NTIA Report 94-311

A Survey of Relative Spectrum Efficiency of Mobile Voice Communication Systems

R.J. Matheson

U.S. DEPARTMENT OF COMMERCE
Ronald H. Brown, Secretary

Larry Irving, Assistant Secretary
for Communications and Information

July 1994

This Page Intentionally Left Blank

This Page Intentionally Left Blank

Preface

Certain commercial radio systems are identified in this paper as examples of mobile voice communication systems. In no case does such identification imply recommendation or endorsement by the National Telecommunication and Information Administration, nor does it imply that the system is necessarily the best available for the purpose. Neither does the comparison of any systems imply that those systems provide an equal level of performance, or that any other conclusions regarding the relative suitability or desirability of such systems for any purposes should be drawn from this study. Although substantial care was taken to verify characteristics of systems discussed in this report, we make no representation that such system descriptions are necessarily correct.

This Page Intentionally Left Blank

This Page Intentionally Left Blank

CONTENTS

FIGURES

TABLES

A SURVEY OF RELATIVE SPECTRUM EFFICIENCY OF MOBILE VOICE COMMUNICATION SYSTEMS

Robert J. Matheson[1]

This paper provides definitions of spectrum efficiency for general communication systems, then simplifies the definitions so that the spectrum efficiency of several contemporary mobile radio systems can be easily compared. A variety of systems currently in use and proposed for near-term deployment are compared to analog FM dispatch radio. The calculations show a ratio of 1 million between the most efficient and least efficient of the systems considered. Based on these calculations, a comparison can be made between various technologies and their ability to deliver communication systems with a very high spectrum efficiency.

Key words: cellular radio, frequency reuse, mobile radio, PCS, spectrum efficiency, spectrum policy, trunked radio systems, vocoders.

1. INTRODUCTION

In the 1992 Bill authorizing the FY93 budget for the National Telecommunications and Information Administration (NTIA), the Congress included a requirement for NTIA to produce a "plan for Federal agencies with existing mobile radio systems to use more spectrum-efficient technologies that are at least as spectrum-efficient and cost-effective as readily available commercial mobile radio systems." In 1991 the Federal Communications Commission began a several year study called the mobile radio "Refarming", designed to improve the efficiency of non-Federal mobile radio systems so that the current severe crowding could be alleviated without the allocation of large amounts of additional spectrum for mobile radio. Since the Federal and non-Federal frequency managers are both wrestling with the problem of mobile spectrum efficiency, it seems appropriate to examine the theoretical foundations of spectrum efficiency, as well as to examine some aspects of spectrum efficiency of existing systems.

The "spectrum efficiency" described in this report has been described in past reports as "technical spectrum efficiency." This means that we are describing only the "technical" aspects of spectrum efficiency, without regard to the very important practical considerations of cost, reliability, or even minimum suitability to perform a particular function. Thus, a comparison of the spectrum efficiency of two systems should not be expected to show which system is "better." Instead, the comparison of spectrum efficiency should be expected to show only which system has the greater technical spectrum efficiency, as defined by the equations within this paper. For example, in a later section of this paper, the spectrum efficiencies of an air traffic control (ATC) channel and a cellular telephone are compared. Although the cellular telephone is much more efficient according to the definitions established here, there is no intention to imply that cellular telephones are "better" than air traffic control channels. The two systems perform much different functions, in greatly different environments, meeting very different constraints.

---

[1]The author is with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, CO 80303

Similarly, for systems that are less obviously different from each other than cellular phones and ATC channels, the reader should remember that the comparison is based only on the narrow criterion of technical spectrum efficiency. Before any conclusions can be drawn regarding which system is better for a particular function, many other factors need to be considered. A partial list of these other factors might include equipment cost, infrastructure cost, technical complexity, reliability, time needed for deployment, robustness under particular circumstances, time to repair, size, weight, power consumption, required transmission distance, message privacy/encryption, over-the-air distribution of encryption keys, ease of operation, spectrum availability, cost of operation, full duplex operation, group call capabilities, compliance with existing protocols, required customer density, expandability/flexibility, legal constraints, compatibility with other systems, biological hazards, consumer preferences, Government regulations, development costs, historical precedents, priority access, operating environment/terrain, etc., etc. These other factors are outside the scope of this report, as are the more general conclusions that might be drawn if they were present.

