemented in the base-station; antenna diversity and equalization have to be realized in the portable.

17.4.1 Propagation Considerations

In comparison to a FDMA system supporting the same user data rate, the transmitted data rate in a TDMA system is larger by a factor equal to the number of users sharing the frequency band. This factor is eight in the pan-European global system for mobile communications (GSM) and three in the advanced mobile phone service (D-AMPS) system. Thus, the symbol duration is reduced by the same factor and severe intersymbol interference results, at least in the cellular environment.

To illustrate, consider the earlier example where each user transmits 25 K symbols per second. Assuming eight users per frequency band leads to a symbol duration of 5 μs. Even in the cordless application with delay spreads of up to 1 μs, an equalizer may be useful to

combat the resulting interference between adjacent symbols. In cellular systems, however, the delay spread of up to 20 μs introduces severe intersymbol interference spanning up to 5 symbol periods. As the delay spread often exceeds the symbol duration, the channel can be classified as frequency selective, emphasizing the observation that the channel affects different spectral components differently.

The intersymbol interference in cellular TDMA systems can be so severe that linear equalizers are insufficient to overcome its negative effects. Instead, more powerful, nonlinear decision feedback or maximum-likelihood sequence estimation equalizers must be employed [9]. Furthermore, all of these equalizers require some information about the channel impulse response that must be estimated from the received signal by means of an embedded training sequence. Clearly, the training sequence carries no user data and, thus, wastes valuable bandwidth.

In general, receivers for cellular TDMA systems will be fairly complex. On the positive side of the argument, however, the frequency selective nature of the channel provides some built-in diversity that makes transmission more robust to channel fading. The diversity stems from the fact that the multipath components of the received signal can be resolved at a resolution roughly equal to the symbol duration, and the different multipath components can be combined by the equalizer during the demodulation of the signal. To further improve robustness to channel fading, coding and interleaving, slow frequency hopping and antenna diversity can be employed as discussed in connection with FDMA.

17.4.2 Initial Channel Assignment

In both FDMA and TDMA systems, channels should not be assigned to a mobile on a permanent basis. A fixed assignment strategy would either be extremely wasteful of precious bandwidth or highly susceptible to cochannel interference. Instead, channels must be assigned on demand. Clearly, this implies the existence of a separate uplink channel on which mobiles can notify the base station of their need for a traffic channel. This uplink channel is referred to as the **random-access channel** because of the type of strategy used to regulate access to it.

The successful procedure for establishing a call that originates from the mobile station is outlined in Fig. 17.3. The mobile initiates the procedure by transmitting a request on the random-access channel. Since this channel is shared by all users in range of the base station, a random access protocol, like the ALOHA protocol, has to be employed to resolve possible collisions. Once the base station has received the mobile's request, it responds with an immediate assignment message that directs the mobile to tune to a dedicated control channel for the ensuing call setup. Upon completion of the call setup negotiation, a traffic channel, i.e., a frequency in FDMA systems or a time slot in TDMA systems, is assigned by the base station and all future communication takes place on that channel. In the case of a mobile-terminating call request, the sequence of events is preceded by a paging message alerting the base station of the call request.

17.4.3 Example 17.2: GSM

Named after the organization that created the system standards (Groupe Speciale Mobile) this pan-European digital cellular system has been deployed in Europe since the early 1990s [6]. GSM uses combined TDMA and FDMA with frequency-division duplex for access. Carriers are spaced at 200 kHz and support eight TDMA time slots each. For the uplink the frequency band 890–915 MHz is allocated, whereas the downlink uses the band 935–960 MHz. Each time slot is of duration 577 μs, which corresponds to 156.26-b periods,

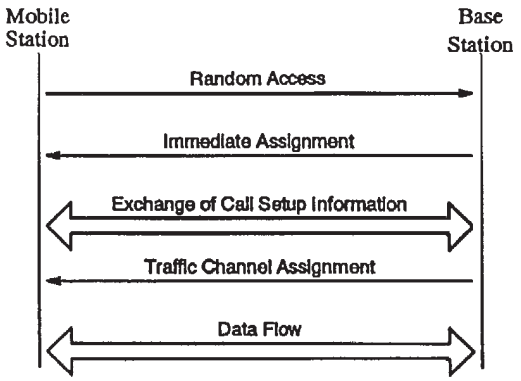


FIGURE 17.3 Mobile-originating call establishment.

including a guard time of 8.25-b periods. Eight consecutive time slots form a GSM frame of duration 4.62 ms.

The GSM modulation is Gaussian minimum shift keying with time-bandwidth product of 0.3, i.e., the modulator bandpass has a cutoff frequency of 0.3 times the bit rate. At the bit rate of 270.8 kb/s, severe intersymbol interference arises in the cellular environment. To facilitate coherent detection, a 26-b training sequence is embedded into every time slot. Time diversity is achieved by interleaving over 8 frames for speech signals and 20 frames for data communication. Sophisticated error-correction coding with varying levels of protection for different outputs of the speech coder is provided. Note that the round-trip delay introduced by the interleaver is on the order of 80 ms for speech signals. GSM provides slow frequency hopping as a further mechanism to improve the efficiency of the interleaver.

17.4.4 Further Remarks

In cellular systems, such as GSM or the North-American D-AMPS, TDMA is combined with FDMA. Different frequencies are used in neighboring cells to provide orthogonal signalling without the need for tight synchronization of base stations. Furthermore, channel assignment can then be performed in each cell individually. Within a cell, one or more frequencies are shared by users in the time domain.

From an implementation standpoint TDMA systems have the advantage that common radio and signal processing equipment at the base station can be shared by users communicating on the same frequency. A somewhat more subtle advantage of TDMA systems arises from the possibility of monitoring surrounding base stations and frequencies for signal quality to support mobile assisted handovers.

17.5 Code Division Multiple Access

CDMA systems employ wideband signals with good cross-correlation properties [7]. That means the output of a filter matched to one user's signal is small when a different user's signal is input. A large body of work exists on spreading sequences that lead to signal sets with small cross correlations [10]. Because of their noise-like appearance such sequences are often referred to as pseudonoise (PN) sequences, and because of their wideband nature CDMA systems are often called spread-spectrum systems.

Spectrum spreading can be achieved mainly in two ways: through frequency hopping as

explained earlier or through direct sequence spreading. In direct sequence spread spectrum, a high-rate, antipodal pseudorandom spreading sequence modulates the transmitted signal such that the bandwidth of the resulting signal is roughly equal to the rate of the spreading sequence. The cross correlation of the signals is then largely determined by the cross-correlation properties of the spreading signals. Clearly, CDMA signals overlap in both time and frequency domains but are separable based on their spreading waveforms.

An immediate consequence of this observation is that CDMA systems do not require tight synchronization between users as do TDMA systems. By the same token, frequency planning and management are not required as frequencies are reused throughout the coverage area.

17.5.1 Propagation Considerations

Spread spectrum is well suited for wireless communication systems because of its built-in frequency diversity. As discussed before, in cellular systems the delay spread measures several microseconds and, hence, the coherence bandwidth of the channel is smaller than 1 MHz. Spreading rates can be chosen to exceed the coherence bandwidth such that the channel becomes frequency selective, i.e., different spectral components are affected unequally by the channel and only parts of the signal are affected by fades. Expressing the same observation in time domain terms, multipath components are resolvable at a resolution equal to the chip period and can be combined coherently, for example, by means of a RAKE receiver [8]. An estimate of the channel impulse response is required for the coherent combination of multipath components. This estimate can be gained from a training sequence or by means of a so-called pilot signal.

Even for cordless telephone systems, operating in environments with submicrosecond delay spread and corresponding coherence bandwidths of a few megahertz, the spreading rate can be chosen large enough to facilitate multipath diversity. If the combination of multipath components already described is deemed too complex, a simpler, but less powerful, form of diversity can be used that decorrelates only the strongest received multipath component and relies on the suppression of other path components by the matched filter.

17.5.2 Multiple-Access Interference

If it is possible to control the relative timing of the transmitted signals, such as on the down-link, the transmitted signals can be made perfectly orthogonal, and if the channel only adds white Gaussian noise, matched filter receivers are optimal for extracting a signal from the superposition of waveforms. If the channel is dispersive because of multipath, the signals arriving at the receiver will no longer be orthogonal and will introduce some multiple-access interference, i.e., signal components from other signals that are not rejected by the matched filter.

On the uplink, extremely tight synchronization between users to within a fraction of a chip period, which is defined as the inverse of the spreading rate, is generally not possible, and measures to control the impact of multiple-access interference must be taken. Otherwise, the near–far problem, i.e., the problem of very strong undesired users' signals overwhelming the weaker signal of the desired user, can severely decrease performance. Two approaches are proposed to overcome the near–far problem: power control with soft handovers and multiuser detection.

Power control attempts to ensure that signals from all mobiles in a cell arrive at the base station with approximately equal power levels. To be effective, power control must be accurate to within about 1 dB and fast enough to compensate for channel fading. For a

mobile moving at 55 mph and transmitting at 1 GHz, the Doppler bandwidth is approximately 100 Hz. Hence, the channel changes its characteristic drastically about 100 times per second and on the order of 1000 b/s must be sent from base station to mobile for power control purposes. As different mobiles may be subject to vastly different fading and shadowing conditions, a large dynamic range of about 80 dB must be covered by power control. Notice, that power control on the downlink is really only necessary for mobiles that are about equidistant from two base stations, and even then neither the update rate nor the dynamic range of the uplink is required.

The interference problem that arises at the cell boundaries where mobiles are within range of two or more base stations can be turned into an advantage through the idea of soft handover. On the downlink, all base stations within range can transmit to the mobile, which in turn can combine the received signals to achieve some gain from the antenna diversity. On the uplink, a similar effect can be obtained by selecting the strongest received signal from all base stations that received a user's signal. The base station that receives the strongest signal will also issue power control commands to minimize the transmit power of the mobile. Note, however, that soft handover requires fairly tight synchronization between base stations, and one of the advantages of CDMA over TDMA is lost.

Multiuser detection is still an emerging technique. It is probably best used in conjunction with power control. The fundamental idea behind this technique is to model multiple-access interference explicitly and devise receivers that reject or cancel the undesired signals. A variety of techniques have been proposed ranging from optimum maximum-likelihood sequence estimation via multistage schemes, reminiscent of decision feedback algorithms, to linear decorrelating receivers. An excellent survey of the theory and practice of multiuser detection is given by [11].

17.5.3 Further Remarks

CDMA systems work well in conjunction with frequency division duplexing. This arrangement decouples the power control problem on the uplink and downlink, respectively.

Signal quality enhancing methods, such as time diversity through coding and interleaving, can be applied just as with the other access methods. In spread spectrum systems, however, coding can be built into the spreading process, avoiding the loss of bandwidth associated with error protection. Additionally, CDMA lends itself naturally to the exploitation of speech pauses that make up more than half the time of a connection. If no signals are transmitted during such pauses, then the instantaneous interference level is reduced and the total number of users supportable by the system can be approximately doubled.

17.6 Comparison and Outlook

The question of which of the access methods is best does not have a single answer. Based on the preceding discussion FDMA is only suited for applications such as cordless telephone with very small cells and submicrosecond delay spreads. In cellular systems and for most versions of personal communication systems, the choice reduces to TDMA vs. CDMA.

In terms of complexity, TDMA receivers require adaptive, nonlinear equalizers when operating in environments with large delay spreads. CDMA systems, in turn, need RAKE receivers and sophisticated power control algorithms. In the future, some form of multiple-access interference rejection is likely to be implemented as well. Time synchronization is required in both systems, albeit for different reasons. The additional complexity for coding and interleaving is comparable for both access methods.

An often quoted advantage of CDMA systems is the fact that the performance will degrade gracefully as the load increases. In TDMA systems, in turn, requests will have to be blocked once all channels in a cell are in use. Hence, there is a hard limit on the number of channels per cell. There are proposals for extended TDMA systems, however, that incorporate reassignment of channels during speech pauses. Not only would such extended TDMA systems match the advantage of the exploitation of speech pauses of CDMA systems, they would also lead to a soft limit on the system capacity. The extended TDMA proposals would implement the statistical multiplexing of the user data, e.g., by means of the packet reservation multiple access protocol [4]. The increase in capacity depends on the acceptable packet loss rate; in other words, small increases in the load lead to small increases in the packet loss probability.

Many comparisons in terms of capacity between TDMA and CDMA can be found in the recent literature. Such comparisons, however, are often invalidated by making assumptions that favor one access method over the other. An important exception constitutes the recent paper by Wyner [12]. Under a simplified model that nevertheless captures the essence of cellular systems, he computes the Shannon capacity. Highlights of his results include the following.

- TDMA is distinctly suboptimal in cellular systems.
- When the signal-to-noise-ratio is large, CDMA appears to achieve twice the capacity of TDMA.
- Multiuser detectors are essential to realize near-optimum performance in CDMA systems.
- Intercell interference in CDMA systems has a detrimental effect when the signal-to-noise ratio is large, but it can be exploited via diversity combining to increase capacity when the signal-to-noise ratio is small.

More research along this avenue is necessary to confirm the validity of the results. In particular, incorporation of realistic channel models into the analysis is required. However, this work represents a substantial step towards quantifying capacity increases achievable with CDMA.

Defining Terms

Capacity: Shannon originally defined capacity as the maximum data rate which permits error-free communication in a given environment. A looser interpretation is normally employed in wireless communication systems. Here capacity denotes the traffic density supported by the system under consideration normalized with respect to bandwidth and coverage area.

Code-division multiple access (CDMA): Systems use signals with very small cross-correlations to facilitate sharing of the broadcast radio channel. Correlators are used to extract the desired user's signal while simultaneously suppressing interfering, parallel transmissions.

Duplexing: Refers to the exchange of messages in both directions of a connection.

Frequency-division multiple access (FDMA): Simultaneous access to the radio channel is facilitated by assigning nonoverlapping frequency bands to different users.

Multiple access: Denotes the support of simultaneous transmissions over a shared communication channel.

Random-access channel: This uplink control channel is used by mobiles to request assignment of a traffic channel. A random access protocol is employed to arbitrate access to this channel.

Time-division multiple access (TDMA): Systems assign nonoverlapping time slots to different users in a round-robin fashion.

References

[1] Bertsekas, D. and Gallager, R., *Data Networks,* Prentice-Hall, Englewood Cliffs, NJ, 1987.

[2] Cox, D.C., Wireless network access for personal communications, *IEEE Comm. Mag.,* 96–115, 1992.

[3] Falconer, D.D., Adachi, F., and Gudmundson, B. Time division multiple access methods for wireless personal communications. *IEEE Comm. Mag.,* 33(1), 50–57, 1995.

[4] Goodman, D., Trends in cellular and cordless communications. *IEEE Comm. Mag.,* 31–40, 1991a.

[5] Goodman, D.J., Second generation wireless information networks. *IEEE Trans. on Vehicular Tech.,* 40(2), 366–374, 1991b.

[6] Hodges, M.R.L., The GSM radio interface. *Br. Telecom Tech. J.,* 8(1), 31–43, 1990.

[7] Kohno, R., Meidan, R., and Milstein, L.B., Spread spectrum access methods for wireless communications. *IEEE Comm. Mag.,* 33(1), 58, 1995.

[8] Proakis, J.G., *Digital Communications.* 2nd ed., McGraw-Hill, New York, 1989.

[9] Proakis, J.G., Adaptive equalization for TDMA digital mobile radio. *IEEE Trans. on Vehicular Tech.,* 40(2), 333–341, 1991.

[10] Sarwate, D.V. and Pursley, M.B., Crosscorrelation properties of pseudorandom and related sequences. *Proceedings of the IEEE,* 68(5), 593–619, 1980.

[11] Verdu, S., Multi-user detection. In *Advances in Statistical Signal Processing—Vol. 2: Signal Detection,* JAI Press, Greenwich, CT, 1992.

[12] Wyner, A.D., Shannon-theoretic approach to a Gaussian cellular multiple-access channel. *IEEE Trans. on Information Theory,* 40(6), 1713–1727, 1994.

Further Information

Several of the IEEE publications, including the *Transactions on Communications, Journal on Selected Areas in Communications, Transactions on Vehicular Technology, Communications Magazine,* and *Personal Communications Magazine* contain articles the on subject of access methods on a regular basis.

18

Rayleigh Fading Channels[1]

18.1 Introduction 18-2
18.2 The Challenge of a Fading Channel 18-2
18.3 Mobile-Radio Propagation: Large-Scale Fading and Small-Scale Fading 18-3
Large-Scale Fading: Path-Loss Mean and Standard Deviation • Small-Scale Fading: Statistics and Mechanisms
18.4 Signal Time-Spreading Viewed in the Time-Delay Domain: Figure 18.1, Block 7—The Multipath Intensity Profile 18-10
Degradation Categories due to Signal Time-Spreading Viewed in the Time Delay Domain
18.5 Signal Time-Spreading Viewed in the Frequency Domain: Figure 18.1, Block 10—The Spaced-Frequency Correlation Function 18-13
Degradation Categories due to Signal Time-Spreading Viewed in the Frequency Domain
18.6 Typical Examples of Flat Fading and Frequency-Selective Fading Manifestations 18-16
18.7 Time Variance Viewed in the Time Domain: Figure 18.1, Block 13—The Spaced-Time Correlation Function .. 18-17
The Concept of Duality • Degradation Categories due to Time Variance Viewed in the Time Domain
18.8 Time Variance Viewed in the Doppler-Shift Domain: Figure 18.1, Block 16—The Doppler Power Spectrum 18-19
18.9 Analogy Between Spectral Broadening in Fading Channels and Spectral Broadening in Digital Signal Keying 18-21
18.10 Degradation Categories due to Time Variance, Viewed in the Doppler-Shift Domain 18-23
18.11 Mitigation Methods 18-24
Mitigation to Combat Frequency-Selective Distortion • Mitigation to Combat Fast-Fading Distortion • Mitigation to Combat Loss in SNR
18.12 Summary of the Key Parameters Characterizing Fading Channels 18-30
Fast-Fading Distortion: Example #1 • Frequency-Selective Fading Distortion: Example #2 • Fast-Fading and Frequency-Selective Fading Distortion: Example #3
18.13 The Viterbi Equalizer as Applied to GSM 18-32
18.14 The Rake Receiver Applied to Direct-Sequence Spread-Spectrum (DS/SS) Systems 18-34
18.15 Conclusion 18-35
References 18-37

Bernard Sklar
Communications Engineering Services

0-8493-8597-0/99/$0.00+$.50

© 1999 by CRC Press LLC

18.1 Introduction

When the mechanisms of fading channels were first modeled in the 1950s and 1960s, the ideas were primarily applied to over-the-horizon communications covering a wide range of frequency bands. The 3–30 MHz high-frequency (HF) band is used for ionospheric communications, and the 300 MHz–3 GHz ultra-high-frequency (UHF) and 3–30 GHz super-high-frequency (SHF) bands are used for tropospheric scatter. Although the fading effects in a mobile radio system are somewhat different from those in ionospheric and tropospheric channels, the early models are still quite useful to help characterize fading effects in mobile digital communication systems. This chapter addresses Rayleigh fading, primarily in the UHF band, that affects mobile systems such as cellular and personal communication systems (PCS). The chapter itemizes the fundamental fading manifestations, types of degradation, and methods to mitigate the degradation. Two particular mitigation techniques are examined: the Viterbi equalizer implemented in the Global System for Mobile Communication (GSM), and the Rake receiver used in CDMA systems built to meet Interim Standard-95 (IS-95).

18.2 The Challenge of a Fading Channel

In the study of communication systems, the classical (ideal) additive-white-Gaussian-noise (AWGN) channel, with statistically independent Gaussian noise samples corrupting data samples free of intersymbol interference (ISI), is the usual starting point for understanding basic performance relationships. The primary source of performance degradation is thermal noise generated in the receiver. Often, external interference received by the antenna is more significant than the thermal noise. This external interference can sometimes be characterized as having a broadband spectrum and quantified by a parameter called antenna temperature [1]. The thermal noise usually has a flat power spectral density over the signal band and a zero-mean Gaussian voltage probability density function (pdf). When modeling practical systems, the next step is the introduction of bandlimiting filters. The filter in the transmitter usually serves to satisfy some regulatory requirement on spectral containment. The filter in the receiver often serves the purpose of a classical "matched filter" [2] to the signal bandwidth. Due to the bandlimiting and phase-distortion properties of filters, special signal design and equalization techniques may be required to mitigate the filter-induced ISI.

If a radio channel's propagating characteristics are not specified, one usually infers that the signal attenuation vs. distance behaves as if propagation takes place over ideal free space. The model of free space treats the region between the transmit and receive antennas as being free of all objects that might absorb or reflect radio frequency (RF) energy. It also assumes that, within this region, the atmosphere behaves as a perfectly uniform and nonabsorbing medium. Furthermore, the earth is treated as being infinitely far away from the propagating signal (or, equivalently, as having a reflection coefficient that is negligible). Basically, in this idealized free-space model, the attenuation of RF energy between the transmitter and receiver behaves according to an inverse-square law. The received power expressed in terms of transmitted power is attenuated by a factor, $L_s(d)$, where this factor is called **path loss** or **free space loss.** When the receiving antenna is isotropic, this factor

[1] A version of this chapter has appeared as two papers in the *IEEE Communications Magazine*, September 1997, under the titles "Rayleigh Fading Channels in Mobile Digital Communication Systems, Part I: Characterization" and "Part II: Mitigation."

is expressed as [1]:

$$L_s(d) = \left(\frac{4\pi d}{\lambda}\right)^2 \tag{18.1}$$

In Eq. (18.1), d is the distance between the transmitter and the receiver, and λ is the wavelength of the propagating signal. For this case of idealized propagation, received signal power is very predictable.

For most practical channels, where signal propagation takes place in the atmosphere and near the ground, the free-space propagation model is inadequate to describe the channel and predict system performance. In a wireless mobile communication system, a signal can travel from transmitter to receiver over multiple reflective paths; this phenomenon is referred to as **multipath propagation.** The effect can cause fluctuations in the received signal's amplitude, phase, and angle of arrival, giving rise to the terminology **multipath fading.** Another name, **scintillation,** having originated in radio astronomy, is used to describe the multipath fading caused by physical changes in the propagating medium, such as variations in the density of ions in the ionospheric layers that reflect high frequency (HF) radio signals. Both names, fading and scintillation, refer to a signal's random fluctuations or fading due to multipath propagation. The main difference is that scintillation involves mechanisms (e.g., ions) that are much smaller than a wavelength. The end-to-end modeling and design of systems that mitigate the effects of fading are usually more challenging than those whose sole source of performance degradation is AWGN.

18.3 Mobile-Radio Propagation: Large-Scale Fading and Small-Scale Fading

Figure 18.1 represents an overview of fading channel manifestations. It starts with two types of fading effects that characterize mobile communications: large-scale fading and small-scale fading. Large-scale fading represents the average signal power attenuation or the path loss due to motion over large areas. In Fig. 18.1, the large-scale fading manifestation is shown in blocks 1, 2, and 3. This phenomenon is affected by prominent terrain contours (e.g., hills, forests, billboards, clumps of buildings, etc.) between the transmitter and receiver. The receiver is often represented as being "shadowed" by such prominences. The statistics of large-scale fading provide a way of computing an estimate of path loss as a function of distance. This is described in terms of a mean-path loss (nth-power law) and a log-normally distributed variation about the mean. Small-scale fading refers to the dramatic changes in signal amplitude and phase that can be experienced as a result of small changes (as small as a half-wavelength) in the spatial separation between a receiver and transmitter. As indicated in Fig. 18.1, blocks 4, 5, and 6, small-scale fading manifests itself in two mechanisms, namely, time-spreading of the signal (or signal dispersion) and time-variant behavior of the channel. For mobile-radio applications, the channel is time-variant because motion between the transmitter and receiver results in propagation path changes. The rate of change of these propagation conditions accounts for the fading rapidity (rate of change of the fading impairments). Small-scale fading is also called **Rayleigh fading** because if the multiple reflective paths are large in number and there is no line-of-sight signal component, the envelope of the received signal is statistically described by a Rayleigh pdf. When there is a dominant nonfading signal component present, such as a line-of-sight propagation path, the small-scale fading envelope is described by a Rician pdf [3]. A mobile radio roaming over a large area must process signals that experience both types of fading: small-scale fading superimposed on large-scale fading.

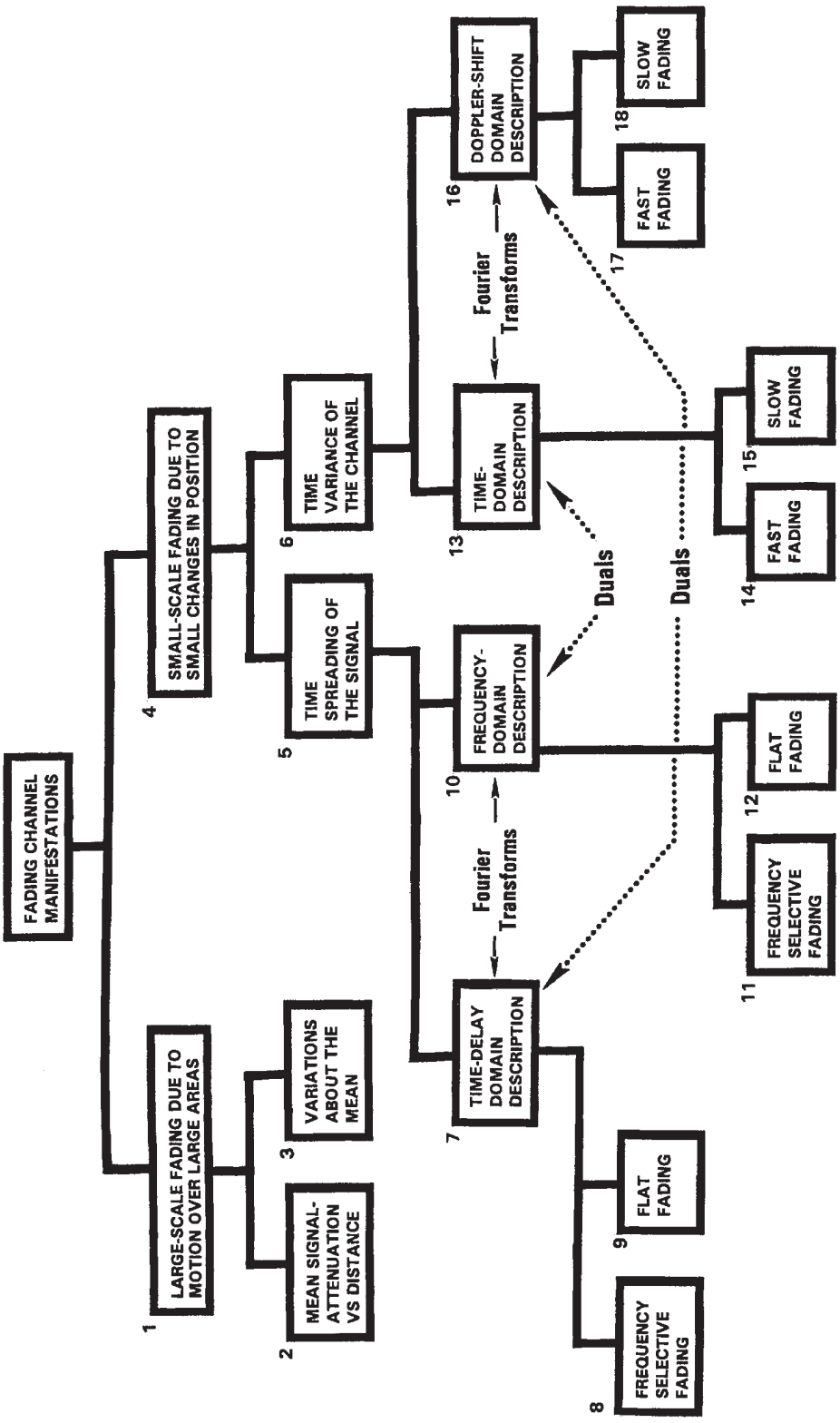


FIGURE 18.1 Fading channel manifestations.

There are three basic mechanisms that impact signal propagation in a mobile communication system. They are reflection, diffraction, and scattering [3].

- Reflection occurs when a propagating electromagnetic wave impinges upon a smooth surface with very large dimensions compared to the RF signal wavelength (λ).
- Diffraction occurs when the radio path between the transmitter and receiver is obstructed by a dense body with large dimensions compared to λ, causing secondary waves to be formed behind the obstructing body. Diffraction is a phenomenon that accounts for RF energy travelling from transmitter to receiver without a line-of-sight path between the two. It is often termed **shadowing** because the diffracted field can reach the receiver even when shadowed by an impenetrable obstruction.
- Scattering occurs when a radio wave impinges on either a large rough surface or any surface whose dimensions are on the order of λ or less, causing the reflected energy to spread out (scatter) in all directions. In an urban environment, typical signal obstructions that yield scattering are lampposts, street signs, and foliage.

Figure 18.1 may serve as a table of contents for the sections that follow. We will examine the two manifestations of small-scale fading: signal time-spreading (signal dispersion) and the time-variant nature of the channel. These examinations will take place in two domains: time and frequency, as indicated in Fig. 18.1, blocks 7, 10, 13, and 16. For signal dispersion, we categorize the fading degradation types as being frequency-selective or frequency-nonselective (flat), as listed in blocks 8, 9, 11, and 12. For the time-variant manifestation, we categorize the fading degradation types as fast-fading or slow-fading, as listed in blocks 14, 15, 17, and 18. The labels indicating Fourier transforms and duals will be explained later.

Figure 18.2 illustrates the various contributions that must be considered when estimating path loss for a link budget analysis in a cellular application [4]. These contributions are:

- Mean path loss as a function of distance, due to large-scale fading
- Near-worst-case variations about the mean path loss (typically 6–10 dB) or large-scale fading margin
- Near-worst-case Rayleigh or small-scale fading margin (typically 20–30 dB)

In Fig. 18.2, the annotations " $\approx$ 1–2% " indicate a suggested area (probability) under the tail of each pdf as a design goal. Hence, the amount of margin indicated is intended to provide adequate received signal power for approximately 98–99% of each type of fading variation (large- and small-scale).

A received signal, is generally described in terms of a transmitted signal $s(t)$ convolved with the impulse response of the channel $h_c(t)$. Neglecting the degradation due to noise, we write:

$$r(t) = s(t) * h_c(t) \tag{18.2}$$

where * denotes convolution. In the case of mobile radios, $r(t)$ can be partitioned in terms of two component random variables, as follows [5]:

$$r(t) = m(t) \times r_0(t) \tag{18.3}$$

where $m(t)$ is called the large-scale-fading component, and $r_0(t)$ is called the small-scale-fading component. $m(t)$ is sometimes referred to as the **local mean** or **log-normal fading**

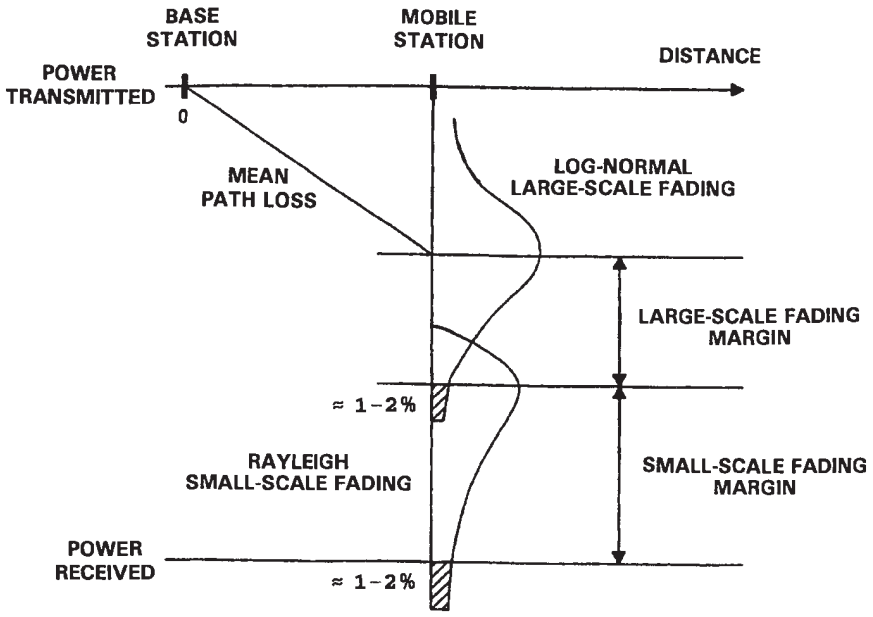


FIGURE 18.2 Link-budget considerations for a fading channel.

because the magnitude of $m(t)$ is described by a log-normal pdf (or, equivalently, the magnitude measured in decibels has a Gaussian pdf). $r_0(t)$ is sometimes referred to as multipath or Rayleigh fading. Figure 18.3 illustrates the relationship between large-scale and small-scale fading. In Fig. 18.3(a), received signal power $r(t)$ vs. antenna displacement (typically in units of wavelength) is plotted for the case of a mobile radio. Small-scale fading superimposed on large-scale fading can be readily identified. The typical antenna displacement between the small-scale signal nulls is approximately a half wavelength. In Fig. 18.3(b), the large-scale fading or local mean, $m(t)$, has been removed in order to view the small-scale fading, $r_0(t)$, about some average constant power.

In the sections that follow, we enumerate some of the details regarding the statistics and mechanisms of large-scale and small-scale fading.

18.3.1 Large-Scale Fading: Path-Loss Mean and Standard Deviation

For the mobile radio application, Okumura [6] made some of the earlier comprehensive path-loss measurements for a wide range of antenna heights and coverage distances. Hata [7] transformed Okumura's data into parametric formulas. For the mobile radio application, the mean path loss, $\overline{L_p}(d)$, as a function of distance, d, between the transmitter and receiver is proportional to an nth-power of d relative to a reference distance d_0 [3].

$$\overline{L_p}(d) \propto \left(\frac{d}{d_0}\right)^n \tag{18.4}$$

$\overline{L_p}(d)$ is often stated in decibels, as shown below.

$$\overline{L_p}(d)\ (\text{dB}) = L_s\,(d_0)\ (\text{dB}) + 10n \log\left(\frac{d}{d_0}\right) \tag{18.5}$$

The reference distance d_0, corresponds to a point located in the far field of the antenna. Typically, the value of d_0 is taken to be 1 km for large cells, 100 m for microcells, and 1 m

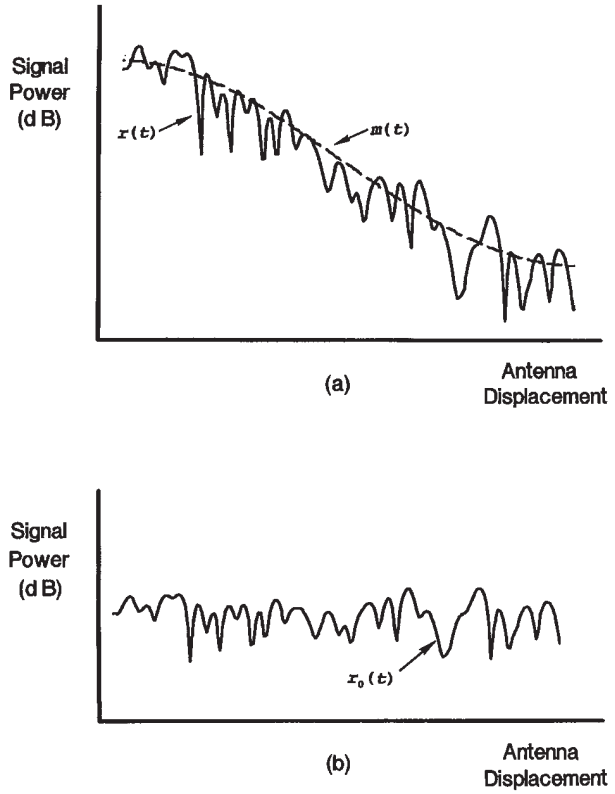


FIGURE 18.3 Large-scale fading and small-scale fading.

for indoor channels. $\overline{L_p}(d)$ is the average path loss (over a multitude of different sites) for a given value of d. Linear regression for a minimum mean-squared estimate (MMSE) fit of $\overline{L_p}(d)$ vs. d on a log-log scale (for distances greater than d_0) yields a straight line with a slope equal to 10 n dB/decade. The value of the exponent n depends on the frequency, antenna heights, and propagation environment. In free space, $n = 2$, as seen in Eq. (18.1). In the presence of a very strong guided wave phenomenon (like urban streets), n can be lower than 2. When obstructions are present, n is larger. The path loss $L_s(d_0)$ to the reference point at a distance d_0 from the transmitter is typically found through field measurements or is calculated using the free-space path loss given by Eq. (18.1). Figure 18.4 shows a scatter plot of path loss vs. distance for measurements made in several German cities [8]. Here, the path loss has been measured relative to the free-space reference measurement at $d_0 = 100$ m. Also shown are straight-line fits to various exponent values.

The path loss vs. distance expressed in Eq. (18.5) is an average, and therefore not adequate to describe any particular setting or signal path. It is necessary to provide for variations about the mean since the environment of different sites may be quite different for similar transmitter-receiver separations. Figure 18.4 illustrates that path-loss variations can be quite large. Measurements have shown that for any value of d, the path loss $L_p(d)$ is a random variable having a log-normal distribution about the mean distant-dependent value $\overline{L_p}(d)$ [9]. Thus, path loss $L_p(d)$ can be expressed in terms of $\overline{L_p}(d)$ plus a random variable X_σ, as follows [3].

$$L_p(d)\ (\text{dB}) = L_s\,(d_0)\ (\text{dB}) + 10\,n\,\log_{10}\left(\frac{d}{d_0}\right) + X_\sigma\ (\text{dB}) \tag{18.6}$$

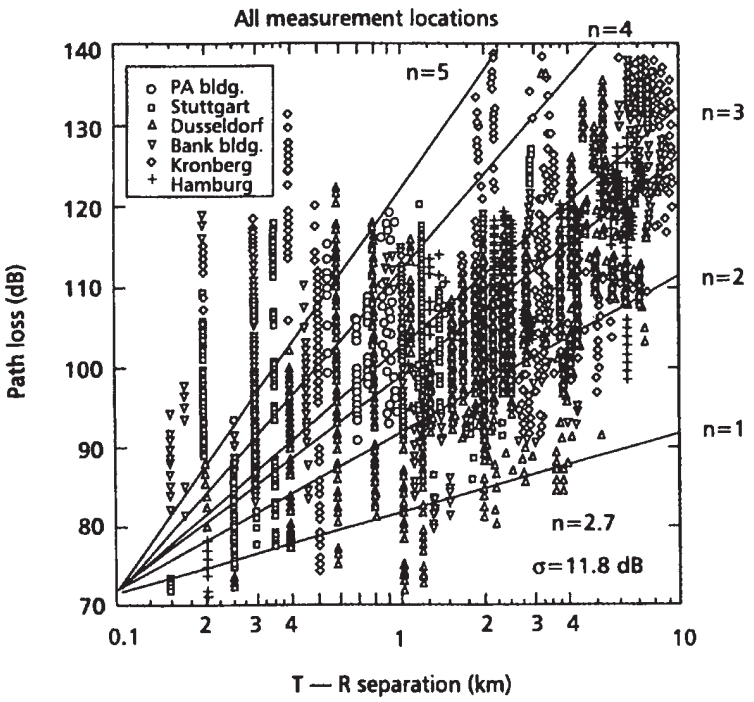


FIGURE 18.4 Path loss vs. distance measured in several German cities.

where X_σ denotes a zero-mean, Gaussian random variable (in decibels) with standard deviation σ (also in decibels). X_σ is site and distance dependent. The choice of a value for X_σ is often based on measurements; it is not unusual for X_σ to take on values as high as 6–10 dB or greater. Thus, the parameters needed to statistically describe path loss due to large-scale fading for an arbitrary location with a specific transmitter-receiver separation are:

- The reference distance d_0
- The path-loss exponent n
- The standard deviation σ of X_σ

There are several good references dealing with the measurement and estimation of propagation path loss for many different applications and configurations [3], [7]–[11].

18.3.2 Small-Scale Fading: Statistics and Mechanisms

When the received signal is made up of multiple reflective rays plus a significant line-of-sight (nonfaded) component, the envelope amplitude due to small-scale fading has a Rician pdf, and is referred to as **Rician fading** [3]. The nonfaded component is called the **specular component.** As the amplitude of the specular component approaches zero, the Rician pdf approaches a Rayleigh pdf, expressed as:

$$p(r) = \left\{ \begin{array}{ll} \dfrac{r}{\sigma^2} \exp\left[-\dfrac{r^2}{2\sigma^2}\right] & \text{for } r \geq 0 \\ 0 & \text{otherwise} \end{array} \right\} \tag{18.7}$$

where r is the envelope amplitude of the received signal, and $2\sigma^2$ is the predetection mean power of the multipath signal. The Rayleigh faded component is sometimes called the **random, scatter,** or **diffuse component.** The Rayleigh pdf results from having no specular component of the signal; thus for a single link it represents the pdf associated with the worst case of fading per mean received signal power. For the remainder of this chapter, it will be assumed that loss of signal-to-noise ratio (SNR) due to fading follows the Rayleigh model described. It will also be assumed that the propagating signal is in the UHF band, encompassing present-day cellular and personal communications services (PCS) frequency allocations—nominally 1 GHz and 2 GHz, respectively.

As indicated in Fig. 18.1, blocks 4, 5, and 6, small-scale fading manifests itself in two mechanisms:

- Time-spreading of the underlying digital pulses within the signal
- A time-variant behavior of the channel due to motion (e.g., a receive antenna on a moving platform).

Figure 18.5 illustrates the consequences of both manifestations by showing the response of a multipath channel to a narrow pulse vs. delay, as a function of antenna position (or time, assuming a constant velocity of motion). In Fig. 18.5, we distinguish between

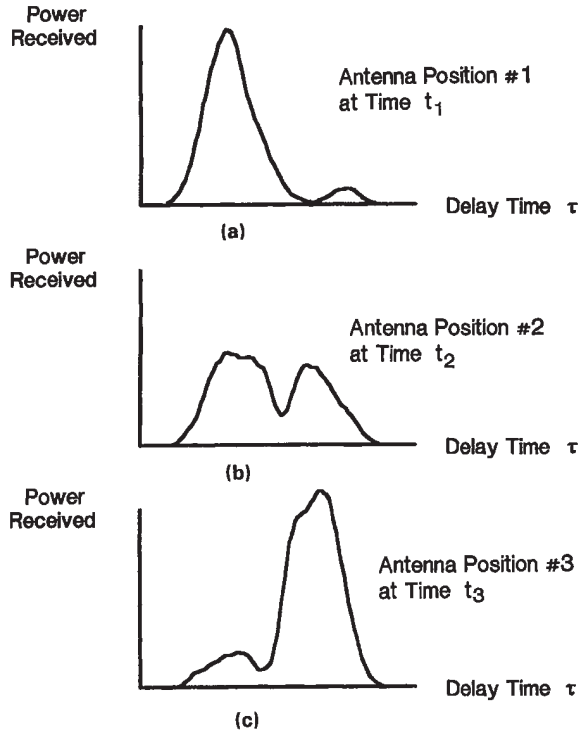


FIGURE 18.5 Response of a multipath channel to a narrow pulse vs. delay, as a function of antenna position.

two different time references—delay time τ and transmission or observation time t. Delay time refers to the time-spreading manifestation which results from the fading channel's nonoptimum impulse response. The transmission time, however, is related to the antenna's motion or spatial changes, accounting for propagation path changes that are perceived as

the channel's time-variant behavior. Note that, for constant velocity, as is assumed in Fig. 18.5, either antenna position or transmission time can be used to illustrate this time-variant behavior. Figures 18.5(a)–(c) show the sequence of received pulse-power profiles as the antenna moves through a succession of equally spaced positions. Here, the interval between antenna positions is 0.4 λ, where λ is the wavelength of the carrier frequency. For each of the three cases shown, the response-pattern differs significantly in the delay time of the largest signal component, the number of signal copies, their magnitudes, and the total received power (area) in the received power profile. Figure 18.6 summarizes these two small-scale fading mechanisms, the two domains (time or time-delay and frequency or Doppler shift) for viewing each mechanism and the degradation categories each mechanism can exhibit. Note that any mechanism characterized in the time domain can be characterized equally well in the frequency domain. Hence, as outlined in Fig. 18.6, the time-spreading mechanism will be characterized in the time-delay domain as a multipath delay spread and in the frequency domain as a channel coherence bandwidth. Similarly, the time-variant mechanism will be characterized in the time domain as a channel coherence time and in the Doppler-shift (frequency) domain as a channel fading rate or Doppler spread. These mechanisms and their associated degradation categories will be examined in greater detail in the sections that follow.