Section 2 contains definitions of spectrum efficiency, including some ways to simplify these definitions to allow rough estimates of comparative spectrum efficiency to be made. Section 3 contains descriptions of some mobile communications systems used by the Federal Government and other non-Federal users, covering a range of traditional and proposed systems. The descriptions of mobile radio systems in this section are included mainly to more fully describe the fairly terse system descriptions in Section 4. Section 4 applies the simplified definitions of relative spectrum efficiency to many of the types of mobile systems described in Section 3, comparing them to conventional analog FM mobile radios. This section shows a very wide range of spectrum efficiency (more than 1,000,000:1) between the least efficient and the most efficient systems.

Section 5 summarizes the findings of Section 4. Section 6 uses the summary of spectrum efficiency to propose some directions in which future mobile systems might move.

The approach taken in this paper relies on the use of convenient approximations and easy assumptions, which may cause concern for the accuracy of the results. Although the results admittedly may suffer a higher degree of uncertainty because of the approximations and assumptions, they do not substantially affect the conclusions of the paper, and they permit many systems to be compared easily, even while lacking detailed information on the technical and deployment parameters of many systems.

2. DEFINITION OF SPECTRUM EFFICIENCY

This section defines several quantities related to spectrum efficiency[2]. Some of these quantities are difficult to define and even more difficult to calculate for real systems. Therefore, we will also define some relative measures, in which some problematic terms may cancel out, allowing useful calculations to be more easily made. Finally, we will step back and look at the communications process as a whole, including some

---

[2]These definitions closely follow the definitions in CCIR Report 662-3 [1]. They have been expanded to include some alternative definitions that more closely match the functional requirements of a wider variety of radio systems.

factors that might not be considered by a communications engineer, but which greatly affect the use of communications systems.

2.1 General Definitions

We will first define spectrum utilization, i.e., the amount of the radio spectrum that is being used in a particular situation. Next, we will define spectrum utilization efficiency, which is the ratio of the amount of communications achieved per the amount of spectrum space used. Since one major use of spectrum utilization efficiency information is in comparing the efficiencies of two systems, we will also define relative spectrum efficiency.

2.1.1 Spectrum Utilization

A radio system operates at a particular frequency (actually across a particular bandwidth), at a given location, and at a particular time. At frequencies sufficiently close to that operating frequency, other nearby radio systems may not be able to operate without causing interference or receiving interference. However, the range of a radio system is not infinite; outside of some distance another radio system could operate on the same frequency without causing interference or receiving it. Finally, some radio systems are not used all of the time; they will not cause or receive interference when they are not being used. Therefore, there is a time factor associated with the radio system.

We define spectrum utilization as the product of the bandwidth, the geometrical (geographic) space, and the time denied to other potential users. The measure is

$$U = B \times S \times T, \tag{1}$$

where U is spectrum utilization,
B is the radio bandwidth,
S is geometric space, and
T is time.

The formula for utilization is a general conceptual formula that will need to be made specific when it is applied to an actual case. It can be quite difficult to apply this concept to a given system, partly because the mathematics become very detailed and partly because there are a number of judgements/assumptions that must be made. The formula represents no specific number until the assumptions are made, and there is no obvious set of default values or even "ideal" values that can be used. Much of the remaining discussion about the definitions in the remainder of Section 2 will deal with the problem of making reasonable choices for a specific system.

Transmitters and receivers both use spectrum space. Transmitters use spectrum space by denying the use of that space to nearby receivers (other than the intended receiver), which would receive interference from the transmitter. This space is called "transmitter-denied space," or simply "transmitter space."

Receivers use spectrum space by denying the use of that space to additional transmitters (assuming that the receiver is entitled to protection from interference). A transmitter operating within that space would cause interference to the receiver's intended operation. This space is called "receiver-denied space," or simply "receiver space."