18.4 Signal Time-Spreading Viewed in the Time-Delay Domain: Figure 18.1, Block 7—The Multipath Intensity Profile

A simple way to model the fading phenomenon was introduced by Bello [13] in 1963; he proposed the notion of wide-sense stationary uncorrelated scattering (WSSUS). The model treats signal variations arriving with different delays as uncorrelated. It can be shown [4, 13] that such a channel is effectively WSS in both the time and frequency domains. With such a model of a fading channel, Bello was able to define functions that apply for all time and all frequencies. For the mobile channel, Fig. 18.7 contains four functions that make up this model [4], [13]–[16]. We will examine these functions, starting with Fig. 18.7(a) and proceeding counter-clockwise toward Fig. 18.7(d).

In Fig. 18.7(a), a **multipath-intensity profile,** $S(\tau)$ vs. time delay τ is plotted. Knowledge of $S(\tau)$ helps answer the question, "For a transmitted impulse, how does the average received power vary as a function of time delay, τ?" The term "time delay" is used to refer to the excess delay. It represents the signal's propagation delay that exceeds the delay of the first signal arrival at the receiver. For a typical wireless radio channel, the received signal usually consists of several discrete multipath components, sometimes referred to as fingers. For some channels, such as the tropospheric scatter channel, received signals are often seen as a continuum of multipath components [14, 16]. For making measurements of the multipath intensity profile, wideband signals (impulses or spread spectrum) need to be used [16]. For a single transmitted impulse, the time, T_m, between the first and last received component represents the **maximum excess delay,** during which the multipath signal power falls to some threshold level below that of the strongest component. The threshold level might be chosen at 10 dB or 20 dB below the level of the strongest component. Note, that for an ideal system (zero excess delay), the function $S(\tau)$ would consist of an ideal impulse with weight equal to the total average received signal power.

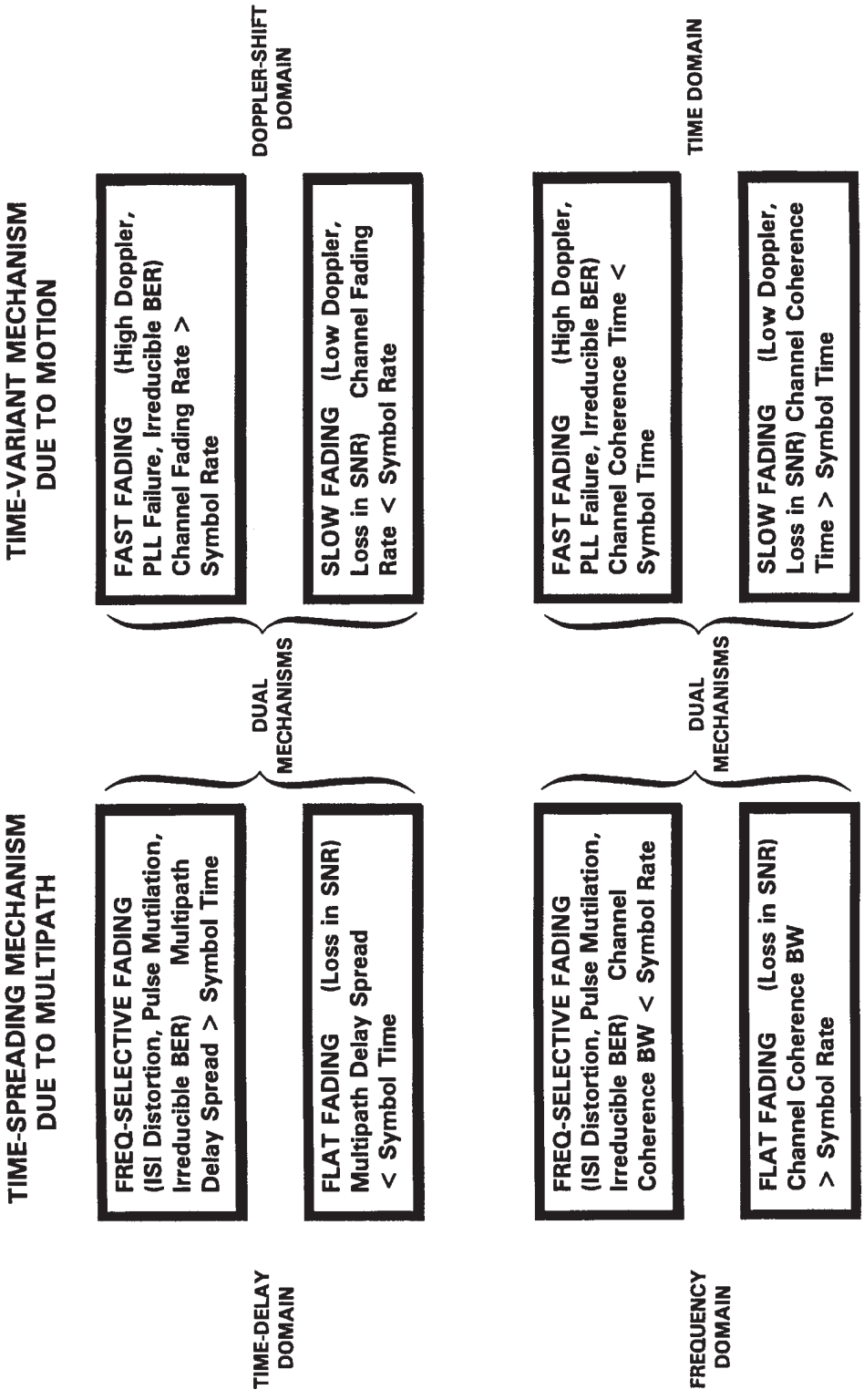


FIGURE 18.6 Small-scale fading: mechanisms, degradation categories, and effects.

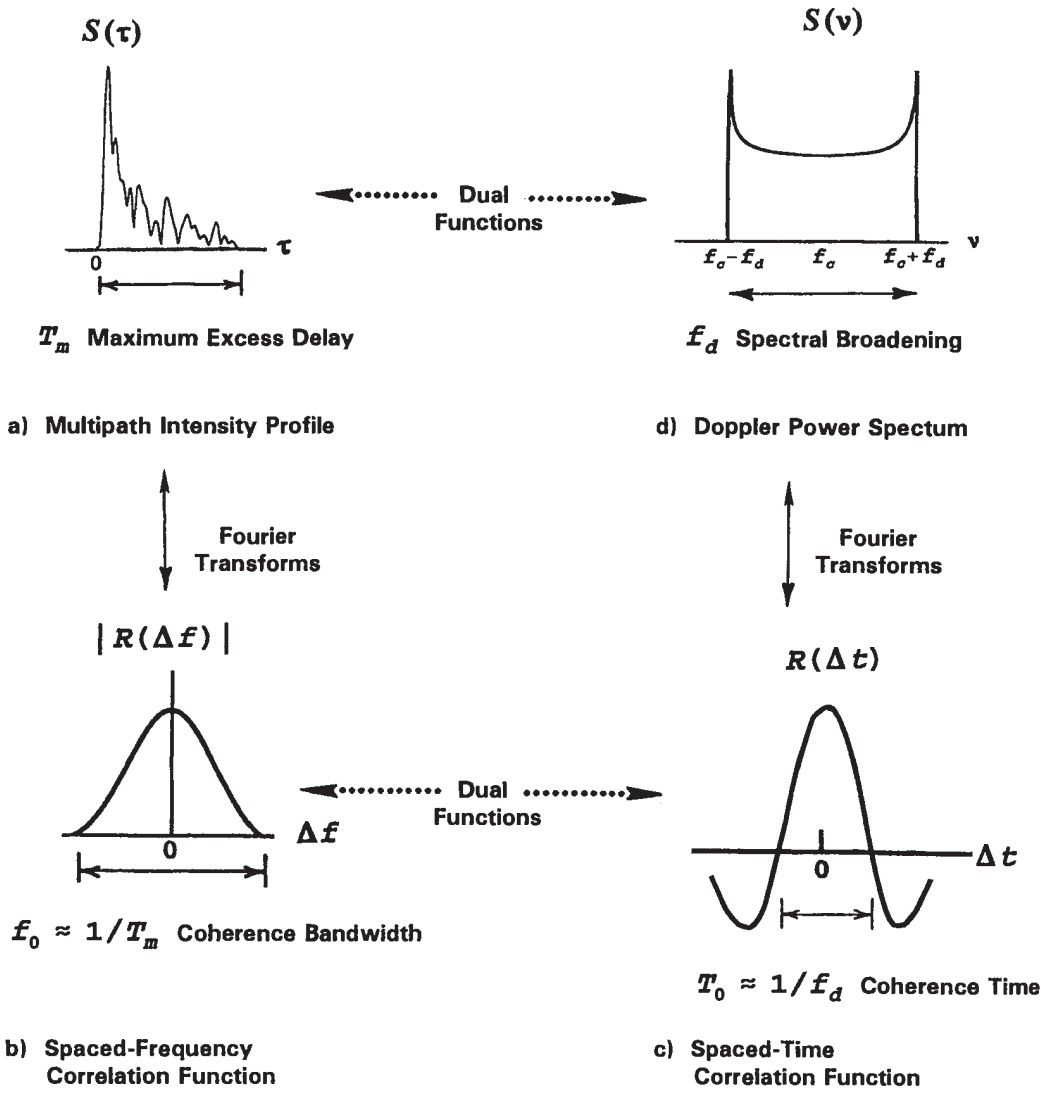


FIGURE 18.7 Relationships among the channel correlation functions and power density functions.

18.4.1 Degradation Categories due to Signal Time-Spreading Viewed in the Time-Delay Domain

In a fading channel, the relationship between maximum excess delay time, T_m, and symbol time, T_s, can be viewed in terms of two different degradation categories, **frequency-selective fading** and **frequency nonselective** or **flat fading,** as indicated in Fig. 18.1, blocks 8 and 9, and Fig. 18.6. A channel is said to exhibit frequency-selective fading if $T_m > T_s$. This condition occurs whenever the received multipath components of a symbol extend beyond the symbol's time duration. Such multipath dispersion of the signal yields the same kind of ISI distortion that is caused by an electronic filter. In fact, another name for this category of fading degradation is **channel-induced ISI.** In the case of frequency-selective fading, mitigating the distortion is possible because many of the multi-

path components are resolvable by the receiver. Later, several such mitigation techniques are described.

A channel is said to exhibit frequency nonselective or flat fading if $T_m < T_s$. In this case, all of the received multipath components of a symbol arrive within the symbol time duration; hence, the components are not resolvable. Here, there is no channel-induced ISI distortion, since the signal time spreading does not result in significant overlap among neighboring received symbols. There is still performance degradation since the unresolvable phasor components can add up destructively to yield a substantial reduction in SNR. Also, signals that are classified as exhibiting flat fading can sometimes experience frequency-selective distortion. This will be explained later when viewing degradation in the frequency domain, where the phenomenon is more easily described. For loss in SNR due to flat fading, the mitigation technique called for is to improve the received SNR (or reduce the required SNR). For digital systems, introducing some form of signal diversity and using error-correction coding is the most efficient way to accomplish this.

18.5 Signal Time-Spreading Viewed in the Frequency Domain: Figure 18.1, Block 10—The Spaced-Frequency Correlation Function

A completely analogous characterization of signal dispersion can begin in the frequency domain. In Fig. 18.7(b), the function $|R(\Delta f)|$ is seen, designated a **spaced-frequency correlation function;** it is the Fourier transform of $S(\tau)$. $R(\Delta f)$ represents the correlation between the channel's response to two signals as a function of the frequency difference between the two signals. It can be thought of as the channel's frequency transfer function. Therefore, the time-spreading manifestation can be viewed as if it were the result of a filtering process. Knowledge of $R(\Delta f)$ helps answer the question, "What is the correlation between received signals that are spaced in frequency $\Delta f = f_1 - f_2$?" $R(\Delta f)$ can be measured by transmitting a pair of sinusoids separated in frequency by Δf, cross-correlating the two separately received signals, and repeating the process many times with ever-larger separation Δf. Therefore, the measurement of $R(\Delta f)$ can be made with a sinusoid that is swept in frequency across the band of interest (a wideband signal). The **coherence bandwidth,** f_0 , is a statistical measure of the range of frequencies over which the channel passes all spectral components with approximately equal gain and linear phase. Thus, the coherence bandwidth represents a frequency range over which frequency components have a strong potential for amplitude correlation. That is, a signal's spectral components in that range are affected by the channel in a similar manner, as for example, exhibiting fading or no fading. Note that f_0 and T_m are reciprocally related (within a multiplicative constant). As an approximation, it is possible to say that

$$f_0 \approx \frac{1}{T_m} \tag{18.8}$$

The maximum excess delay, T_m, is not necessarily the best indicator of how any given system will perform on a channel because different channels with the same value of T_m can exhibit very different profiles of signal intensity over the delay span. A more useful measurement of delay spread is most often characterized in terms of the root mean squared (rms) delay spread, σ_τ, where

$$\sigma_\tau = \sqrt{\overline{\tau^2} - (\overline{\tau})^2} \tag{18.9}$$

$\overline{\tau}$ is the mean excess delay, $(\overline{\tau})^2$ is the mean squared, $\overline{\tau^2}$ is the second moment, and σ_τ is the square root of the second central moment of $S(\tau)$ [3].

An exact relationship between coherence bandwidth and delay spread does not exist, and must be derived from signal analysis (usually using Fourier techniques) of actual signal dispersion measurements in particular channels. Several approximate relationships have been described. If coherence bandwidth is defined as the frequency interval over which the channel's complex frequency transfer function has a correlation of at least 0.9, the coherence bandwidth is approximately [17]

$$f_0 \approx \frac{1}{50\sigma_\tau} \tag{18.10}$$

For the case of a mobile radio, an array of radially uniformly spaced scatterers, all with equal-magnitude reflection coefficients but independent, randomly occurring reflection phase angles [18, 19] is generally accepted as a useful model for urban surroundings. This model is referred to as the **dense-scatterer channel model.** With the use of such a model, coherence bandwidth has similarly been defined [18] for a bandwidth interval over which the channel's complex frequency transfer function has a correlation of at least 0.5 to be

$$f_0 = \frac{0.276}{\sigma_\tau} \tag{18.11}$$

The ionospheric-effects community employs the following definition

$$f_0 = \frac{1}{2\pi\sigma_\tau} \tag{18.12}$$

A more popular approximation of f_0 corresponding to a bandwidth interval having a correlation of at least 0.5 is [3]

$$f_0 \approx \frac{1}{5\sigma_\tau} \tag{18.13}$$

18.5.1 Degradation Categories due to Signal Time-Spreading Viewed in the Frequency Domain

A channel is referred to as frequency-selective if $f_0 < 1/T_s \approx W$, where the symbol rate $1/T_s$ is nominally taken to be equal to the signal bandwidth W. In practice, W may differ from $1/T_s$ due to system filtering or data modulation type (quaternary phase shift keying, QPSK, minimum shift keying, MSK, etc.) [21]. Frequency-selective fading distortion occurs whenever a signal's spectral components are not all affected equally by the channel. Some of the signal's spectral components, falling outside the coherence bandwidth, will be affected differently (independently) compared to those components contained within the coherence bandwidth. This occurs whenever $f_0 < W$ and is illustrated in Fig. 18.8(a).

Frequency-nonselective or flat fading degradation occurs whenever $f_0 > W$. Hence, all of the signal's spectral components will be affected by the channel in a similar manner (e.g., fading or no fading); this is illustrated in Fig. 18.8(b). Flat-fading does not introduce channel-induced ISI distortion, but performance degradation can still be expected due to the loss in SNR whenever the signal is fading. In order to avoid channel-induced ISI distortion, the channel is required to exhibit flat fading by insuring that

$$f_0 > W \approx \frac{1}{T_s} \tag{18.14}$$

Hence, the channel coherence bandwidth f_0 sets an upper limit on the transmission rate that can be used without incorporating an equalizer in the receiver.

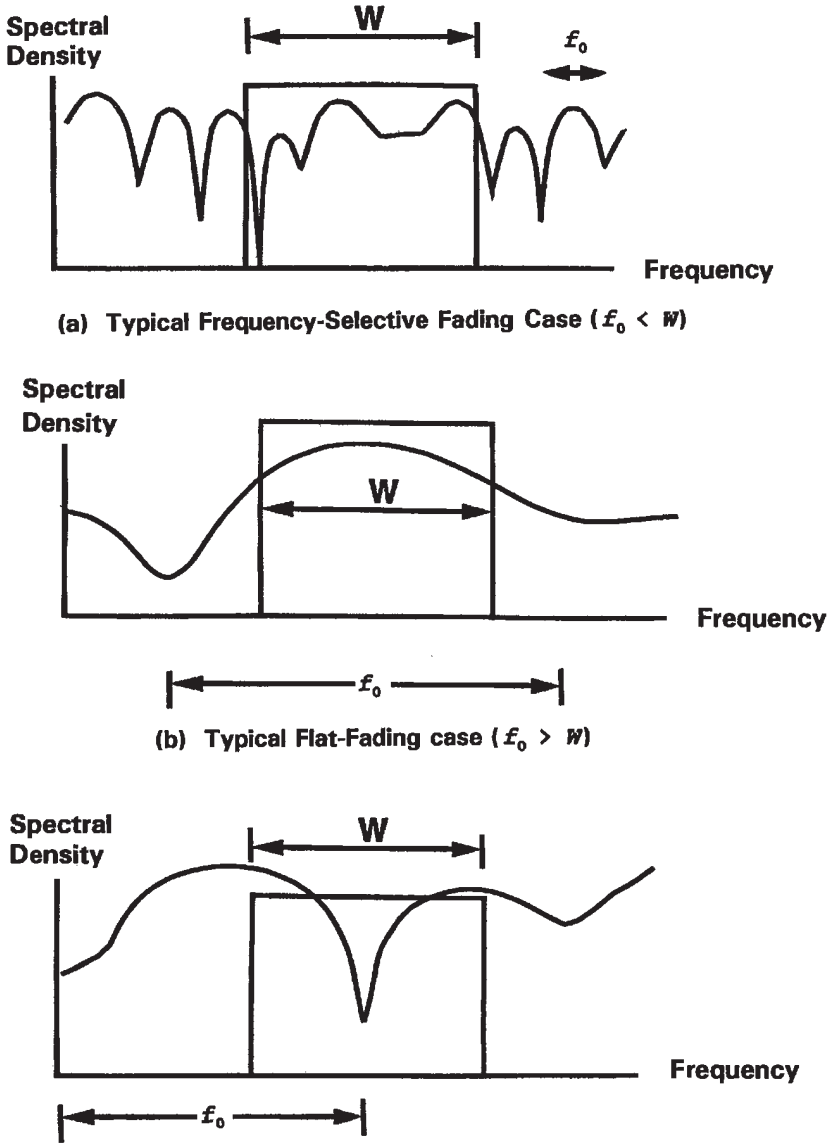


FIGURE 18.8 Relationships between the channel frequency-transfer function and a signal with bandwidth W.

For the flat-fading case, where $f_0 > W$ (or $T_m < T_s$), Fig. 18.8(b) shows the usual flat-fading pictorial representation. However, as a mobile radio changes its position, there will be times when the received signal experiences frequency-selective distortion even though $f_0 > W$. This is seen in Fig. 18.8(c), where the null of the channel's frequency transfer function occurs at the center of the signal band. Whenever this occurs, the baseband pulse will be especially mutilated by deprivation of its DC component. One consequence of the loss of DC (zero mean value) is the absence of a reliable pulse peak on which to establish the timing synchronization, or from which to sample the carrier phase carried by the pulse [18].

Thus, even though a channel is categorized as flat fading (based on rms relationships), it can still manifest frequency-selective fading on occasions. It is fair to say that a mobile radio channel, classified as having flat-fading degradation, cannot exhibit flat fading all of the time. As f_0 becomes much larger than W (or T_m becomes much smaller than T_s), less time will be spent in conditions approximating Fig. 18.8(c). By comparison, it should be clear that in Fig. 18.8(a) the fading is independent of the position of the signal band, and frequency-selective fading occurs all the time, not just occasionally.

18.6 Typical Examples of Flat Fading and Frequency-Selective Fading Manifestations

Figure 18.9 shows some examples of flat fading and frequency-selective fading for a direct-sequence spread-spectrum (DS/SS) system [20, 22]. In Fig. 18.9, there are three plots of the

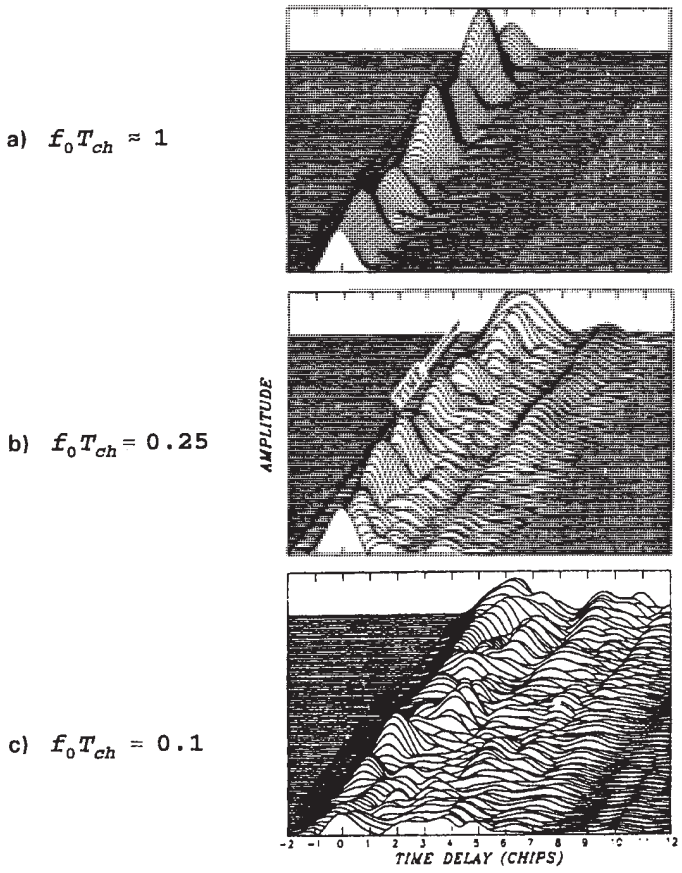


FIGURE 18.9 DS/SS Matched-filter output time-history examples for three levels of channel conditions, where T_{ch} is the time duration of a chip.

output of a pseudonoise (PN) code correlator vs. delay as a function of time (transmission or observation time). Each amplitude vs. delay plot is akin to $S(\tau)$ vs. τ shown in Fig. 18.7(a).

The key difference is that the amplitudes shown in Fig. 18.9 represent the output of a correlator; hence, the waveshapes are a function not only of the impulse response of the channel, but also of the impulse response of the correlator. The delay time is expressed in units of chip durations (chips), where the chip is defined as the spread-spectrum minimal-duration keying element. For each plot, the observation time is shown on an axis perpendicular to the amplitude vs. time-delay plane. Figure 18.9 is drawn from a satellite-to-ground communications link exhibiting scintillation because of atmospheric disturbances. However, Fig. 18.9 is still a useful illustration of three different channel conditions that might apply to a mobile radio situation. A mobile radio that moves along the observation-time axis is affected by changing multipath profiles along the route, as seen in the figure. The scale along the observation-time axis is also in units of chips. In Fig. 18.9(a), the signal dispersion (one "finger" of return) is on the order of a chip time duration, T_{ch}. In a typical DS/SS system, the spread-spectrum signal bandwidth is approximately equal to $1/T_{ch}$; hence, the normalized coherence bandwidth f_0T_{ch} of approximately unity in Fig. 18.9(a) implies that the coherence bandwidth is about equal to the spread-spectrum bandwidth. This describes a channel that can be called frequency-nonselective or slightly frequency-selective. In Fig. 18.9(b), where $f_0T_{ch} = 0.25$, the signal dispersion is more pronounced. There is definite interchip interference, and the coherence bandwidth is approximately equal to 25% of the spread-spectrum bandwidth. In Fig. 18.9(c), where $f_0T_{ch} = 0.1$, the signal dispersion is even more pronounced, with greater interchip-interference effects, and the coherence bandwidth is approximately equal to 10% of the spread-spectrum bandwidth. The channels of Figs. 18.9(b) and (c) can be categorized as moderately and highly frequency-selective, respectively, with respect to the basic signalling element, the chip. Later, we show that a DS/SS system operating over a frequency-selective channel at the chip level does not necessarily experience frequency-selective distortion at the symbol level.

18.7 Time Variance Viewed in the Time Domain: Figure 18.1, Block 13—The Spaced-Time Correlation Function

Until now, we have described signal dispersion and coherence bandwidth, parameters that describe the channel's time-spreading properties in a local area. However, they do not offer information about the time-varying nature of the channel caused by relative motion between a transmitter and receiver, or by movement of objects within the channel. For mobile-radio applications, the channel is time variant because motion between the transmitter and receiver results in propagation-path changes. Thus, for a transmitted continuous wave (CW) signal, as a result of such motion, the radio receiver sees variations in the signal's amplitude and phase. Assuming that all scatterers making up the channel are stationary, then whenever motion ceases, the amplitude and phase of the received signal remain constant; that is, the channel appears to be time invariant. Whenever motion begins again, the channel appears time variant. Since the channel characteristics are dependent on the positions of the transmitter and receiver, time variance in this case is equivalent to spatial variance.

Figure 18.7(c) shows the function $R(\Delta t)$, designated the **spaced-time correlation function;** it is the autocorrelation function of the channel's response to a sinusoid. This function specifies the extent to which there is correlation between the channel's response to a sinusoid sent at time t_1 and the response to a similar sinusoid sent at time t_2, where $\Delta t = t_2 - t_1$. The **coherence time,** T_0, is a measure of the expected time duration over which the channel's response is essentially invariant. Earlier, we made measurements of signal dispersion and coherence bandwidth by using wideband signals. Now, to measure the time-variant nature

of the channel, we use a narrowband signal. To measure $R(\Delta t)$ we can transmit a single sinusoid ($\Delta f = 0$) and determine the autocorrelation function of the received signal. The function $R(\Delta t)$ and the parameter T_0 provide us with knowledge about the fading rapidity of the channel. Note that for an ideal **time-invariant channel** (e.g., a mobile radio exhibiting no motion at all), the channel's response would be highly correlated for all values of Δt, and $R(\Delta t)$ would be a constant function. When using the dense-scatterer channel model described earlier, with constant velocity of motion, and an unmodulated CW signal, the normalized $R(\Delta t)$ is described as

$$R(\Delta t) = J_0\left(kV\Delta t\right) \tag{18.15}$$

where $J_0(\cdot)$ is the zero-order Bessel function of the first kind, V is velocity, $V\Delta t$ is distance traversed, and $k = 2\pi/\lambda$ is the free-space phase constant (transforming distance to radians of phase). Coherence time can be measured in terms of either time or distance traversed (assuming some fixed velocity of motion). Amoroso described such a measurement using a CW signal and a dense-scatterer channel model [18]. He measured the statistical correlation between the combination of received magnitude and phase sampled at a particular antenna location x_0, and the corresponding combination sampled at some displaced location $x_0 + \zeta$, with displacement measured in units of wavelength λ. For a displacement ζ of 0.38λ between two antenna locations, the combined magnitudes and phases of the received CW are statistically uncorrelated. In other words, the state of the signal at x_0 says nothing about the state of the signal at $x_0 + \zeta$. For a given velocity of motion, this displacement is readily transformed into units of time (coherence time).

18.7.1 The Concept of Duality

Two operators (functions, elements, or systems) are dual when the behavior of one with reference to a time-related domain (time or time-delay) is identical to the behavior of the other with reference to the corresponding frequency-related domain (frequency or Doppler shift).

In Fig. 18.7, we can identify functions that exhibit similar behavior across domains. For understanding the fading channel model, it is useful to refer to such functions as duals. For example, $R(\Delta f)$ in Fig. 18.7(b), characterizing signal dispersion in the frequency domain, yields knowledge about the range of frequency over which two spectral components of a received signal have a strong potential for amplitude and phase correlation. $R(\Delta t)$ in Fig. 18.7(c), characterizing fading rapidity in the time domain, yields knowledge about the span of time over which two received signals have a strong potential for amplitude and phase correlation. We have labeled these two correlation functions as duals. This is also noted in Fig. 18.1 as the duality between blocks 10 and 13, and in Fig. 18.6 as the duality between the time-spreading mechanism in the frequency domain and the time-variant mechanism in the time domain.

18.7.2 Degradation Categories due to Time Variance Viewed in the Time Domain

The time-variant nature of the channel or fading rapidity mechanism can be viewed in terms of two degradation categories as listed in Fig. 18.6: **fast fading** and **slow fading.** The terminology "fast fading" is used for describing channels in which $T_0 < T_s$, where T_0 is the channel coherence time and T_s is the time duration of a transmission symbol. Fast fading describes a condition where the time duration in which the channel behaves in a correlated

manner is short compared to the time duration of a symbol. Therefore, it can be expected that the fading character of the channel will change several times during the time that a symbol is propagating, leading to distortion of the baseband pulse shape. Analogous to the distortion previously described as channel-induced ISI, here distortion takes place because the received signal's components are not all highly correlated throughout time. Hence, fast fading can cause the baseband pulse to be distorted, resulting in a loss of SNR that often yields an irreducible error rate. Such distorted pulses cause synchronization problems (failure of phase-locked-loop receivers), in addition to difficulties in adequately defining a matched filter.

A channel is generally referred to as introducing slow fading if $T_0 > T_s$. Here, the time duration that the channel behaves in a correlated manner is long compared to the time duration of a transmission symbol. Thus, one can expect the channel state to virtually remain unchanged during the time in which a symbol is transmitted. The propagating symbols will likely not suffer from the pulse distortion described above. The primary degradation in a slow-fading channel, as with flat fading, is loss in SNR.

18.8 Time Variance Viewed in the Doppler-Shift Domain: Figure 18.1, Block 16—The Doppler Power Spectrum

A completely analogous characterization of the time-variant nature of the channel can begin in the Doppler-shift (frequency) domain. Figure 18.7(d) shows a **Doppler power spectral density,** $S(v)$, plotted as a function of Doppler-frequency shift, v. For the case of the dense-scatterer model, a vertical receive antenna with constant azimuthal gain, a uniform distribution of signals arriving at all arrival angles throughout the range $(0, 2\pi)$, and an unmodulated CW signal, the signal spectrum at the antenna terminals is [19]

$$S(v) = \frac{1}{\pi f_d \sqrt{1 - \left(\frac{v - f_c}{f_d}\right)^2}} \tag{18.16}$$

The equality holds for frequency shifts of v that are in the range $\pm f_d$ about the carrier frequency f_c and would be zero outside that range. The shape of the RF Doppler spectrum described by Eq. (18.16) is classically bowl-shaped, as seen in Fig. 18.7(d). Note that the spectral shape is a result of the dense-scatterer channel model. Equation (18.16) has been shown to match experimental data gathered for mobile radio channels [23]; however, different applications yield different spectral shapes. For example, the dense-scatterer model does not hold for the indoor radio channel; the channel model for an indoor area assumes $S(v)$ to be a flat spectrum [24].

In Fig. 18.7(d), the sharpness and steepness of the boundaries of the Doppler spectrum are due to the sharp upper limit on the Doppler shift produced by a vehicular antenna traveling among the stationary scatterers of the dense scatterer model. The largest magnitude (infinite) of $S(v)$ occurs when the scatterer is directly ahead of the moving antenna platform or directly behind it. In that case the magnitude of the frequency shift is given by

$$f_d = \frac{V}{\lambda} \tag{18.17}$$

where V is relative velocity and λ is the signal wavelength. f_d is positive when the transmitter and receiver move toward each other and negative when moving away from each other. For scatterers directly broadside of the moving platform, the magnitude of the frequency

shift is zero. The fact that Doppler components arriving at exactly 0° and 180° have an infinite power spectral density is not a problem, since the angle of arrival is continuously distributed and the probability of components arriving at exactly these angles is zero [3, 19].

$S(v)$ is the Fourier transform of $R(\Delta t)$. We know that the Fourier transform of the autocorrelation function of a time series is the magnitude squared of the Fourier transform of the original time series. Therefore, measurements can be made by simply transmitting a sinusoid (narrowband signal) and using Fourier analysis to generate the power spectrum of the received amplitude [16]. This Doppler power spectrum of the channel yields knowledge about the spectral spreading of a transmitted sinusoid (impulse in frequency) in the Doppler-shift domain. As indicated in Fig. 18.7, $S(v)$ can be regarded as the dual of the multipath intensity profile, $S(\tau)$, since the latter yields knowledge about the time spreading of a transmitted impulse in the time-delay domain. This is also noted in Fig. 18.1 as the duality between blocks 7 and 16, and in Fig. 18.6 as the duality between the time-spreading mechanism in the time-delay domain and the time-variant mechanism in the Doppler-shift domain.

Knowledge of $S(v)$ allows us to glean how much spectral broadening is imposed on the signal as a function of the rate of change in the channel state. The width of the Doppler power spectrum is referred to as the **spectral broadening** or **Doppler spread,** denoted by f_d, and sometimes called the **fading bandwidth** of the channel. Equation (18.16) describes the Doppler frequency shift. In a typical multipath environment, the received signal arrives from several reflected paths with different path distances and different angles of arrival, and the Doppler shift of each arriving path is generally different from that of another path. The effect on the received signal is seen as a Doppler spreading or spectral broadening of the transmitted signal frequency, rather than a shift. Note that the Doppler spread, f_d, and the coherence time, T_0, are reciprocally related (within a multiplicative constant). Therefore, we show the approximate relationship between the two parameters as

$$T_0 \approx \frac{1}{f_d} \tag{18.18}$$

Hence, the Doppler spread f_d or $1/T_0$ is regarded as the typical **fading rate** of the channel. Earlier, T_0 was described as the expected time duration over which the channel's response to a sinusoid is essentially invariant. When T_0 is defined more precisely as the time duration over which the channel's response to a sinusoid has a correlation of at least 0.5, the relationship between T_0 and f_d is approximately [4]

$$T_0 \approx \frac{9}{16\pi f_d} \tag{18.19}$$

A popular "rule of thumb" is to define T_0 as the geometric mean of Eqs. (18.18) and (18.19). This yields

$$T_0 = \sqrt{\frac{9}{16\pi f_d^2}} = \frac{0.423}{f_d} \tag{18.20}$$

For the case of a 900 MHz mobile radio, Fig. 18.10 illustrates the typical effect of Rayleigh fading on a signal's envelope amplitude vs. time [3]. The figure shows that the distance traveled by the mobile in the time interval corresponding to two adjacent nulls (small-scale fades) is on the order of a half-wavelength ($\lambda/2$) [3]. Thus, from Fig. 18.10 and Eq. (18.17), the time (approximately, the coherence time) required to traverse a distance $\lambda/2$ when traveling at a constant velocity, V, is:

$$T_0 \approx \frac{\lambda/2}{V} = \frac{0.5}{f_d} \tag{18.21}$$

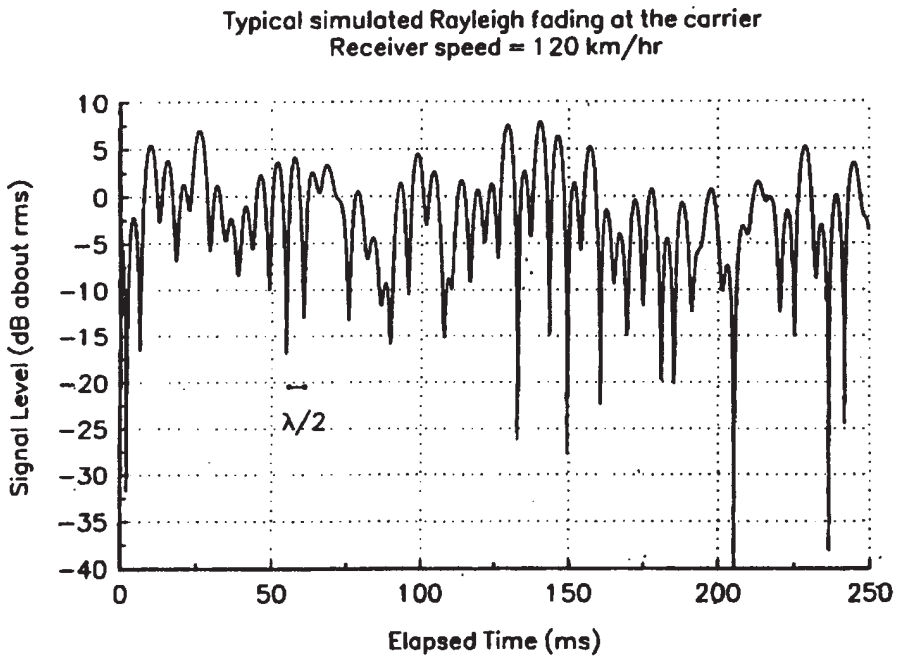


FIGURE 18.10 A typical Rayleigh fading envelope at 900 MHz.

Thus, when the interval between fades is taken to be $\lambda/2$, as in Fig. 18.10, the resulting expression for T_0 in Eq. (18.21) is quite close to the rule-of-thumb shown in Eq. (18.20). Using Eq. (18.21), with the parameters shown in Fig. 18.10 (velocity = 120 km/hr, and carrier frequency = 900 MHz), it is straightforward to compute that the coherence time is approximately 5 ms and the Doppler spread (channel fading rate) is approximately 100 Hz. Therefore, if this example represents a voice-grade channel with a typical transmission rate of 10^4 symbols/s, the fading rate is considerably less than the symbol rate. Under such conditions, the channel would manifest slow-fading effects. Note that if the abscissa of Fig. 18.10 were labeled in units of wavelength instead of time, the figure would look the same for any radio frequency and any antenna speed.

18.9 Analogy Between Spectral Broadening in Fading Channels and Spectral Broadening in Digital Signal Keying

Help is often needed in understanding why spectral broadening of the signal is a function of fading rate of the channel. Figure 18.11 uses the keying of a digital signal (such as amplitude-shift-keying or frequency-shift-keying) to illustrate an analogous case. Figure 18.11(a) shows that a single tone, $\cos 2\pi f_c t$ $(-\infty < t < \infty)$ that exists for all time is characterized in the frequency domain in terms of impulses (at $\pm f_c$). This frequency domain representation is ideal (i.e., zero bandwidth), since the tone is pure and neverending. In practical applications, digital signalling involves switching (keying) signals on and off at a required rate. The keying operation can be viewed as multiplying the infinite-duration tone in Fig. 18.11(a) by an ideal rectangular (switching) function in Fig. 18.11(b). The frequency-domain description of the ideal rectangular function is of the form $(\sin f)/f$. In Fig. 18.11(c), the result of the multiplication yields a tone, $\cos 2\pi f_c t$, that is time-duration limited in the interval

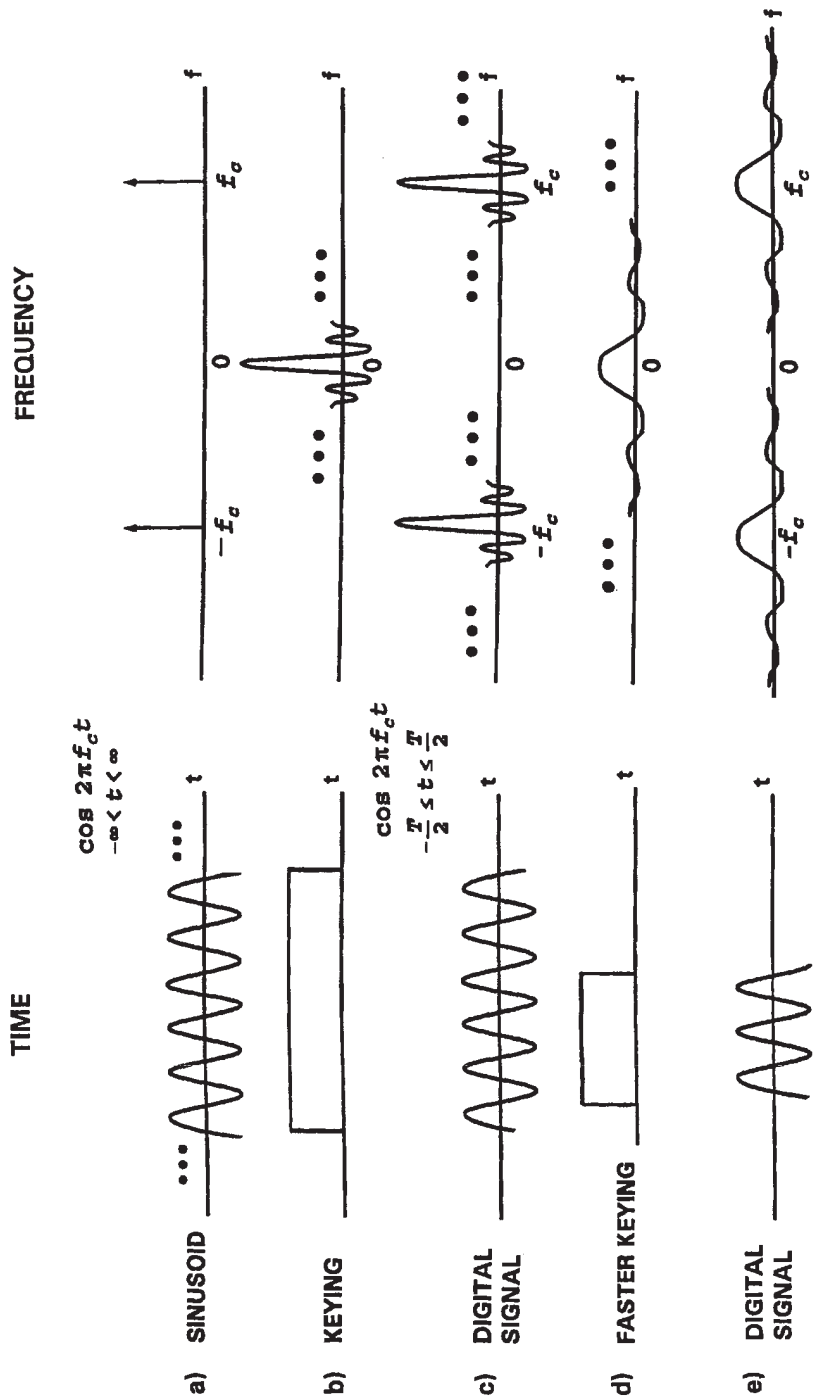


FIGURE 18.11 Analogy between spectral broadening in fading and spectral broadening in keying a digital signal.

$-T/2 < t < T/2$. The resulting spectrum is obtained by convolving the spectral impulses in part (a) with the $(\sin f)/f$ function in part (b), yielding the broadened spectrum in part (c). It is further seen that, if the signalling occurs at a faster rate characterized by the rectangle of shorter duration in part (d), the resulting spectrum of the signal in part (e) exhibits greater spectral broadening. The changing state of a fading channel is somewhat analogous to the keying on and off of digital signals. The channel behaves like a switch, turning the signal "on" and "off." The greater the rapidity of the change in the channel state, the greater the spectral broadening of the received signals. The analogy is not exact because the on and off switching of signals may result in phase discontinuities, but the typical multipath-scatterer environment induces phase-continuous effects.

18.10 Degradation Categories due to Time Variance, Viewed in the Doppler-Shift Domain

A channel is referred to as fast fading if the symbol rate, $1/T_s$ (approximately equal to the signalling rate or bandwidth W) is less than the fading rate, $1/T_0$ (approximately equal to f_d); that is, fast fading is characterized by

$$W < f_d \tag{18.22a}$$

or

$$T_s > T_0 \tag{18.22b}$$

Conversely, a channel is referred to as slow fading if the signalling rate is greater than the fading rate. Thus, in order to avoid signal distortion caused by fast fading, the channel must be made to exhibit slow fading by insuring that the signalling rate must exceed the channel fading rate. That is

$$W > f_d \tag{18.23a}$$

or

$$T_s < T_0 \tag{18.23b}$$

In Eq. (18.14), it was shown that due to signal dispersion, the coherence bandwidth, f_0, sets an upper limit on the signalling rate which can be used without suffering frequency-selective distortion. Similarly, Eq. (18.23) shows that due to Doppler spreading, the channel fading rate, f_d, sets a lower limit on the signalling rate that can be used without suffering fast-fading distortion. For HF communicating systems, when teletype or Morse-coded messages were transmitted at a low data rate, the channels were often fast fading. However, most present-day terrestrial mobile-radio channels can generally be characterized as slow fading.

Equation (18.23) doesn't go far enough in describing what we desire of the channel. A better way to state the requirement for mitigating the effects of fast fading would be that we desire $W \gg f_d$ (or $T_s \ll T_0$). If this condition is not satisfied, the random frequency modulation (FM) due to varying Doppler shifts will limit the system performance significantly. The Doppler effect yields an irreducible error rate that cannot be overcome by

simply increasing E_b/N_0 [25]. This irreducible error rate is most pronounced for any modulation that involves switching the carrier phase. A single specular Doppler path, without scatterers, registers an instantaneous frequency shift, classically calculated as $f_d = V/\lambda$. However, a combination of specular and multipath components yields a rather complex time dependence of instantaneous frequency which can cause much larger frequency swings than $\pm V/\lambda$ when detected by an instantaneous frequency detector (a nonlinear device) [26]. Ideally, coherent demodulators that lock onto and track the information signal should suppress the effect of this FM noise and thus cancel the impact of Doppler shift. However, for large values of f_d, carrier recovery becomes a problem because very wideband (relative to the data rate) phase-lock loops (PLLs) need to be designed. For voice-grade applications with bit-error rates of 10^{-3} to 10^{-4}, a large value of Doppler shift is considered to be on the order of $0.01 \times W$. Therefore, to avoid fast-fading distortion and the Doppler-induced irreducible error rate, the signalling rate should exceed the fading rate by a factor of 100 to 200 [27]. The exact factor depends on the signal modulation, receiver design, and required error-rate [3], [26]–[29]. Davarian [29] showed that a frequency-tracking loop can help lower, but not completely remove, the irreducible error rate in a mobile system when using differential minimum-shift keyed (DMSK) modulation.

18.11 Mitigation Methods

Figure 18.12, subtitled "The Good, The Bad, and The Awful," highlights three major performance categories in terms of bit-error probability, P_B, vs. E_b/N_0. The leftmost exponentially-shaped curve represents the performance that can be expected when using any nominal modulation type in AWGN. Observe that with a reasonable amount of E_b/N_0, good performance results. The middle curve, referred to as the **Rayleigh limit,** shows the performance degradation resulting from a loss in SNR that is characteristic of flat fading or slow fading when there is no line-of-sight signal component present. The curve is a function of the reciprocal of E_b/N_0 (an inverse-linear function), so for reasonable values of SNR, performance will generally be "bad." In the case of Rayleigh fading, parameters with overbars are often introduced to indicate that a mean is being taken over the "ups" and "downs" of the fading experience. Therefore, one often sees such bit-error probability plots with mean parameters denoted by $\overline{P_B}$ and $\overline{E_b}/N_0$. The curve that reaches an irreducible level, sometimes called an **error floor,** represents "awful" performance, where the bit-error probability can approach the value of 0.5. This shows the severe distorting effects of frequency-selective fading or fast fading.

If the channel introduces signal distortion as a result of fading, the system performance can exhibit an irreducible error rate; when larger than the desired error rate, no amount of E_b/N_0 will help achieve the desired level of performance. In such cases, the general approach for improving performance is to use some form of mitigation to remove or reduce the distortion. The mitigation method depends on whether the distortion is caused by frequency-selective fading or fast fading. Once the distortion has been mitigated, the P_B vs. E_b/N_0 performance should have transitioned from the "awful" bottoming out curve to the merely "bad" Rayleigh limit curve. Next, we can further ameliorate the effects of fading and strive to approach AWGN performance by using some form of diversity to provide the receiver with a collection of uncorrelated samples of the signal, and by using a powerful error-correction code.

In Fig. 18.13, several mitigation techniques for combating the effects of both signal distortion and loss in SNR are listed. Just as Figs. 18.1 and 18.6 serve as a guide for characterizing fading phenomena and their effects, Fig. 18.13 can similarly serve to describe mitigation

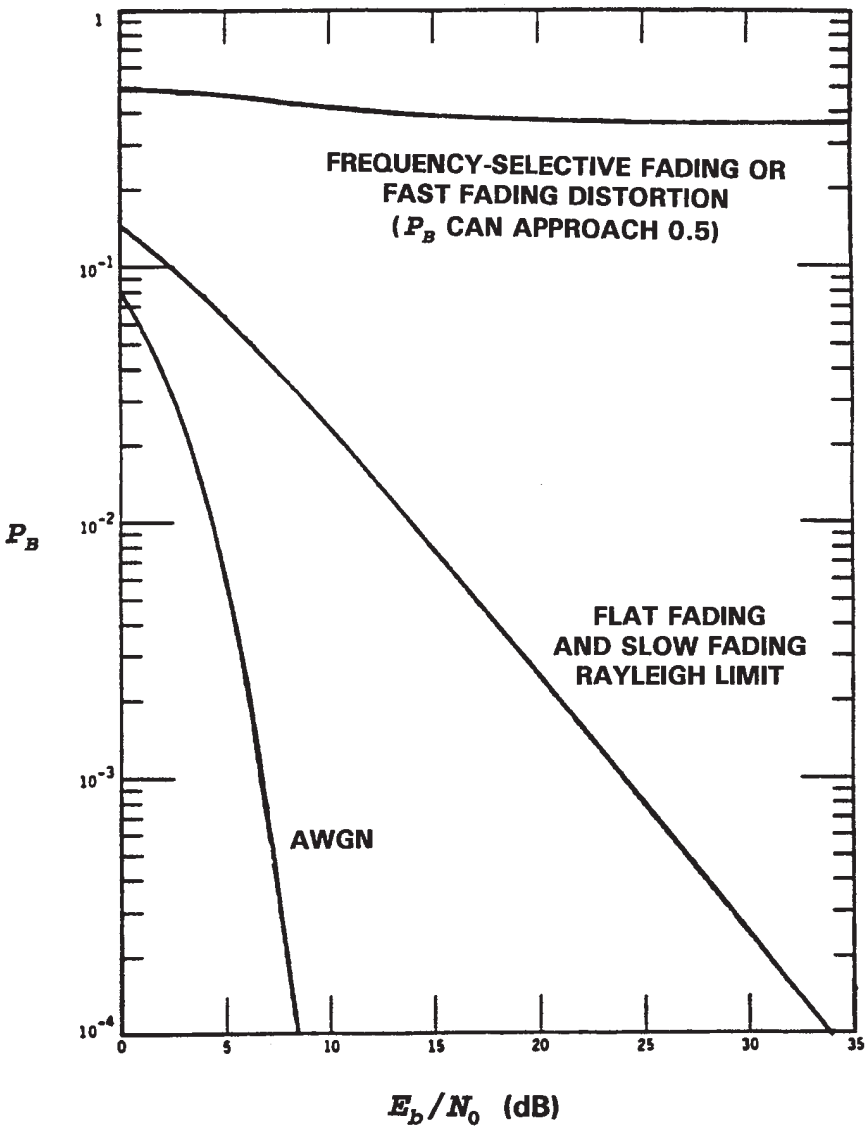


FIGURE 18.12 Error performance: The good, the bad, and the awful.

methods that can be used to ameliorate the effects of fading. The mitigation approach to be used should follow two basic steps: first, provide distortion mitigation; second, provide diversity.

18.11.1 Mitigation to Combat Frequency-Selective Distortion

- Equalization can compensate for the channel-induced ISI that is seen in frequency-selective fading. That is, it can help move the operating point from the error-performance curve that is "awful" in Fig. 18.12 to the one that is "bad." The process of equalizing the ISI involves some method of gathering the dispersed symbol energy back together into its original time interval. In effect, equaliza-

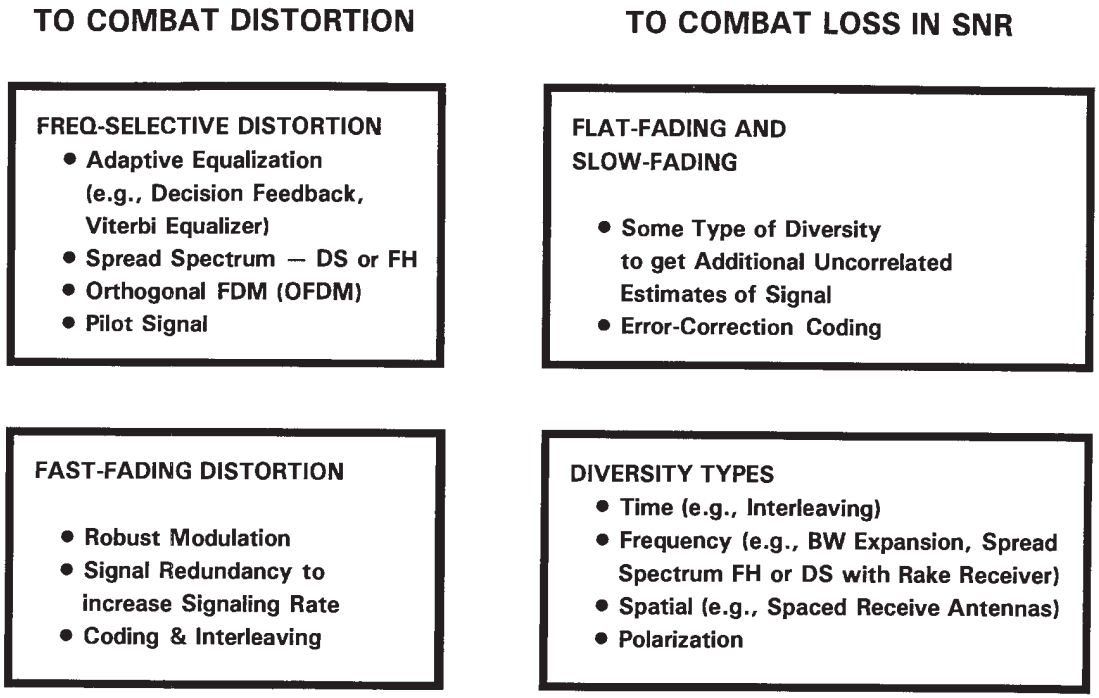


FIGURE 18.13 Basic mitigation types.

tion involves insertion of a filter to make the combination of channel and filter yield a flat response with linear phase. The phase linearity is achieved by making the equalizer filter the complex conjugate of the time reverse of the dispersed pulse [30]. Because in a mobile system the channel response varies with time, the equalizer filter must also change or adapt to the time-varying channel. Such equalizer filters are, therefore, called adaptive equalizers. An equalizer accomplishes more than distortion mitigation; it also provides diversity. Since distortion mitigation is achieved by gathering the dispersed symbol's energy back into the symbol's original time interval so that it doesn't hamper the detection of other symbols, the equalizer is simultaneously providing each received symbol with energy that would otherwise be lost.

- The decision feedback equalizer (DFE) has a feedforward section that is a linear transversal filter [30] whose length and tap weights are selected to coherently combine virtually all of the current symbol's energy. The DFE also has a feedback section which removes energy that remains from previously detected symbols [14], [30]–[32]. The basic idea behind the DFE is that once an information symbol has been detected, the ISI that it induces on future symbols can be estimated and subtracted before the detection of subsequent symbols.
- The maximum-likelihood sequence estimation (MLSE) equalizer tests all possible data sequences (rather than decoding each received symbol by itself) and chooses the data sequence that is the most probable of the candidates. The MLSE equalizer was first proposed by Forney [33] when he implemented the equalizer using the Viterbi decoding algorithm [34]. The MLSE is optimal in the sense that it minimizes the probability of a sequence error. Because the Viterbi decoding algorithm is the way in which the MLSE equalizer is typically implemented, the equalizer is often referred to as the **Viterbi equalizer.** Later in this chapter, we illustrate the adaptive equalization performed in the Global System for Mobile Communications (GSM) using the Viterbi equalizer.
- Spread-spectrum techniques can be used to mitigate frequency-selective ISI distortion because the hallmark of any spread-spectrum system is its capability to reject interference, and ISI is a type of interference. Consider a direct-sequence spread-spectrum (DS/SS) binary phase shift keying (PSK) communication channel comprising one direct path and one reflected path. Assume that the propagation from transmitter to receiver results in a multipath wave that is delayed by τ_k compared to the direct wave. If the receiver is synchronized to the waveform arriving via the direct path, the received signal, $r(t)$, neglecting noise, can be expressed as

$$r(t) = Ax(t)g(t)\cos(2\pi f_c t) + \alpha Ax(t-\tau_k)\, g(t-\tau_k)\cos(2\pi f_c t + \Theta) \qquad (18.24)$$

where $x(t)$ is the data signal, $g(t)$ is the pseudonoise (PN) spreading code, and τ_k is the differential time delay between the two paths. The angle Θ is a random phase, assumed to be uniformly distributed in the range $(0, 2\pi)$, and α is the attenuation of the multipath signal relative to the direct path signal. The receiver multiplies the incoming $r(t)$ by the code $g(t)$. If the receiver is synchronized to the direct path signal, multiplication by the code signal yields

$$Ax(t)g^2(t)\cos(2\pi f_c t) + \alpha Ax(t-\tau_k)\, g(t)g(t-\tau_k)\cos(2\pi f_c t + \Theta) \qquad (18.25)$$

where $g^2(t) = 1$, and if τ_k is greater than the chip duration, then,

$$\left| \int g^*(t) g\left(t - \tau_k\right) dt \right| \ll \int g^*(t) g(t) dt \tag{18.26}$$

over some appropriate interval of integration (correlation), where * indicates complex conjugate, and τ_k is equal to or larger than the PN chip duration. Thus, the spread spectrum system effectively eliminates the multipath interference by virtue of its code-correlation receiver. Even though channel-induced ISI is typically transparent to DS/SS systems, such systems suffer from the loss in energy contained in all the multipath components not seen by the receiver. The need to gather up this lost energy belonging to the received chip was the motivation for developing the Rake receiver [35]–[37]. The Rake receiver dedicates a separate correlator to each multipath component (finger). It is able to coherently add the energy from each finger by selectively delaying them (the earliest component gets the longest delay) so that they can all be coherently combined.

- Earlier, we described a channel that could be classified as flat fading, but occasionally exhibits frequency-selective distortion when the null of the channel's frequency transfer function occurs at the center of the signal band. The use of DS/SS is a good way to mitigate such distortion because the wideband SS signal would span many lobes of the selectively faded frequency response. Hence, a great deal of pulse energy would then be passed by the scatterer medium, in contrast to the nulling effect on a relatively narrowband signal [see Fig. 18.8(c)] [18].
- Frequency-hopping spread-spectrum (FH/SS) can be used to mitigate the distortion due to frequency-selective fading, provided the hopping rate is at least equal to the symbol rate. Compared to DS/SS, mitigation takes place through a different mechanism. FH receivers avoid multipath losses by rapid changes in the transmitter frequency band, thus avoiding the interference by changing the receiver band position before the arrival of the multipath signal.
- Orthogonal frequency-division multiplexing (OFDM) can be used in frequency-selective fading channels to avoid the use of an equalizer by lengthening the symbol duration. The signal band is partitioned into multiple subbands, each one exhibiting a lower symbol rate than the original band. The subbands are then transmitted on multiple orthogonal carriers. The goal is to reduce the symbol rate (signalling rate), $W \approx 1/T_s$, on each carrier to be less than the channel's coherence bandwidth f_0. OFDM was originally referred to as Kineplex. The technique has been implemented in the U.S. in mobile radio systems [38], and has been chosen by the European community under the name Coded OFDM (COFDM), for high-definition television (HDTV) broadcasting [39].
- Pilot signal is the name given to a signal intended to facilitate the coherent detection of waveforms. Pilot signals can be implemented in the frequency domain as an in-band tone [40], or in the time domain as a pilot sequence, which can also provide information about the channel state and thus improve performance in fading [41].

18.11.2 Mitigation to Combat Fast-Fading Distortion

- For fast fading distortion, use a robust modulation (noncoherent or differentially coherent) that does not require phase tracking, and reduce the detector integration time [20].

- Increase the symbol rate, $W \approx 1/T_s$, to be greater than the fading rate, $f_d \approx 1/T_0$, by adding signal redundancy.
- Error-correction coding and interleaving can provide mitigation because instead of providing more signal energy, a code reduces the required E_b/N_0. For a given E_b/N_0, with coding present, the error floor will be lowered compared to the uncoded case.
- An interesting filtering technique can provide mitigation in the event of fast-fading distortion and frequency-selective distortion occurring simultaneously. The frequency-selective distortion can be mitigated by the use of an OFDM signal set. Fast fading, however, will typically degrade conventional OFDM because the Doppler spreading corrupts the orthogonality of the OFDM subcarriers. A polyphase filtering technique [42] is used to provide time-domain shaping and duration extension to reduce the spectral sidelobes of the signal set and thus help preserve its orthogonality. The process introduces known ISI and adjacent channel interference (ACI) which are then removed by a post-processing equalizer and canceling filter [43].

18.11.3 Mitigation to Combat Loss in SNR

After implementing some form of mitigation to combat the possible distortion (frequency-selective or fast fading), the next step is to use some form of diversity to move the operating point from the error-performance curve labeled as "bad" in Fig. 18.12 to a curve that approaches AWGN performance. The term "diversity" is used to denote the various methods available for providing the receiver with uncorrelated renditions of the signal. Uncorrelated is the important feature here, since it would not help the receiver to have additional copies of the signal if the copies were all equally poor. Listed below are some of the ways in which diversity can be implemented.

- Time diversity—Transmit the signal on L different time slots with time separation of at least T_0. Interleaving, often used with error-correction coding, is a form of time diversity.
- Frequency diversity—Transmit the signal on L different carriers with frequency separation of at least f_0. Bandwidth expansion is a form of frequency diversity. The signal bandwidth, W, is expanded to be greater than f_0, thus providing the receiver with several independently fading signal replicas. This achieves frequency diversity of the order $L = W/f_0$. Whenever W is made larger than f_0, there is the potential for frequency-selective distortion unless we further provide some mitigation such as equalization. Thus, an expanded bandwidth can improve system performance (via diversity) only if the frequency-selective distortion the diversity may have introduced is mitigated.
- Spread spectrum is a form of bandwidth expansion that excels at rejecting interfering signals. In the case of direct-sequence spread-spectrum (DS/SS), it was shown earlier that multipath components are rejected if they are delayed by more than one chip duration. However, in order to approach AWGN performance, it is necessary to compensate for the loss in energy contained in those rejected components. The Rake receiver (described later) makes it possible to coherently combine the energy from each of the multipath components arriving along different paths. Thus, used with a Rake receiver, DS/SS modulation can be said to achieve path diversity. The Rake receiver is needed in phase-coherent reception,

but in differentially coherent bit detection, a simple delay line (one bit long) with complex conjugation will do the trick [44].

- Frequency-hopping spread-spectrum (FH/SS) is sometimes used as a diversity mechanism. The GSM system uses slow FH (217 hops/s) to compensate for those cases where the mobile user is moving very slowly (or not at all) and happens to be in a spectral null.
- Spatial diversity is usually accomplished through the use of multiple receive antennas, separated by a distance of at least 10 wavelengths for a base station (much less for a mobile station). Signal processing must be employed to choose the best antenna output or to coherently combine all the outputs. Systems have also been implemented with multiple spaced transmitters; an example is the Global Positioning System (GPS).
- Polarization diversity [45] is yet another way to achieve additional uncorrelated samples of the signal.
- Any diversity scheme may be viewed as a trivial form of repetition coding in space or time. However, there exist techniques for improving the loss in SNR in a fading channel that are more efficient and more powerful than repetition coding. Error-correction coding represents a unique mitigation technique, because instead of providing more signal energy it reduces the required E_b/N_0 in order to accomplish the desired error performance. Error-correction coding coupled with interleaving [20], [46]–[51] is probably the most prevalent of the mitigation schemes used to provide improved performance in a fading environment.

18.12 Summary of the Key Parameters Characterizing Fading Channels

We summarize the conditions that must be met so that the channel does not introduce frequency-selective distortion and fast-fading distortion. Combining the inequalities of Eqs. (18.14) and (18.23), we obtain

$$f_0 > W > f_d \tag{18.27a}$$

or

$$T_m < T_s < T_0 \tag{18.27b}$$

In other words, we want the channel coherence bandwidth to exceed our signalling rate, which in turn should exceed the fading rate of the channel. Recall that without distortion mitigation, f_0 sets an upper limit on signalling rate, and f_d sets a lower limit on it.

18.12.1 Fast-Fading Distortion: Example #1

If the inequalities of Eq. (18.27) are not met and distortion mitigation is not provided, distortion will result. Consider the fast-fading case where the signalling rate is less than the channel fading rate, that is,

$$f_0 > W < f_d \tag{18.28}$$

Mitigation consists of using one or more of the following methods. (See Fig. 18.13).

- Choose a modulation/demodulation technique that is most robust under fast-fading conditions. That means, for example, avoiding carrier recovery with PLLs since the fast fading could keep a PLL from achieving lock conditions.
- Incorporate sufficient redundancy so that the transmission symbol rate exceeds the channel fading rate. As long as the transmission symbol rate does not exceed the coherence bandwidth, the channel can be classified as flat fading. However, even flat-fading channels will experience frequency-selective distortion whenever a channel null appears at the band center.

Since this happens only occasionally, mitigation might be accomplished by adequate error-correction coding and interleaving.

- The above two mitigation approaches should result in the demodulator operating at the Rayleigh limit [20] (see Fig. 18.12). However, there may be an irreducible floor in the error-performance vs. E_b/N_0 curve due to the FM noise that results from the random Doppler spreading. The use of an in-band pilot tone and a frequency-control loop can lower this irreducible performance level.
- To avoid this error floor caused by random Doppler spreading, increase the signalling rate above the fading rate still further (100–200 $\times$ fading rate) [27]. This is one architectural motive behind time-division multiple access (TDMA) mobile systems.
- Incorporate error-correction coding and interleaving to lower the floor and approach AWGN performance.

18.12.2 Frequency-Selective Fading Distortion: Example #2

Consider the frequency-selective case where the coherence bandwidth is less than the symbol rate; that is,

$$f_0 < W > f_d \tag{18.29}$$

Mitigation consists of using one or more of the following methods. (See Fig. 18.13).

- Since the transmission symbol rate exceeds the channel-fading rate, there is no fast-fading distortion. Mitigation of frequency-selective effects is necessary. One or more of the following techniques may be considered:
- Adaptive equalization, spread spectrum (DS or FH), OFDM, pilot signal. The European GSM system uses a midamble training sequence in each transmission time slot so that the receiver can learn the impulse response of the channel. It then uses a Viterbi equalizer (explained later) for mitigating the frequency-selective distortion.
- Once the distortion effects have been reduced, introduce some form of diversity and error-correction coding and interleaving in order to approach AWGN performance. For direct-sequence spread-spectrum (DS/SS) signalling, the use of a Rake receiver (explained later) may be used for providing diversity by coherently combining multipath components that would otherwise be lost.

18.12.3 Fast-Fading and Frequency-Selective Fading Distortion: Example #3

Consider the case where the coherence bandwidth is less than the signalling rate, which in turn is less than the fading rate. The channel exhibits both fast-fading and frequency-

selective fading which is expressed as

$$f_0 < W < f_d \tag{18.30a}$$

or

$$f_0 < f_d \tag{18.30b}$$

Recalling from Eq. (18.27) that f_0 sets an upper limit on signalling rate and f_d sets a lower limit on it, this is a difficult design problem because, unless distortion mitigation is provided, the maximum allowable signalling rate is (in the strict terms of the above discussion) less than the minimum allowable signalling rate. Mitigation in this case is similar to the initial approach outlined in example #1.

- Choose a modulation/demodulation technique that is most robust under fast-fading conditions.
- Use transmission redundancy in order to increase the transmitted symbol rate.
- Provide some form of frequency-selective mitigation in a manner similar to that outlined in example #2.
- Once the distortion effects have been reduced, introduce some form of diversity and error-correction coding and interleaving in order to approach AWGN performance.

18.13 The Viterbi Equalizer as Applied to GSM

Figure 18.14 shows the GSM time-division multiple access (TDMA) frame, having a duration of 4.615 ms and comprising 8 slots, one assigned to each active mobile user. A normal transmission burst occupying one slot of time contains 57 message bits on each side

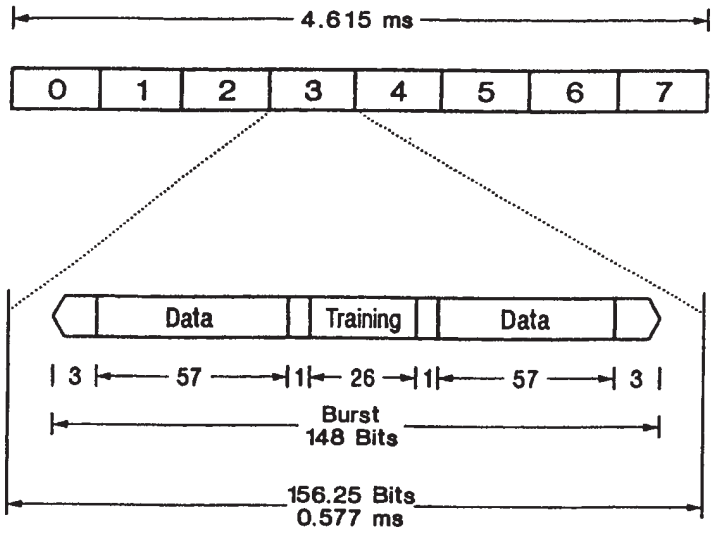


FIGURE 18.14 The GSM TDMA frame and time-slot containing a normal burst.

of a 26-bit midamble called a **training** or **sounding sequence.** The slot-time duration is 0.577 ms (or the slot rate is 1733 slots/s). The purpose of the midamble is to assist the receiver in estimating the impulse response of the channel in an adaptive way (during the time duration of each 0.577 ms slot). In order for the technique to be effective, the fading behavior of the channel should not change appreciably during the time interval of one slot. In other words, there should not be any fast-fading degradation during a slot time when the receiver is using knowledge from the midamble to compensate for the channel's fading behavior. Consider the example of a GSM receiver used aboard a high-speed train, traveling at a constant velocity of 200 km/hr (55.56 m/s). Assume the carrier frequency to be 900 MHz, (the wavelength is $\lambda = 0.33$ m). From Eq. (18.21), we can calculate that a half-wavelength is traversed in approximately the time (coherence time)

$$T_0 \approx \frac{\lambda/2}{V} \approx 3 \text{ ms} \tag{18.31}$$

Therefore, the channel coherence time is over 5 times greater than the slot time of 0.577 ms. The time needed for a significant change in fading behavior is relatively long compared to the time duration of one slot. Note, that the choices made in the design of the GSM TDMA slot time and midamble were undoubtedly influenced by the need to preclude fast fading with respect to a slot-time duration, as in this example.

The GSM symbol rate (or bit rate, since the modulation is binary) is 271 kilosymbols/s and the bandwidth is $W = 200$ kHz. If we consider that the typical rms delay spread in an urban environment is on the order of $\sigma_\tau = 2\mu s$, then using Eq. (18.13) the resulting coherence bandwidth is $f_0 \approx 100$ kHz. It should therefore be apparent that since $f_0 < W$, the GSM receiver must utilize some form of mitigation to combat frequency-selective distortion. To accomplish this goal, the Viterbi equalizer is typically implemented.

Figure 18.15 illustrates the basic functional blocks used in a GSM receiver for estimating the channel impulse response, which is then used to provide the detector with channel-corrected reference waveforms [52]. In the final step, the Viterbi algorithm is used to

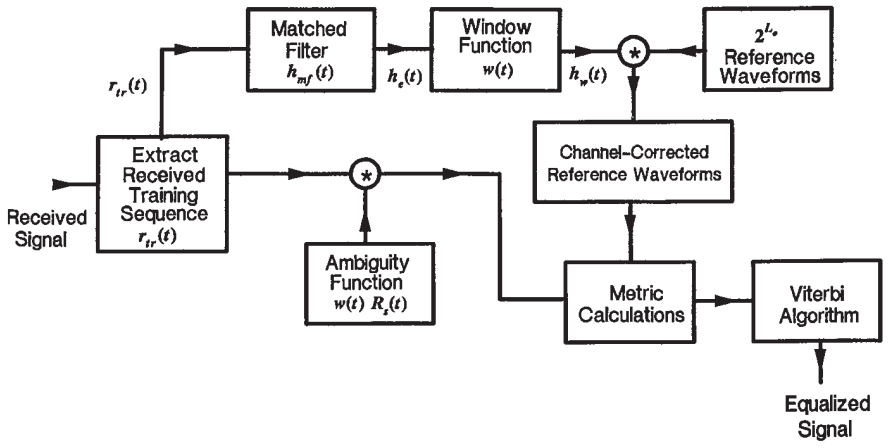


FIGURE 18.15 The Viterbi equalizer as applied to GSM.

compute the MLSE of the message. As stated in Eq. (18.2), a received signal can be described in terms of the transmitted signal convolved with the impulse response of the channel, $h_c(t)$. We show this below, using the notation of a received training sequence,

$r_{tr}(t)$, and the transmitted training sequence, $s_{tr}(t)$, as follows:

$$r_{tr}(t) = s_{tr}(t) * h_c(t) \tag{18.32}$$

where * denotes convolution. At the receiver, $r_{tr}(t)$ is extracted from the normal burst and sent to a filter having impulse response, $h_{mf}(t)$, that is matched to $s_{tr}(t)$. This matched filter yields at its output an estimate of $h_c(t)$, denoted $h_e(t)$, developed from Eq. (18.32) as follows.

$$\begin{aligned} h_e(t) &= r_{tr}(t) * h_{mf}(t) \\ &= s_{tr}(t) * h_c(t) * h_{mf}(t) \\ &= R_s(t) * h_c(t) \end{aligned} \tag{18.33}$$

where $R_s(t)$ is the autocorrelation function of $s_{tr}(t)$. If $R_s(t)$ is a highly peaked (impulse-like) function, then $h_e(t) \approx h_c(t)$.

Next, using a windowing function, $w(t)$, we truncate $h_e(t)$ to form a computationally affordable function, $h_w(t)$. The window length must be large enough to compensate for the effect of typical channel-induced ISI. The required observation interval L_0 for the window can be expressed as the sum of two contributions. The interval of length L_{CISI} is due to the controlled ISI caused by Gaussian filtering of the baseband pulses, which are then MSK modulated. The interval of length L_C is due to the channel-induced ISI caused by multipath propagation; therefore, L_0 can be written as

$$L_0 = L_{CISI} + L_C \tag{18.34}$$

The GSM system is required to provide mitigation for distortion due to signal dispersions of approximately 15–20 μs. The bit duration is 3.69 μs. Thus, the Viterbi equalizer used in GSM has a memory of 4–6 bit intervals. For each L_0-bit interval in the message, the function of the Viterbi equalizer is to find the most likely L_0-bit sequence out of the 2^{L_0} possible sequences that might have been transmitted. Determining the most likely L_0-bit sequence requires that 2^{L_0} meaningful reference waveforms be created by modifying (or disturbing) the 2^{L_0} ideal waveforms in the same way that the channel has disturbed the transmitted message. Therefore, the 2^{L_0} reference waveforms are convolved with the windowed estimate of the channel impulse response, $h_w(t)$ in order to derive the disturbed or channel-corrected reference waveforms. Next, the channel-corrected reference waveforms are compared against the received data waveforms to yield metric calculations. However, before the comparison takes place, the received data waveforms are convolved with the known windowed autocorrelation function $w(t)R_s(t)$, transforming them in a manner comparable to that applied to the reference waveforms. This filtered message signal is compared to all possible 2^{L_0} channel-corrected reference signals, and metrics are computed as required by the Viterbi decoding algorithm (VDA). The VDA yields the maximum likelihood estimate of the transmitted sequence [34].

18.14 The Rake Receiver Applied to Direct-Sequence Spread-Spectrum (DS/SS) Systems

Interim Specification 95 (IS-95) describes a DS/SS cellular system that uses a Rake receiver [35]–[37] to provide path diversity. In Fig. 18.16, five instances of chip transmissions corresponding to the code sequence 1 0 1 1 1 are shown, with the transmission or observation times labeled t_{-4} for the earliest transmission and t_0 for the latest. Each abscissa

shows three "fingers" of a signal that arrive at the receiver with delay times τ_1, τ_2, and τ_3. Assume that the intervals between the t_i transmission times and the intervals between the τ_i delay times are each one chip long. From this, one can conclude that the finger arriving at the receiver at time t_{-4}, with delay τ_3, is time coincident with two other fingers, namely the fingers arriving at times t_{-3} and t_{-2} with delays τ_2 and τ_1, respectively. Since, in this example, the delayed components are separated by exactly one chip time, they are *just* resolvable. At the receiver, there must be a sounding device that is dedicated to estimating the τ_i delay times. Note that for a terrestrial mobile radio system, the fading rate is relatively slow (milliseconds) or the channel coherence time large compared to the chip time ($T_0 > T_{ch}$). Hence, the changes in τ_i occur slowly enough so that the receiver can readily adapt to them.

Once the τ_i delays are estimated, a separate correlator is dedicated to processing each finger. In this example, there would be three such dedicated correlators, each one processing a delayed version of the same chip sequence 1 0 1 1 1. In Fig. 18.16, each correlator receives chips with power profiles represented by the sequence of fingers shown along a diagonal line. Each correlator attempts to match these arriving chips with the same PN code, similarly delayed in time. At the end of a symbol interval (typically there may be hundreds or thousands of chips per symbol), the outputs of the correlators are coherently combined, and a symbol detection is made. At the chip level, the Rake receiver resembles an equalizer, but its real function is to provide diversity.

The interference-suppression nature of DS/SS systems stems from the fact that a code sequence arriving at the receiver merely one chip time late, will be approximately orthogonal to the particular PN code with which the sequence is correlated. Therefore, any code chips that are delayed by one or more chip times will be suppressed by the correlator. The delayed chips only contribute to raising the noise floor (correlation sidelobes). The mitigation provided by the Rake receiver can be termed path diversity, since it allows the energy of a chip that arrives via multiple paths to be combined coherently. Without the Rake receiver, this energy would be transparent and therefore lost to the DS/SS system. In Fig. 18.16, looking vertically above point τ_3, it is clear that there is interchip interference due to different fingers arriving simultaneously. The spread-spectrum processing gain allows the system to endure such interference at the chip level. No other equalization is deemed necessary in IS-95.

18.15 Conclusion

In this chapter, the major elements that contribute to fading in a communication channel have been characterized. Figure 18.1 was presented as a guide for the characterization of fading phenomena. Two types of fading, large-scale and small-scale, were described. Two manifestations of small-scale fading (signal dispersion and fading rapidity) were examined, and the examination involved two views, time and frequency. Two degradation categories were defined for dispersion: frequency-selective fading and flat-fading. Two degradation categories were defined for fading rapidity: fast and slow. The small-scale fading degradation categories were summarized in Fig. 18.6. A mathematical model using correlation and power density functions was presented in Fig. 18.7. This model yields a nice symmetry, a kind of "poetry" to help us view the Fourier transform and duality relationships that describe the fading phenomena. Further, mitigation techniques for ameliorating the effects of each degradation category were treated, and these techniques were summarized in Fig. 18.13. Finally, mitigation methods that have been implemented in two system types, GSM and CDMA systems meeting IS-95, were described.

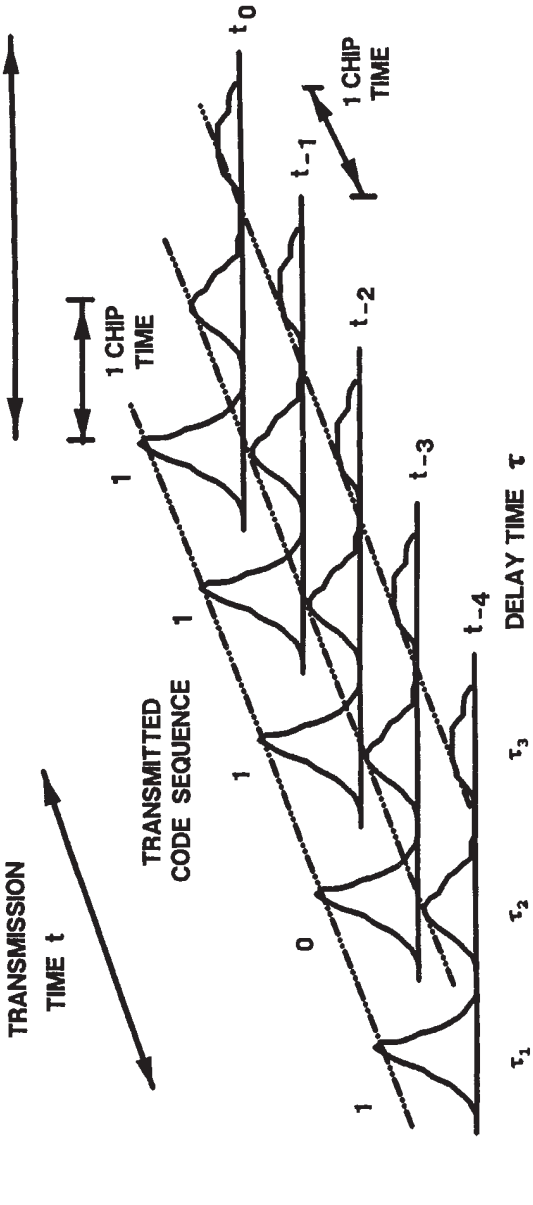


FIGURE 18.16 Example of received chips seen by a 3-finger rake receiver.

References

[1] Sklar, B., *Digital Communications: Fundamentals and Applications,* Prentice-Hall, Englewood Cliffs, NJ, Ch. 4, 1988.

[2] Van Trees, H.L., *Detection, Estimation, and Modulation Theory, Part I,* John Wiley & Sons, New York, Ch. 4, 1968.

[3] Rappaport, T.S., *Wireless Communications,* Prentice-Hall, Upper Saddle River, New Jersey, Chs. 3 and 4, 1996.

[4] Greenwood, D. and Hanzo, L., Characterisation of Mobile Radio Channels, *Mobile Radio Communications,* Steele, R., Ed., Pentech Press, London, Ch. 2, 1994.

[5] Lee, W.C.Y., Elements of cellular mobile radio systems, *IEEE Trans. Vehicular Technol.,* V-35(2), 48–56, May 1986.

[6] Okumura, Y. et al., Field strength and its variability in VHF and UHF land mobile radio service, *Rev. Elec. Comm. Lab.,* 16(9-10), 825–873, 1968.

[7] Hata, M., Empirical formulæ for propagation loss in land mobile radio services, *IEEE Trans. Vehicular Technol.,* VT-29(3), 317–325, 1980.

[8] Seidel, S.Y. et al., Path loss, scattering and multipath delay statistics in four European cities for digital cellular and microcellular radiotelephone, *IEEE Trans. Vehicular Technol.,* 40(4), 721–730, Nov. 1991.

[9] Cox, D.C., Murray, R., and Norris, A., 800 MHz Attenuation measured in and around suburban houses, *AT&T Bell Laboratory Technical Journal,* 673(6), 921–954, Jul.-Aug. 1984.

[10] Schilling, D.L. et al., Broadband CDMA for personal communications systems, *IEEE Commun. Mag.,* 29(11), 86–93, Nov. 1991.

[11] Andersen, J.B., Rappaport, T.S., and Yoshida, S., Propagation measurements and models for wireless communications channels, *IEEE Commun. Mag.,* 33(1), 42–49, Jan. 1995.

[12] Amoroso, F., Investigation of signal variance, bit error rates and pulse dispersion for DSPN signalling in a mobile dense scatterer ray tracing model, *Intl. J. Satellite Commun.,* 12, 579–588, 1994.

[13] Bello, P.A., Characterization of randomly time-variant linear channels, *IEEE Trans. Commun. Syst.,* 360–393, Dec. 1963.

[14] Proakis, J.G., *Digital Communications,* McGraw-Hill, New York, Ch. 7, 1983.

[15] Green, P.E., Jr., Radar astronomy measurement techniques, *MIT Lincoln Laboratory,* Lexington, MA, Tech. Report No. 282, Dec. 1962.

[16] Pahlavan, K. and Levesque, A.H., *Wireless Information Networks,* John Wiley & Sons, New York, Chs. 3 and 4, 1995.

[17] Lee, W.Y.C., *Mobile Cellular Communications,* McGraw-Hill, New York, 1989.

[18] Amoroso, F., Use of DS/SS signalling to mitigate Rayleigh fading in a dense scatterer environment, *IEEE Personal Commun.,* 3(2), 52–61, Apr. 1996.

[19] Clarke, R.H., A statistical theory of mobile radio reception, *Bell Syst. Tech. J.,* 47(6), 957–1000, Jul.-Aug. 1968.

[20] Bogusch, R.L., *Digital Communications in Fading Channels: Modulation and Coding,* Mission Research Corp., Santa Barbara, California, Report No. MRC-R-1043, Mar. 11, 1987.