For most communication systems, receiver and transmitter spectrum spaces are both important. For some systems, however, one space may be much more important than the other. Radio astronomy, for example, utilizes only receiver space. The presence of a radio astronomy receiver denies the use of nearby spectrum space to other transmitters. Expressed differently, a radio astronomy receiver is surrounded by "receiver-denied" spectrum space. There is no corresponding practical transmitter space, since the "transmitters" are located at astronomical distances from our area of concern. An ISM[3] transmitter, on the other hand, may be surrounded only by "transmitter-denied" spectrum space, where other receivers may experience interference.

The process of calculating the amount of spectrum space used can be formidable. Propagation models and terrain databases may be needed to calculate coverage. Transmitter emission characteristics and receiver susceptibility characteristics must be used to determine bandwidth factors. In the case of mobile systems, transmission time statistics and transmitter/receiver location statistics must be known. Transmitter and receiver antenna patterns are important for many systems. Some choices will have to be made about defining the tolerable level of interference that can be received by particular types of systems. Finally, these calculations must be made for every combination of receivers and transmitters within a given operating area. Even if fast computers are used for the calculations, database limitations and other practical matters will require that simplifying assumptions be used when performing the computations. Even if the complexity problem were solved, there would remain a problem of selecting the appropriate values for many parameters. There is a wide range of defensible choices for many of these parameters, which can significantly affect the answers and severely challenge the credibility of the results.

2.1.2 Spectrum Utilization Efficiency

Spectrum utilization efficiency is defined as the ratio of information transferred to the amount of spectrum utilization. The measure is

$$Q = I/U = I/(B \times S \times T) \qquad (2)$$

where
Q is spectrum utilization efficiency,
I is the rate of information transfer,
U is the amount of spectrum utilization = spectrum space used
B is bandwidth,
S is geometric space, and
T is time.

[3]ISM: Industrial, scientific, and medical transmitters are controlled by Part 18 of the FCC rules. These systems include transmitters used for various processes like RF welding of plastics, microwave ovens, medical diathermy, etc.

Since the intent of using the spectrum is to transfer information, the spectrum utilization efficiency is a technical measure of how efficiently one is using the spectrum. (There are, of course, many additional factors that must be taken into account besides efficiency—such as cost, reliability, availability of equipment, special operational requirements, etc.). The difficulty of calculating a value for U (spectrum utilization) has been discussed in the previous section, leaving only the problems associated with choosing realistic assumptions about I, the amount of information transferred by the system.

For some digital systems, the rate at which information is transferred can be easily described in terms of a bit rate. For other systems, there can be problems determining exactly how much data is being transferred. How much data is there in a TV image? What if the image does not change from frame to frame? How much less information is in a poor TV image than a good one? What is the information rate of a radar or a flood warning alarm system. Does the absence of an airplane on a radar screen convey an equal amount of information as the presence of one? How much information is the flood warning alarm transmitting when there is no flood? Although these questions suggest some difficultly in assigning a number to the information rate, a numerical value will need to be established before a value for Q can be computed.

Even for a "simple" voice system, there can be considerable disagreement about the information rate of the system. If one speaks 3 words per second using a basic vocabulary of 8000 words, the basic data rate of a voice channel would be 39 bits/s (assuming that each word in an 8000-word vocabulary could be designated by a specific 13-bit code). On the other hand, Nyquist sampling on a 3-kHz audio channel using 8-bit quantization would require a data rate of at least 48,000 bits/s. Both answers (and everything in between) can be defended rigorously on solid technical grounds. Thus, arbitrary (but reasonable and necessary) assumptions could lead to a 1200:1 range of answers for the information rate of a voice system.

Similar orders-of-magnitude differences affect assumptions used to calculate the amount of spectrum space used by a system. The effect of such assumptions on the numerical values of spectrum efficiency is sufficient to cause some distrust in numerical values from any such calculations made on practical systems, especially if the numbers have been derived by different investigators using (slightly?) different assumptions. It would be risky to use such calculations to compare the spectrum efficiencies of two systems, when those two efficiencies are expected to be approximately equal (perhaps differing by a factor of 10 or less).