[21] Amoroso, F., The bandwidth of digital data signals, *IEEE Commun. Mag.,* 18(6), 13–24, Nov. 1980.

[22] Bogusch, R.L. et al., Frequency selective propagation effects on spread-spectrum receiver tracking, *Proc. IEEE,* 69(7), 787–796, Jul. 1981.

[23] Jakes, W.C., Ed., *Microwave Mobile Communications,* John Wiley & Sons, New York, 1974.

[24] *Joint Technical Committee of Committee T1 R1P1.4 and TIA TR46.3.3/TR45.4.4 on Wireless Access,* Draft Final Report on RF Channel Characterization, Paper No. JTC(AIR)/94.01.17-238R4, Jan. 17, 1994.

[25] Bello, P.A. and Nelin, B.D., The influence of fading spectrum on the binary error probabilities of incoherent and differentially coherent matched filter receivers, *IRE Trans. Commun. Syst.,* CS-10, 160–168, Jun. 1962.

[26] Amoroso, F., Instantaneous frequency effects in a Doppler scattering environment, *IEEE International Conference on Communications,* 1458–1466, Jun. 7–10, 1987.

[27] Bateman, A.J. and McGeehan, J.P., Data transmission over UHF fading mobile radio channels, *IEEE Proc.,* 131, Pt. F(4), 364–374, Jul. 1984.

[28] Feher, K., *Wireless Digital Communications,* Prentice-Hall, Upper Saddle River, NJ, 1995.

[29] Davarian, F., Simon, M., and Sumida, J., DMSK: A Practical 2400-bps Receiver for the Mobile Satellite Service, Jet Propulsion Laboratory Publication 85-51 (MSAT-X Report No. 111), Jun. 15, 1985.

[30] Rappaport, T.S., *Wireless Communications,* Prentice-Hall, Upper Saddle River, NJ, Ch. 6, 1996.

[31] Bogusch, R.L., Guigliano, F.W., and Knepp, D.L., Frequency-selective scintillation effects and decision feedback equalization in high data-rate satellite links, *Proc. IEEE,* 71(6), 754–767, Jun. 1983.

[32] Qureshi, S.U.H., Adaptive equalization, *Proc. IEEE,* 73(9), 1340–1387, Sept. 1985.

[33] Forney, G.D., The Viterbi algorithm, *Proc. IEEE,* 61(3), 268–278, Mar. 1978.

[34] Sklar, B., *Digital Communications: Fundamentals and Applications,* Prentice-Hall, Englewood Cliffs, NJ, Ch. 6, 1988.

[35] Price, R. and Green, P.E., Jr., A communication technique for multipath channels, *Proc. IRE,* 555–570, Mar. 1958.

[36] Turin, G.L., Introduction to spread-spectrum antimultipath techniques and their application to urban digital radio, *Proc. IEEE,* 68(3), 328–353, Mar. 1980.

[37] Simon, M.K., Omura, J.K., Scholtz, R.A., and Levitt, B.K., *Spread Spectrum Communications Handbook,* McGraw-Hill, New York, 1994.

[38] Birchler, M.A. and Jasper, S.C., A 64 kbps Digital Land Mobile Radio System Employing M-16QAM, *Proceedings of the 1992 IEEE Intl. Conference on Selected Topics in Wireless Communications,* Vancouver, British Columbia, 158–162, Jun. 25–26, 1992.

[39] Sari, H., Karam, G., and Jeanclaude, I., Transmission techniques for digital terrestrial TV broadcasting, *IEEE Commun. Mag.,* 33(2), 100–109, Feb. 1995.

[40] Cavers, J.K., The performance of phase locked transparent tone-in-band with symmetric phase detection, *IEEE Trans. Commun.,* 39(9), 1389–1399, Sept. 1991.

[41] Moher, M.L. and Lodge, J.H., TCMP—A modulation and coding strategy for Rician fading channel, *IEEE J. Selected Areas Commun.,* 7(9), 1347–1355, Dec. 1989.

[42] Harris, F., On the Relationship Between Multirate Polyphase FIR Filters and Windowed, Overlapped FFT Processing, *Proceedings of the Twenty Third Annual Asilomar Conference on Signals, Systems, and Computers,* Pacific Grove, California, 485–488, Oct. 30 to Nov. 1, 1989.

[43] Lowdermilk, R.W. and Harris, F., Design and Performance of Fading Insensitive Orthogonal Frequency Division Multiplexing (OFDM) using Polyphase Filtering Techniques, *Proceedings of the Thirtieth Annual Asilomar Conference on Signals, Systems, and Computers,* Pacific Grove, California, Nov. 3–6, 1996.

[44] Kavehrad, M. and Bodeep, G.E., Design and experimental results for a direct-sequence spread-spectrum radio using differential phase-shift keying modulation for indoor wireless communications, *IEEE JSAC,* SAC-5(5), 815–823, Jun. 1987.

[45] Hess, G.C., *Land-Mobile Radio System Engineering,* Artech House, Boston, 1993.

[46] Hagenauer, J. and Lutz, E., Forward error correction coding for fading compensation in mobile satellite channels, *IEEE JSAC*, SAC-5(2), 215–225, Feb. 1987.

[47] McLane, P.I. et al., PSK and DPSK trellis codes for fast fading, shadowed mobile satellite communication channels, *IEEE Trans. Commun.*, 36(11), 1242–1246, Nov. 1988.

[48] Schlegel, C. and Costello, D.J., Jr., Bandwidth efficient coding for fading channels: code construction and performance analysis, *IEEE JSAC*, 7(9), 1356–1368, Dec. 1989.

[49] Edbauer, F., Performance of interleaved trellis-coded differential 8–PSK modulation over fading channels, *IEEE J. Selected Areas Commun.*, 7(9), 1340–1346, Dec. 1989.

[50] Soliman, S. and Mokrani, K., Performance of coded systems over fading dispersive channels, *IEEE Trans. Commun.*, 40(1), 51–59, Jan. 1992.

[51] Divsalar, D. and Pollara, F., Turbo Codes for PCS Applications, *Proc. ICC'95*, Seattle, Washington, 54–59, Jun. 18–22, 1995.

[52] Hanzo, L. and Stefanov, J., The Pan-European Digital Cellular Mobile Radio System—known as GSM, *Mobile Radio Communications*. Steele, R., Ed., Pentech Press, London, Ch. 8, 1992.

19

Space-Time Processing

19.1 Introduction 19-1
19.2 The Space-Time Wireless Channel 19-1
Multipath Propagation • Space-Time Channel Model
19.3 Signal Models 19-3
Signal Model at Base Station (Reverse Link) • Signal Model at Mobile (Forward Link) • Discrete Time Signal Model • Signal-Plus-Interference Model
19.4 ST Receive Processing (Base) 19-6
Receive Channel Estimation (Base) • Multiuser ST Receive Algorithms • Single-User ST Receive Algorithms
19.5 ST Transmit Processing (Base) 19-8
Transmit Channel Estimation (Base) • ST Transmit Processing • Forward Link Processing at the Mobile
19.6 Summary 19-9
Defining Terms 19-9
References 19-10

Arogyaswami J. Paulraj
Stanford University

19.1 Introduction

Mobile radio signal processing includes modulation and demodulation, channel coding and decoding, equalization and diversity. Current cellular modems mainly use temporal signal processing. Use of spatio-temporal signal processing can improve average signal power, mitigate fading, and reduce cochannel and intersymbol interference. This can significantly improve the capacity, coverage, and quality of wireless networks.

A space-time processing radio operates simultaneously on multiple antennas by processing signal samples both in space and time. In receive, space-time (ST) processing can increase array gain, spatial and temporal diversity and reduce cochannel interference and intersymbol interference. In transmit, the spatial dimension can enhance array gain, improve diversity, reduce generation of cochannel and inter-symbol interference.

19.2 The Space-Time Wireless Channel

19.2.1 Multipath Propagation

Multipath scattering gives rise to a number of propagation effects described below.

0-8493-8597-0/99/$0.00+$.50

© 1999 by CRC Press LLC

Scatterers Local to Mobile

Scattering local to the mobile is caused by buildings/other scatterers in the vicinity of the mobile (a few tens of meters). Mobile motion and local scattering give rise to Doppler spread which causes time-selective fading. For a mobile traveling at 65 mph, the Doppler spread is about 200 Hz in the 1900 MHz band. While local scatterers contribute to Doppler spread, the delay spread they contribute is usually insignificant because of the small scattering radius. Likewise, the angle spread induced at the base station is also small.

Remote Scatterers

The emerging wavefront from the local scatterers may then travel directly to the base or may be scattered toward the base by remote dominant scatterers, giving rise to specular multipaths. These remote scatterers can be either terrain features or high-rise building complexes. Remote scattering can cause significant delay and angle spreads.

Scatterers Local to Base

Once these multiple wavefronts reach the base station, they may be scattered further by local structures such as buildings or other structures in the vicinity of the base. Such scattering will be more pronounced for low elevation and below roof-top antennas. Scattering local to the base can cause severe angle spread which in turn causes space-selective fading. See Figure 19.1 for a depiction of different types of scattering.

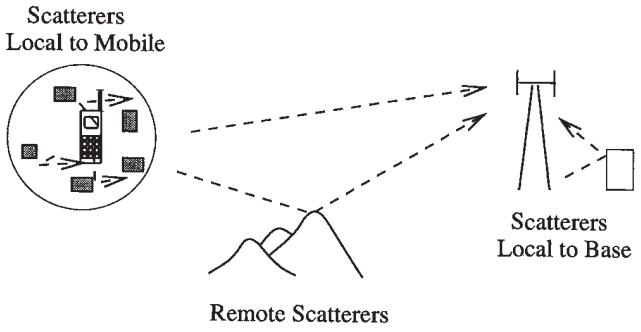


FIGURE 19.1 Multipath propagation in macrocells.

The forward link channel is affected in similar ways by these scatterers, but in a reverse order.

19.2.2 Space-Time Channel Model

The effect of delay, Doppler and angle spreads makes the channel selective in frequency, time, and space. Figure 19.2 shows plots of the frequency response at each branch of a four-antenna receiver operating with a 200 Khz bandwidth. We can see that the channel is highly frequency-selective since the delay spread reaches 10 to 15 μs. Also, an angle spread of 30° causes variations in the channel from antenna to antenna. The channel variation in time depends upon the Doppler spread. As expected, the plots show negligible channel variation between adjacent time slots, despite the high velocity of the mobile (100 kph). Use of longer time slots such as in IS-136 will result in significant channel variations over the slot period. Therefore, space-time processing should address the effect of the three spreads on the signal.

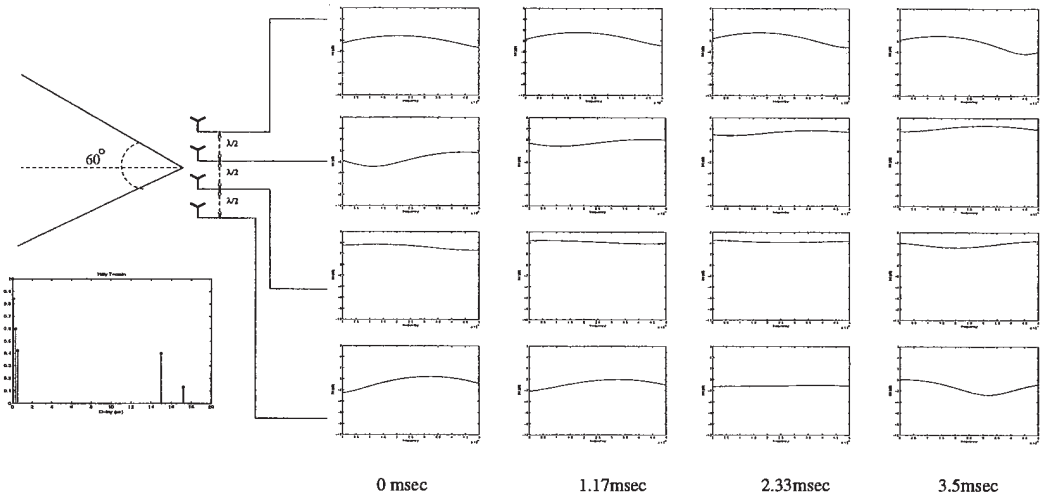


FIGURE 19.2 ST channel.

19.3 Signal Models

We develop signal models for nonspread modulation used in time division multiple access (TDMA) systems.

19.3.1 Signal Model at Base Station (Reverse Link)

We assume that antenna arrays are used at the base station only and that the mobile has a single omni antenna. The mobile transmits a channel coded and modulated signal which does not incorporate any spatial (or indeed any special temporal) processing. See Figure 19.3.

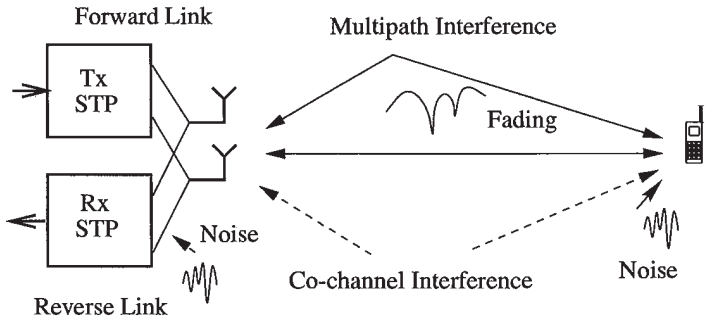


FIGURE 19.3 ST Processing Model.

The baseband signal $x_i(t)$ received by the base station at the ith element of an m element antenna array is given by

$$x_i(t) = \sum_{l=1}^{L} a_i(\theta_l)\, \alpha_l^R(t) u(t - \tau_l) + n_i(t) \tag{19.1}$$

where L is the number of multipaths, $a_i(\theta_l)$ is the response of the ith element for the lth

path from direction θ_l, $\alpha_l^R(t)$ is the complex path fading, τ_l is the path delay, $n_i(t)$ is the additive noise and $u(\cdot)$ is the transmitted signal that depends on the modulation waveform and the information data stream.

For a linear modulation, the baseband transmitted signal is given by

$$u(t) = \sum_k g(t - kT)s(k) \tag{19.2}$$

where $g(\cdot)$ is the pulse shaping waveform and $s(k)$ represents the information bits.

In the above model we have assumed that the inverse signal bandwidth is large compared to the travel time across the array. Therefore, the complex envelopes of the signals received by different antennas from a given path are identical except for phase and amplitude differences that depend on the path angle-of-arrival, array geometry and the element pattern. This angle-of-arrival dependent phase and amplitude response at the ith element is $a_i(\theta_l)$.

We collect all the element responses to a path arriving from angle θ_l into an m-dimensional vector, called the *array response vector* defined as

$$\begin{aligned} \mathbf{a}(\theta_l) &= [a_1(\theta_l)\, a_2(\theta_l) \ldots a_m(\theta_l)]^T \\ \mathbf{x}(t) &= \sum_{l=1}^{L} \mathbf{a}(\theta_l)\, \alpha_l^R(t) u(t - \tau_l) + \mathbf{n}(t) \end{aligned} \tag{19.3}$$

where $\mathbf{x}(t)$ and $\mathbf{n}(t)$ are m-dimensional complex vectors. The fading $|\alpha^R(t)|$ is Rayleigh or Rician distributed depending on the propagation model.

19.3.2 Signal Model at Mobile (Forward Link)

In this model, the base station transmits different signals from each antenna with a defined relationship between them. In the case of a two element array, some examples of transmitted signals $u_i(t), i = 1, 2$ can be: (a) delay diversity: $u_2(t) = u_1(t - T)$ where T is the symbol period; (b) Doppler diversity: $u_2(t) = u_1(t)e^{j\omega t}$ where ω is differential carrier offset; (c) beamforming: $u_2(t) = w_2 u_1(t)$ where w_2 is complex scalar; and (d) space-time coding: $u_1(t) = \sum_k g(t - kT)s^1(k), u_2(t) = \sum_k g(t - kT)s^2(k)$ where $s^1(k)$ and $s^2(k)$ are related to the symbol sequence $s(k)$ through coding.

The received signal at the mobile is then given by

$$x(t) = \sum_{i=1}^{m} \sum_{l=1}^{L} a_i(\theta_l)\, \alpha_l^F(t) u_i(t - \tau_l) + n(t) \tag{19.4}$$

where the path delay τ_l and angle parameters θ_l are the same as those of the reverse link. $\alpha_l^F(t)$ is the complex fading on the forward link. In (fast) **TDD** systems $\alpha_l^F(t)$ will be identical to the reverse link complex fading $\alpha_l^R(t)$. In a **FDD** system $\alpha_l^F(t)$ and $\alpha_l^R(t)$ will usually have the same statistics but will in general be uncorrelated with each other. We assume $a_i(\theta_l)$ is the same for both links. This is only approximately true in FDD systems.

If simple beamforming alone is used in transmit, the signals radiated from the antennas are related by a complex scalar and result in a directional transmit beam which may selectively couple into the multipath environment and differentially scale the power in each path.

The signal received by the mobile in this case can be written as

$$x(t) = \sum_{l=1}^{L} \mathbf{w}^H \mathbf{a}(\theta_l)\, \alpha_l^F(t) u(t - \tau_l) + n(t) \tag{19.5}$$

where $\mathbf{w}$ is the beamforming vector.

19.3.3 Discrete Time Signal Model

The channel model described above uses physical path parameters such as path gain, delay, and angle of arrival. In practice these are not known and the discrete time received signal uses a more convenient discretized "symbol response" channel model.

We derive a discrete-time signal model at the base station antenna array. Let the continuous-time output from the receive antenna array $\mathbf{x}(t)$ be sampled at the symbol rate at instants $t = t_o + kT$. Then the vector array output may be written as

$$\mathbf{x}(k) = \mathbf{H}^R \mathbf{s}(k) + n(k) \tag{19.6}$$

where $\mathbf{H}^R$ is the reverse link symbol response channel (a $m \times N$ matrix) that captures the effects of the array response, symbol waveform and path fading. m is the number of antennas, N is the channel length in symbol periods and $\mathbf{n}(k)$ the sampled vector of additive noise. Note that $\mathbf{n}(k)$ may be colored in space and time, as discussed later. $\mathbf{H}^R$ is assumed to be time invariant. $\mathbf{s}(k)$ is a vector of N consecutive elements of the data sequence and is defined as

$$\mathbf{s}(k) = \begin{bmatrix} s(k) \\ \vdots \\ s(k-N+1) \end{bmatrix} \tag{19.7}$$

Note that we have assumed a sampling rate of one sample per symbol. Higher sampling rates may be used. Also, $\mathbf{H}^R$ is given by

$$\mathbf{H}^R = \sum_{l=1}^{L} \mathbf{a}(\theta_l)\, \alpha_l^R \mathbf{g}^T(\tau_l) \tag{19.8}$$

where $\mathbf{g}(\tau_l)$ is a vector defined by T spaced sampling of the pulse shaping function $g(\cdot)$ with an offset of τ_l.

Likewise the forward discrete signal model at the mobile is given by

$$x(k) = \sum_{i=1}^{m} \mathbf{h}_i^F \mathbf{s}(k) + n(k) \tag{19.9}$$

where $\mathbf{h}_i^F$ is a $1 \times N$ composite channel from the symbol sequence via the ith antenna to the mobile receiver which includes the effect transmit ST processing at the base station.

In the case of two antenna delay diversity, $\mathbf{h}_i^F$ is given by

$$\mathbf{h}_1^F = \sum_{l=1}^{L} a_1(\theta_l)\, \alpha_l^F \mathbf{g}(\tau_l) \tag{19.10}$$

and

$$\mathbf{h}_2^F = \sum_{l=1}^{L} a_2(\theta_l)\, \alpha_l^F \mathbf{g}(\tau_l - T) \tag{19.11}$$

If spatial beamforming alone is used, the signal model becomes

$$x(k) = \sum_{l=1}^{L} \mathbf{w}^H \mathbf{H}^F \mathbf{s}(k) + n(k) \tag{19.12}$$

where $\mathbf{H}^F$ is the *intrinsic* forward (F) channel given by

$$\mathbf{H}^F = \begin{bmatrix} \mathbf{h}_1^F \\ \mathbf{h}_2^F \end{bmatrix} = \sum_{l=1}^{L} \mathbf{a}(\theta_l)\, \alpha_l^F \mathbf{g}^T(\tau_l) \tag{19.13}$$

19.3.4 Signal-Plus-Interference Model

The overall received signal-plus-interference-and-noise model at the base station antenna array can be written as

$$\mathbf{x}(k) = \mathbf{H}_s^R \mathbf{s}_s(k) + \sum_{q=1}^{Q-1} \mathbf{H}_q^R \mathbf{s}_q(k) + \mathbf{n}(k) \tag{19.14}$$

where $\mathbf{H}_s^R$ and $\mathbf{H}_q^R$ are channels for signal and **CCI**, respectively, while $\mathbf{s}_s$ and $\mathbf{s}_q$ are the corresponding data sequences. Note that Eq. (19.14) appears to suggest that the signal and interference are baud synchronous. However, this can be relaxed and the time offsets can be absorbed into the channel $\mathbf{H}_q^R$.

Similarly, the signal at the mobile can also be extended to include CCI. Note that in this case, the source of interference is from other base stations (in TDMA) and the channel is between the interfering base station and the desired mobile. It is often convenient to handle signals in blocks. Therefore, we may collect M consecutive snapshots of $\mathbf{x}(\cdot)$ corresponding to time instants $k, \ldots, k+M-1$, (and dropping subscripts for a moment), we get

$$\mathbf{X}(k) = \mathbf{H}^R \mathbf{S}(k) + \mathbf{N}(k) \tag{19.15}$$

where $\mathbf{X}(k)$, $\mathbf{S}(k)$ and $\mathbf{N}(k)$ are defined appropriately. Similarly the received signal at the mobile in the forward link has a block representation using a row vector.

19.4 ST Receive Processing (Base)

The base station receiver receives the signals from the array antenna which consist of the signals from the desired mobile and the cochannel signals along with associated intersymbol interference and fading. The task of the receiver is to maximize signal power and mitigate fading, CCI and **ISI**. There are two broad approaches for doing this—one is multiuser detection wherein we demodulate both the cochannel and desired signals jointly, the other is to cancel CCI. The structure of the receiver depends on the nature of the channel estimates available and the tolerable receiver complexity. There are a number of options and we discuss only a few salient cases. Before discussing the receiver processing, we discuss how receiver channel is estimated.

19.4.1 Receive Channel Estimation (Base)

In many mobile communications standards, such as GSM and IS-54, explicit training signals are inserted inside the TDMA data bursts.

Let $\mathbf{T}$ be the training sequence arranged in a matrix form ($\mathbf{T}$ is arranged to be a Toeplitz matrix). Then, during the training burst, the received data is given by

$$\mathbf{X} = \mathbf{H}^R \mathbf{T} + \mathbf{N} \tag{19.16}$$

Clearly $\mathbf{H}^R$ can be estimated using least squares as

$$\mathbf{H}^R = \mathbf{X}\mathbf{T}^\dagger \tag{19.17}$$

where $\mathbf{T}^\dagger = \mathbf{T}^H \left(\mathbf{T}\mathbf{T}^H\right)^{-1}$.

The use of training consumes spectrum resource. In GSM, for example, about 20% of the bits are dedicated to training. Moreover, in rapidly varying mobile channels, we may have to retrain frequently, resulting in even poorer spectral efficiency. There is, therefore, increased interest in blind methods that can estimate a channel without an explicit training signal.

19.4.2 Multiuser ST Receive Algorithms

In multiuser (MU) algorithms, we address the problem of jointly demodulating the multiple signals. Recall the received signal is given by

$$\mathbf{X} = \mathbf{H}^R\mathbf{S} + \mathbf{N} \tag{19.18}$$

where $\mathbf{H}^R$ and $\mathbf{S}$ are suitably defined to include multiple users and are of dimensions $m \times NQ$ and $NQ \times M$, respectively.

If the channels for all the arriving signals are known, then we jointly demodulate all the user data sequences using multiuser **maximum likelihood sequence estimation** (MLSE). Starting with the data model in Eq. (19.18), we can then search for multiple user data sequences that minimize the ML cost function

$$\min_{\mathbf{S}} \left\|\mathbf{X} - \mathbf{H}^R\mathbf{S}\right\|_F^2 \tag{19.19}$$

The multiuser MLSE will have a large number of states in the trellis. Efficient techniques for implementing this complex receiver are needed. Multiuser MLSE detection schemes outperform all other receivers.

19.4.3 Single-User ST Receive Algorithms

In this scheme we only demodulate the desired user and cancel the CCI. Therefore, after CCI cancellation we can use MLSE receivers to handle diversity and ISI. In this method there is potential conflict between CCI mitigation and diversity maximization. We are forced to allocate the available degrees of freedom (antennas) to the competing requirements.

One approach is to cancel CCI by a space-time filter followed by an MLSE receiver to handle ISI. We do this by reformulating the MLSE criterion to arrive at a joint solution for the ST-MMSE filter and the effective channel for the scalar MLSE.

Another approach is to use a ST-MMSE receiver to handle both CCI and ISI. In a space-time filter (equalizer-beamformer), $\mathbf{W}$ has the following form

$$\mathbf{W}(k) = \begin{bmatrix} w_{11}(k) & \cdots & w_{1M}(k) \\ \vdots & \cdots & \vdots \\ w_{m1}(k) & \cdots & w_{mM}(k) \end{bmatrix} \tag{19.20}$$

In order to obtain a convenient formulation for the space-time filter output, we introduce the quantities $W(k)$ and $X(k)$ as follows

$$\begin{aligned} X(k) &= vec\left(\mathbf{X}(k)\right) \quad (mM \times 1) \\ W(k) &= vec\left(\mathbf{W}(k)\right) \quad (mM \times 1) \end{aligned} \tag{19.21}$$

where the operator $vec(\cdot)$ is defined as:

$$vec\left([\mathbf{v}_1 \cdots \mathbf{v}_M]\right) = \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_M \end{bmatrix}$$

The ST-MMSE filter chooses the filter weights to achieve the minimum mean square error. The ST-MMSE filter takes the familiar form

$$W = \mathbf{R}_{XX}^{-1} \overline{H^R} \tag{19.22}$$

where $\overline{H^R}$ is one column of $vec\ (\mathbf{H}^R)$. In ST-MMSE the CCI and spatial diversity conflict for the spatial degrees of freedom. Likewise, temporal diversity and ISI cancellation conflict for the temporal degrees of freedom.

19.5 ST Transmit Processing (Base)

The goal in ST transmit processing is to maximize the average signal power and diversity at the receiver as well as minimize cochannel generation to other mobiles. Note that the base station transmission cannot directly affect the CCI seen by its intended mobile. In transmit the space-time processing needs channel knowledge, but since it is carried out prior to transmission and, therefore, before the signal encounters the channel, this is different from the reverse link where the space-time processing is carried out after the channel has affected the signal. Note that the mobile receiver will, of course, need to know the channel for signal demodulation, but since it sees the signal after transmission through the channel, it can estimate the forward link channel using training signals transmitted from the individual transmitter antennas.

19.5.1 Transmit Channel Estimation (Base)

The transmit channel estimation at the base of the vector forward channel can be done via feedback by use of reciprocity principles. In a TDD system, if the duplexing time is small compared to the coherence time of the channel, both channels are the same and the base-station can use its estimate of the reverse channel as the forward channel; i.e., $\mathbf{H}^F = \mathbf{H}^R$, where $\mathbf{H}^R$ is the reverse channel (we have added superscript R to emphasize the receive channel). In FDD systems, the forward and reverse channels can potentially be very different. This arises from differences in instantaneous complex path gains $\alpha^R \neq \alpha^F$. The other channel components $\mathbf{a}(\theta_l)$ and $\mathbf{g}(\tau_l)$ are very nearly equal.

A direct approach to estimating the forward channel is to feed back the signal from the mobile unit and then estimate the channel. We can do this by transmitting orthogonal training signals through each base station antenna. We can feed back from the mobile to the base the received signal for each transmitted signal and thus estimate the channel.

19.5.2 ST Transmit Processing

The primary goals at the transmitter are to maximize diversity in the link and to reduce CCI generation to other mobiles. The diversity maximization depends on the inherent diversity at the antenna array and cannot be created at the transmitter. The role of ST processing is limited to maximizing the exploitability of this diversity at the receiver. This usually leads to use of orthogonal or near orthogonal signalling at each antenna: $\int u_1(t)\ u_2(t)\ dt \approx\ 0$.

Orthogonality ensures that the transmitted signals are separable at the mobile which can now combine these signals after appropriate weighting to attain maximum diversity.

In order to minimize CCI, our goal is to use the beamforming vector $\mathbf{w}$ to steer the radiated energy and therefore minimize the interference at the other mobiles while maximizing the signal level at one's own mobile. Note that the CCI at the reference mobile is not controlled by its own base station but is generated by other base stations. Reducing CCI at one's own mobile requires the cooperation of the other base stations.

Therefore we choose $\mathbf{w}$ such that

$$\max_{\mathbf{w}} \frac{E(\mathbf{w}^H \mathbf{H}^F \mathbf{s}(k)\mathbf{s}(k)^H \mathbf{H}^{FH}\mathbf{w})}{\sum_{q=1}^{Q-1} \mathbf{w}^H \mathbf{H}_q^F \mathbf{H}_q^{FH} \mathbf{w}} \tag{19.23}$$

where $Q-1$ is the number of susceptible outer cell mobiles. $\mathbf{H}_q^F$ is the channel from the base station to the qth outer cell mobile. In order to solve the above equation, we need to know the forward link channel $\mathbf{H}^F$ to the reference mobile and $\mathbf{H}_q^F$ to cochannel mobiles. In general, such complete channel knowledge may not be available and suboptimum receivers must be designed. Furthermore, we need to find a receiver that harmonizes maximization of diversity and reduction of CCI. Use of transmit ST processing affects $\mathbf{H}^F$ and thus can be incorporated.

19.5.3 Forward Link Processing at the Mobile

The mobile will receive the composite signal from all the base station transmit antennas and will need to demodulate the signal to estimate the symbol sequence. In doing so it usually needs to estimate the individual channels from each base station antenna to itself. This is usually done via the use of training signals on each transmit antenna. Note that as the number of transmit antennas increases, there is a greater burden of training requirements. The use of transmit ST processing reduces the CCI power observed by the mobile as well enhances the diversity available.

19.6 Summary

Use of space-time processing can significantly improve average signal power, mitigate fading, and reduce cochannel and intersymbol interference in wireless networks. This can in turn result in significantly improved capacity, coverage, and quality of wireless networks.

In this chapter we have discussed applications of ST processing to TDMA systems. The applications to CDMA systems follow similar principles, but differences arise due to the nature of the signal and interference models.

Defining Terms

ISI: Intersymbol intereference is caused by multipath propagation where one symbol interferes with other symbols.

CCI: Cochannel interference arises from neighboring cells where the frequency channel is reused.

Maximum Likelihood Sequence Estimation: A technique for channel equalization based on determining the best symbol sequence that matches the received signal.

References

[1] Lindskog, E. and Paulraj, A., A taxonomy of space-time signal processing, *IEE Trans. Radar and Sonar,* 25–31, Feb. 1998.

[2] Ng, B.C. and Paulraj, A., Space-time processing for PCS, *IEEE PCS Magazine,* 5(1), 36–48, Feb. 1998.

[3] Paulraj, A. and Papadias, C.B., Space-time processing for wireless communications, *IEEE Signal Processing Magazine,* 14(5), 49–83, Nov. 1997.

[4] Paulraj, A., Papadias, C., Reddy, V.U., and Van der Veen, A., *A Review of Space-Time Signal Processing for Wireless Communications,* in *Signal Processing for Wireless Communications,* V. Poor, Ed., Prentice Hall, 179–210, Dec. 1997.

20

Location Strategies for Personal Communications Services

Ravi Jain
Bell Communications Research

Yi-Bing Lin
Bell Communications Research

Seshadri Mohan[1]
Bell Communications Research

20.1 Introduction 20-2
20.2 An Overview of PCS 20-3
Aspects of Mobility—Example 20.1 • A Model for PCS
20.3 IS-41 Preliminaries 20-6
Terminal/Location Registration • Call Delivery
20.4 Global System for Mobile Communications 20-9
Architecture • User Location Strategy
20.5 Analysis of Database Traffic Rate for IS-41 and GSM . 20-12
The Mobility Model for PCS Users • Additional Assumptions • Analysis of IS-41 • Analysis of GSM
20.6 Reducing Signalling During Call Delivery 20-15
20.7 Per-User Location Caching 20-16
20.8 Caching Threshold Analysis 20-17
20.9 Techniques for Estimating Users' LCMR 20-21
The Running Average Algorithm • The Reset-K Algorithm • Comparison of the LCMR Estimation Algorithms
20.10 Discussion 20-27
Conditions When Caching Is Beneficial • Alternative Network Architectures • LCMR Estimation and Caching Policy
20.11 Conclusions 20-30
Acknowledgment 20-31
References 20-31

[1]Address correspondence to: Seshadri Mohan, MCC-1A216B, Bellcore, 445 South St, Morristown, NJ 07960; Phone: 973-829-5160, Fax: 973-829-5888, e-mail: smohan@bellcore.com.

©1996 by Bell Communications Research, Inc. Used with permission. The material in this chapter appeared originally in the following IEEE publications: S. Mohan and R. Jain. 1994. Two user location strategies for personal communications services, *IEEE Personal Communications: The Magazine of Nomadic Communications and Computing*, pp. 42–50, Feb., and R. Jain, C.N. Lo, and S. Mohan. 1994. A caching strategy to reduce network impacts of PCS, J-SAC Special Issue on Wireless and Mobile Networks, Aug.

20.1 Introduction

The vision of nomadic personal communications is the ubiquitous availability of services to facilitate exchange of information (voice, data, video, image, etc.) between nomadic end users independent of time, location, or access arrangements. To realize this vision, it is necessary to locate users that move from place to place. The strategies commonly proposed are two-level hierarchical strategies, which maintain a system of mobility databases, home location registers (HLR) and visitor location resisters (VLR), to keep track of user locations. Two standards exist for carrying out two-level hierarchical strategies using HLRs and VLRs. The standard commonly used in North America is the EIA/TIA Interim Standard 41 (IS 41) [6] and in Europe the Global System for Mobile Communications (GSM) [15, 18]. In this chapter, we refer to these two strategies as *basic* location strategies.

We introduce these two strategies for locating users and provide a tutorial on their usage. We then analyze and compare these basic location strategies with respect to load on mobility databases and signalling network. Next we propose an auxiliary strategy, called the *per-user caching* or, simply, the *caching* strategy, that augments the basic location strategies to reduce the signalling and database loads.

The outline of this chapter is as follows. In Section 20.2 we discuss different forms of mobility in the context of personal communications services (PCS) and describe a reference model for a PCS architecture. In Sections 20.3 and 20.4, we describe the user location strategies specified in the IS-41 and GSM standards, respectively, and in Section 20.5, using a simple example, we present a simplified analysis of the database loads generated by each strategy. In Section 20.6, we briefly discuss possible modifications to these protocols that are likely to result in significant benefits by either reducing query and update rate to databases or reducing the signalling traffic or both. Section 20.7 introduces the caching strategy followed by an analysis in the next two sections. This idea attempts to exploit the spatial and temporal locality in calls received by users, similar to the idea of exploiting locality of file access in computer systems [20]. A feature of the caching location strategy is that it is useful only for certain classes of PCS users, those meeting certain call and mobility criteria. We encapsulate this notion in the definition of the user's call-to-mobility ratio (CMR), and local CMR (LCMR), in Section 20.8. We then use this definition and our PCS network reference architecture to quantify the costs and benefits of caching and the threshold LCMR for which caching is beneficial, thus characterizing the classes of users for which caching should be applied. In Section 20.9 we describe two methods for estimating users' LCMR and compare their effectiveness when call and mobility patterns are fairly stable, as well as when they may be variable. In Section 20.10, we briefly discuss alternative architectures and implementation issues of the strategy proposed and mention other auxiliary strategies that can be designed. Section 20.11 provides some conclusions and discussion of future work.

The choice of platforms on which to realize the two location strategies (IS-41 and GSM) may vary from one service provider to another. In this paper, we describe a possible realization of these protocols based on the advanced intelligent network (AIN) architecture (see [2, 5]), and signalling system 7 (SS7). It is also worthwhile to point out that several strategies have been proposed in the literature for locating users, many of which attempt to reduce the signalling traffic and database loads imposed by the need to locate users in PCS.

20.2 An Overview of PCS

This section explains different aspects of mobility in PCS using an example of two nomadic users who wish to communicate with each other. It also describes a reference model for PCS.

20.2.1 Aspects of Mobility—Example 20.1

PCS can involve two possible types of mobility, terminal mobility and personal mobility, that are explained next.

Terminal Mobility: This type of mobility allows a terminal to be identified by a unique terminal identifier independent of the point of attachment to the network. Calls intended for that terminal can therefore be delivered to that terminal regardless of its network point of attachment. To facilitate terminal mobility, a network must provide several functions, which include those that locate, identify, and validate a terminal and provide services (e.g., deliver calls) to the terminal based on the location information. This implies that the network must store and maintain the location information of the terminal based on a unique identifier assigned to that terminal. An example of a terminal identifier is the IS-41 EIA/TIA cellular industry term mobile identification number (MIN), which is a North American Numbering Plan (NANP) number that is stored in the terminal at the time of manufacture and cannot be changed. A similar notion exists in GSM (see Section 20.4).

Personal Mobility: This type of mobility allows a PCS user to make and receive calls independent of both the network point of attachment and a specific PCS terminal. This implies that the services that a user has subscribed to (stored in that user's service profile) are available to the user even if the user moves or changes terminal equipment. Functions needed to provide personal mobility include those that identify (authenticate) the end user and provide services to an end user independent of both the terminal and the location of the user. An example of a functionality needed to provide personal mobility for voice calls is the need to maintain a user's location information based on a unique number, called the universal personal telecommunications (UPT) number, assigned to that user. UPT numbers are also NANP numbers. Another example is one that allows end users to define and manage their service profiles to enable users to tailor services to suit their needs. In Section 20.4, we describe how GSM caters to personal mobility via smart cards.

For the purposes of the example that follows, the terminal identifiers (TID) and UPT numbers are NANP numbers, the distinction being TIDs address terminal mobility and UPT numbers address personal mobility. Though we have assigned two different numbers to address personal and terminal mobility concerns, the same effect could be achieved by a single identifier assigned to the terminal that varies depending on the user that is currently utilizing the terminal. For simplicity we assume that two different numbers are assigned.

Figure 20.1 illustrates the terminal and personal mobility aspects of PCS, which will be explained via an example. Let us assume that users Kate and Al have, respectively, subscribed to PCS services from PCS service provider (PSP) A and PSP B. Kate receives the UPT number, say, 500 111 4711, from PSP A. She also owns a PCS terminal with TID 200 777 9760. Al too receives his UPT number 500 222 4712 from PSP B, and he owns a PCS terminal with TID 200 888 5760. Each has been provided a personal identification number (PIN) by their respective PSP when subscription began. We assume that the two PSPs have subscribed to PCS access services from a certain network provider such as, for example, a local exchange carrier (LEC). (Depending on the capabilities of the PSPs, the access services provided may vary. Examples of access services include translation of UPT

number to a routing number, terminal and personal registration, and call delivery. Refer to Bellcore, [3], for further details). When Kate plugs in her terminal to the network, or when she activates it, the terminal registers itself with the network by providing its TID to the network. The network creates an entry for the terminal in an appropriate database, which, in this example, is entered in the terminal mobility database (TMDB) A. The entry provides a mapping of her terminal's TID, 200 777 9760, to a routing number (RN), RN1. All of these activities happen without Kate being aware of them. After activating her terminal, Kate registers herself at that terminal by entering her UPT number (500 111 4711) to inform the network that all calls to her UPT number are to be delivered to her at the terminal. For security reasons, the network may want to authenticate her and she may be prompted to enter her PIN number into her terminal. (Alternatively, if the terminal is equipped with a smart card reader, she may enter her smart card into the reader. Other techniques, such as, for example, voice recognition, may be employed). Assuming that she is authenticated, Kate has now registered herself. As a result of personal registration by Kate, the network creates an entry for her in the personal mobility database (PMDB) A that maps her UPT number to the TID of the terminal at which she registered. Similarly, when Al activates his terminal and registers himself, appropriate entries are created in TMDB B and PMDB B. Now Al wishes to call Kate and, hence, he dials Kate's UPT number (500 111 4711). The network carries out the following tasks.

1. The switch analyzes the dialed digits and recognizes the need for AIN service, determines that the dialed UPT number needs to be translated to a RN by querying PMDB A and, hence, it queries PMDB A.
2. PMDB A searches its database and determines that the person with UPT number 500 111 4711 is currently registered at terminal with TID 200 777 9760.
3. PMDB A then queries TMDB A for the RN of the terminal with TID 200 777 9760. TMDB A returns the RN (RN1).
4. PMDB A returns the RN (RN1) to the originating switch.
5. The originating switch directs the call to the switch RN1, which then alerts Kate's terminal. The call is completed when Kate picks up her terminal.

Kate may take her terminal wherever she goes and perform registration at her new location. From then on, the network will deliver all calls for her UPT number to her terminal at the new location. In fact, she may actually register on someone else's terminal too. For example, suppose that Kate and Al agree to meet at Al's place to discuss a school project they are working on together. Kate may register herself on Al's terminal (TID 200 888 9534). The network will now modify the entry corresponding to 4711 in PMDB A to point to B 9534. Subsequent calls to Kate will be delivered to Al's terminal.

The scenario given here is used only to illustrate the key aspects of terminal and personal mobility; an actual deployment of these services may be implemented in ways different from those suggested here. We will not discuss personal registration further. The analyses that follow consider only terminal mobility but may easily be modified to include personal mobility.