2.1.3 Relative Spectrum Efficiency

Systems that perform the same function can be compared using the concept of relative spectrum utilization efficiency. We define the quantity of "relative efficiency" as the ratio of two spectrum utilization efficiencies:

$$R = Q_a/Q_b, \tag{3}$$

where R = relative efficiency = ratio of spectrum utilization efficiencies,
Q_a = spectrum efficiency of system A,
Q_b = spectrum efficiency of system B.

Although the use of R may seem to have doubled the computational problems, it often makes the computation much easier. Since many of the same difficult-to-compute quantities are part of Q_a and Q_b, they cancel out without ever having to be calculated. For example, if system A and system B perform the same function, they must cause the same amount of information to be transferred (even if the transfer mechanisms are completely different). This means that the "I_a" and "I_b" terms are equal and cancel out in the computations, avoiding the associated conceptual and calculation problems.

In the specific case of voice channels, we assume that each system carries the same amount of data, but we needn't define exactly how much data that is. The purist might complain that one voice channel has a better signal-to-noise ratio, less decoding delay, more natural sounding speech, etc.—i.e., that one voice channel is not equal to another. Part of the answer to the objection is to insist that the systems actually have to perform the same functions. A high-fidelity studio-to-transmitter audio link meets different functional specifications than a police mobile radio. The other part of the answer is that even if voice channels don't meet exactly the same specifications, they will still be much more alike than the 1200:1 differences that were encountered in the assumptions needed to calculate absolute information transfer rates.

Similarly, many of the difficult calculations that are part of the spectrum utilization factors may cancel out. This often leaves a simple set of ratios to be calculated, e.g. system A is essentially identical to system B, except system B uses twice the bandwidth, system A has 3 times the coverage area, etc.

It is tempting to over-simplify the calculation of R, by pretending that a single change does not cause any other changes. For example, we might compare two systems that are "identical," except that one uses 16-QAM modulation and the other uses 32-QAM. We might simply note that the 32-QAM system would require half the bandwidth and declare it to have twice the efficiency of the 16-QAM. A closer examination might show that the 32-QAM system requires a better signal-to-noise ratio and signal-to interference ratios to operate, requiring more transmitter power and making it less tolerant of interference, etc. This requires adjustments to other parameters besides the bandwidth. This situation is typical, but calculations of relative efficiency are still usually much simpler than calculations of absolute efficiency.

The ratio of the two computed efficiencies is often more useful than the absolute values of the two efficiencies, since the efficiencies don't relate to anything familiar. (Is a spectrum utilization efficiency of 0.79 Mbits/s/MHz/m^2 good or bad? Or what?) In contrast, the relative spectrum efficiency will show, for example, that system A is 3.5 times as efficient as system B.

When several systems providing the same service are compared, it is useful to compare all of them to a single reference system providing the same service. The system selected as the reference system can be

a. the most efficient system that can be practically built, or
b. a system that can be easily defined and understood, or
c. a system that is widely used—a de facto industry standard.

In Section 4, we compare the efficiency of several types of mobile systems with an analog FM dispatch mobile radio, since this is the most widely used mobile system and its characteristics are well known.

2.2 Comments on the Definition of Spectrum Efficiency

This section discusses some alternative definitions of spectrum efficiency that may be more suitable for certain types of systems. Section 2.2.1 contains alternative definitions of service provided or information transferred (I). Section 2.2.2 provides a discussion of alternative definitions of spectrum space consumed.

2.2.1 Alternative Definitions for Information Transferred

In the definition of spectrum efficiency (Equation 2), the definition of I (I = information transfer rate) contains only a factor for the amount of data transferred. It does not include factors for the distance the data travelled or the number of places the data was received (or could potentially have been received). These are important constraints on the definition of spectrum efficiency; they are responsible for the high efficiency calculated for short-path systems. Since this definition implies the types of systems that are considered efficient, it is reasonable to ask whether alternative definitions should be considered.

A generic radio system often performs two separate functions: transport and access. The transport function includes moving data from one geographic location to another. The access function provides the user a convenient way to get access to data that is already close by. The role of a point-to-point microwave system is mainly transport. The role of a cordless telephone is mostly access. Traditional radio systems like broadcasting stations often include both roles of access and transport. Today, the transport role is being more completely filled with optical fiber, while wireless technologies are being optimized to fill the access role.