20.2.2 A Model for PCS

Figure 20.2 illustrates the reference model used for the comparative analysis. The model assumes that the HLR resides in a service control point (SCP) connected to a regional signal transfer point (RSTP). The SCP is a storehouse of the AIN service logic, i.e., functionality used to perform the processing required to provide advanced services, such as speed call-

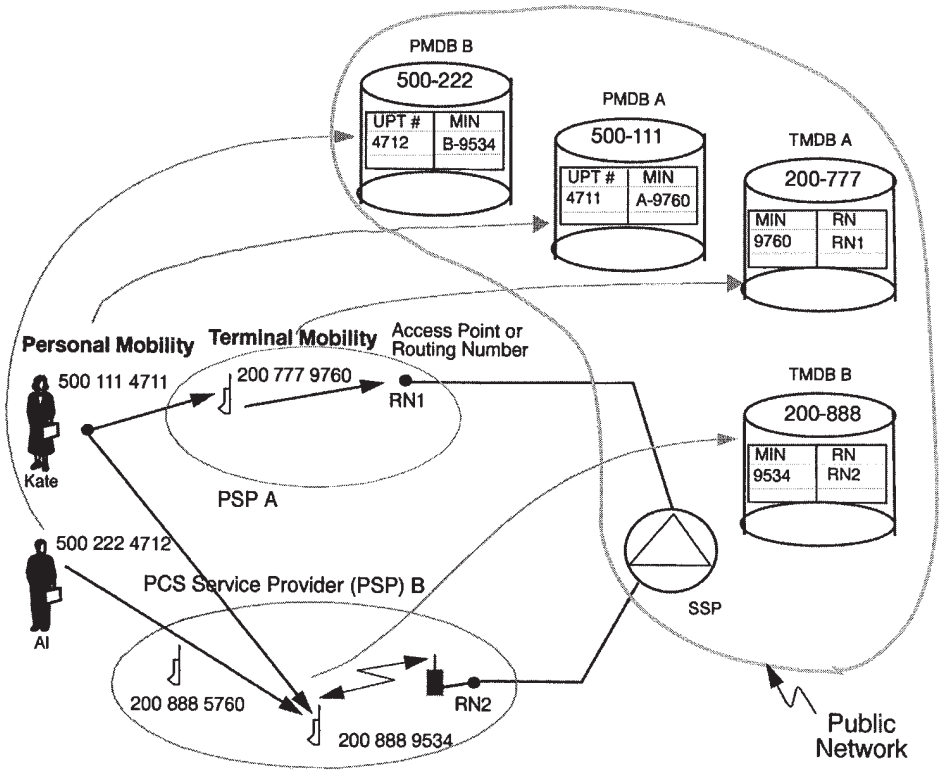


FIGURE 20.1 Illustrating terminal and personal mobility.

ing, outgoing call screening, etc., in the AIN architecture (see Bellcore, [2] and Berman and Brewster, [5]). The RSTP and the local STP (LSTP) are packet switches, connected together by various links such A links or D links, that perform the signalling functions of the SS7 network. Such functions include, for example, global title translation for routing messages between the AIN switching system, which is also referred to as the service switching point (SSP), and SCP and IS-41 messages [6]. Several SSPs may be connected to an LSTP.

The reference model in Fig. 20.2 introduces several terms which are explained next. We have tried to keep the terms and discussions fairly general. Wherever possible, however, we point to equivalent cellular terms from IS-41 or GSM.

For our purposes, the geographical area served by a PCS system is partitioned into a number of radio port coverage areas (or cells, in cellular terms) each of which is served by a radio port (or, equivalently, base station) that communicates with PCS terminals in that cell. A registration area (also known in the cellular world as location area) is composed of a number of cells. The base stations of all cells in a registration area are connected by wireline links to a mobile switching center (MSC). We assume that each registration area is served by a single VLR. The MSC of a registration area is responsible for maintaining and accessing the VLR and for switching between radio ports. The VLR associated with a registration area is responsible for maintaining a subset of the user information contained in the HLR.

Terminal registration process is initiated by terminals whenever they move into a new registration area. The base stations of a registration area periodically broadcast an identifier associated with that area. The terminals periodically compare an identifier they have stored

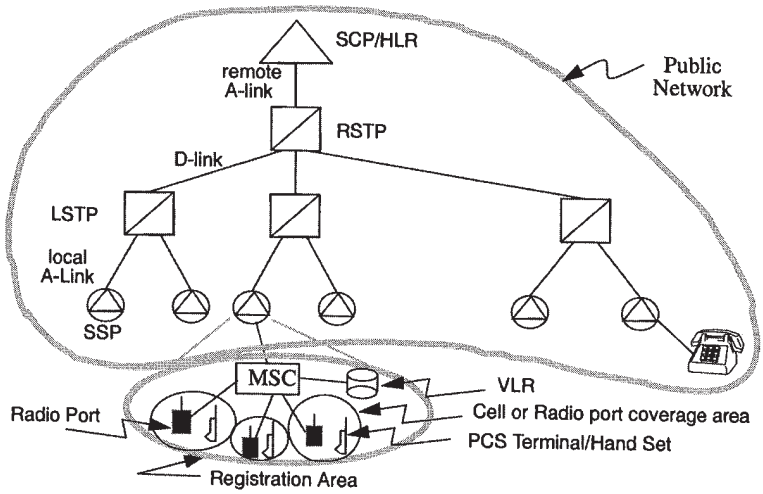


FIGURE 20.2 Example of a reference model for a PCS.

with the identifier to the registration area being broadcast. If the two identifiers differ, the terminal recognizes that it has moved from one registration area to another and will, therefore, generate a registration message. It also replaces the previous registration area identifier with that of the new one. Movement of a terminal within the same registration area will not generate registration messages. Registration messages may also be generated when the terminals are switched on. Similarly, messages are generated to deregister them when they are switched off.

PCS services may be provided by different types of commercial service vendors. Bellcore, [3] describes three different types of PSPs and the different access services that a public network may provide to them. For example, a PSP may have full network capabilities with its own switching, radio management, and radio port capabilities. Certain others may not have switching capabilities, and others may have only radio port capabilities. The model in Fig. 20.2 assumes full PSP capabilities. The analysis in Section 20.5 is based on this model and modifications may be necessary for other types of PSPs.

It is also quite possible that one or more registration areas may be served by a single PSP. The PSP may have one or more HLRs for serving its service area. In such a situation users that move within the PSP's serving area may generate traffic to the PSP's HLR (not shown in Fig. 20.2) but not to the network's HLR (shown in Fig. 20.2). In the interest of keeping the discussions simple, we have assumed that there is one-to-one correspondence between SSPs and MSCs and also between MSCs, registration areas, and VLRs. One impact of locating the SSP, MSC, and VLR in separate physical sites connected by SS7 signalling links would be to increase the required signalling message volume on the SS7 network. Our model assumes that the messages between the SSP and the associated MSC and VLR do not add to signalling load on the public network. Other configurations and assumptions could be studied for which the analysis may need to be suitably modified. The underlying analysis techniques will not, however, differ significantly.

20.3 IS-41 Preliminaries

We now describe the message flow for call origination, call delivery, and terminal registration, sometimes called location registration, based on the IS-41 protocol. This protocol is described in detail in EIA/TIA, [6]. Only an outline is provided here.

20.3.1 Terminal/Location Registration

During IS-41 registration, signalling is performed between the following pairs of network elements:

- New serving MSC and the associated database (or VLR)
- New database (VLR) in the visited area and the HLR in the public network
- HLR and the VLR in former visited registration area or the old MSC serving area.

Figure 20.3 shows the signalling message flow diagram for IS-41 registration activity, focusing only on the essential elements of the message flow relating to registration; for details of variations from the basic registration procedure, see Bellcore, [3].

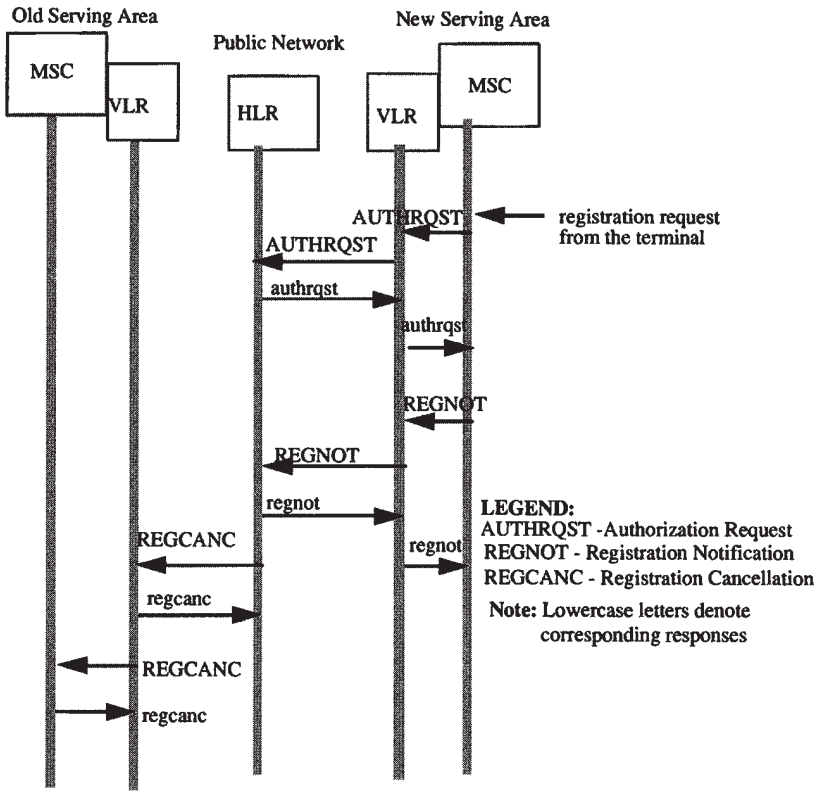


FIGURE 20.3 Signalling flow diagram for registration in IS-41.

The following steps describe the activities that take place during registration.

1. Once a terminal enters a new registration area, the terminal sends a registration request to the MSC of that area.
2. The MSC sends an authentication request (AUTHRQST) message to its VLR to authenticate the terminal, which in turn sends the request to the HLR. The HLR sends its response in the authrqst message.

3. Assuming the terminal is authenticated, the MSC sends a registration notification (REGNOT) message to its VLR.
4. The VLR in turn sends a REGNOT message to the HLR serving the terminal. The HLR updates the location entry corresponding to the terminal to point to the new serving MSC/VLR. The HLR sends a response back to the VLR, which may contain relevant parts of the user's service profile. The VLR stores the service profile in its database and also responds to the serving MSC.
5. If the user/terminal was registered previously in a different registration area, the HLR sends a registration cancellation (REGCANC) message to the previously visited VLR. On receiving this message, the VLR erases all entries for the terminal from the record and sends a REGCANC message to the previously visited MSC, which then erases all entries for the terminal from its memory.

The protocol shows authentication request and registration notification as separate messages. If the two messages can be packaged into one message, then the rate of queries to HLR may be cut in half. This does not necessarily mean that the total number of messages are cut in half.

20.3.2 Call Delivery

The signalling message flow diagram for IS-41 call delivery is shown in Fig. 20.4. The following steps describe the activities that take place during call delivery.

1. A call origination is detected and the number of the called terminal (for example, MIN) is received by the serving MSC. Observe that the call could have originated from within the public network from a wireline phone or from a wireless terminal in an MSC/VLR serving area. (If the call originated within the public network, the AIN SSP analyzes the dialed digits and sends a query to the SCP.)
2. The MSC determines the associated HLR serving the called terminal and sends a location request (LOCREQ) message to the HLR.
3. The HLR determines the serving VLR for that called terminal and sends a routing address request (ROUTEREQ) to the VLR, which forwards it to the MSC currently serving the terminal.
4. Assuming that the terminal is idle, the serving MSC allocates a temporary identifier, called a temporary local directory number (TLDN), to the terminal and returns a response to the HLR containing this information. The HLR forwards this information to the originating SSP/MSC in response to its LOCREQ message.
5. The originating SSP requests call setup to the serving MSC of the called terminal via the SS7 signalling network using the usual call setup protocols.

Similar to the considerations for reducing signalling traffic for location registration, the VLR and HLR functions could be united in a single logical database for a given serving area and collocated; further, the database and switch can be integrated into the same piece of physical equipment or be collocated. In this manner, a significant portion of the messages exchanged between the switch, HLR and VLR as shown in Fig. 20.4 will not contribute to signalling traffic.

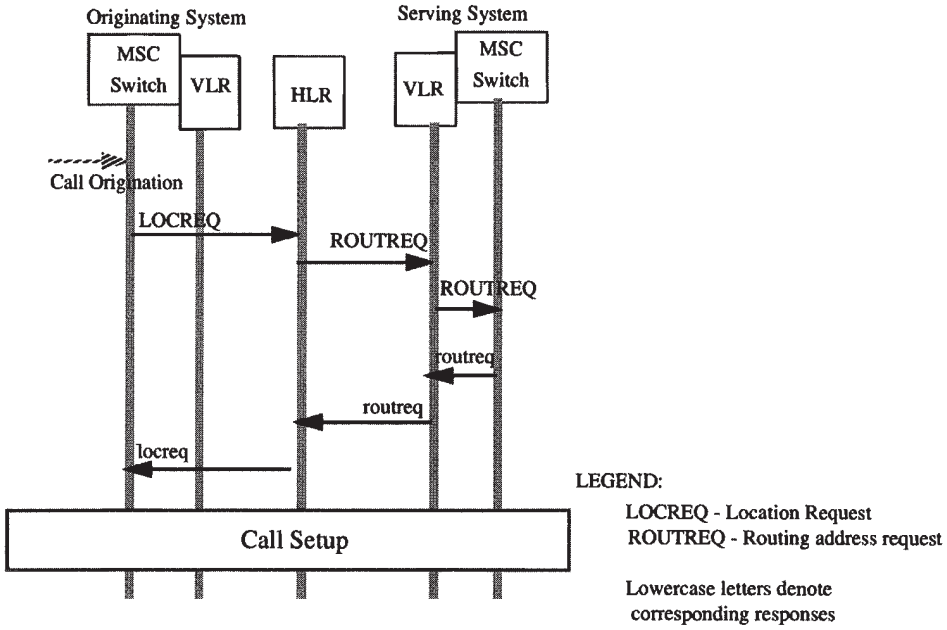


FIGURE 20.4 Signalling flow diagram for call delivery in IS-41.

20.4 Global System for Mobile Communications

In this section we describe the user location strategy proposed in the European Global System for Mobile Communications (GSM) standard and its offshoot, digital cellular system 1800 (DCS1800). There has recently been increased interest in GSM in North America, since it is possible that early deployment of PCS will be facilitated by using the communication equipment already available from European manufacturers who use the GSM standard. Since the GSM standard is relatively unfamiliar to North American readers, we first give some background and introduce the various abbreviations. The reader will find additional details in Mouley and Pautet, [18]. For an overview on GSM, refer to Lycksell, [15].

The abbreviation GSM originally stood for Groupe Special Mobile, a committee created within the pan-European standardization body Conference Europeenne des Posts et Telecommunications (CEPT) in 1982. There were numerous national cellular communication systems and standards in Europe at the time, and the aim of GSM was to specify a uniform standard around the newly reserved 900-MHz frequency band with a bandwidth of twice 25 MHz. The phase 1 specifications of this standard were frozen in 1990. Also in 1990, at the request of the United Kingdom, specification of a version of GSM adapted to the 1800-MHz frequency, with bandwidth of twice 75 MHz, was begun. This variant is referred to as DCS1800; the abbreviation GSM900 is sometimes used to distinguish between the two variations, with the abbreviation GSM being used to encompass both GSM900 and DSC1800. The motivation for DCS1800 is to provide higher capacities in densely populated urban areas, particularly for PCS. The DCS1800 specifications were frozen in 1991, and by 1992 all major GSM900 European operators began operation.

At the end of 1991, activities concerning the post-GSM generation of mobile communications were begun by the standardization committee, using the name universal mobile telecommunications system (UMTS) for this effort. In 1992, the name of the standardization committee was changed from GSM to special mobile group (SMG) to distinguish it from the 900-MHz system itself, and the term GSM was chosen as the commercial trademark

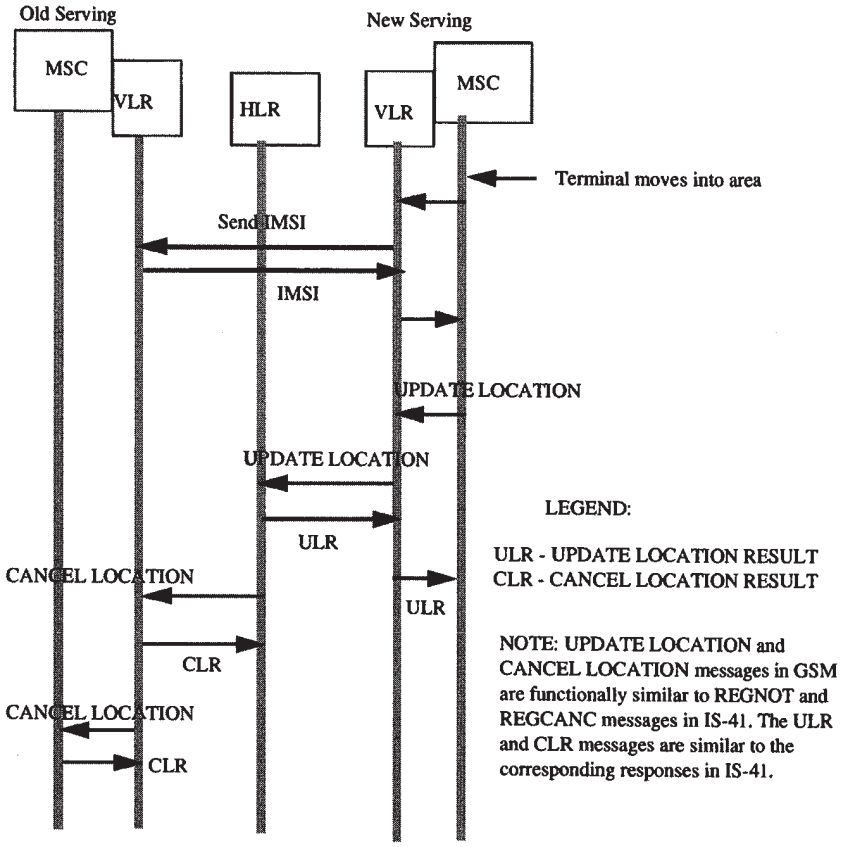


FIGURE 20.5 Flow diagram for registration in GSM.

of the European 900-MHz system, where GSM now stands for global system for mobile communications.

The GSM standard has now been widely adopted in Europe and is under consideration in several other non-European countries, including the United Arab Emirates, Hong Kong, and New Zealand. In 1992, Australian operators officially adopted GSM.

20.4.1 Architecture

In this section we describe the GSM architecture, focusing on those aspects that differ from the architecture assumed in the IS-41 standard.

A major goal of the GSM standard was to enable users to move across national boundaries and still be able to communicate. It was considered desirable, however, that the operational network within each country be operated independently. Each of the operational networks is called a public land mobile network (PLMN) and its commercial coverage area is confined to the borders of one country (although some radio coverage overlap at national boundaries may occur), and each country may have several competing PLMNs.

A GSM customer subscribes to a single PLMN called the home PLMN, and subscription information includes the services the customer subscribes to. During normal operation, a user may elect to choose other PLMNs as their service becomes available (either as the user moves or as new operators enter the marketplace). The user's terminal [GSM calls the terminal a mobile station (MS)] assists the user in choosing a PLMN in this case, either

presenting a list of possible PLMNs to the user using explicit names (e.g., DK Sonofon for the Danish PLMN) or choosing automatically based on a list of preferred PLMNs stored in the terminal's memory. This PLMN selection process allows users to choose between the services and tariffs of several competing PLMNs. Note that the PLMN selection process differs from the cell selection and handoff process that a terminal carries out automatically without any possibility of user intervention, typically based on received radio signal strengths and, thus, requires additional intelligence and functionality in the terminal.

The geographical area covered by a PLMN is partitioned into MSC serving areas, and a registration area is constrained to be a subset of a single MSC serving area. The PLMN operator has complete freedom to allocate cells to registration areas. Each PLMN has, logically speaking, a single HLR, although this may be implemented as several physically distributed databases, as for IS-41. Each MSC also has a VLR, and a VLR may serve one or several MSCs. As for IS-41, it is interesting to consider how the VLR should be viewed in this context. The VLR can be viewed as simply a database off loading the query and signalling load on the HLR and, hence, logically tightly coupled to the HLR or as an ancillary processor to the MSC. This distinction is not academic; in the first view, it would be natural to implement a VLR as serving several MSCs, whereas in the second each VLR would serve one MSC and be physically closely coupled to it. For GSM, the MSC implements most of the signalling protocols, and at present all switch manufacturers implement a combined MSC and VLR, with one VLR per MSC [18].

A GSM mobile station is split in two parts, one containing the hardware and software for the radio interface and the other containing subscribers-specific and location information, called the subscriber identity module (SIM), which can be removed from the terminal and is the size of a credit card or smaller. The SIM is assigned a unique identity within the GSM system, called the international mobile subscriber identity (IMSI), which is used by the user location strategy as described the next subsection. The SIM also stores authentication information, services lists, PLMN selection lists, etc., and can itself be protected by password or PIN.

The SIM can be used to implement a form of large-scale mobility called SIM roaming. The GSM specifications standardize the interface between the SIM and the terminal, so that a user carrying his or her SIM can move between different terminals and use the SIM to personalize the terminal. This capability is particularly useful for users who move between PLMNs which have different radio interfaces. The user can use the appropriate terminal for each PLMN coverage area while obtaining the personalized facilities specified in his or her SIM. Thus, SIMs address personal mobility. In the European context, the usage of two closely related standards at different frequencies, namely, GSM900 and DCS1800, makes this capability an especially important one and facilitates interworking between the two systems.

20.4.2 User Location Strategy

We present a synopsis of the user location strategy in GSM using call flow diagrams similar to those used to describe the strategy in IS-41.

In order to describe the registration procedure, it is first useful to clarify the different identifiers used in this procedure. The SIM of the terminal is assigned a unique identity, called the IMSI, as already mentioned. To increase confidentiality and make more efficient use of the radio bandwidth, however, the IMSI is not normally transmitted over the radio link. Instead, the terminal is assigned a temporary mobile subscriber identity (TMSI) by the VLR when it enters a new registration area. The TMSI is valid only within a given registration area and is shorter than the IMSI. The IMSI and TMSI are identifiers that are

internal to the system and assigned to a terminal or SIM and should not be confused with the user's number that would be dialed by a calling party; the latter is a separate number called the mobile subscriber integrated service digital network (ISDN) number (MSISDN), and is similar to the usual telephone number in a fixed network.

We now describe the procedure during registration. The terminal can detect when it has moved into the cell of a new registration area from the system information broadcast by the base station in the new cell. The terminal initiates a registration update request to the new base station; this request includes the identity of the old registration area and the TMSI of the terminal in the old area. The request is forwarded to the MSC, which, in turn, forwards it to the new VLR. Since the new VLR cannot translate the TMSI to the IMSI of the terminal, it sends a request to the old VLR to send the IMSI of the terminal corresponding to that TMSI. In its response, the old VLR also provides the required authentication information. The new VLR then initiates procedures to authenticate the terminal. If the authentication succeeds, the VLR uses the IMSI to determine the address of the terminal's HLR.

The ensuing protocol is then very similar to that in IS-41, except for the following differences. When the new VLR receives the registration affirmation (similar to regnot in IS-41) from the HLR, it assigns a new TMSI to the terminal for the new registration area. The HLR also provides the new VLR with all relevant subscriber profile information required for call handling (e.g., call screening lists, etc.) as part of the affirmation message. Thus, in contrast with IS-41, authentication and subscriber profile information are obtained from both the HLR and old VLR and not just the HLR.

The procedure for delivering calls to mobile users in GSM is very similar to that in IS-41. The sequence of messages between the caller and called party's MSC/VLRs and the HLR is identical to that shown in the call flow diagrams for IS-41, although the names, contents and lengths of messages may be different and, hence, the details are left out. The interested reader is referred to Mouly and Pautet, [18], or Lycksell, [15], for further details.

20.5 Analysis of Database Traffic Rate for IS-41 and GSM

In the two subsections that follow, we state the common set of assumptions on which we base our comparison of the two strategies.

20.5.1 The Mobility Model for PCS Users

In the analysis that follows in the IS-41 analysis subsection, we assume a simple mobility model for the PCS users. The model, which is described in [23], assumes that PCS users carrying terminals are moving at an average velocity of v and their direction of movement is uniformly distributed over $[0, 2\pi]$. Assuming that the PCS users are uniformly populated with a density of ρ and the registration area boundary is of length L, it has been shown that the rate of registration area crossing R is given by

$$R = \frac{\rho v L}{\pi} \tag{20.1}$$

Using Eq. (20.1), we can calculate the signalling traffic due to registration, call origination, and delivery. We now need a set of assumptions so that we may proceed to derive the traffic rate to the databases using the model in Fig. 20.2.

20.5.2 Additional Assumptions

The following assumptions are made in performing the analysis.

- 128 total registration areas
- Square registration area size: $(7.575 \text{ km})^2 = 57.5 \text{ km}^2$, with border length $L = 30.3$ km
- Average call origination rate = average call termination (delivery) rate = 1.4/h/terminal
- Mean density of mobile terminals = $\rho = 390/\text{km}^2$
- Total number of mobile terminals = $128 \times 57.4 \times 390 = 2.87 \times 10^6$
- Average call origination rate = average call termination (delivery) rate = 1.4/h/terminal
- Average speed of a mobile, $v = 5.6$ km/h
- Fluid flow mobility model

The assumptions regarding the total number of terminals may also be obtained by assuming that a certain public network provider serves 19.15×10^6 users and that 15% (or 2.87×10^6) of the users also subscribe to PCS services from various PSPs.

Note that we have adopted a simplified model that ignores situations where PCS users may turn their handsets on and off that will generate additional registration and deregistration traffic. The model also ignores wireline registrations. These activities will increase the total number of queries and updates to HLR and VLRs.

20.5.3 Analysis of IS-41

Using Eq. (20.1) and the parameter values assumed in the preceding subsection, we can compute the traffic due to registration. The registration traffic is generated by mobile terminals moving into a new registration area, and this must equal the mobile terminals moving out of the registration area, which per second is

$$R_{\text{reg, VLR}} = \frac{390 \times 30.3 \times 5.6}{3600\pi} = 5.85$$

This must also be equal to the number of deregistrations (registration cancellations),

$$R_{\text{dereg, VLR}} = 5.85$$

The total number of registration messages per second arriving at the HLR will be

$$R_{\text{reg, HLR}} = R_{\text{reg, VLR}} \times \text{ total No. of registration areas} = 749$$

The HLR should, therefore, be able to handle, roughly, 750 updates per second. We observe from Fig. 20.3 that authenticating terminals generate as many queries to VLR and HLR as the respective number of updates generated due to registration notification messages.

The number of queries that the HLR must handle during call origination and delivery can be similarly calculated. Queries to HLR are generated when a call is made to a PCS user. The SSP that receives the request for a call, generates a location request (LOCREQ) query to the SCP controlling the HLR. The rate per second of such queries must be equal to the rate of calls made to PCS users.

This is calculated as

$$\begin{aligned} R_{\text{CallDeliv, HLR}} &= \text{call rate per user} \times \text{total number of users} \\ &= \frac{1.4 \times 2.87 \times 10^5}{3600} \\ &= 1116 \end{aligned}$$

For calls originated from a mobile terminal by PCS users, the switch authenticates the terminal by querying the VLR. The rate per second of such queries is determined by the rate of calls originating in an SSP serving area, which is also a registration area (RA). This is given by

$$R_{\text{CallOrig, VLR}} = \frac{1116}{128} = 8.7$$

This is also the number of queries per second needed to authenticate terminals of PCS users to which calls are delivered:

$$R_{\text{CallDeliv, VLR}} = 8.7$$

Table 20.1 summarizes the calculations.

TABLE 20.1 IS-41 Query and Update Rates to HLR and VLR

Activity	HLR Updates/s	VLR Updates/s	HLR Queries/s	VLR queries/s
Mobility-related activities at registration	749	5.85	749	5.85
Mobility-related activities at deregistration		5.85		
Call origination				8.7
Call delivery			1116	8.7
Total (per RA)	5.85	11.7	14.57	23.25
Total (Network)	749	1497.6	1865	2976

20.5.4 Analysis of GSM

Calculations for query and update rates for GSM may be performed in the same manner as for IS-41, and they are summarized in Table 20.2. The difference between this table and Table 20.1 is that in GSM the new serving VLR does not query the HLR separately in order to authenticate the terminal during registration and, hence, there are no HLR queries during registration. Instead, the entry (749 queries) under HLR queries in Table 20.1, corresponding to mobility-related authentication activity at registration, gets equally divided between the 128 VLRs. Observe that with either protocol the total database traffic rates are conserved, where the total database traffic for the entire network is given by the sum of all of the entries in the last row total (Network), i.e.,

$$\text{HLR updates} + \text{VLR updates} + \text{HLR queries} + \text{VLR queries}$$

From Tables 20.1 and 20.2 we see that this quantity equals 7087.

The conclusion is independent of any variations we may provide to the assumptions in earlier in the section. For example, if the PCS penetration (the percentage of the total users

TABLE 20.2 GSM Query and Update Rates to HLR and VLR

Activity	HLR Updates/s	VLR Updates/s	HLR Queries/s	VLR Queries/s
Mobility-related activities at registration	749	5.85		11.7
Mobility-related activities at deregistration		5.85		
Call origination				8.7
Call delivery			1116	8.7
Total (per VLR)	749	11.7	1116	29.1
Total (Network)	749	1497.6	1116	3724.8

subscribing to PCS services) were to increase from 15 to 30%, all of the entries in the two tables will double and, hence, the total database traffic generated by the two protocols will still be equal.

20.6 Reducing Signalling During Call Delivery

In the preceding section, we provided a simplified analysis of some scenarios associated with user location strategies and the associated database queries and updates required. Previous studies [13, 16] indicate that the signalling traffic and database queries associated with PCS due to user mobility are likely to grow to levels well in excess of that associated with a conventional call. It is, therefore, desirable to study modifications to the two protocols that would result in reduced signalling and database traffic. We now provide some suggestions.

For both GSM and IS-41, delivery of calls to a mobile user involves four messages: from the caller's VLR to the called party's HLR, from the HLR to the called party's VLR, from the called party's VLR to the HLR, and from the HLR to the caller's VLR. The last two of these messages involve the HLR, whose role is to simply relay the routing information provided by the called party's VLR to the caller's VLR. An obvious modification to the protocol would be to have the called VLR directly send the routing information to the calling VLR. This would reduce the total load on the HLR and on signalling network links substantially. Such a modification to the protocol may not be easy, of course, due to administrative, billing, legal, or security concerns. Besides, this would violate the query/response model adopted in IS-41, requiring further analysis.

A related question which arises is whether the routing information obtained from the called party's VLR could instead be stored in the HLR. This routing information could be provided to the HLR, for example, whenever a terminal registers in a new registration area. If this were possible, two of the four messages involved in call delivery could be eliminated. This point was discussed at length by the GSM standards body, and the present strategy was arrived at. The reason for this decision was to reduce the number of temporary routing numbers allocated by VLRs to terminals in their registration area. If a temporary routing number (TLDN in IS-41 or MSRN in GSM) is allocated to a terminal for the whole duration of its stay in a registration area, the quantity of numbers required is much greater than if a number is assigned on a per-call basis. Other strategies may be employed to reduce signalling and database traffic via intelligent paging or by storing user's mobility behavior in user profiles (see, for example, Tabbane, [22]). A discussion of these techniques is beyond the scope of the paper.

20.7 Per-User Location Caching

The basic idea behind per-user location caching is that the volume of SS7 message traffic and database accesses required in locating a called subscriber can be reduced by maintaining local storage, or cache, of user location information at a switch. At any switch, location caching for a given user should be employed only if a large number of calls originate for that user from that switch, relative to the user's mobility. Note that the cached information is kept at the switch from which calls originate, which may or may not be the switch where the user is currently registered.

Location caching involves the storage of location pointers at the originating switch; these point to the VLR (and the associated switch) where the user is currently registered. We refer to the procedure of locating a PCS user a *FIND* operation, borrowing the terminology from Awerbuch and Peleg, [1]. We define a basic *FIND,* or *BasicFIND*( ), as one where the following sequence of steps takes place.

1. The incoming call to a PCS user is directed to the nearest switch.
2. Assuming that the called party is not located within the immediate RA, the switch queries the HLR for routing information.
3. The HLR contains a pointer to the VLR in whose associated RA the subscriber is currently situated and launches a query to that VLR.
4. The VLR, in turn, queries the MSC to determine whether the user terminal is capable of receiving the call (i.e., is idle) and, if so, the MSC returns a routable address (TLDN in IS-41) to the VLR.
5. The VLR relays the routing address back to the originating switch via the HLR.

At this point, the originating switch can route the call to the destination switch. Alternately, *BasicFIND*( ) can be described by pseudocode as follows. (We observe that a more formal method of specifying PCS protocols may be desirable).

```
BasicFIND( ){
    Call to PCS user is detected at local switch;
    if called party is in same RA then return;
    Switch queries called party's HLR;
    Called party's HLR queries called party's current VLR, V;
    V returns called party's location to HLR;
    HLR returns location to calling switch;
}
```

In the *FIND* procedure involving the use of location caching, or *CacheFIND*( ), each switch contains a local memory (cache) that stores location information for subscribers. When the switch receives a call origination (from either a wire-line or wireless caller) directed to a PCS subscriber, it first checks its cache to see if location information for the called party is maintained. If so, a query is launched to the pointed VLR; if not, *BasicFIND*( ), as just described, is followed. If a cache entry exists and the pointed VLR is queried, two situations are possible. If the user is still registered at the RA of the pointed VLR (i.e., we have a *cache hit*), the pointed VLR returns the user's routing address. Otherwise, the pointed VLR returns a *cache miss.*

```
CacheFIND( ){
    Call to PCS user is detected at local switch;
```

```
        if called is in same RA then return;
        if there is no cache entry for called user
        then invoke BasicFIND( ) and return;
        Switch queries the VLR, V, specified in the cache entry;
        if called is at V, then
            V returns called party's location to calling switch;
        else {
            V returns "miss" to calling switch;
            Calling switch invokes BasicFIND( );
        }
}
```

When a cache hit occurs we save one query to the HLR [a VLR query is involved in both *CacheFIND*( ) and *BasicFIND*( )], and we also save traffic along some of the signalling links; instead of four message transmissions, as in *BasicFIND*( ), only two are needed. In steady-state operation, the cached pointer for any given user is updated only upon a miss.

Note that the *BasicFIND*( ) procedure differs from that specified for roaming subscribers in the IS-41 standard EIA/TIA, [6]. In the IS-41 standard, the second line in the *BasicFIND*( ) procedure is omitted, i.e., every call results in a query of the called user's HLR. Thus, in fact, the procedure specified in the standard will result in an even higher network load than the *BasicFIND*( ) procedure specified here. To make a fair assessment of the benefits of *CacheFIND*( ), however, we have compared it against *BasicFIND*( ). Thus, the benefits of *CacheFIND*( ) investigated here depend specifically on the use of caching and not simply on the availability of user location information at the local VLR.

20.8 Caching Threshold Analysis

In this section we investigate the classes of users for which the caching strategy yields net reductions in signalling traffic and database loads. We characterize classes of users by their CMR. The CMR of a user is the average number of calls to a user per unit time, divided by the average number of times the user changes registration areas per unit time. We also define a LCMR, which is the average number of calls to a user from a given originating switch per unit time, divided by the average number of times the user changes registration areas per unit time.

For each user, the amount of savings due to caching is a function of the probability that the cached pointer correctly points to the user's location and increases with the user's LCMR. In this section we quantify the minimum value of LCMR for caching to be worthwhile. This caching threshold is parameterized with respect to costs of traversing signalling network elements and network databases and can be used as a guide to select the subset of users to whom caching should be applied. The analysis in this section shows that estimating user's LCMRs, preferably dynamically, is very important in order to apply the caching strategy. The next section will discuss methods for obtaining this estimate.

From the pseudocode for *BasicFIND*( ), the signalling network cost incurred in locating a PCS user in the event of an incoming call is the sum of the cost of querying the HLR (and receiving the response), and the cost of querying the VLR which the HLR points to (and receiving the response). Let

α = cost of querying the HLR and receiving a response

β = cost of querying the pointed VLR and receiving a response

Then, the cost of *BasicFIND( )* operation is

$$C_B = \alpha + \beta \tag{20.2}$$

To quantify this further, assume costs for traversing various network elements as follows.

A_l = cost of transmitting a location request or response message on A link between SSP and LSTP
D = cost of transmitting a location request on response message or D link
A_r = cost of transmitting a location request or response message on A link between RSTP and SCP
L = cost of processing and routing a location request or response message by LSTP
R = cost of processing and routing a location request or response message by RSTP
H_Q = cost of a query to the HLR to obtain the current VLR location
V_Q = cost of a query to the VLR to obtain the routing address

Then, using the PCS reference network architecture (Fig. 80.2),

$$\alpha = 2(A_l + D + A_r + L + R) + H_Q \tag{20.3}$$
$$\beta = 2(A_l + D + A_r + L + R) + V_Q \tag{20.4}$$

From Eqs. (20.2)–(20.4)

$$C_B = 4(A_l + D + A_r + L + R) + H_Q + V_Q \tag{20.5}$$

We now calculate the cost of *CacheFIND( )*. We define the *hit ratio* as the relative frequency with which the cached pointer correctly points to the user's location when it is consulted. Let

p = cache hit ratio
C_H = cost of the *CacheFIND( )* procedure when there is a hit
C_M = cost of the *CacheFIND( )* procedure when there is a miss

Then the cost of *CacheFIND( )* is

$$C_C = p\,C_H + (1-p)C_M \tag{20.6}$$

For *CacheFIND( )*, the signalling network costs incurred in locating a user in the event of an incoming call depend on the hit ratio as well as the cost of querying the VLR, which is stored in the cache; this VLR query may or may not involve traversing the RSTP. In the following, we say a VLR is a *local* VLR if it is served by the same LSTP as the originating switch, and a *remote* VLR otherwise. Let

q = Prob (VLR in originating switch's cache is a local VLR)
δ = cost of querying a local VLR
ϵ = cost of querying a remote VLR
η = cost of updating the cache upon a miss

Then,

$$\delta = 4A_l + 2L + V_Q \tag{20.7}$$
$$\epsilon = 4(A_l + D + L) + 2R + V_Q \tag{20.8}$$
$$C_H = q\delta + (1-q)\epsilon \tag{20.9}$$

Since updating the cache involves an operation to a fast local memory rather than a database operation, we shall assume in the following that $\eta = 0$. Then,

$$C_M = C_H + C_B = q\delta + (1-q)\epsilon + \alpha + \beta \tag{20.10}$$

From Eqs. (20.6), (20.9) and (20.10) we have

$$C_C = \alpha + \beta + \epsilon - p(\alpha + \beta) + q(\delta - \epsilon) \tag{20.11}$$

For net cost savings we require $C_C < C_B$, or that the hit ratio exceeds a *hit ratio threshold* p_T, derived using Eqs. (20.6), (20.9), and (20.2),

$$p > p_T = \frac{C_H}{C_B} = \frac{\epsilon + q(\delta - \epsilon)}{\alpha + \beta} \tag{20.12}$$

$$= \frac{4A_l + 4D + 4L + 2R + V_Q - q(4D + 2L + 2R)}{4A_l + 4D + 4A_r + 4L + 4R + H_Q + V_Q} \tag{20.13}$$

Equation (20.13) specifies the hit ratio threshold for a user, evaluated at a given switch, for which local maintenance of a cached location entry produces cost savings. As pointed out earlier, a given user's hit ratio may be location dependent, since the rates of calls destined for that user may vary widely across switches.

The hit ratio threshold in Eq. (20.13) is comprised of heterogeneous cost terms, i.e., transmission link utilization, packet switch processing, and database access costs. Therefore, numerical evaluation of the hit ratio threshold requires either detailed knowledge of these individual quantities or some form of simplifying assumptions. Based on the latter approach, the following two possible methods of evaluation may be employed.

1. Assume one or more cost terms dominate, and simplify Eq. (20.13) by setting the remaining terms to zero.
2. Establish a common unit of measure for all cost terms, for example, *time delay.* In this case, A_l, A_r, and D may represent transmission delays of fixed transmission speed (e.g., 56 kb/s) signalling links, L and R may constitute the sum of queueing and service delays of packet switches (i.e., STPs), and H_Q and V_Q the transaction delays for database queries.

In this section we adopt the first method and evaluate Eq. (20.13) assuming a single term dominates. (In Section 20.9 we present results using the second method). Table 20.3 shows the hit ratio threshold required to obtain net cost savings, for each case in which one of the cost terms is dominant.

In Table 20.3 we see that if the cost of querying a VLR or of traversing a local A link is the dominant cost, caching for users who may move is never worthwhile, regardless of users' call reception and mobility patterns. This is because the caching strategy essentially distributes the functionality of the HLR to the VLRs. Thus, the load on the VLR and the local A link is always increased, since any move by a user results in a cache miss. On the other hand, for a fixed user (or telephone), caching is always worthwhile. We also observe that if the remote A links or HLR querying are the bottlenecks, caching is worthwhile even for users with very low hit ratios.

As a simple average-case calculation, consider the net network benefit of caching when HLR access and update is the performance bottleneck. Consider a scenario where $u = 50\%$ of PCS users receive $c = 80\%$ of their calls from $s = 5$ RAs where their hit ratio $p > 0$, and $s' = 4$ of the SSPs at those RAs contain sufficiently large caches. Assume that caching

TABLE 20.3 Minimum Hit Ratios and LCMRs for Various Individual Dominant Signalling Network Cost Terms

Dominant Cost Term	Hit ratio Threshold, p_T	LCMR Threshold, $LCMR_T$	LCMR Threshold ($q = 0.043$)	LCMR Threshold ($q = 0.25$)
A_l	1	∞	∞	∞
A_r	0	0	0	0
D	$1 - q$	$1/q - 1$	22	3
L	$1 - q/2$	$2/q - 1$	45	7
R	$1 - q/2$	$2/q - 1$	45	7
H_Q	0	0	0	0
V_Q	1	∞	∞	∞

is applied only to this subset of users and to no other users. Suppose that the average hit ratio for these users is $p = 80\%$, so that 80% of the HLR accesses for calls to these users from these RA are avoided. Then the net saving in the accesses to the system's HLR is $H = (u\,c\,s'\,p)/s = 25\%$.

We discuss other quantities in Table 20.3 next. It is first useful to relate the cache hit ratio to users' calling and mobility patterns directly via the LCMR. Doing so requires making assumptions about the distribution of the user's calls and moves. We consider the steady state where the incoming call stream from an SSP to a user is a Poisson process with arrival rate λ, and the time that the user resides in an RA has a general distribution $F(t)$ with mean $1/\mu$. Thus,

$$LCMR = \frac{\lambda}{\mu} \tag{20.14}$$

Let t be the time interval between two consecutive calls from the SSP to the user and t_1 be the time interval between the first call and the time when the user moves to a new RA. From the random observer property of the arrival call stream [7], the hit ratio is

$$p = \; Pr[t < t_1] = \int_{t=0}^{\infty} \lambda e^{-\lambda t} \int_{t_1=t}^{\infty} \mu\,[1 - F\,(t_1)]\;\mathrm{d}t_1\,\mathrm{d}t$$

If $F(t)$ is an exponential distribution, then

$$p = \frac{\lambda}{\lambda + \mu} \tag{20.15}$$

and we can derive the *LCMR threshold*, the minimum LCMR required for caching to be beneficial assuming incoming calls are a Poisson process and intermove times are exponentially distributed,

$$LCMR_T = \frac{p_T}{1 - p_T} \tag{20.16}$$

Equation (20.16) is used to derive LCMR thresholds assuming various dominant costs terms, as shown in Table 20.3.

Several values for $LCMR_T$ in Table 20.3 involve the term q, i.e., the probability that the pointed VLR is a local VLR. These values may be numerically evaluated by simplifying assumptions. For example, assume that all of the SSPs in the network are uniformly distributed amongst l LSTPs. Also, assume that all of the PCS subscribers are uniformly distributed in location across all SSPs and that each subscriber exhibits the same incoming call rate at every SSP. Under those conditions, q is simply $1/l$. Consider the case of the

public switched telephone network. Given that there are a total of 160 local access transport area (LATA) across the 7 Regional Bell Operating Company (RBOC) regions [4], the average number of LATAs, or l, is 160/7 or 23. Table 20.3 shows the results with $q = 1/l$ in this case.

We observe that the assumption that all users receive calls uniformly from all switches in the network is extremely conservative. In practice, we expect that user call reception patterns would display significantly more locality, so that q would be larger and the LCMR thresholds required to make caching worthwhile would be smaller. It is also worthwhile to consider the case of a RBOC region with PCS deployed in a few LATA only, a likely initial scenario, say, 4 LATAs. In either case the value of q would be significantly higher; Table 20.3 shows the LCMR threshold when $q = 0.25$.

It is possible to quantify the net costs and benefits of caching in terms of signalling network impacts in this way and to determine the hit ratio and LCMR threshold above which users should have the caching strategy applied. Applying caching to users whose hit ratio and LCMR is below this threshold results in net increases in network impacts. It is, thus, important to estimate users' LCMRs accurately. The next section discusses how to do so.

20.9 Techniques for Estimating Users' LCMR

Here we sketch some methods of estimating users' LCMR. A simple and attractive policy is to not estimate these quantities on a per-user basis at all. For instance, if the average LCMR over all users in a PCS system is high enough (and from Table 20.3, it need not be high depending on which network elements are the dominant costs), then caching could be used at every SSP to yield net system-wide benefits. Alternatively, if it is known that at any given SSP the average LCMR over all users is high enough, a cache can be installed at that SSP. Other variations can be designed.

One possibility for deciding about caching on a per-user basis is to maintain information about a user's calling and mobility pattern at the HLR and to download it periodically to selected SSPs during off-peak hours. It is easy to envision numerous variations on this idea.

In this section we investigate two possible techniques for estimating LCMR on a per-user basis when caching is to be deployed. The first algorithm, called the *running average* algorithm, simply maintains a running average of the hit ratio for each user. The second algorithm, called the *reset-K* algorithm, attempts to obtain a measure of the hit ratio over the recent history of the user's movements. We describe the two algorithms next and evaluate their effectiveness using a stochastic analysis taking into account user calling and mobility patterns.

20.9.1 The Running Average Algorithm

The running average algorithm maintains, for every user that has a cache entry, the running average of the hit ratio. A running count is kept of the number of calls to a given user, and, regardless of the *FIND* procedure used to locate the user, a running count of the number of times that the user was at the same location for any two consecutive calls; the ratio of these numbers provides the measured running average of the hit ratio. We denote the measured running average of the hit ratio by p_M; in steady state, we expect that $p_M = p$. The user's previous location as stored in the cache entry is used only if the running average of the hit ratio p_M is greater than the cache hit threshold p_T. Recall that the cache scheme

outperforms the basic scheme if $p > p_T = C_H/C_B$. Thus, in steady state, the running average algorithm will outperform the basic scheme when $p_M > p_T$.

We consider, as before, the steady state where the incoming call stream from an SSP to a user is a Poisson process with arrival rate λ, and the time that the user resides in an RA has an exponential distribution with mean $1/\mu$. Thus $LCMR = \lambda/\mu$ [Eq. (20.14)] and the location tracking cost at steady state is

$$C_C = \begin{cases} p_M C_H + (1 - p_M) C_B, & p_M > p_T \\ C_B, & \text{otherwise} \end{cases} \tag{20.17}$$

Figure 20.6 plots the cost ratio C_C/C_B from Eq. (20.17) against $LCMR$. (This corresponds to assigning uniform units to all cost terms in Eq. (20.13), i.e., the second evaluation method as discussed in Section 20.8. Thus, the ratio C_C/C_B may represent the percentage reduction in user location time with the caching strategy compared to the basic strategy.) The figure indicates that in the steady state, the caching strategy with the running average algorithm for estimating LCMR can significantly outperform the basic scheme if $LCMR$ is sufficiently large. For instance with $LCMR \sim 5$, caching can lead to cost savings of 20–60% over the basic strategy.

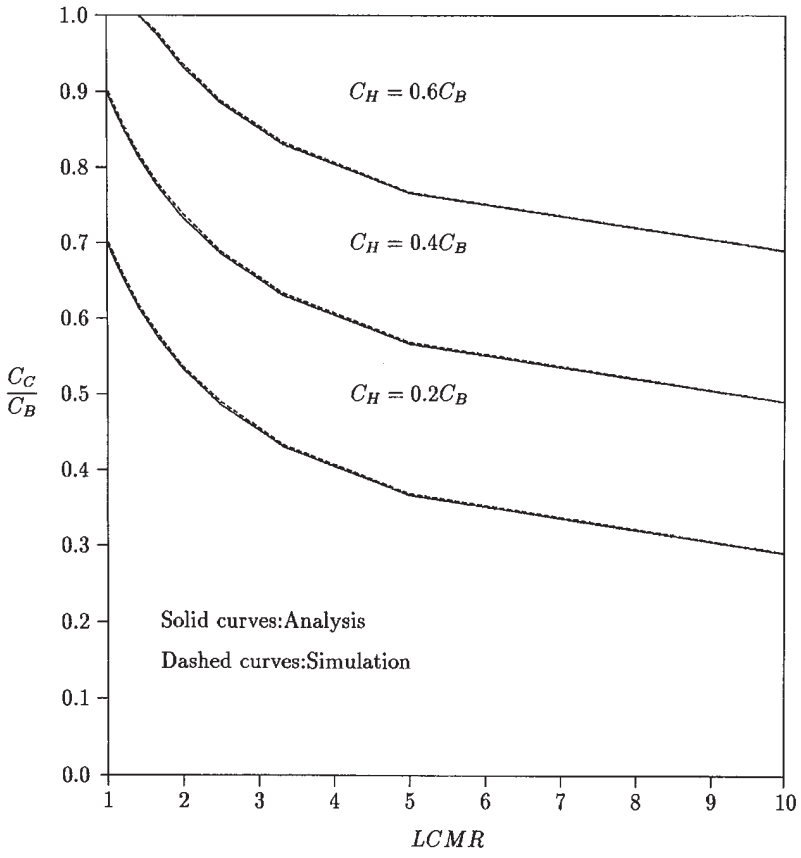


FIGURE 20.6 The location tracking cost for the running average algorithm.

Equation (20.17) (cf., solid curves in Fig. 20.6) is validated against a simple Monte Carlo simulation (cf., dashed curves in Fig. 20.6). In the simulation, the confidence interval for the 95% confidence level of the output measure C_C/C_B is within 3% of the mean value. This simulation model will later be used to study the running average algorithm when the mean of the movement distribution changes from time to time [which cannot be modeled by using Eq. (20.17)].

One problem with the running average algorithm is that the parameter p is measured from the entire past history of the user's movement, and the algorithm may not be sufficiently dynamic to adequately reflect the recent history of the user's mobility patterns.

20.9.2 The Reset-K Algorithm

We may modify the running average algorithm such that p is measured from the recent history. Define every K incoming calls as a *cycle*. The modified algorithm, which is referred to as the reset-K algorithm, counts the number of cache hits n in a cycle. If the measured hit ratio for a user, $p_M = n/K \geq p_T$, then the cache is enabled for that user, and the cached information is always used to locate the user in the next cycle. Otherwise, the cache is disabled for that user, and the basic scheme is used. At the beginning of a cycle, the cache hit count is reset, and a new p_M value is measured during the cycle.

To study the performance of the reset-K algorithm, we model the number of cache misses in a cycle by a Markov process. Assume as before that the call arrivals are a Poisson process with arrival rate λ and the time period the user resides in an RA has an exponential distribution with mean $1/\mu$. A pair (i, j), where $i > j$, represents the state that there are j cache misses before the first i incoming phone calls in a cycle. A pair $(i, j)^*$, where $i \geq j \geq 1$, represents the state that there are $j - 1$ cache misses before the first i incoming phone calls in a cycle, and the user moves between the ith and the $i + 1$ phone calls. The difference between (i, j) and $(i, j)^*$ is that if the Markov process is in the state (i, j) and the user moves, then the process moves into the state $(i, j + 1)^*$. On the other hand, if the process is in state $(i, j)^*$ when the user moves, the process remains in $(i, j)^*$ because at most one cache miss occurs between two consecutive phone calls.

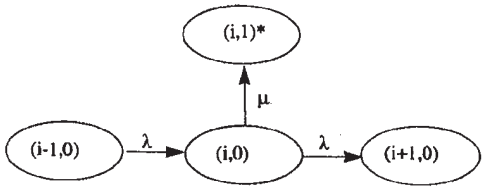
Figure 20.7(a) illustrates the transitions for state $(i, 0)$ where $2 < i < K+1$. The Markov process moves from $(i - 1, 0)$ to $(i, 0)$ if a phone call arrives before the user moves out. The rate is λ. The process moves from $(i, 0)$ to $(i, 1)^*$ if the user moves to another RA before the $i + 1$ call arrival. Let $\pi(i, j)$ denote the probability of the process being in state (i, j). Then the transition equation is

$$\pi(i, 0) = \frac{\lambda}{\lambda + \mu}\pi(i - 1, 0), \qquad 2 < i < K + 1 \tag{20.18}$$

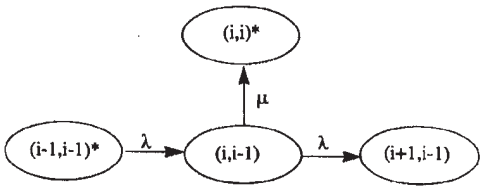
Figure 20.7(b) illustrates the transitions for state $(i, i - 1)$ where $1 < i < K + 1$. The only transition into the state $(i, i - 1)$ is from $(i - 1, i - 1)^*$, which means that the user always moves to another RA after a phone call. [Note that there can be no state $(i - 1, i - 1)$ by definition and, hence, no transition from such a state.] The transition rate is λ. The process moves from $(i, i - 1)$ to $(i, i)^*$ with rate μ, and moves to $(i + 1, i - 1)$ with rate λ. Let $\pi^*(i, j)$ denote the probability of the process being in state $(i, j)^*$. Then the transition equation is

$$\pi(i, i - 1) = \frac{\lambda}{\lambda + \mu}\pi^*(i - 1, i - 1), \qquad 1 < i < K + 1 \tag{20.19}$$

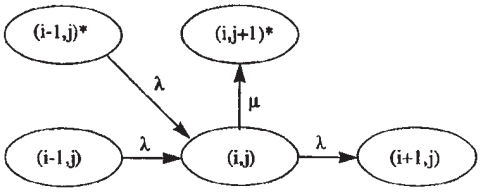
Figure 20.7(c) illustrates the transitions for state (i, j) where $2 < i < K + 1, 0 < j < i - 1$. The process may move into state (i, j) from two states $(i - 1, j)$ and $(i - 1, j)^*$ with rate



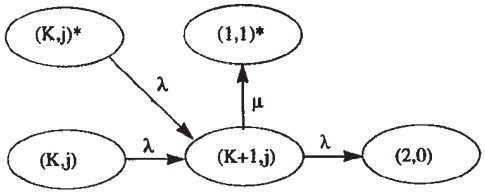
(a) Transitions for state (i,0) (2 < i < K+1)



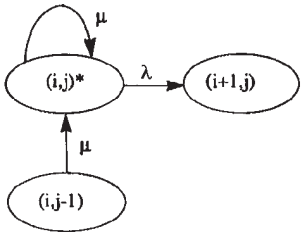
(b) Transitions for state (i,i-1)(1 < i < K+1)



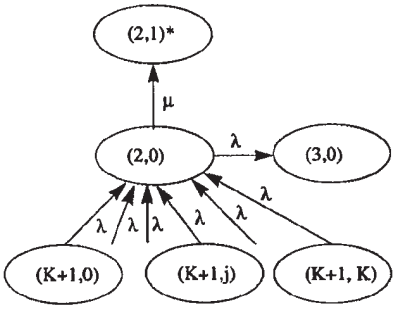
(c) Transitions for state (i,j) (2< i < K+1, 0 < j < i-1)



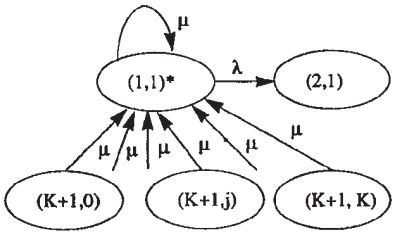
(d) Transitions for state (K+1,j) (0 < j < K+1)



(e) Transitions for state (i,j)* (0 < j ≤ i, 1 < i < K



(f) Transitions for state (2,0)



(g) Transitions for state (1,1)*

FIGURE 20.7 State transitions.

λ, respectively. The process moves from (i,j) to $(i,j+1)^*$ or $(i+1,j)$ with rates μ and λ, respectively. The transition equation is

$$\pi(i,j) = \frac{\lambda}{\lambda+\mu}\left[\pi(i-1,j)+\pi^*(i-1,j)\right], \quad 2<i<K+1, \qquad 0<j<i-1 \tag{20.20}$$

Figure 20.7(d) illustrates the transitions for state $(K+1,j)$ where $0<j<K+1$. Note that if a phone call arrives when the process is in (K,j) or $(K,j)^*$, the system enters a new cycle (with rate λ), and we could represent the new state as (1, 0). In our model, we introduce the state $(K+1,j)$ instead of (1, 0), where

$$\sum_{0\le j\le K} \pi(K+1,j) = \pi(1,0)$$

so that the hit ratio, and thus the location tracking cost, can be derived [see Eq. (20.25)]. The process moves from $(K+1,j)$ [i.e., (1, 0)] to $(1,1)^*$ with rate μ if the user moves before the next call arrives. Otherwise, the process moves to (2, 0) with rate λ. The transition equation is

$$\pi(K+1,j) = \frac{\lambda}{\lambda+\mu}\left[\pi(K,j)+\pi^*(K,j)\right], \qquad 0<j<K+1 \tag{20.21}$$

For $j=0$, the transition from $(K,j)^*$ to $(K+1,0)$ should be removed in Fig. 20.7(d) because the state $(K,0)^*$ does not exist. The transition equation for $(K+1,0)$ is given in Eq. (20.18). Figure 20.7(e) illustrates the transitions for state $(i,j)^*$ where $0<j<i$, $1<i<K+1$. The process can only move to $(i,j)^*$ from $(i,j-1)$ (with rate μ). From the definition of $(i,j)^*$, if the user moves when the process is in $(i,j)^*$, the process remains in $(i,j)^*$ (with rate μ). Otherwise, the process moves to $(i+1,j)$ with rate λ. The transition equation is

$$\pi^*(i,j) = \frac{\mu}{\lambda}\pi(i,j-1), \qquad 0<j\le i, \qquad 1<i<K+1, \qquad i\ge 2 \tag{20.22}$$

The transitions for (2, 0) are similar to the transitions for $(i,0)$ except that the transition from (1, 0) is replaced by $(K+1,0),\ldots,(K+1,K)$ [cf., Fig. 20.7(f)]. The transition equation is

$$\pi(2,0) = \frac{\lambda}{\lambda+\mu}\left[\sum_{0\le j\le K}\pi(K+1,j)\right] \tag{20.23}$$

Finally, the transitions for $(1,1)^*$ is similar to the transitions for $(i,j)^*$ except that the transition from (1, 0) is replaced by $(K+1,0),\ldots,(K+1,K)$ [cf., Fig. 20.7(g)]. The transition equation is

$$\pi^*(1,1) = \frac{\mu}{\lambda}\left[\sum_{0\le j\le K}\pi(K+1,j)\right] \tag{20.24}$$

Suppose that at the beginning of a cycle, the process is in state $(K+1,j)$, then it implies that there are j cache misses in the previous cycle. The cache is enabled if and only if

$$p_M \ge p_T = \frac{C_H}{C_B} \Rightarrow 1-\frac{j}{K}\ge\frac{C_H}{C_B} \Rightarrow 0\le j\le\left\lceil K\left(1-\frac{C_H}{C_B}\right)\right\rceil$$

Thus, the probability that the measured hit ratio $p_M < p_T$ in the previous cycle is

$$Pr\left[p_M < p_T\right] = \frac{\sum_{\lceil k[1-(C_H/C_B)]\rceil < j \le K} \pi(K+1,j)}{\sum_{0 \le j \le K} \pi(K+1,j)}$$

and the location tracking cost for the reset-K algorithm is

$$C_C = C_B Pr\left[p_M < p_T\right] + (1 - Pr\left[p_M < p_T\right]) \times \left\{ \sum_{0 \le j \le K} \left(\frac{(K-j)C_H}{K} + \frac{j(C_H + C_B)}{K} \right) \left[\frac{\pi(K+1,j)}{\sum_{0 \le i \le K} \pi(K+1,i)} \right] \right\} \quad (20.25)$$

The first term Eq. (20.25) represents the cost incurred when caching is disabled because the hit ratio threshold exceeds the hit ratio measured in the previous cycle. The second term is the cost when the cache is enabled and consists of two parts, corresponding to calls during which hits occur and calls during which misses occur. The ratio in square brackets is the conditional probability of being in state $\pi(K+1,j)$ during the current cycle.

The numerical computation of $\pi(K+1,j)$ can be done as follows. First, compute $a_{i,j}$ and $b_{i,j}$ where $\pi(i,j) = a_{i,j}\pi^*(1,1)$ and $\pi^*(i,j) = b_{i,j}\pi^*(1,1)$. Note that $a_{i,j} = 0 (b_{i,j} = 0)$ if $\pi(i,j)[\pi^*(i,j)]$ is not defined in Eqs. (20.18)–(20.24). Since

$$\sum_{i,j} [\pi(i,j) + \pi^*(i,j)] = 1$$

we have

$$\pi^*(1,1) = \frac{1}{\sum_{i,j} (a_{i,j} + b_{i,j})}$$

and $\pi(K+1,j)$ can be computed and the location tracking cost for the reset-K algorithm is obtained using Eq. (20.25).

The analysis is validated by a Monte Carlo simulation. In the simulation, the confidence interval for the 98% confidence level of the output measure C_C/C_B is within 3% of the mean value. Figure 20.8 plots curves for Eq. (20.25) (the solid curves) against the simulation experiments (the dashed curves) for $K = 20$ and $C_H = 0.5C_B$ and $0.3C_B$, respectively. The figure indicates that the analysis is consistent with the simulation model.

20.9.3 Comparison of the LCMR Estimation Algorithms

If the distributions for the incoming call process and the user movement process never change, then we would expect the running average algorithm to outperform the reset-K algorithm (especially when K is small) because the measured hit ratio p_M in the running average algorithm approaches the true hit ratio value p in the steady state. Surprisingly, the performance for the reset-K algorithm is roughly the same as the running average algorithm even if K is as small as 10. Figure 20.9 plots the location tracking costs for the running average algorithm and the reset-K algorithm with different K values.

The figure indicates that in steady state, when the distributions for the incoming call process and the user movement process never change, the running average algorithm outperforms reset K, and a large value of K outperforms a small K but the differences are insignificant.

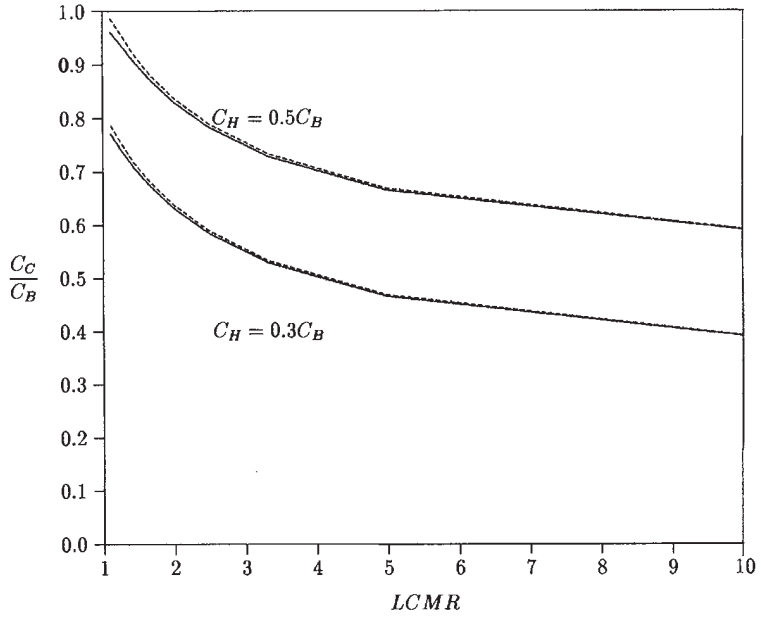


FIGURE 20.8 The location tracking costs for the reset-K algorithm; $K = 20$.

If the distributions for the incoming call process or the user movement process change from time to time, we expect that the reset-K algorithm outperforms the running average algorithm. We have examined this proposition experimentally. In the experiments, 4000 incoming calls are simulated. The call arrival rate changes from 0.1 to 1.0, 0.3, and then 5.0 for every 1000 calls (other sequences have been tested and similar results are observed). For every data point, the simulation is repeated 1000 times to ensure that the confidence interval for the 98% confidence level of the output measure C_C/C_B is within 3% of the mean value. Figure 20.10 plots the location tracking costs for the two algorithms for these experiments. By changing the distributions of the incoming call process, we observe that the reset-K algorithm is better than the running average algorithm for all C_H/C_B values.

20.10 Discussion

In this section we discuss aspects of the caching strategy presented here. Caching in PCS systems raises a number of issues not encountered in traditional computer systems, particularly with respect to architecture and locality in user call and mobility patterns. In addition, several variations in our reference assumptions are possible for investigating the implementation of the caching strategies. Here we sketch some of the issues involved.

20.10.1 Conditions When Caching Is Beneficial

We summarize the conditions for which the auxiliary strategies are worthwhile, under the assumptions of our analysis.

The caching strategy is very promising when the HLR update (or query load) or the remote A link is the performance bottleneck, since a low $LCMR$ ($LCMR > 0$) is required. For caching, the total database load and signalling network traffic is reduced whenever there is a cache hit. In addition, load and traffic is redistributed from the HLR and higher level SS7 network elements (RSTP, D links) to the VLRs and lower levels where excess network

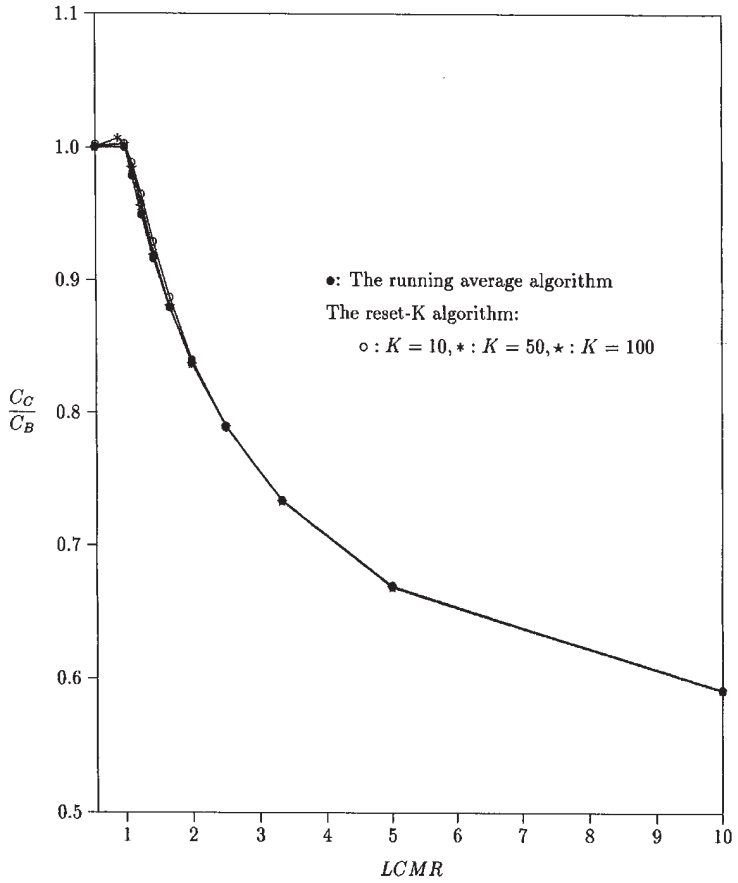


FIGURE 20.9 The location tracking costs for the running average algorithm and the reset-K algorithm; $C_H = 0.5C_B$.

capacity may be more likely to exist. If the VLR is the performance bottleneck, the caching strategy is not promising, unless the VLR capacity is upgraded.

The benefits of the caching strategy depend on user call and mobility patterns when the D link, RSTP, and LSTP are the performance bottlenecks. We have used a Poisson call arrival model and exponential intermove time to estimate this dependence. Under very conservative assumptions, for caching to be beneficial requires relatively high $LCMR$ (25–50); we expect that in practice this threshold could be lowered significantly (say, $LCMR > 7$). Further experimental study is required to estimate the amount of locality in user movements for different user populations to investigate this issue further. It is possible that for some classes of users data obtained from active badge location system studies (e.g., Fishman and Mazer, [8]) could be useful. In general, it appears that caching could also potentially provide benefits to some classes of users even when the D link, the RSTP, or the LSTP are the bottlenecks.

We observe that more accurate models of user calling and mobility patterns are required to help resolve the issues raised in this section. We are currently engaged in developing theoretical models for user mobility and estimating their effect on studies of various aspects of PCS performance [10].

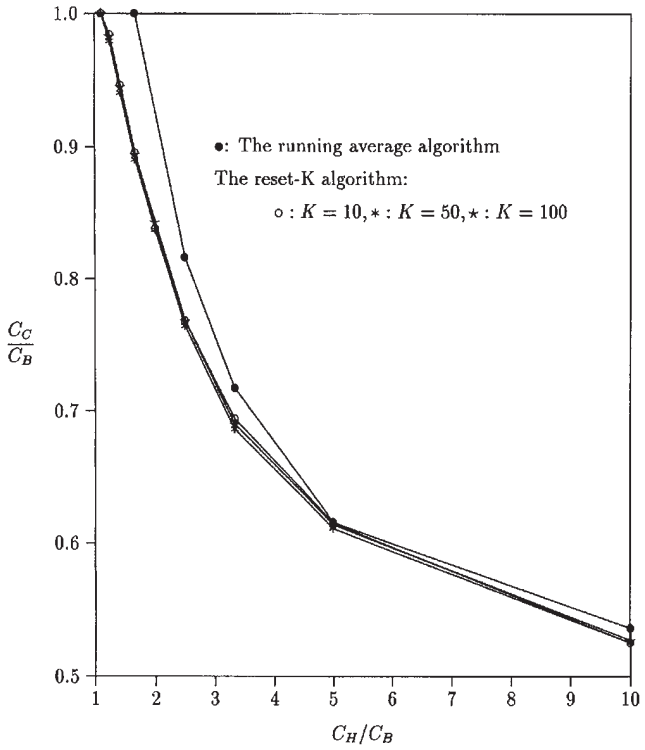


FIGURE 20.10 Comparing the running average algorithm and the reset-K algorithm under unstable call traffic.

20.10.2 Alternative Network Architectures

The reference architecture we have assumed (Fig. 20.2) is only one of several possible architectures. It is possible to consider variations in the placement of the HLR and VLR functionality, (e.g., placing the VLR at a local SCP associated with the LSTP instead of at the SSP), the number of SSPs served by an LSTP, the number of HLRs deployed, etc. It is quite conceivable that different regional PCS service providers and telecommunications companies will deploy different signalling network architectures, as well as placement of databases for supporting PCS within their serving regions [19]. It is also possible that the number and placement of databases in a network will change over time as the number of PCS users increases.

Rather than consider many possible variations of the architecture, we have selected a reference architecture to illustrate the new auxiliary strategies and our method of calculating their costs and benefits. Changes in the architecture may result in minor variations in our analysis but may not significantly affect our qualitative conclusions.

20.10.3 LCMR Estimation and Caching Policy

It is possible that for some user populations estimating the LCMR may not be necessary, since they display a relatively high-average LCMR. For some populations, as we have shown in Section 20.9, obtaining accurate estimates of user LCMR in order to decide whether or not to use caching can be important in determining the net benefits of caching.

In general, schemes for estimating the LCMR range from static to dynamic and from distributed to centralized. We have presented two simple distributed algorithms for estimating

LCMR, based on a long-range and short-range running calculation; the former is preferable if the call and mobility pattern of users is fairly Tuning the amount of history that is used to determine whether caching should be employed for a particular user is an obvious area for further study but is outside the scope of this chapter.

An alternative approach is to utilize some user-supplied information, by requesting profiles of user movements (e.g., see Tabbane, [22] and to integrate this with the caching strategy. A variation of this approach is to use some domain knowledge about user populations and their characteristics.

A related issue is that of cache size and management. In practice it is likely that the monetary cost of deploying a cache may limit its size. In that case, cache entries may not be maintained for some users; selecting these users carefully is important to maximize the benefits of caching. Note that the cache hit ratio threshold cannot necessarily be used to determine which users have cache entries, since it may be useful to maintain cache entries for some users even though their hit ratios have temporarily fallen below the threshold. A simple policy that has been found to be effective in computer systems in the least recently used (LRU) policy [20] in which cache entries that have been least recently used are discarded; LRU may offer some guidance in this context.

20.11 Conclusions

We began this chapter with an overview of the nuances of PCS, such as personal and terminal mobility, registration, deregistration, call delivery, etc. A tutorial was then provided on the two most common strategies for locating users in PCS, in North American interim standard IS-41 and the Pan-European standard GSM. A simplified analysis of the two standards was then provided to show the reader the extent to which database and signalling traffic is likely to be generated by PCS services. Suggestions were then made that are likely to result in reduced traffic.

Previous studies [12, 13, 14, 16] of PCS-related network signalling and data management functionalities suggest a high level of utilization of the signalling network in supporting call and mobility management activities for PCS systems. Motivated by the need to evolve location strategies to reduce signalling and database loads, we then presented an auxiliary strategy, called per-user caching, to augment the basic user location strategy proposed in standards [6, 18].

Using a reference system architecture for PCS, we quantified the criteria under which the caching strategy produces reductions in the network signalling and database loads in terms of users' LCMRs. We have shown that, if the HLR or the remote A link in an SS7 architecture is the performance bottleneck, caching is useful regardless of user call and mobility patterns. If the D link or STPs are the performance bottlenecks, caching is potentially beneficial for large classes of users, particularly if they display a degree of locality in their call reception patterns. Depending on the numbers of PCS users who meet these criteria, the system-wide impacts of these strategies could be significant. For instance, for users with $LCMR \sim 5$ and stable call and move patterns, caching can result in cost reduction of 20–60% over the basic user location strategy *BasicFIND*( ) under our analysis. Our results are conservative in that the *BasicFIND*( ) procedure we have used for comparison purposes already reduces the network impacts compared to the user location strategy specified in PCS standards such as IS-41.

We have also investigated in detail two simple on-line algorithms for estimating users' LCMRs and examined the call and mobility patterns for which each would be useful. The algorithms allow a system designer to tune the amount of history used to estimate a users'

LCMR and, hence, to attempt to optimize the benefits due to caching. The particular values of cache hit ratios and LCMR thresholds will change with variations in the way the PCS architecture and the caching strategy is implemented, but our general approach can still be applied. There are several issues deserving further study with respect to deployment of the caching strategy, such as the effect of alternative PCS architectures, integration with other auxiliary strategies such as the use of user profiles, and effective cache management policies.

Recently, we have augmented the work reported in this paper by a simulation study in which we have compared the caching and basic user location strategies [9]. The effect of using a time-based criterion for enabling use of the cache has also been considered [11]. We have proposed elsewhere, for users with low CMRs, an auxiliary strategy involving a system of forwarding pointers to reduce the signalling traffic and database loads [10], a description of which is beyond the scope of this chapter.

Acknowledgment

We acknowledge a number of our colleagues in Bellcore who have reviewed several previous papers by the authors and contributed to improving the clarity and readability of this work.

References

[1] Awerbuch, B. and Peleg, D., Concurrent online tracking of mobile users. In *Proc. SIGCOMM Symp. Comm. Arch. Prot.*, Oct. 1991.

[2] Bellcore., Advanced intelligent network release 1 network and operations plan, Issue 1. Tech. Rept. SR-NPL-001623. Bell Communications Research, Morristown, NJ, Jun. 1991.

[3] Bellcore., Personal communications services (PCS) network access services to PCS providers, Special Report SR-TSV-002459, Bell Communications Research, Morristown, NJ, Oct. 1993a.

[4] Bellcore., Switching system requirements for interexchange carrier interconnection using the integrated services digital network user part (ISDNUP). Tech. Ref. TR-NWT-000394. Bell Communications Research. Morristown, NJ, Dec. 1992c.

[5] Berman, R.K. and Brewster, J.H., Perspectives on the AIN architecture. *IEEE Comm. Mag.*, 1(2), 27–32, 1992.

[6] Electronic Industries Association/Telecommunications Industry Association., Cellular radio telecommunications intersystem operations. Tech. Rept. IS-41. Rev. B. Jul. 1991.

[7] Feller, W., *An Introduction to Probability Theory and Its Applications.* John Wiley & Sons, New York, 1966.

[8] Fishman, N. and Mazer, M., Experience in deploying an active badge system. In *Proc. Globecom Workshop on Networking for Pers. Comm. Appl.*, Dec. 1992.

[9] Harjono, H., Jain, R., and Mohan, S., Analysis and simulation of a cache-based auxiliary location strategy for PCS. In *Proc. IEEE Conf. Networks Pers. Comm.*, 1994.

[10] Jain, R. and Lin Y.-B., An auxiliary user location strategy employing forwarding pointers to reduce network impacts of PCS. *ACM Journal on Wireless Info. Networks (WINET)*, 1(2), 1995.

[11] Lin, Y.-B., Determining the user locations for personal communications networks. *IEEE Trans. Vehic. Tech.*, 466–473, Aug. 1994.

[12] Lo, C., Mohan, S., and Wolff, R., A comparison of data management alternatives for personal communications applications. Second Bellcore Symposium on Performance Modeling, SR-TSV-002424, Bell Communications Research, Morristown, NJ, Nov. 1993.

[13] Lo, C.N., Wolff, R.S., and Bernhardt, R.C., An estimate of network database transaction volume to support personal communications services. In *Proc. Intl. Conf. Univ. Pers. Comm.*, 1992.

[14] Lo, C. and Wolff, R., Estimated network database transaction volume to support wireless personal data communications applications. In *Proc. Intl. Conf. Comm.*, May 1993.

[15] Lycksell, E., GSM system overview. Tech. Rept. Swedish Telecom. Admin., Jan. 1991.

[16] Meier-Hellstern, K. and Alonso, E., The use of SS7 and GSM to support high density personal communications. In *Proc. Intl. Conf. Comm.*, 1992.

[17] Mohan, S. and Jain, R., Two user location strategies for PCS. *IEEE Pers. Comm. Mag.*, Premiere issue. 42–50, Feb. 1994.

[18] Mouly, M. and Pautet, M.B., *The GSM System for Mobile Communications*. M. Mouly, 49 rue Louise Bruneau, Palaiseau, France, 1992.

[19] Russo, P., Bechard, K., Brooks, E., Corn, R.L., Honig, W.L., Gove, R., and Young, J., In rollout in the United States. *IEEE Comm. Mag.*, 56–63, Mar. 1993.

[20] Silberschatz, A. and Peterson, J., *Operating Systems Concepts*. Addison-Wesley, Reading, MA, 1988.

[21] Tabbane, S., Comparison between the alternative location strategy (AS) and the classical location strategy (CS). Tech. Rept. Rutgers Univ. WINLAB. Rutgers, NJ, Jul. 1992.

[22] Tabbane, S., Evaluation of an alternative location strategy for future high density wireless communications systems. Tech. Rept. WINALAB-TR-51, Rutgers Univ. WINLAB. Rutgers, NJ, Jan. 1993.

[23] Thomas, R., Gilbert, H., and Mazziotto, G., Influence of the mobile station on the performance of a radio mobile cellular network. In *Proc. 3rd Nordic Seminar*. Sep. 1988.

21

Cell Design Principles

21.1 Introduction 21-1
21.2 Cellular Principles 21-2
21.3 Performance Measures and System Requirements 21-2
21.4 System Expansion Techniques 21-3
21.5 Basic Design Steps 21-4
21.6 Traffic Engineering 21-5
21.7 Cell Coverage 21-6
Propagation Model • Base Station Coverage • Application Examples
21.8 Interference 21-10
Adjacent Channel Interference • Cochannel Interference
21.9 Conclusions 21-13
Defining Terms 21-14
References 21-14
Further Information 21-15

Michel Daoud Yacoub
University of Campinas

21.1 Introduction

Designing a cellular network is a challenging task that invites engineers to exercise all of their knowledge in telecommunications. Although it may not be necessary to work as an expert in all of the fields, the interrelationship among the areas involved impels the designer to naturally search for a deeper understanding of the main phenomena. In other words, the time for segregation, when radio engineers and traffic engineers would not talk to each other, at least through a common vocabulary, is probably gone.

A great many aspects must be considered in a cellular network planning. The main ones include the following.

Radio Propagation: Here the topography and the morphology of the terrain, the urbanization factor and the clutter factor of the city, and some other aspects of the target geographical region under investigation will constitute the input data for the radio coverage design.

Frequency Regulation and Planning: In most countries there is a centralized organization, usually performed by a government entity, regulating the assignment and use of the radio spectrum. The frequency planning within the assigned spectrum should then be made so that interferences are minimized and the traffic demand is satisfied.

Modulation: As far as analog systems are concerned, the narrowband FM is widely used due to its remarkable performance in the presence of fading. The North American

0-8493-8597-0/99/$0.00+$.50

© 1999 by CRC Press LLC

Digital Cellular Standard IS-54 proposes the $\pi/4$ differential quadrature phase-shift keying ($\pi/4$ DQPSK) modulation, whereas the Global Standard for Mobile Communications (GSM) establishes the use of the Gaussian minimum-shift keying (GMSK).

Antenna Design: To cover large areas and for low-traffic applications omnidirectional antennas are recommended. Some systems at their inception may have these characteristics, and the utilization of omnidirectional antennas certainly keeps the initial investment low. As the traffic demand increases, the use of some sort of capacity enhancement technique to meet the demand, such as replacing the omnidirectional by directional antennas, is mandatory.

Transmission Planning: The structure of the channels, both for signalling and voice, is one of the aspects to be considered in this topic. Other aspects include the performance of the transmission components (power capacity, noise, bandwidth, stability, etc.) and the design or specification of transmitters and receivers.

Switching Exchange: In most cases this consists of adapting the existing switching network for mobile radio communications purposes.

Teletraffic: For a given grade of service and number of channels available, how many subscribers can be accommodated into the system? What is the proportion of voice and signalling channels?

Software Design: With the use of microprocessors throughout the system there are software applications in the mobile unit, in the base station, and in the switching exchange.

Other aspects, such as human factors, economics, etc., will also influence the design.

This chapter outlines the aspects involving the basic design steps in cellular network planning. Topics, such as traffic engineering, cell coverage, and interference, will be covered, and application examples will be given throughout the section so as to illustrate the main ideas. We start by recalling the basic concepts including *cellular principles, performance measures and system requirements,* and *system expansion techniques.*

21.2 Cellular Principles

The basic idea of the cellular concept is *frequency reuse* in which the same set of channels can be reused in different geographical locations sufficiently apart from each other so that *cochannel interference* be within tolerable limits. The set of channels available in the system is assigned to a group of *cells* constituting the *cluster.* Cells are assumed to have a *regular hexagonal* shape and the number of cells per cluster determines the *repeat pattern.* Because of the hexagonal geometry only certain repeat patterns can tessellate. The number N of cells per cluster is given by

$$N = i^2 + ij + j^2 \tag{21.1}$$

where i and j are integers. From Eq. (21.1) we note that the clusters can accommodate only certain numbers of cells such as $1, 3, 4, 7, 9, 12, 13, 16, 19, 21, \ldots$, the most common being 4 and 7. The number of cells per cluster is intuitively related with system capacity as well as with transmission quality. The fewer cells per cluster, the larger the number of channels per cell (higher traffic carrying capacity) and the closer the cocells (potentially more cochannel interference). An important parameter of a cellular layout relating these entities is the D/R ratio, where D is the distance between cocells and R is the cell radius. In a hexagonal geometry it is found that

$$D/R = \sqrt{3N} \tag{21.2}$$

21.3 Performance Measures and System Requirements

Two parameters are intimately related with the grade of service of the cellular systems: carrier-to-cochannel interference ratio and blocking probability.

A high carrier-to-cochannel interference ratio in connection with a low-blocking probability is the desirable situation. This can be accomplished, for instance, in a large cluster with a low-traffic condition. In such a case the required grade of service can be achieved, although the resources may not be efficiently utilized. Therefore, a measure of efficiency is of interest. The **spectrum efficiency** η_s expressed in erlang per square meter per hertz, yields a measure of how efficiently space, frequency, and time are used, and it is given by

$$\eta_s = \frac{\text{number of reuses}}{\text{coverage area}} \times \frac{\text{number of channels}}{\text{bandwidth available}} \times \frac{\text{time the channel is busy}}{\text{total time of the channel}}$$

Another measure of interest is the **trunking efficiency** in which the number of subscribers per channel is obtained as a function of the number of channels per cell for different values of blocking probability. As an example, assume that a cell operates with 40 channels and that the mean blocking probability is required to be 5%. Using the erlang-B formula (refer to the Traffic Engineering section of this chapter), the traffic offered is calculated as 34.6 erlang. If the traffic per subscriber is assumed to be 0.02 erl, a total of $34.6/0.02 = 1730$ subscribers in the cell is found. In other words, the trunking efficiency is $1730/40 = 43.25$ subscribers per channel in a 40-channel cell. Simple calculations show that the trunking efficiency decreases rapidly when the number of channels per cell falls below 20.

The basic specifications require cellular services to be offered with a fixed telephone network quality. Blocking probability should be kept below 2%. As for the transmission aspect, the aim is to provide good quality service for 90% of the time. Transmission quality concerns the following parameters:

- Signal-to-cochannel interference (S/I_c) ratio
- Carrier-to-cochannel interference ratio (C/I_c)
- Signal plus noise plus distortion-to-noise plus distortion ($SINAD$) ratio
- Signal-to-noise (S/N) ratio
- Adjacent channel interference selectivity (ACS)

The S/I_c is a subjective measure, usually taken to be around 17 dB. The corresponding C/I_c depends on the modulation scheme. For instance, this is around 8 dB for 25-kHz FM, 12 dB for 12.5-kHz FM, and 7 dB for GMSK, but the requirements may vary from system to system. A common figure for $SINAD$ is 12 dB for 25-kHz FM. The minimum S/N requirement is 18 dB, whereas ACS is specified to be no less than 70 dB.

21.4 System Expansion Techniques

The obvious and most common way of permitting more subscribers into the network is by allowing a system performance degradation but within acceptable levels. The question is how to objectively define what is acceptable. In general, the subscribers are more likely to tolerate a poor quality service rather than not having the service at all. Some alternative expansion techniques, however, do exist that can be applied to increase the system capacity. The most widely known are as follows.

Adding New Channels: In general, when the system is set up not all of the channels need be used, and growth and expansion can be planned in an orderly manner by utilizing the channels that are still available.

Frequency Borrowing: If some cells become more overloaded than others, it may be possible to reallocate channels by transferring frequencies so that the traffic demand can be accommodated.

Change of Cell Pattern: Smaller clusters can be used to allow more channels to attend a bigger traffic demand at the expense of a degradation of the transmission quality.