The selected definition (Equation 2) of spectrum efficiency is consistent with the new paradigms in communications, where radio mainly furnishes a convenient means to access data. If data can be easily transported anywhere by non-radio means (e.g., optical fiber), there is little additional benefit achieved by using radio to transport data over a large distance. Therefore, this definition of spectrum efficiency—which we will call the "access" version—is proper in omitting the "transport" role of radio communications.

There are, however, many radio systems that are designed to perform the transport role. For these systems, one may wish to use a definition of spectrum efficiency that recognizes the transport role. For example, consider the spectrum efficiency of point-to-point microwave systems. The major function of such systems is to transport data over long distances, and the definition of spectrum efficiency should include a factor that includes the distance over which the data is transported. In this case the definition for Q_T (transport version) would be

$$Q_T = I \times D / U, \tag{4}$$

where Q_T is spectrum utilization efficiency (transport version),
I is the rate of information transfer,
D is the distance over which the information is transported,
U is the amount of spectrum utilization = spectrum space used.

A broadcasting application requires another definition of spectrum efficiency, since neither the identity, quantity, nor location of the users is known and the system cannot be expected to provide specific links to these unknown users. For broadcasting, the object is to provide a "coverage area," within which all potential or actual listener/viewers can be served. In this case, the linear factor D should be replaced with an area factor A, matching the benefit of having a larger coverage area. This yields

$$Q_C = I \times A / U, \quad (5)$$

where Q_C is spectrum utilization efficiency (coverage version),
I is the amount of information transferred,
A is the area of coverage, and
U is the amount of spectrum utilization = spectrum space used.

These examples illustrate the need to choose the appropriate definition of spectrum efficiency, as applied to a particular radio system. The access version (equation 2) is the proper version when the basic objective is to establish a communications link, including most mobile, wireless, and PCS applications. Equation 4 (transport version) is the definition that applies to radio systems whose chief purpose is to transmit a signal over a long distance, such as long-haul microwave. Equation 5 (coverage version) contains the proper definition for radio systems whose purpose is to provide services to all locations in a specified coverage area—a typical broadcasting or radar application.

It may be unclear which definition applies in a given circumstance. In a mobile service, the access version would normally apply. Consider, however, a mobile system that allows "group call" operations (where all members of a selected group must be able to listen to all traffic on a given channel). If users in a selected call group are scattered over a large area, it may be necessary to transmit a message throughout every cell of a large multi-cell system. This situation begins to look suspiciously like a broadcasting operation, which would be described using the area version of the definition.

The crucial difference is whether the system can identify the locations of users earmarked to receive group call messages. If the system knows who and where the users are, one could imagine a system optimally configured to provide the required links while using minimum spectrum space. Therefore, the access definition would apply. On the other hand, if the service is such that the system cannot not know who or where the users are, there is no possible way to ensure communications with the required users, except to provide sufficient signal over a required coverage area. Under this circumstance, the "coverage" definition would apply.

Similar arguments can be applied to a long-range mobile radio system, where it is required that the radio signal travels a long distance. A mobile system would normally use the access definition. However, in the case of a long-range HF communication system, the mobile receiver may be 3000 km away, in an area where there is no other communications infrastructure. It would be reasonable to wonder whether this radio is being used more to provide access or to provide transport. One sufficient, but not necessary, criterion for the use of the transport definition: use the transport definition if the total distance requires the use of multiple hops because a single hop cannot span the distance. Thus, a long-haul microwave system with four hops would use the transport definition.

2.2.2 Alternative Definitions for Spectrum Space Utilized.

The definition for spectrum utilization (Equation 1) is U = Bandwidth x Space x Time. Although space is nominally a 3-dimensional quantity (i.e., volume), we often assume that space is constrained to a thin layer near the surface of the earth, since that is the only volume which will usually contain transmitters or receivers. If the thickness of this layer is less than r, where r is the typical frequency reuse distance, there is no chance of frequency reuse along the vertical dimension. Accordingly, we would assume that the volume of interest has the properties of a surface and varies as r^2, where r is the radius of a circle on the surface of the earth. For the case where r is very small, however, the geometry may become 3-dimensional, and the amount of space used will vary as r^3. This situation occurs, for example, in an urban skyscraper environment, where frequencies may also be reused on a vertical basis (e.g. every 10th floor). The use of an r^3 factor would give an even stronger theoretical advantage to short path systems. In most of the systems considered in this paper, however, r is sufficiently large that the use of r^2 is appropriate.