Cell Splitting: By reducing the size of the cells, more cells per area, and consequently more channels per area, are used with a consequent increase in traffic capacity. A radius reduction by a factor of f reduces the coverage area and increases the number of base stations by a factor of f^2. Cell splitting usually takes place at the midpoint of the congested areas and is so planned in order that the old base stations are kept.

Sectorization: A cell is divided into a number of sectors, three and six being the most common arrangements, each of which is served by a different set of channels and illuminated by a directional antenna. The sector, therefore, can be considered as a new cell. The base stations can be located either at the center or at the corner of the cell. The cells in the first case are referred to as center-excited cells and in the second as corner-excited cells. Directional antennas cut down the cochannel interference, allowing the cocells to be more closely spaced. Closer cell spacing implies smaller D/R ratio, corresponding to smaller clusters, i.e., higher capacity.

Channel Allocation Algorithms: The efficient use of channels determines the good performance of the system and can be obtained by different channel assignment techniques. The most widely used algorithm is based on fixed allocation. Dynamic allocation strategies may give better performance but are very dependent on the traffic profile and are usually difficult to implement.

21.5 Basic Design Steps

Engineering a cellular system to meet the required objectives is not a straightforward task. It demands a great deal of information, such as market demographics, area to be served, traffic offered, and other data not usually available in the earlier stages of system design. As the network evolves, additional statistics will help the system performance assessment and replanning. The main steps in a cellular system design are as follows.

Definition of the Service Area: In general, the responsibility for this step of the project lies on the operating companies and constitutes a tricky task, because it depends on the market demographics and, consequently, on how much the company is willing to invest.

Definition of the Traffic Profile: As before, this step depends on the market demographics and is estimated by taking into account the number of potential subscribers within the service area.

Choice of Reuse Pattern: Given the traffic distribution and the interference requirements a choice of the reuse pattern is carried out.

Location of the Base Stations: The location of the first base station constitutes an important step. A significant parameter to be taken into account in this is the relevance of the region to be served. The base station location is chosen so as to be at the center of or as close as possible to the target region. Data, such as available infrastructure and land, as well as local regulations are taken into consideration in this step. The cell radius is defined as a function of the traffic distribution. In urban areas, where the traffic is more heavily concentrated, smaller cells are chosen so as to attend the demand with the available

channels. In suburban and in rural areas, the radius is chosen to be large because the traffic demand tends to be small. Once the placement of the first base station has been defined, the others will be accommodated in accordance with the repeat pattern chosen.

Radio Coverage Prediction: Given the topography and the morphology of the terrain, a radio prediction algorithm, implemented in the computer, can be used to predict the signal strength in the geographic region. An alternative to this relies on field measurements with the use of appropriate equipment. The first option is usually less costly and is widely used.

Design Checkup: At this point it is necessary to check whether or not the parameters with which the system has been designed satisfy the requirements. For instance, it may be necessary to re-evaluate the base station location, the antenna height, etc., so that better performance can be attained.

Field Measurements: For a better tuning of the parameters involved, field measurements (radio survey) should be included in the design. This can be carried out with transmitters and towers provisionally set up at the locations initially defined for the base station.

The cost assessment may require that a redesign of the system should be carried out.

21.6 Traffic Engineering

The starting point for engineering the traffic is the knowledge of the required grade of service. This is usually specified to be around 2% during the busy hour. The question lies on defining the busy hour. There are usually three possible definitions: (1) busy hour at the busiest cell, (2) system busy hour, and (3) system average over all hours.

The estimate of the subscriber usage rate is usually made on a demographic basis from which the traffic distribution can be worked out and the cell areas identified. Given the repeat pattern (cluster size), the cluster with the highest traffic is chosen for the initial design. The traffic A in each cell is estimated and, with the desired blocking probability $E(A, M)$, the erlang-B formula as given by Eq. (21.3) is used to determine the number of channels per cell, M

$$E(M, A) = \frac{A^M/M!}{\sum_{i=0}^{M} A^i/i!} \tag{21.3}$$

In case the total number of available channels is not large enough to provide the required grade of service, the area covered by the cluster should be reduced in order to reduce the traffic per cell. In such a case, a new study on the interference problems must be carried out. The other clusters can reuse the same channels according to the reuse pattern. Not all channels need be provided by the base stations of those cells where the traffic is supposedly smaller than that of the heaviest loaded cluster. They will eventually be used as the system grows.

The traffic distribution varies in time and space, but it is commonly bell shaped. High concentrations are found in the city center during the rush hour, decreasing toward the outskirts. After the busy hour and toward the end of the day, this concentration changes as the users move from the town center to their homes. Note that because of the mobility of the users handoffs and roaming are always occurring, reducing the channel holding times in the cell where the calls are generated and increasing the traffic in the cell where the mobiles travel. Accordingly, the erlang-B formula is, in fact, a rough approximation used to model the traffic process in this ever-changing environment. A full investigation of the traffic performance in such a dynamic system requires all of the phenomena to be taken into

account, making any traffic model intricate. Software simulation packages can be used so as to facilitate the understanding of the main phenomena as well as to help system planning. This is a useful alternative to the complex modeling, typically present in the analysis of cellular networks, where closed-form solutions are not usually available

On the other hand, conventional traffic theory, in particular, the erlang-B formula, is a handy tool widely used in cellular planning. At the inception of the system the calculations are carried out based on the best available traffic estimates, and the system capacity is obtained by grossly exaggerating the calculated figures. With the system in operation some adjustments must be made so that the requirements are met.

The approach just mentioned assumes the simplest channel assignment algorithm: the fixed allocation. It has the maximum spatial efficiency in channel reuse, since the channels are always assigned at the minimum reuse distance. Moreover, because each cell has a fixed set of channels, the channel assignment control for the calls can be distributed among the base stations.

The main problem of fixed allocation is its inability to deal with the alteration of the traffic pattern. Because of the mobility of the subscribers, some cells may experience a sudden growth in the traffic offered, with a consequent deterioration of the grade of service, whereas other cells may have free channels that cannot be used by the congested cells.

A possible solution for this is the use of dynamic channel allocation algorithms in which the channels are allocated on a demand basis There is an infinitude of strategies using the dynamic assignment principles, but they are usually complex to implement. An interim solution can be exercised if the change of the traffic pattern is predictable. For instance, if a region is likely to have an increase of the traffic on a given day (say, a football stadium on a match day), a mobile base station can be moved toward such a region in order to alleviate the local base.

Another specific solution uses the traffic available at the boundary between cells that may well communicate with more than one base station. In this case, a call that is blocked in its own cell can be directed to the neighboring cell to be served by its base station. This strategy, called *directed retry,* is known to substantially improve the traffic capacity. On the other hand, because channels with marginally acceptable transmission quality may be used, an increase in the interference levels, both for adjacent channel and cochannel, can be expected. Moreover, subscribers with radio access only to their own base will experience an increase in blocking probability.

21.7 Cell Coverage

The propagation of energy in a mobile radio environment is strongly influenced by several factors, including the natural and artificial relief, propagation frequency, antenna heights, and others. A precise characterization of the signal variability in this environment constitutes a hard task. Deterministic methods, such as those described by the *free space, plane earth,* and *knife-edge diffraction* propagation models, are restricted to very simple situations. They are useful, however, in providing the basic mechanisms of propagation. Empirical methods, such as those proposed by many researchers (e.g., [1, 4, 5, 8]; and others), use curves and/or formulas based on field measurements, some of them including deterministic solutions with various correction factors to account for the propagation frequency, antenna height, polarization, type of terrain, etc. Because of the random characteristics of the mobile radio signal, however, a single deterministic treatment of this signal will certainly lead the problem to a simplistic solution. Therefore, we may treat the signal on a statistical basis and interpret the results as random events occurring with a given

probability. The cell coverage area is then determined as the proportion of locations where the received signal is greater than a certain threshold considered to be satisfactory.

Suppose that at a specified distance from the base station the *mean signal strength* is considered to be known. Given this we want to determine the cell radius such that the mobiles experience a received signal above a certain threshold with a stipulated probability. The mean signal strength can be determined either by any of the prediction models or by field measurements. As for the statistics of the mobile radio signal, five distributions are widely accepted today: lognormal, Rayleigh, Suzuki [11], Rice, and Nakagami. The lognormal distribution describes the variation of the mean signal level (large-scale variations) for points having the same transmitter–receiver antennas separation, whereas the other distributions characterize the instantaneous variations (small-scale variations) of the signal. In the calculations that follow we assume a lognormal environment. The other environments can be analyzed in a like manner; although this may not be of interest if some sort of diversity is implemented, because then the effects of the small-scale variations are minimized.

21.7.1 Propagation Model

Define m_w and k as the mean powers at distances x and x_0, respectively, such that

$$m_w = k\left(\frac{x}{x_0}\right)^{-\alpha} \tag{21.4}$$

where α is the path loss coefficient. Expressed in decibels, $M_w = 10\log m_w$, $K = 10\log k$ and

$$M_w = K - 10\alpha\log\left(\frac{x}{x_0}\right) \tag{21.5}$$

Define the received power as $w = v^2/2$, where v is the received envelope. Let $p(W)$ be the probability density function of the received power W, where $W = 10\log w$. In a lognormal environment, v has a lognormal distribution and

$$p(W) = \frac{1}{\sqrt{2\pi}\sigma_w}\exp\left(-\frac{(W - M_w)^2}{2\sigma_w^2}\right) \tag{21.6}$$

where M_W is the mean and σ_w is the standard deviation, all given in decibels. Define w_T and $W_T = 10\log w_T$ as the threshold above which the received signal is considered to be satisfactory. The probability that the received signal is below this threshold is its *probability distribution function* $P(W_T)$, such that

$$P(W_T) = \int_{-\infty}^{W_T} p(W)\,\mathrm{d}W = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left[\frac{(W_T - M_W)^2}{2\sigma_w^2}\right] \tag{21.7}$$

where erf( ) is the error function defined as

$$\mathrm{erf}(y) = \frac{2}{\sqrt{\pi}}\int_0^y \exp\left(-t^2\right)\,\mathrm{d}t \tag{21.8}$$

21.7.2 Base Station Coverage

The problem of estimating the cell area can be approached in two different ways. In the first approach, we may wish to determine the proportion β of locations at x_0 where the received

signal power w is above the threshold power w_T. In the second approach, we may estimate the proportion μ of the circular area defined by x_0 where the signal is above this threshold. In the first case, this proportion is averaged over the perimeter of the circumference (cell border); whereas in the second approach, the average is over the circular area (cell area).

The proportion β equals the probability that the signal at x_0 is greater than this threshold. Hence,

$$\beta = \text{prob}\,(W \geq W_T) = 1 - P\,(W_T) \tag{21.9}$$

Using Eqs. (21.5) and (21.7) in Eq. (21.9) we obtain

$$\beta = \frac{1}{2} - \frac{1}{2}\text{erf}\left[\frac{W_T - K + 10\alpha \log\,(x/x_0)}{\sqrt{2}\sigma_w}\right] \tag{21.10}$$

This probability is plotted in Fig. 21.1, for $x = x_0$ (cell border).

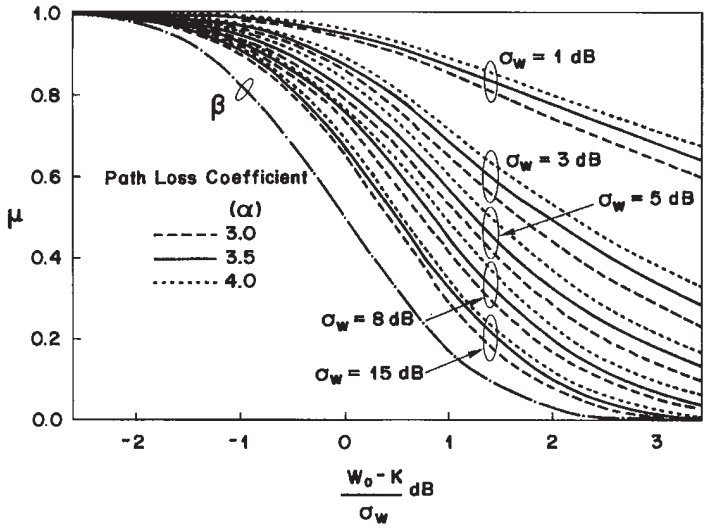


FIGURE 21.1 Proportion of locations where the received signal is above a given threshold; the dashdot line corresponds to the β approach and the other lines to the μ approach.

Let prob$(W \geq W_T)$ be the probability of the received power W being above the threshold W_T within an infinitesimal area $\text{d}S$. Accordingly, the proportion μ of locations within the circular area S experiencing such a condition is

$$\mu = \frac{1}{S}\int_S [1 - P\,(W_T)]\,\text{d}S \tag{21.11}$$

where $S = \pi r^2$ and $\text{d}S = x\,\text{d}x\text{d}\theta$. Note that $0 \leq x \leq x_0$ and $0 \leq \theta \leq 2\pi$. Therefore, solving for $\text{d}\theta$, we obtain

$$\mu = 2\int_0^1 u\beta\,\text{d}u \tag{21.12}$$

where $u = x/x_0$ is the normalized distance.

Inserting Eq. (21.10) in Eq. (21.12) results in

$$\mu = 0.5\left\{1 + \text{erf}\,(a) + \exp\left(\frac{2ab+1}{b^2}\right)\left[1 - \text{erf}\left(\frac{ab+1}{b}\right)\right]\right\} \tag{21.13}$$

where $a = (K - W_T)/\sqrt{2}\sigma_w$ and $b = 10\alpha \log(e)/\sqrt{2}\sigma_w$.

These probabilities are plotted in Fig. 21.1 for different values of standard deviation and path loss coefficients.

21.7.3 Application Examples

From the theory that has been developed it can be seen that the parameters affecting the probabilities β and μ for cell coverage are the path loss coefficient α, the standard deviation σ_w, the required threshold W_T, and a certain power level K, measured or estimated at a given distance from the base station.

The applications that follow are illustrated for two different standard deviations: $\sigma_w = 5$ dB and $\sigma_w = 8$ dB. We assume the path loss coefficient to be $\alpha = 4$ (40 dB/decade), the mobile station receiver sensitivity to be -116 dB (1 mW), and the power level estimated at a given distance from the base station as being that at the cell border, $K = -102$ dB (1 mW). The receiver is considered to operate with a *SINAD* of 12 dB for the specified sensitivity. Assuming that cochannel interference levels are negligible and given that a signal-to-noise ratio S/N of 18 dB is required, the threshold W_T will be -116 dB (1 mW) + (18 − 12) dB (1 mW) = -110 dB (1 mW).

Three cases will be explored as follows.

Case 1: We want to estimate the probabilities β and μ that the received signal exceeds the given threshold 1) at the border of the cell, probability β and 2) within the area delimited by the cell radius, probability μ.

Case 2: It may be interesting to estimate the cell radius x_0 such that the received signal be above the given threshold with a given probability (say 90%) (1) at the perimeter of the cell and (2) within the cell area. This problem implies the calculation of the mean signal strength K at the distance x_0 (the new cell border) of the base station. Given K and given that at a distance x_0 (the former cell radius) the mean signal strength M_w is known [note that in this case $M_w = -102$ dB (1 mW)], the ratio x_0/x can be estimated.

Case 3: To fulfill the coverage requirement, rather than calculating the new cell radius, as in Case 2, a signal strength at a given distance can be estimated such that a proportion of the locations at this distance, proportion β, or within the area delimited by this distance, proportion μ, will experience a received signal above the required threshold. This corresponds to calculating the value of the parameter K already carried out in Case 2 for the various situations.

The calculation procedures are now detailed for $\sigma_w = 5$ dB. Results are also shown for $\sigma_w = 8$ dB.

Case 1: Using the given parameters we obtain $(W_T - K)/\sigma_w = -1.6$. With this value in Fig. 21.1, we obtain the probability that the received signal exceeds -116 dB (1 mW) for $S/N = 18$ dB given that at the cell border the mean signal power is -102 dB (1 mW) given in Table 21.1.

TABLE 21.1 Case 1 Coverage Probability

Standard Deviation, dB	β Approach (Border Coverage), %	μ Approach (Area Coverage), %
5	97	100
8	84	95

Note, from Table 21.1, that the signal at the cell border exceeds the receiver sensitivity with 97% probability for $\sigma_w = 5$ dB and with 84% probability for $\sigma_w = 8$ dB. If, on the other hand, we are interested in the area coverage rather than in the border coverage, then these figures change to 100% and 95%, respectively.

Case 2: From Fig. 21.1, with $\beta = 90\%$ we find $(W_T - K)/\sigma_w = -1.26$. Therefore, $K = -103.7$ dB (1 mW). Because $M_w - K = -10\alpha \log(x/x_0)$, then $x_0/x = 1.10$. Again, from Fig. 21.1, with $\mu = 90\%$ we find $(W_T - K)/\sigma_w = -0.48$, yielding $K = -107.6$ dB (1 mW). Because $M_w - K = -10\alpha \log(x/x_0)$, then $x_0/x = 1.38$. These results are summarized in Table 21.2, which shows the normalized radius of a cell where the received signal power is above -116 dB (1 mW) with 90% probability for $S/N = 18$ dB, given that at a reference distance from the base station (the cell border) the received mean signal power is -102 dB (1 mW).

TABLE 21.2 Case 2 Normalized Radius

Standard Deviation, dB	β Approach (Border Coverage)	μ Approach (Area Coverage)
5	1.10	1.38
8	0.88	1.27

Note, from Table 21.2, that in order to satisfy the 90% requirement at the cell border the cell radius can be increased by 10% for $\sigma_w = 5$ dB. If, on the other hand, for the same standard deviation the 90% requirement is to be satisfied within the cell area, rather than at the cell border, a substantial gain in power is achieved. In this case, the cell radius can be increased by a factor of 1.38. For $\sigma_w = 8$ dB and 90% coverage at the cell border, the cell radius should be reduced to 88% of the original radius. For area coverage, an increase of 27% of the cell radius is still possible.

Case 3: The values of the mean signal power K are taken from Case 2 and shown in Table 21.3, which shows the signal power at the cell border such that 90% of the locations will experience a received signal above -116 dB for $S/N = 18$ dB.

TABLE 21.3 Case 3 Signal Power

Standard Deviation dB	β Approach (Border Coverage), dB (1 mW)	μ Approach (Area Coverage), dB (1 mW)
5	-103.7	-107.6
8	-99.8	-106.2

21.8 Interference

Radio-frequency interference is one of the most important issues to be addressed in the design, operation, and maintenance of mobile communication systems. Although both intermodulation and intersymbol interferences also constitute problems to account for in system planning, a mobile radio system designer is mainly concerned about adjacent-channel and cochannel interferences.

21.8.1 Adjacent Channel Interference

Adjacent-channel interference occurs due to equipment limitations, such as frequency instability, receiver bandwidth, filtering, etc. Moreover, because channels are kept very close to each other for maximum spectrum efficiency, the random fluctuation of the signal, due to fading and near–far effect, aggravates this problem.

Some simple, but efficient, strategies are used to alleviate the effects of adjacent channel interference. In narrowband systems, the total frequency spectrum is split into two halves so that the reverse channels, composing the uplink (mobile to base station) and the forward channels, composing the downlink (base station to mobile), can be separated by half of the spectrum. If other services can be inserted between the two halves, then a greater frequency separation, with a consequent improvement in the interference levels, is accomplished. Adjacent channel interference can also be minimized by avoiding the use of adjacent channels within the same cell. In the same way, by preventing the use of adjacent channels in adjacent cells a better performance is achieved. This strategy, however, is dependent on the cellular pattern. For instance, if a seven-cell cluster is chosen, adjacent channels are inevitably assigned to adjacent cells.

21.8.2 Cochannel Interference

Undoubtedly the most critical of all interferences that can be engineered by the designer in cellular planning is cochannel interference. It arises in mobile radio systems using cellular architecture because of the frequency reuse philosophy.

A parameter of interest to assess the system performance in this case is the carrier-to-cochannel interference ratio C/I_c. The ultimate objective of estimating this ratio is to determine the reuse distance and, consequently, the repeat pattern. The C/I_c ratio is a random variable, affected by random phenomena such as (1) location of the mobile, (2) fading, (3) cell site location, (4) traffic distribution, and others. In this subsection we shall investigate the **outage probability,** i.e., the probability of failing to achieve adequate reception of the signal due to cochannel interference. This parameter will be indicated by $p(CI)$. As can be inferred, this is intrinsically related to the repeat pattern.

Cochannel interference will occur whenever the wanted signal does not simultaneously exceed the minimum required signal level s_0 and the n interfering signals, $i_1, i_2, \ldots, i_n$, by some protection ratio r. Consequently, the conditional outage probability, given n interferers, is

$$\begin{aligned} p\,(CI\,|\,n) \;&=\; 1 - \int_{s_0}^{\infty} p(s) \int_0^{s/r} p\,(i_1) \int_0^{(s/r)-i_1} p\,(i_2) \cdots \\ &\quad \times \int_0^{(s/r)-i_1-\cdots-i_{n-1}} p\,(i_n)\; \mathrm{d}i_n \cdots \mathrm{d}i_2\, \mathrm{d}i_1\, \mathrm{d}s \end{aligned} \tag{21.14}$$

The total outage probability can then be evaluated by

$$p(CI) = \sum_n p\,(CI\,|\,n)\, p(n) \tag{21.15}$$

where $p(n)$ is the distribution of the number of active interferers.

In the calculations that follow we shall assume an interference-only environment, i.e., $s_0 = 0$, and the signals to be Rayleigh faded. In such a fading environment the probability density function of the signal-to-noise ratio x is given by

$$p(x) = \frac{1}{x_m} \exp\left(-\frac{x}{x_m}\right) \tag{21.16}$$

where x_m is the mean signal-to-noise ratio. Note that $x = s$ and $x_m = s_m$ for the wanted signal, and $x = i_j$ and $x_m = i_{mj}$ for the interfering signal j, with s_m and i_{mj} being the mean of s and i_j, respectively.

By using the density of Eq. (21.16) in Eq. (21.14) we obtain

$$p(CI \mid n) = \sum_{j=1}^{n} \prod_{k=1}^{j} \frac{z_k}{1 + z_k} \tag{21.17}$$

where $z_k = r s_m / i_{mk}$

If the interferers are assumed to be equal, i.e., $z_k = z$ for $k = 1, 2, \ldots, n$, then

$$p(CI \mid n) = 1 - \left(\frac{z}{1+z}\right)^n \tag{21.18}$$

Define $Z = 10 \log z$, $S_m = 10 \log s_m$, $I_m = 10 \log i_m$, and $R_r = 10 \log r$. Then, $Z = S_m - (I_m + R_r)$. Equation (21.18) is plotted in Fig. 21.2 as a function of Z for $n = 1$ and $n = 6$, for the situation in which the interferers are equal.

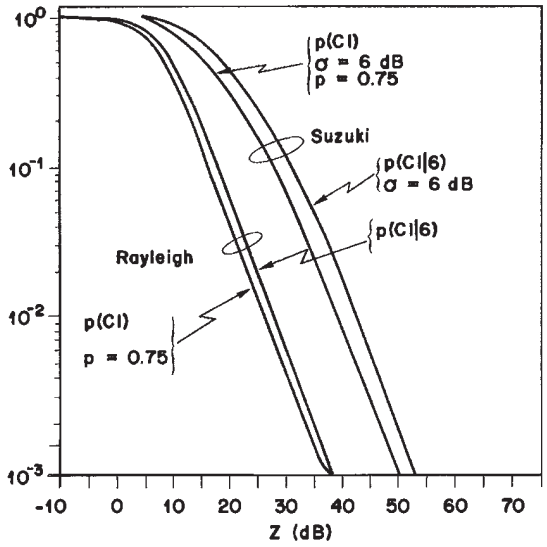


FIGURE 21.2 Conditional and unconditional outage probability for $n = 6$ interferes in a Rayleigh environment and in a Suzuki environment with $\sigma = 6$ dB.

If the probability of finding an interferer active is p, the distribution of active interferers is given by the binomial distribution. Considering the closest surrounding cochannels to be the most relevant interferers we then have six interferers. Thus

$$p(n) = \binom{6}{n} p^n (1-p)^{6-n} \tag{21.19}$$

For equal capacity cells and an evenly traffic distribution system, the probability p is approximately given by

$$p = \sqrt[M]{B} \tag{21.20}$$

where B is the blocking probability and M is the number of channels in the cell.

Now Eqs. (21.20), (21.19), and (21.18) can be combined into Eq. (21.15) and the outage probability is estimated as a function of the parameter Z and the channel occupancy p. This is shown in Fig. 21.2 for $p = 75\%$ and $p = 100\%$.

A similar, but much more intricate, analysis can be carried out for the other fading environments. Note that in our calculations we have considered only the situation in which both the wanted signal and the interfering signals experience Rayleigh fading. For a more complete analysis we may assume the wanted signal to fade differently from the interfering signals, leading to a great number of possible combinations. A case of interest is the investigation of the influence of the standard deviation in the outage probability analysis. This is illustrated in Fig. 21.2 for the Suzuki (lognormal plus Rayleigh) environment with $\sigma = 6$ dB.

Note that by definition the parameter z is a function of the carrier-to-cochannel interference ratio, which, in turn, is a function of the reuse distance. Therefore, the outage probability can be obtained as a function of the cluster size, for a given protection ratio.

The ratio between the mean signal power s_m and the mean interfering power i_m equals the ratio between their respective distances d_s and d_i such that

$$\frac{s_m}{i_m} = \left(\frac{d_s}{d_i}\right)^{-\alpha} \tag{21.21}$$

where α is the path loss coefficient. Now, (1) let D be the distance between the wanted and interfering base stations, and (2) let R be the cell radius. The cochannel interference worst case occurs when the mobile is positioned at the boundary of the cell, i.e., $d_s = R$ and $d_i = D - R$. Then,

$$\frac{i_m}{s_m} = \left(\frac{D}{R} - 1\right)^{-\alpha} \tag{21.22a}$$

or, equivalently,

$$S_m - I_m = 10\alpha \log\left(\frac{D}{R} - 1\right) \tag{21.22b}$$

In fact, $S_m - I_m = Z + R_r$. Therefore,

$$Z + R_r = 10\alpha \log\left(\sqrt{3N} - 1\right) \tag{21.23}$$

With Eq. (21.23) and the curves of Fig. 21.2, we can compare some outage probabilities for different cluster sizes. The results are shown in Table 21.4 where we have assumed a protection ratio $R_r = 0$ dB. The protection ratio depends on the modulation scheme and varies typically from 8 dB (25-kHz FM) to 20 dB [single sideband (SSB) modulation].

Note, from Table 21.4, that the standard deviation has a great influence in the calculations of the outage probability.

21.9 Conclusions

The interrelationship among the areas involved in a cellular network planning is substantial. Vocabularies belonging to topics, such as radio propagation, frequency planning and regulation, modulation schemes, antenna design, transmission, teletraffic, and others, are common to all cellular engineers.

TABLE 21.4 Probability of Cochannel Interference in Different Cell Clusters

		Outage Probability, %			
		Rayleigh		Suzuki $\sigma = 6$ dB	
N	$Z + R$, dB	$p = 75\%$	$p = 100\%$	$p = 75\%$	$p = 100\%$
1	−4.74	100	100	100	100
3	10.54	31	40	70	86
4	13.71	19	26	58	74
7	19.40	4.7	7	29	42
12	24.46	1	2.1	11	24
13	25.19	0.9	1.9	9	22

Designing a cellular network to meet system requirements is a challenging task which can only be partially and roughly accomplished at the design desk. Field measurements play an important role in the whole process and constitute an essential step used to tune the parameters involved.

Defining Terms

Outage probability: The probability of failing to achieve adequate reception of the signal due to, for instance, cochannel interference.

Spectrum efficiency: A measure of how efficiently space, frequency, and time are used. It is expressed in erlang per square meter per hertz.

Trunking efficiency: A function relating the number of subscribers per channel and the number of channels per cell for different values of blocking probability.

References

[1] Egli, J., Radio above 40 Mc over irregular terrain. *Proc. IRE.*, 45(10), 1383–1391, 1957.

[2] Hata, M., Empirical formula for propagation loss in land-mobile radio services. *IEEE Trans. Vehicular Tech.*, VT-29, 317–325, 1980.

[3] Ho, M.J. and Stüber, G.L., Co-channel interference of microcellular systems on shadowed Nakagami fading channels. *Proc. IEEE Vehicular Tech. Conf.*, 568–571, 1993.

[4] Ibrahim, M.F. and Parsons, J.D., Signal strength prediction in built-up areas, Part I: median signal strength. *Proc. IEEE.* Pt. F. (130), 377–384, 1983.

[5] Lee, W.C.Y., *Mobile Communications Design Fundamentals,* Howard W. Sams, Indianapolis, IN, 1986.

[6] Leonardo, E.J. and Yacoub, M.D., A statistical approach for cell coverage area in land mobile radio systems. Proceedings of the 7th IEE. Conf. on Mobile and Personal Comm., Brighton, UK, 16–20, Dec. 1993a.

[7] Leonardo, E.J. and Yacoub, M.D., (Micro) Cell coverage area using statistical methods. Proceedings of the IEEE Global Telecom. Conf. GLOBECOM'93, Houston, TX, 1227–1231, Dec. 1993b.

[8] Okumura, Y., Ohmori, E., Kawano, T., and Fukuda, K., Field strength and its variability in VHF and UHF land mobile service. *Rev. Elec. Comm. Lab.*, 16, 825–873, Sept.-Oct. 1968.

[9] Reudink, D.O., Large-scale variations of the average signal. In *Microwave Mobile Communications,* 79–131, John Wiley & Sons, New York, 1974.

[10] Sowerby, K.W. and Williamson, A.G., Outage probability calculations for multiple cochannel interferers in cellular mobile radio systems. *IEE Proc.*, Pt. F. 135(3), 208–215, 1988.
[11] Suzuki, H., A statistical model for urban radio propagation. *IEEE Trans. Comm.*, 25(7), 673–680, 1977.

Further Information

The fundamentals of mobile radio engineering in connection with many practical examples and applications as well as an overview of the main topics involved can be found in Yacoub, M.D., *Foundations of Mobile Radio Engineering,* CRC Press, Boca Raton, FL, 1993.

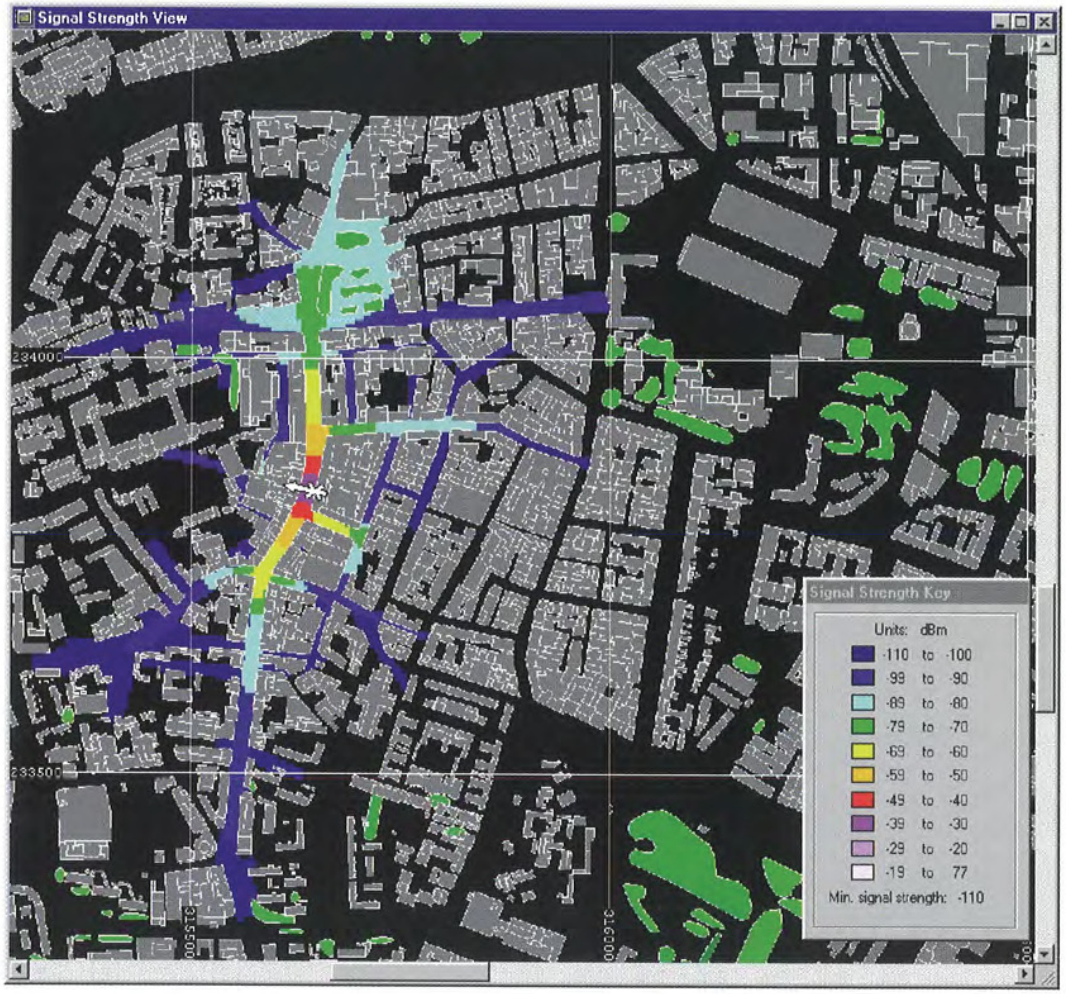


FIGURE 22.4 NP WorkPlace plot of a city street microcell. The map grid size is 500 meters.

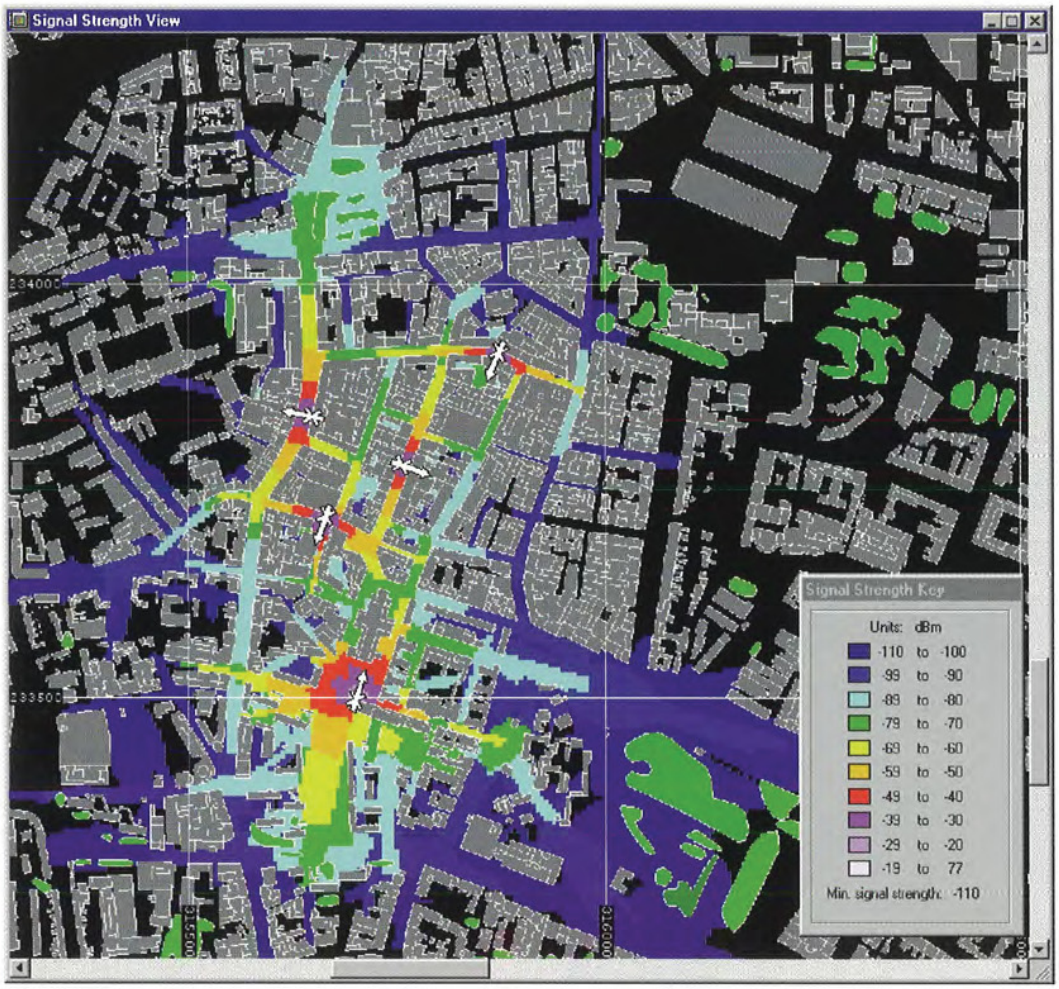


FIGURE 22.5 Cluster of microcells in a city center. The map grid size is 500 meters.

22

Microcellular Radio Communications

22.1 Introducing Microcells 22-1
22.2 Highway Microcells 22-2
Spectral Efficiency of Highway Microcellular Network
22.3 City Street Microcells 22-5
Teletraffic Issues
22.4 Indoor Microcells 22-9
22.5 Microcellular Infrastructure 22-10
Radio over Fiber • Miniaturized Microcellular BSs
22.6 Multiple Access Issues 22-13
22.7 Discussion 22-14
Defining Terms 22-14
References 22-14
Further Information 22-15

Raymond Steele
Southampton University and Multiple Access Communications

22.1 Introducing Microcells

In mobile radio communications an operator will be assigned a specific bandwidth W in which to operate a service. The operator will, in general, not design the mobile equipment, but purchase equipment that has been designed and standardized by others. The performance of this equipment will have a profound effect on the number of subscribers the network can support, as we will show later. Suppose the equipment requires a radio channel of bandwidth B. The operator can therefore fit $N_T = W/B$ channels into the allocated spectrum W.

Communications with mobiles are made from fixed sites, known as base stations (BSs). Clearly, if a mobile travels too far from its BS, the quality of the communications link becomes unacceptable. The perimeter around the BS where acceptable communications occur is called a cell and, hence, the term cellular radio. BSs are arranged so that their radio coverage areas, or cells, overlap, and each BS may be given $N = N_T/M$ channels. This implies that there are M BS and each BS uses a different set of channels.

The number N_T is relatively low, perhaps only 1000. As radio channels cannot operate with 100% utilization, the cluster of BSs or cells has fewer than 1000 simultaneous calls. In order to make the business viable, more users must be supported by the network. This is achieved by repeatedly reusing the channels. Clusters of BSs are tessellated with each cluster using the same N_T channels. This means that there are users in each cluster using

0-8493-8597-0/99/$0.00+$.50

© 1999 by CRC Press LLC

the same frequency band at the same time, and inevitably there will be interference. This interference is known as cochannel interference. Cochannel cells, i.e., cells using the same channels, must be spaced sufficiently far apart for the interference levels to be acceptable. A mobile will therefore receive the wanted signal of power S and a total interference power of I, and the signal-to-interference ratio (SIR) is a key system design parameter.

Suppose we have large cells, a condition that occurs during the initial stages of deploying a network when coverage is important. For a given geographical area G_A we may have only one cluster of seven cells, and this may support some 800 simultaneous calls in our example. As the subscriber base grows, the number of clusters increases to, say, 100 with the area of each cluster being appropriately decreased. The network can now support some 80,000 simultaneous calls in the area G_A. As the number of subscribers continues to expand, we increase the number of clusters. The geographical area occupied by each cluster is now designed to match the number of potential users residing in that area. Consequently, the smallest clusters and, hence, the highest density of channels per area is found in the center of cities. As each cluster has the same number of channels, the smaller the clusters and, therefore, the smaller the cells, the greater the **spectral efficiency** measured in erlang per hertz per square meter. Achieving this higher spectral efficiency requires a concomitant increase in the infrastructure that connects the small cell BSs to their base station controller (BSC). The BSCs are part of the nonradio part of the mobile network that is interfaced with the public switched telephone network (PSTN) or the integrated service digital network (ISDN).

As we make the cells smaller, we change from locating the BS antennas on top of tall buildings or hills, where they produce large cells or macrocells, to the tops of small buildings or the sides of large buildings, where they form minicells, to lamp post elevations, where they form **street microcells.** Each decrease in cell size is accompanied by a reduction in the radiated power levels from the BSs and from the mobiles. As the BS antenna height is lowered, the neighboring buildings and streets increasingly control the radio propagation. This chapter is concerned with microcells and microcellular networks. We commence with the simplest type of microcells, namely, those used for highways.

22.2 Highway Microcells

Since their conception by Steele and Prabhu, [10], many papers have been published on **highway microcells,** ranging from propagation measurements to teletraffic issues [1]–[8], [11]. Figure 22.1 shows the basic concepts for a highway microcellular system having two cells per cluster. The highway is partitioned into contiguous cigar-shaped segments formed by directional antennas. Omnidirectional antennas can be used at junctions, roundabouts, cloverleaf, and other road intersections. The BS antennas are mounted on poles at elevations of some 6–12 m. Figure 22.2 shows received signal levels as a function of the distance d between BS and MS for different roads [1]. The average loss in received signal level, or path loss, is approximately inversely proportional to d^4. The path loss is associated with a slow fading component that is due to the variations in the terrain, the road curvature and cuttings, and the presence of other vehicles.

The curves in the figure are plotted for an 18-element yagi BS antenna having a gain of 15 dB and a front-to-back ratio of 25 dB. In Fig. 22.2 reference is made to junctions on different motorways, e.g., junction 5 on motorway M4. This is because the BS antennas are mounted at these road junctions with the yagi antenna pointing along the highway in order to create a cigar-shaped cell. The flat part of the curve near the BS is due to the MS receiver being saturated by high-signal levels. Notice that the curve related to M25, junction 11,

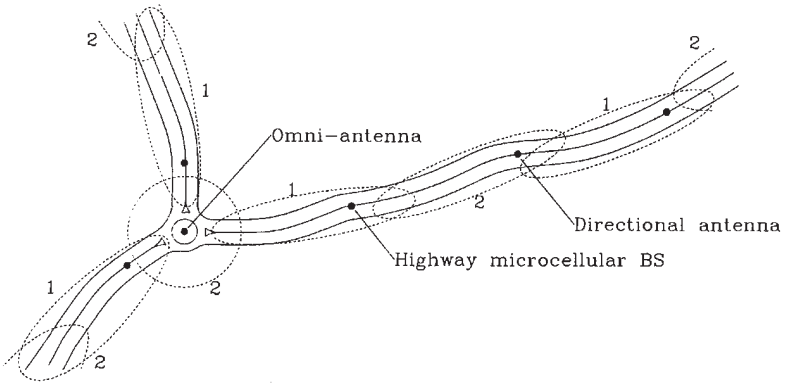


FIGURE 22.1 Highway microcellular clusters. Microcells with the same number use the same frequency set.