It has also been suggested that a "nodal" definition of spectrum usage be used for point-to-point fixed microwave systems. This definition is based on the operational limitations of this service, where many microwave links must converge at a small number of sites, and the ability to add more links at a crowded site limits how densely the band can be used. This definition would be particularly dependent on the narrowness of the antenna beamwidth and freedom from serious sidelobes.

A linear definition of spectrum usage might be used for geosynchronous satellites. Since there is only a single circular orbit line 36,000 km (22,400 mi) above the equator where geosynchronous satellites can be parked, the linear position along this orbit may be a useful simplification of geometrical space.

2.2.3 Ambiguities Resolved by Relative Spectrum Efficiency

No matter what the exact form of the definition for spectrum efficiency, the definition of relative spectrum efficiency will help to keep the answers useful. As long as the systems being compared use the same definition of spectrum efficiency, the value for relative spectrum efficiency will not be affected. This is another instance where the use of relative spectrum efficiency avoids the need to make critical, but ambiguous, choices.

2.3 Factors that Constitute Relative Spectrum Efficiency

The previous section described a quantity called "relative spectrum efficiency." This section builds on that concept, by breaking down a mobile communication system into a number of component factors for relative spectrum efficiency, each of which can independently be compared with various alternative technologies for performing that function. Although there may be other ways to split up the factors that cause spectrum to be used, we have attempted to select factors that are independent of each other—i.e., so that changes in one factor do not cause changes in other factors. These factors are factors in the mathematical sense; doubling any one of them will cause the amount of spectrum required to change by a factor of two. Each of the factors can be considered to be a "relative efficiency" factor, a pure number that can be increased or decreased by some ratio.

We have broken these factors out in order to study what makes the use of mobile radio to be efficient and what makes it to be inefficient, as well as to understand where improvements are most easily possible. These factors will use the present technology as a reference. In particular, we will relate these factors to a conventional FM analog mobile radio system. The factors are listed below, beginning with the most immediate (personal) to the most global (detached). The particular order does not change the interpretation of the factors.

Message efficiency factors:

1. Decision efficiency - Send a message? How big a message? What priority?
2. Mode efficiency - Use radio or not?

Technical efficiency factors:

3. Configuration efficiency - How many frequencies per message?
4. Queuing efficiency - trunked systems, blocking probabilities
5. Compression efficiency - voice and video compression algorithms
6. Modulation efficiency - bandwidth. How many bits/Hz?
7. Range efficiency - frequency reuse, cell size, power control

Spectrum management efficiency factors:

8. Assignment efficiency - separation distance, sharing rules
9. Allocation efficiency - Does allocation bandwidth match needs?

Economic efficiency factor:

10. Resource efficiency - Does total process give maximum service for minimum cost?

There may be some question whether all of these factors belong on the list. Although only the technology factors will be considered in this report, it is useful to note the existence of other factors as well. Even the most efficient radio system uses more radio spectrum to send a message than an administrative system that is organized to allow a message to travel by non-radio means. Which system should be considered the most spectrum-efficient? How does one balance the benefits of an administrative reorganization against the use of more efficient technology to send radio messages? We have developed few tools to evaluate the numerical efficiencies of the non-technical factors. Nevertheless, it would represent a sub-optimal answer to consider only solutions that improve the technical factors.

The ten spectrum efficiency factors listed above will be described in the remaining part of this chapter, especially as they apply to a mobile radio system. The five technical spectrum efficiency factors defined above will be used as a basis for comparing the relative spectrum efficiency of mobile radio systems in section 4. When numerical results are needed, the systems being evaluated will be compared with a reference system. For these studies, a single-channel analog FM simplex dispatch system with 15-kHz channel bandwidth is the selected reference system. In all cases, a number greater than 1 means that the target systems is more efficient (uses less spectrum) than the reference system.

2.3.1 Decision Efficiency

The factor leading to a decision to send a message or not to send one is not generally considered to be part of spectrum management, though it is at the heart of the use of communications. This factor also includes how important it is that the message gets through and the length of the message. Since we are living in what has been called an "information age," it seems likely that there will be a continuing need to send messages. A value greater than 1 means that more messages will have to be sent.

2.3.2 Mode Efficiency

This factor deals with the decision to use a radio-based medium to send the message that was made necessary in Factor 1. Most messages are not sent via radio, but are delivered verbally in person, by mail or memo, by telephone, etc. Until recently, radio was used mainly when speed of delivery was crucial and when wire-based communications (e.g., telephone) were not available. The current emphasis on "wireless" technologies represents a change in mode efficiency, including more frequent decisions to use a radio mode for convenience and quick response time. It is anticipated that many future messages will move via radio modes (including cellular voice and data, PCS, and wireless LAN) instead of the traditional wireline, postal, or personal modes.

Although the "wireless" changes would give a mode efficiency factor of less than 1, it should also be noted that many messages that were previously carried by point-to-point microwave systems are now carried by optical fiber. Thus, the mode factor for many of the microwave services is now typically greater than 1.

It should also be noted that systems that have good mode factors often have poor technical factors. The terse messages sent by air traffic controllers are partly necessary because of the very poor technical spectrum efficiency factors associated with the geometry of air traffic control links. On the other hand, the very high technical efficiency of cellular telephone is needed partly because of the poor decision efficiency and mode efficiency of typical cellular phone users. (In other words, the cellular phone system must be technically efficient, because its customers talk too much.)

Factor 1 and Factor 2 establish the basic requirements for the use of the radio spectrum. Once it has been determined that a message needs to be sent by radio, there are a large number of techniques that can be used to transmit the messages. The selection of these techniques determine how much spectrum is needed to transmit the message. The next five factors are called the technical efficiency factors. These five factors are the major area of consideration for this report.

2.3.3 Configuration Efficiency

This factor relates to the number of radio channels that are needed for a message, compared to a simplex system using one frequency switched between transmit and receive functions. Based on the convention that better spectrum efficiency will correspond to numbers larger than 1, a system requiring two frequencies per voice channel will have a configuration efficiency = 0.5. A system requiring four frequencies per voice channel has a configuration efficiency = 0.25, etc. A full-duplex system requires two frequencies for the full duration of a message (therefore, configuration efficiency = 0.5).

Many mobile radio systems use the half-duplex mode, where one frequency is used for a mobile transmitter and another frequency is used for the mobile receiver. A half-duplex station cannot listen while it is transmitting. Although a half-duplex system uses two frequencies, each frequency is typically used for only half of the duration of a message (configuration efficiency = 1).

A repeated, trunked, half-duplex system requires two frequencies for the duration of the message (configuration efficiency = 0.5). A half-duplex system requiring the use of multiple (N) serial repeaters to get coverage over an area would require 2N frequencies for the duration of a message (configuration efficiency = 1/2N).

2.3.4 Queuing Efficiency

This factor relates to the ability to secure a channel for use when it is needed, and to release the channel at other times. It relates specifically to the use of trunked systems, where the user experiences improved channel availability even when channels are in heavy demand. The queuing efficiency factor is the ratio of the number of users that can be serviced with a required level of channel availability, compared with the number of users that could obtain that level of channel availability with an equal number of channels on a reference system.

If "User X" believes that his need for immediate communications is so important that he can tolerate only a 0.1 percent chance of his message being blocked because the channel is occupied by another user, there are two approaches to meeting his needs. He can insist that he be given the sole use of a channel—or at least a channel that is shared by other users whose total use of the channel is less than 0.1 percent of the time. This approach severely restricts the number of users that can share a channel with User X. If User X is a heavy user of the channel (e.g., 20 percent of the time), it might be felt that this was sufficiently heavy use that no one would object dedicating the channel to the single user. A more frustrating situation occurs when User X makes very light use of the channel, but still requires exclusive use of the channel because of his need for immediate channel availability.