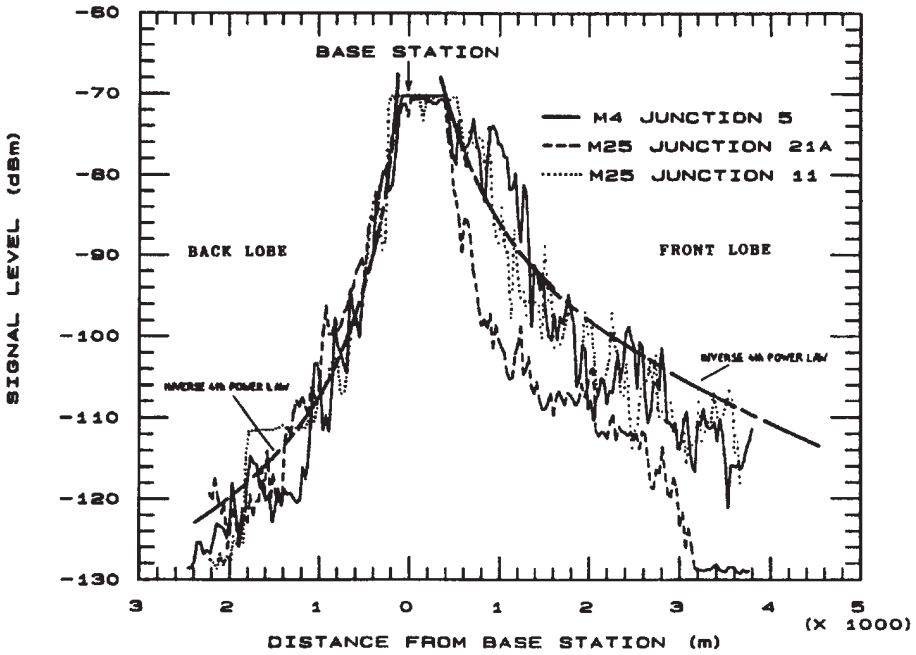


FIGURE 22.2 Overlayed received signal strength profiles of various highway cells including the inverse fourth power law curve for both the front and back antenna lobes. *Source:* Chia et al. 1987. Propagation and bit error ratio measurements for a microcellular system. *IIERE.* 57(6), 5255–5266. With permission.

decreases rapidly with distance when the MS leaves the immediate vicinity of the BS. This is due to the motorway making a sharp turn into a cutting, causing the MS to lose line of sight (LOS) with the BS. Later the path loss exponent is approximately 4. Experiments have shown that using the arrangement just described, with each BS transmitting 16 mW at 900 MHz, 16 kb/s noncoherent frequency shift keying, two-cell clusters could be formed where each cell has a length along the highway ranging from 1 to 2 km. For MSs traveling at 110 km/h the average handover rate is 1.2 per minute [1].

22.2.1 Spectral Efficiency of Highway Microcellular Network

Spectral efficiency is a key system parameter. The higher the efficiency, the greater will be the teletraffic carried by the network for the frequency band assigned by the regulating authorities per unit geographical area. We define the spectral efficiency in mobile radio communications in erlang per hertz per square meter as

$$\eta \triangleq A_{CT}/S_T W \tag{22.1}$$

although erlang per megahertz per square kilometer is often used. In this equation, A_{CT} is the total traffic carried by the microcellular network,

$$A_{CT} = CA_C \tag{22.2}$$

where C is the number of microcells in the network and A_C the carried traffic by each microcellular BS. The total area covered of the tessellated microcells is

$$S_T = CS \tag{22.3}$$

where S is the average area of a microcell, whereas the total bandwidth available is

$$W = MNB \tag{22.4}$$

whose terms M, N, and B were defined in Section 22.1. Substituting Eqs. (22.2)–(22.4) into Eq. (22.1), yields

$$\eta = \frac{\rho}{SMB} \tag{22.5}$$

where

$$\rho = A_C/N \tag{22.6}$$

is the utilization of each BS channel.

If the length of each microcell is L, there are n up lanes and n down lanes, and each vehicle occupies an effective lane length V, which is speed dependent, then the total number of vehicles in a cell is

$$K = 2nL/V \tag{22.7}$$

Given that all vehicles have a mobile terminal, the maximum number of mobiles in a cell is K. In a highway microcell we are not interested in the actual area $S = 2nL$ but in how many vehicles can occupy this area, namely, the effective area K. Notice that K is largest in a traffic jam when all vehicles are stationary and V only marginally exceeds the vehicle length. Given that N is sufficiently large, η is increased when the traffic flow is decreased.

Using fixed channel assignment (FCA) with frequency division multiple access (FDMA) or with time division multiple access (TDMA), the cluster size M can be two. Using dynamic channel assignment (DCA) with TDMA, or code division multiple access (CDMA), causes the spectral efficiency η to be very high because for a given traffic utilization ρ and channel bandwidth B, the S is small (as we are considering microcells), and M may be thought of as 1, or less, due to sectorization. The total traffic A_{CT}, given by Eq. (22.2), is also very high because by making L relatively short, C is accordingly high.

The traffic carried by a microcellular BS is

$$A_C = [\lambda_N (1 - P_{bn}) + \lambda_H (1 - P_{fhm})] \overline{T}_H \tag{22.8}$$

where P_{bn} is the probability of a new call being blocked, P_{fhm} is the probability of handover failure when mobiles enter the microcell while making a call and concurrently no channel is

available, λ_N and λ_H are the new call and handover rates, respectively, and $\overline{T}_H$ is the mean channel holding time of all calls. For the simple case where no channels are exclusively reserved for handovers, $P_{bn} = P_{fhm}$, and

$$A_C = \lambda_T \overline{T}_H (1 - P_{bn}) = A(1 - P_{bn}) \tag{22.9}$$

where

$$\lambda_T = \lambda_N + \lambda_H \tag{22.10}$$

and A is the total offered traffic. The mathematical complexity resides in calculating A and P_{bn}, and the reader is advised to consult El-Dolil, Wong, and Steele, [2], and Steele and Nofal, [8].

Priority schemes have been proposed whereby N channels are available for handover, but only $N - N_h$ for new calls. Thus N_h channels are exclusively reserved for handover [2]. While P_{bn} marginally increases, P_{fhm} decreases by orders of magnitude for the same average number of new calls per sec per microcell. This is important as people prefer to be blocked while attempting to make a call compared to having a call in progress terminated due to no channel being available on handover. An important enhancement is to use an oversailing macrocellular cluster, where each macrocell supports a microcellular cluster. The role of the macrocell is to provide channels to support microcells that are overloaded and to provide communications to users who are in areas not adequately covered by the microcells [2]. When vehicles are in a solid traffic jam, there are no handovers and so N_h should be zero. When traffic is flowing fast, N_h should be high. Accordingly a useful strategy is to make N_h adaptive to the new call and handover rates [9].

22.3 City Street Microcells

We will define a city street microcell as one where the BS antenna is located below the lowest building. As a consequence, the diffraction over the buildings can be ignored, and the heights of the buildings are of no consequence. Roads and their attendant buildings form trenches or canyons through which the mobiles travel. If there is a direct line-of-sight path between the BS and a MS and a ground-reflected path, the received signal level vs BS–MS distance is as shown in Fig. 22.3. Should there be two additional paths from rays reflected from the buildings, then the profile for this four-ray situation is also shown in Fig. 22.3. These theoretical curves show that as the MS travels from the BS the average received signal level is relatively constant and then decreases relatively rapidly. This is a good characteristic as it offers a good signal level within the microcell, and the interference into adjacent microcells falls off rapidly with distance.

In practice there are many paths, but there is often a dominant one. As a consequence the fading is Rician [7]. The Rician distribution approximates to a Gaussian one when the received signal is from a dominant path with the power in the scattered paths being negligible, to a Rayleigh one when there is no dominant path. Macrocells usually have Rayleigh fading, whereas in microcells the fading only occasionally becomes Rayleigh and is more likely to be closer to Gaussian. This means that the depth of the fades in microcells are usually significantly smaller than in macrocells enabling microcellular communications to operate closer to the receiver noise floor without experiencing error bursts and to accommodate higher cochannel interference levels. Because of the small dimensions of the microcells, the delays between the first and last significant paths is relatively small compared to the corresponding delays in macrocells. Consequently, the impulse response is generally shorter in microcells and, therefore, the transmitted bit rate can be significantly higher before in-

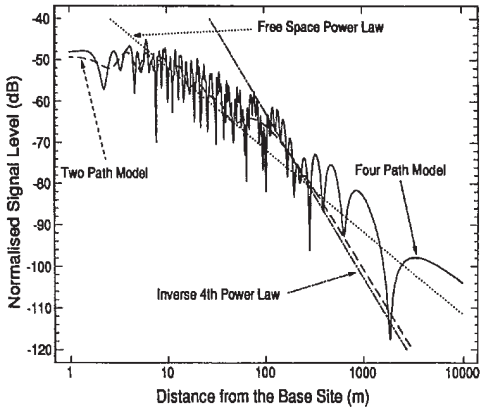


FIGURE 22.3 Signal level profiles for the two- and four-path models. Also shown are the free space and inverse fourth power laws. *Source*: Green. 1990. Radio link design for microcellular system, *British Telecom. Tech. J.*, 8(1), 85–96. With permission.

tersymbol interference is experienced compared to the situation in macrocells. Microcells are, therefore, more spectrally efficient with an enhanced propagation environment.

There are two types of these city street microcells, one for pedestrians and the other for vehicles. In general, there will be more portables carried by pedestrians than mobile stations in cars. Also, as cars travel more quickly than people, their microcells are accordingly larger than for pedestrians. The handover rates for portables and vehicular MS may be similar, and networks must be capable of handling the many handovers per call that may occur. In addition, the time available to complete a handover may be very short compared to those in macrocells.

City street microcells are irregular when the streets are irregular as demonstrated by the NP WorkPlace[1] plot of a BS in a city area displayed in Fig. 22.4. To achieve a contiguous coverage we site the BSs one at a time. Having sited the first BS and located the microcellular boundary along the streets, we locate adjacent BSs such that their boundaries butt with each other along the main streets. Unless many microcellular BSs are deployed, there will be some secondary streets where there will be insufficient signal levels. Those areas that are not covered by the microcellular BS will be accommodated by an oversailing macrocellular BS that services the complete cluster of microcellular BSs. Figure 22.5 shows a cluster of microcells; the oversailing macrocell could be sited outside the area of this figure. We emphasize that total coverage by microcells in a typical city center is difficult to achieve, and it is vital that oversailing macrocells are used to cover these dead spots. The macrocell also facilitates handovers and efficient microcellular channel utilization.

There are important differences between highway microcells and city microcells, which relate to their one- and two-dimensional characteristics. A similar comment applies to street microcells and hexagonal cells. Basically, the buildings have a profound effect on cochannel interference. The buildings shield much of the cochannel interference, and the double regression path loss law of microcells [3] also decreases interference if the break-distance constitutes the notional microcell boundary. City microcellular clusters may have

[1]NP WorkPlace is a propriety software outdoor planning tool developed by Multiple Access Communications Ltd.

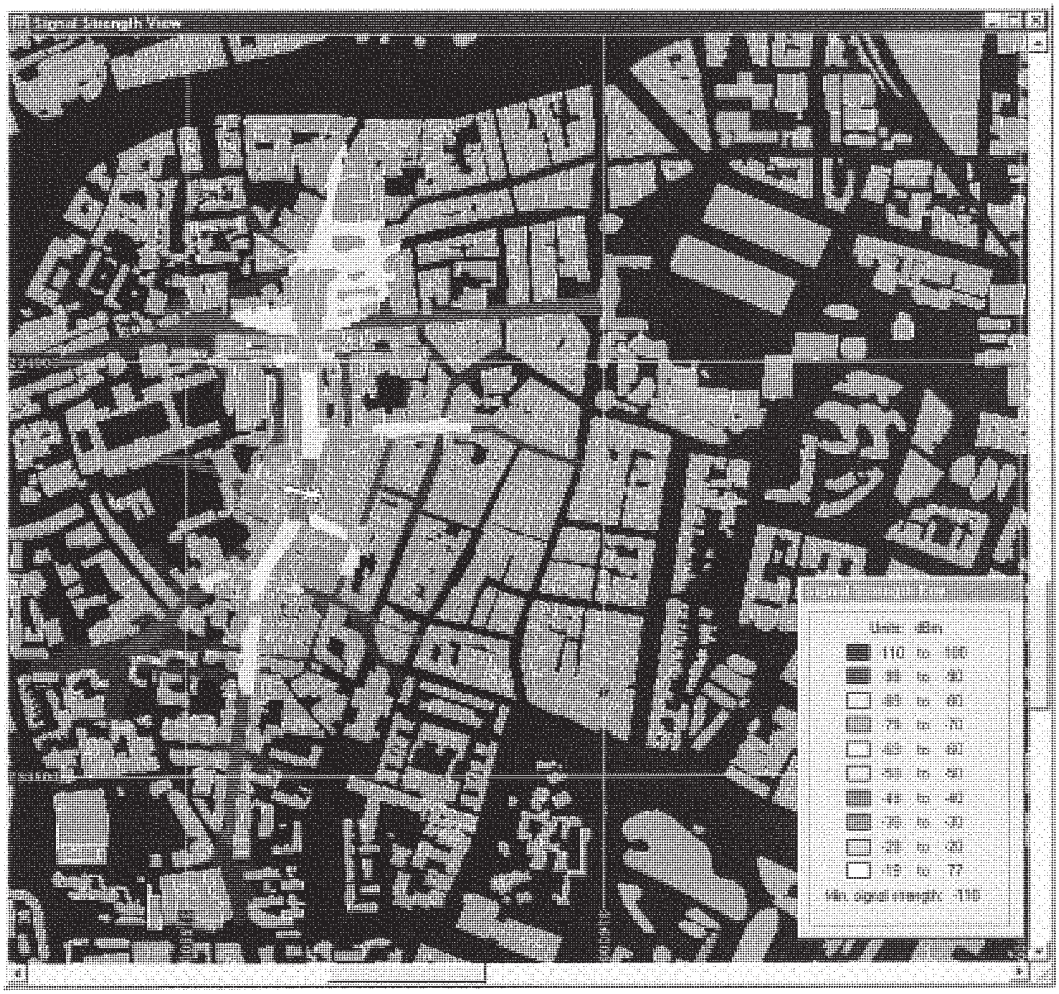


FIGURE 22.4 NP WorkPlace plot of a city street microcell. The map gridsize is 500 meters. The color version of this plot is presented elsewhere in this Handbook.

as few as two microcells, but four is more typical, and in some topologies six or more may be required. The irregularity of city streets means that some signals can find paths through building complexes to give cochannel interference where it is least expected.

22.3.1 Teletraffic Issues

Consider the arrangement where each microcellular cluster is overlaid by a macrocell. The macrocells are also clustered.The arrangement is shown in Fig. 22.6. The total traffic carried is

$$A_{CT} = C_m A_{cm} + C_M A_{CM} \tag{22.11}$$

where C_m and C_M are number of microcells and macrocells in the network, respectively. Each microcellular BS has N channels and carries A_{cm} erlang. The corresponding values

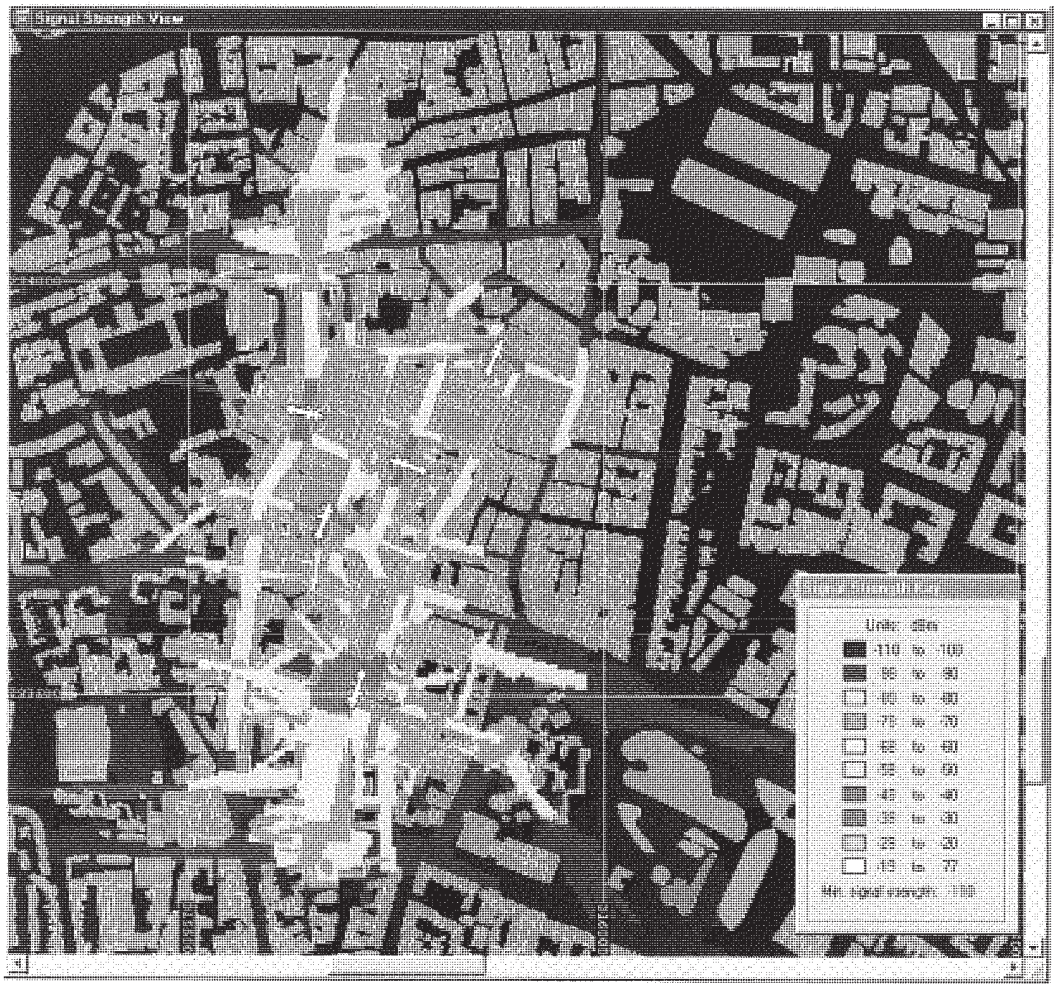


FIGURE 22.5 Cluster of microcells in a city centre. The map gridsize is 500 meters. The color version of this plot is presented elsewhere in this Handbook.

for the macrocellular BSs are N_0 and A_{CM}. The channel utilization for the network is

$$\rho_2 = \frac{C_m A_{cm} + C_M A_{CM}}{C_m N + C_M N_0} = \frac{M A_{cm} + A_{CM}}{MN + N_0} \tag{22.12}$$

where M is the number of microcells per cluster.

The spectral efficiency is found by noting that the total bandwidth is

$$B_T = B_c \left(MN + M_0 N_0\right) \tag{22.13}$$

where B_c is the effective channel bandwidth, and M_0 is the number of macrocells per macrocellular cluster. The traffic carried by a macrocellular cluster and its cluster of microcells is

$$A_M = A_{CM} M_0 + M_0 \left(A_{cm} M\right) \tag{22.14}$$

over an area of

$$S_M = \left(S_m M\right) M_0 \tag{22.15}$$

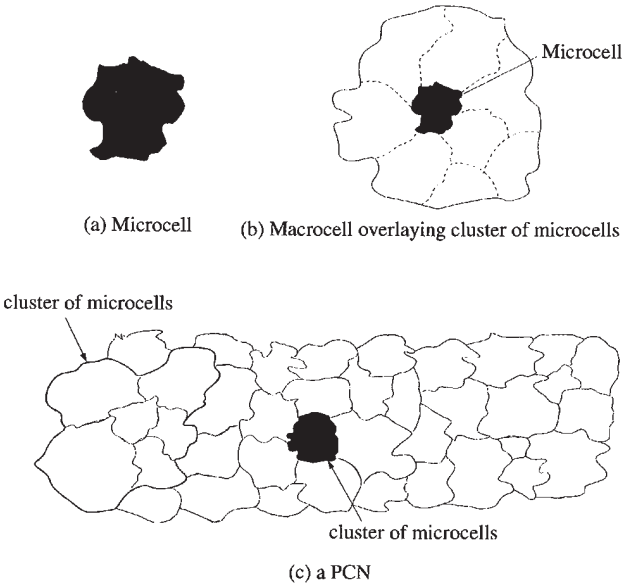


FIGURE 22.6 Microcellular clusters with oversailing macrocellular clusters. Each macrocell is associated with a particular microcellular cluster. *Source*: Steele and Williams. 1993. Third generation PCN and the intelligent multimode mobile portable. *IEE Elect. and Comm. Eng. J.*, 5(3), 147–156. With permission.

where S_m is the area of each microcell.

The spectral efficiency is, therefore,

$$\begin{aligned} \eta &= \frac{A_{CM} + A_{cm}M}{B_c\left(MN + M_0N_0\right)S_mM} \\ &= \frac{\rho_2}{B_cS_mM}\left\{\frac{MN + N_0}{MN + M_0N_0}\right\} \end{aligned} \tag{22.16}$$

We note that by using oversailing macrocells to assist microcells experiencing overloading we are able to operate the microcells at high levels of channel utilization. However, the channel utilization of the macrocells must not be high if the probability of calls forced to terminate due to handover failure is to be minuscule.

22.4 Indoor Microcells

Microcellular BSs may be located within buildings to produce **indoor microcells** whose dimensions may extend from a small office, to part of larger offices, to a complete floor, or to a number of floors. The microcells are box-like for a single office microcell; or may contain many boxes, e.g., when a microcell contains contiguous offices. Furniture, such as bookcases, filing cabinets, and desks may represent large obstacles that may introduce shadowing effects. The signal attenuation through walls, floors, and ceilings may vary dramatically depending on the construction of the building. There is electromagnetic leakage down stairwells and through service ducting, and signals may leave the building and re-enter it after reflection and diffractions from other buildings.

Predicting the path loss in office microcells is, therefore, fraught with difficulties. At the outset it may not be easy to find out the relevant details of the building construction. Even if these are known, an estimation of the attenuation factors for walls, ceilings, and floors

from the building construction is far from simple. Then there is a need to predict the effect of the furniture, effect of doors, the presence of people, and so on. Simple equations have been proposed. For example, the one by Keenan and Motley, [4], who represent the path loss in decibels by

$$PL = L(V) + 20\log_{10} d + n_f a_f + n_w a_w \tag{22.17}$$

where d is the straight line distance between the BS and the MS, a_f and a_w are the attenuation of a floor and a wall, respectively, n_f and n_w are the number of floors and walls along the line d, respectively, and $L(V)$ is a so-called clutter loss, which is frequency dependent. This equation should be used with caution. Researchers have made many measurements and found that even when they use computer ray tracing techniques the results can be considerably disparate. Errors having a standard deviation of 8–12 dB are not unusual at the time of writing. Given the wide range of path loss in mobile communications, however, and the expense of making measurements, particularly when many BS locations are examined, means that there is nevertheless an important role for planning tools to play, albeit their poorer accuracy compared to street microcellular planning tools.

As might be expected, the excess path delays in buildings is relatively small. The maximum delay spread within rooms and corridors may be <200 and 300 ns, respectively [6]. The digital European cordless telecommunication (DECT) indoor system operates at 1152 kb/s without either channel coding or equalization [7]. This means that the median rms values are relatively low, and 25 ns has been measured [6]. When the delay spread becomes too high resulting in bit errors, the DECT system hops the user to a better channel using DCA.

22.5 Microcellular Infrastructure

An important requirement of first and second generation mobile networks is to restrict the number of base stations to achieve sufficient network capacity with an acceptably low probability of blocking. This approach is wise given the cost of base stations and their associated equipment, plus the cost and difficulties in renting sites. It is somewhat reminiscent of the situation faced by early electronic circuit designers who needed to minimize the number of tubes and later the number of discrete transistors in their equipment. It was the introduction of microelectronics that freed the circuit designer. We are now in an analogous situation where we need to free the network designers of the third generation communication networks, allowing the design to have microcellular BSs in the position where they are required, without being concerned if they are rarely used, and knowing that the cost of the microcellular network is a minor one compared to the overall network cost. This approach is equivalent to installing electric lighting where we are not unduly concerned if not all the lights are switched on at any particular time, preferring to be able to provide illumination where and when it is needed.

To realize high-capacity mobile communications we need to design microcellular BSs of negligible costs, of coffee mug dimensions, and with the ability to connect them at the cost of, say, electrical wiring in streets and buildings. Cordless telecommunication (CT) BSs are already of shoe-box size, and companies are designing coffee mug-size versions. The cost of these BSs will be low in mass production, and many BSs will be equivalent in cost to one first generation analog cellular BS. Microcellular BSs could be miniaturized, fully functional BSs achieved by using microelectronic techniques and by exploiting the low-radiated power levels (<10 mW) required. At the other extreme, the microcells could be formed using distribution points (DPs) that only have optical-to-microwave converters, microwave-to-optical converters, and linear amplifiers, with the remainder of the BS at another location. In between the miniaturized, fully functional BSs and the DPs there is a

range of options that depends on how much complexity is built into the microcellular BS and how the intelligence of the network is distributed.

22.5.1 Radio over Fiber

The method of using DPs to form microcells is often referred to as radio over fiber (ROF) [5]. Figure 22.7(a) shows a microcellular BS transmitting to a MS. When the DP concept is evoked, the microcellular BS contains electrical-to-optical (E/O) and optical-to-electrical (O/E) converters as shown in Fig. 22.7(b). The microwave signal that would have been radiated to the mobile is now applied, after suitable attenuation, to a laser transmitter. Essentially, the microwave signal amplitude modulates the laser, and the modulated signal is conveyed over a single-mode optical fiber to the distribution point. O/E conversion ensues followed by power amplification, and the resulting signal is transmitted to the MS. Signals from the MS are low-noise amplified and applied to the laser transmitter in the DP. Optical signals are sent from the DP to the BS where O/E conversion is performed followed by radio reception.

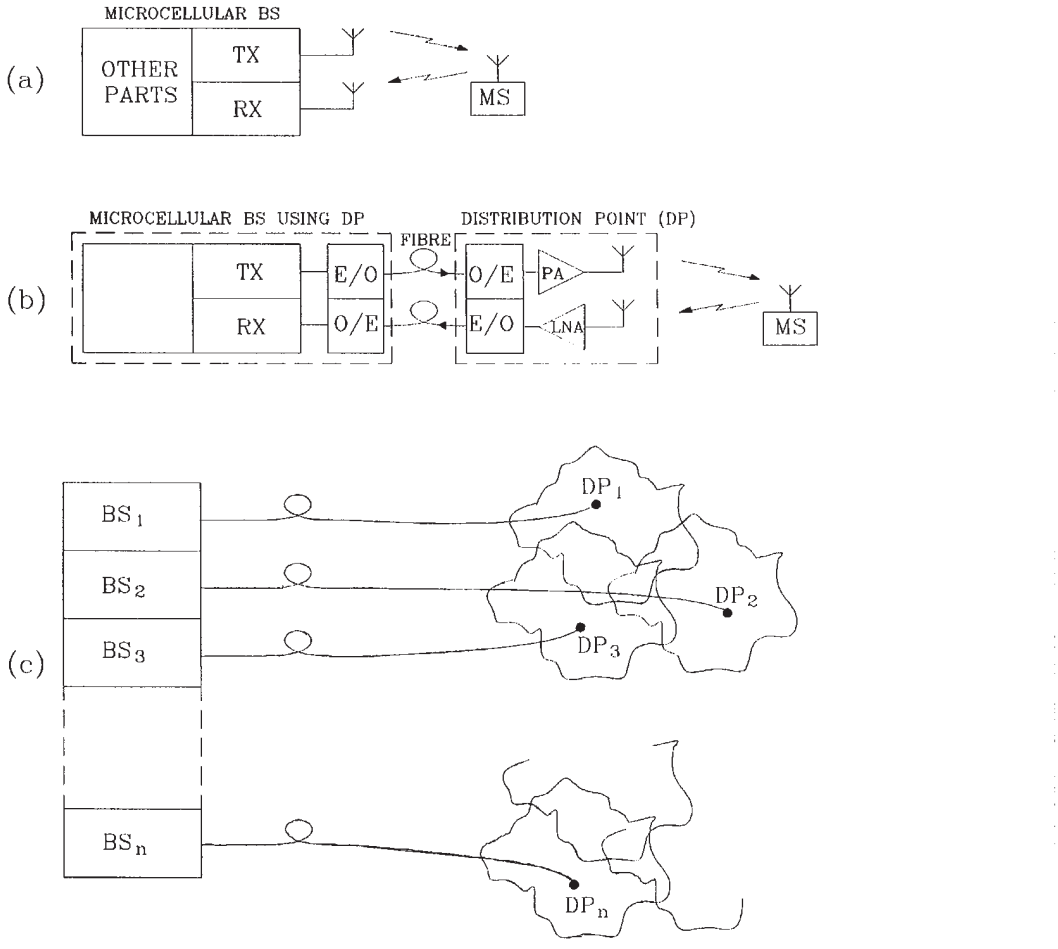


FIGURE 22.7 Creating microcells using DPs: (a) microcellular BS communicating with MS, (b) radio over fiber to distribution point, and (c) microcellular clusters using DPs.

In general, the BS transceiver will be handling multicarrier signals for many mobiles, and the DP will accordingly be transceiving signals with many mobiles whose power levels may be significantly different, even when power control is used. Care must be exercised to avoid serious intermodulation products arising in the optical components.

Figure 22.7(c) shows the cositing of n microcellular BSs for use with DPs. This cositing may be conveniently done at a mobile switching center (MSC). Shown in the figure are the DPs and their irregular shaped overlapping microcells. The DPs can be attached to lamp posts in city streets, using electrical power from the electric light supply and the same ducting as used by the electrical wiring, or local telephone ducting. The DPs can also be attached to the outside of buildings. DPs within buildings may be conveniently mounted on ceilings.

The DP concept allows small, lightweight equipment in the form of DPs to be geographically distributed to form microcells; however, there are problems. The N radio carriers cause intermodulation products (IMPs), which may be reduced by decreasing the depth of amplitude modulation for each radio channel. Unfortunately this also decreases the carrier-to-noise ratio (CNR) and the dynamic range of the link. With TDMA having many channels per carrier, we can decrease the number of radio carriers and make the IMPs more controllable. CDMA is particularly adept at coping with IMPs. The signals arriving from the MSs may have different power levels, in spite of power control. Because of the small size cells, the dynamic range of the signals arriving from MSs having power control should be <20 dB. If not, the power levels of the signals arriving at the DP from the MSs may need to be made approximately similar by individual amplification at the DP. We also must be careful to limit the length of the fiber as optical signals propagate along fibers much more slowly than radio signals propagate in free space. This should not be a problem in microcells, unless the fiber makes many detours before arriving at its DP.

The current cost of lasers is not sufficiently low for the ROF DP technique to be deployed. However, there is research into lasers, which are inherently simple, low cost, robust, and provide narrow line widths. There are also the developments in optoelectronic integrated circuits that may ultimately bring costs down. In addition, wavelength division multiplexing will bring benefits.

22.5.2 Miniaturized Microcellular BSs

The low-radiated power levels used by BSs and MSs in microcells have important ramifications on BS equipment design. Small fork combiners can be used along with linear amplifiers. Even FDMA BSs become simple when the high-radiated power levels are abandoned. It is the changes to the size of the RF components that enables the size of the microcellular BSs to be small.

The interesting question that next arises is, how much baseband signal processing complexity should the microcellular BS have? If the microcellular BSs are connected to nodes in an optical LAN, we can convey the final baseband signals to the BS, leaving the BS with the IF and front-end RF components. This means that processing of the baseband signals will be done at the group station (GS), which may be a MSC connected to the LAN. The GS will, therefore, transcode the signals from the ISDN into a suitable format. Using an optical LAN, however, and with powerful microelectronics, the transcoding and full BS operations could be done at each microcellular BS. Indeed, the microcellular BS may eventually execute many of the operations currently handled by the MSC.

22.6 Multiple Access Issues

There are three basic multiple access methods. Time division multiple access, frequency division multiple access, and spread spectrum multiple access (SSMA). SSMA comes in two versions; frequency-hopping SSMA and discrete-sequence SSMA. The latter is usually referred to as CDMA. There are also many hybrids of these systems. The principles of multiple access are described elsewhere in this book and will not be repeated here. Instead, we will comment on key factors that effect the choice of the multiple access method in microcellular environments.

As a preamble, if we observe the equations for spectral efficiency η, we see that η is inversely proportional to the number of microcells per cluster M. The smallest value of M is unity, where every microcell uses the same frequencies. Under these conditions the SIR will be low. Thus to achieve high η, we need a low value of M, and for an acceptable bit error rate (BER), we require the radio link to be able to operate with low values of SIRs. Because cellular radio operates in an intentional jamming environment, whereas CDMA was conceived to operate in a military environment where jamming by the enemy is expected, CDMA is a most appropriate multiple access method for cellular radio. The CDMA system, IS-95, will operate efficiently in single cell ($M = 1$) clusters where each cell is sectorized.

In highway microcells two-cell clusters can be used with TDMA and FDMA. Street microcells have complex shapes, see Figs. 22.4 and 22.5, and when FCA is used with TDMA and FDMA, there is a danger that high levels of interference will be ducted through streets and cause high-interference levels in a small segment of a microcell. To accommodate this phenomenon, the system must have a rapid handover (HO) capability, with HO to either a different channel at the same BS, to the interfering cell, or to an oversailing macrocell. CDMA is much more adept at handling this situation. The irregularity of street microcells, except in regularly shaped cities, such as midtown Manhattan, suggests that FCA should not be used. If it is, it requires $M \geq 4$. Instead, DCA should be employed. For example, when DCA is used with TDMA we abandon the notion of clusters of microcells. We may arrange for all microcells to have the same frequency set and design the system with accurate power control to contain the cochannel interference, and to encourage the MS to switch to another channel at the current BS or switch to a new BS directly when the SIR becomes below a threshold at either end of the duplex link. The application of DCA increases the capacity and can also contend with the situation where a MS suddenly experiences a rapid peak of cochannel interference during its travels.

The interference levels in CDMA in street microcells is mainly from users within its microcell, rather than from users in other microcells due to the shielding of the buildings. For CDMA to operate efficiently in street microcells, it should increase its chip rate to ensure it can exploit path diversity in its RAKE receiver. By increasing the chip rate, higher data rates can be accommodated and, hence, a greater variety of services. CDMA should be used in a similar way in office microcells. If the chip rate cannot be increased, however, the equipment installer can deploy a distributed antenna system where between each antenna a delay element is introduced. By this means path diversity gains are realized.

TDMA/DCA is appropriate for indoor microcells, where the complexity of the DCA is easier to implement compared to street microcells. FDMA should be considered for indoor microcells where it is well suited to provide high-bit-rate services since the transmitted rate is the same as the source rate. It also benefits from the natural shielding that exists within buildings to contain the cochannel interference and the low-power levels that simplify equipment design.

22.7 Discussion

At the time of writing, microcells are used in cordless telecommunications, where indoor microcells and outdoor telepoint microcells are used. There are very few microcells in cellular systems because there are no commercially available microcellular BSs. Nevertheless, operators have formed microcells using existing macrocellular BSs. Microcellular BSs, however, do exist in manufacturer's laboratories, and their entrance into the market is imminent. When microcells are deployed in large numbers, the vast increase in teletraffic will call for new network topologies and protocols.

In our deliberations we focused on highway microcells, city-street microcells, and indoor microcells. Minicells, where the BS antenna is below most of the buildings but above others, are currently being deployed. We may anticipate the fusion of the types of minicells and microcells. We will have microcells of strange shapes, like city street microcells but in three dimensions. Street microcells may serve the lower floors of buildings and vice versa. Microcells, located in minicell environments, may cover the streets as well as floors in neighboring buildings. We may also anticipate very small microcells, the so-called picocells. Indeed, we will have multicellular networks with multimode radio interfaces. This means that an intelligent multimode terminal with its supporting network will be required [11]. The role of microcells is to carry the high-bit-rate traffic and, hence, support a wide range of services. Our teletraffic equations tell us that microcellular personal communication networks will support orders more teletraffic than current conventional systems. Technology advancements will produce coffee cup size microcellular BSs and facilitate new network architectures that will eventually lead to the widespread concentration of intelligence at the BSs.

Defining Terms

Highway microcells: Segments of a highway having a base station and supporting mobile communications.

Indoor microcells: Small volumes of a building, e.g., an office, having a base station and supporting mobile communications.

Spectral efficiency: Has a special meaning in cellular radio. It is the traffic carried in erlang per hertz (or kilohertz) per area in square meters (or square kilometeres).

Street microcells: Small cells whose shape are determined by the street topology and their buildings. The base station antennas are below the urban skyline.

References

[1] Chia, S.T.S., Steele, R., Green, E., and Baran, A., Propagation and bit error ratio measurements for a microcellular system. *JIERE,* Supplement 57(6), 5255–5266, 1987.

[2] El-Dolil, S.A., Wong, W.C., and Steele, R., Teletraffic performance of highway microcells with overlay macrocell. *IEEE JSAC,* 7(1), 71–78, 1989.

[3] Green, E., Radio link design for microcellular systems. *British Telecom Tech. J.,* 8(1), 85–96, 1990.

[4] Keenan, J.M. and Motley, A.J., Radio coverage in buildings. *British Telecom. Tech. J.,* 8(1), 19–24, 1990.

[5] Merrett, R.P., Cooper, A.J., and Symington, I.C., A cordless access system using radio-over-fiber techniques. IEEE VT-91, 921–924.

[6] Saleh, A.A.M. and Valenzula, R.A., A statistical model for indoor multipath propagation. *IEEE JSAC,* 128–137, Feb. 1987.

[7] Steele, R., *Mobile Radio Communications,* Pentech Press, London, 1992.

[8] Steele, R. and Nofal, M., Teletraffic performance of microcellular personal communication networks. *IEE Proc-I,* 139(4), 448–461, 1992.

[9] Steele, R., Nofal, M., and El-Dolil, S., An adaptive algorithm for variable teletraffic demand in highway microcells. *Electronic Letters,* 26(14), 988–990, 1990.

[10] Steele R. and Prabhu, V.K., Mobile radio cellular structures for high user density and large data rates. In *Proc of the IEE,* Pt. F. (5), 396–404, 1985.

[11] Steele, R. and Williams, J.E.B., Third generation PCN and the intelligent multimode mobile portable, *IEE Elec. and Comm. Eng. J.,* 5(3), 147–156, 1990.

Further Information

The *IEEE Communications Magazine Special Issue* on an update on personal communications, Vol. 30, No. 12, Dec. 1992 provides a good introduction to microcells, particularly the paper by L.J. Greenstein, et al.

23

Fixed and Dynamic Channel Assignment

23.1 Introduction 23-1
23.2 The Resource Assignment Problem 23-1
23.3 Fixed Channel Assignment 23-2
23.4 Enhanced Fixed Channel Assignment 23-3
23.5 Dynamic Channel Assignment 23-4
23.6 CDMA Systems 23-6
23.7 Conclusion 23-7
References 23-7

Bijan Jabbari
George Mason University

23.1 Introduction

One of the important aspects of frequency reuse-based cellular radio as compared to early land mobile telephone systems is the potential for dynamic allocation of channels to traffic demand. This fact had been recognized from the early days of research (e.g., see [18, Chapter 7], [7, 8]) in this field. With the emergence of wireless personal communications and use of microcell with nonuniform traffic, radio resource assignment becomes essential to network operation and largely determines the available spectrum efficiency. The primary reason for this lies in the use of microcell in dense urban areas where distinct differences exist as compared to large cell systems due to radio propagation and fading effects that affect the interference conditions.

In this chapter, we will first review the channel reuse constraint and then describe methods to accomplish the assignment. Subsequently, we will consider variations of fixed channel assignment and discuss dynamic resource assignment. Finally, we will briefly discuss the traffic modeling aspect.

23.2 The Resource Assignment Problem

The resources in a wireless cellular network are derived either in frequency division multiple access (FDMA), time division multiple access (TDMA), or joint frequency–time (MC-TDMA) [1]. In these channel derivation techniques, the frequency reuse concept is used throughout the service areas comprised of cells and microcells. The same channel is used by distinct terminals in different cells, with the only constraint of meeting a given interference

0-8493-8597-0/99/$0.00+$.50

© 1999 by CRC Press LLC

threshold. In spread spectrum multiple access (SSMA) such as the implemented code division multiple access (CDMA) system (IS-95) [24] each subscriber spreads its transmitted signal over the same frequency band by using a pseudorandom sequence simultaneously. As any active channel is influenced by the others, a new channel can be set up only if the overall interference is below a given threshold. Thus, the problem of resource assignment in CDMA relates to transmission power control in forward (base station to mobile terminal) and reverse (mobile terminal to base station) channels. Of course, the problem of power control applies to TDMA and FDMA as well, but not to the extent that it impacts the capacity of CDMA. Here we first focus on time- and frequency-based access methods. We subsequently present some preliminary discussion on spread spectrum CDMA systems.

Fixed channel assignment (FCA) and dynamic channel assignment (DCA) techniques are the two extremes of allocating radio channels to mobile subscribers. For a specific grade of service and quality of transmission, the assignment scheme provides a tradeoff between spectrum utilization and implementation complexity. The performance parameters from a radio resource assignment point of view are interference constraints (quality of transmission link), probability of call blocking (grade of service), and the system capacity (spectrum utilization) described by busy hour erlang traffic that can be carried by the network. In a cellular system, however, there exist other functions, such as handoff and its execution or radio access control. These functions may be facilitated by the use of specific assignment schemes and, therefore, they should be considered in such a tradeoff [14].

The problem of channel assignment can be described as the following: Given a set of channels derived from the specified spectrum, assign the channels and their transmission power such that for every set of assigned channels to cell i, $(C/I)_i > (C/I)_0$. Here, $(C/I)_0$ represents the minimum allowed carrier to interference and $(C/I)_i$ represents carrier to interference at cell i.

23.3 Fixed Channel Assignment

In fixed channel assignment the interference constraints are ensured by a frequency plan independent of the number and location of active mobiles. Each cell is then assigned a fixed number of carriers, dependent on the traffic density and cell size. The corresponding frequency plan remains fixed on a long-term basis. In reconfigurable FCA (sometimes referred to as flexible FCA), however, it is possible to reconfigure the frequency plan periodically in response to near/medium term changes in predicted traffic demand.

In FCA, for a given set of communications system parameters, $(C/I)_0$ relates to a specific quality of transmission link (e.g., probability of bit error or voice quality). This parameter in turn relates to the number of channel sets [18] (or cluster size) given by $K = 1/3\,(D/R)^2$. Thus, the ratio D/R is determined by $(C/I)_0$. Here D is the cochannel reuse distance and R is the cell radius. For example, in the North American cellular system advanced mobile phone service (AMPS), $(C/I)_0 = 18$ dB, which results in $K = 7$ or $D = 4.6R$. Here, we have used a propagation attenuation proportional to the fourth power of the distance. The radius of the cell is determined mainly by the projected traffic density. In Fig. 23.1 a seven cell cluster with frequency sets F1 through F7 (in cells designated A–G) has been illustrated. It is seen that the same set of frequencies is repeated two cells away.

The number of channels for each cell can be determined through the erlang-B formula (for example, see Cooper [6]) by knowing the busy hour traffic and the desired probability of blocking (grade of service). Probability of blocking P_B is related to offered traffic A, and