The second approach is to place User X on a multi-channel trunked system. A trunked system assigns a channel to a user on a call-by-call basis, allowing the user to communicate if any of the channels is available. Even if many of the channels are being used, it is relatively unlikely that they will all be in use. This allows more users to share a given number of channels if they are trunked, compared to the use of the same number of channels on an independent basis. Figure 1 shows the improvement in blocking probability (the chance that the user will be denied the use of a channel)

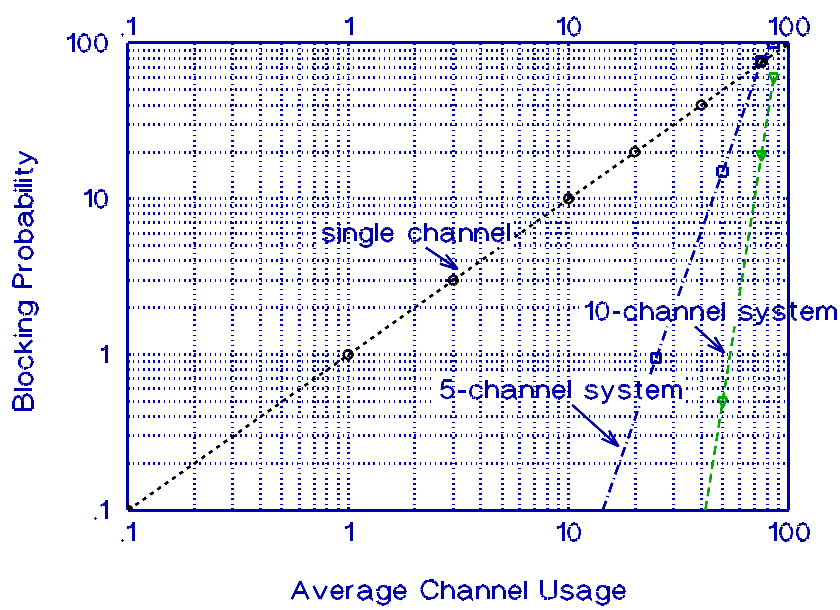

Figure 1. Percentage blocking statistics on trunked systems.

versus average channel usage for 5-channel and 10-channel trunked systems. In the example shown, the system requires a control channel to accept requests from users and to make temporary channel assignments. Therefore the 5-channel system carries traffic on 4 channels, and the 10-channel system carries traffic on 9 channels. If the trunked system is used as a repeater (which is usually the case), each channel will consist of a pair of frequencies.

As shown in Figure 1, a single channel has a blocking probability equal to the average channel usage (i.e., if the channel is being used by someone else, it will block my use of the channel). The blocking probability is less than average usage for 5- and 10-channel trunked systems. For example, if a blocking probability of 1 percent is acceptable, one could tolerate 1 percent average usage in a 1-channel system, 25 percent average usage in a 5-channel system, and 55 percent usage in a 10-channel system. This represents an improvement in the amount of usage that can be obtained from an average channel of 25 times and 55 times, respectively, for the 5-channel and 10-channel systems. The amount of improvement goes down when higher blocking probabilities are tolerated. For 10 percent blocking, improvements of 3.5 and 6 are provided in 5-channel and 10-channel systems, respectively. For very high usage (close to 100 percent), this type of trunked system actually produces poorer performance than a non-trunked system, because the requirement for a control channel occupies a channel that could otherwise carry traffic.

It is difficult to choose a single number to describe the amount of spectrum efficiency benefit one obtains from the use of trunked channels. If one assumed that users demand low blockage probabilities, (e.g., 1 percent), a 10-channel system would provide an improvement of 55 times. On the other hand, in a major catastrophe, a user might be willing to accept a higher blocking probability (possibly 20 percent?). Under these conditions, a 10-channel trunked system could provide service with 75 percent average usage. Thus, an improvement factor of slightly less than 4 would be obtained.

Many users are not as concerned with blocking probability as with "waiting time," the amount of time a user will have to wait until a channel becomes available. Although this is a more complicated problem, depending on user message-length statistics, the same type of improvement is provided by trunked systems. With N channels to chose from, there is N times the probability that at least one channel will become available in a given period of time.

Lacking detailed specifications on the allowable channel blocking probability for a "typical" user, the number of channels in a typical trunked system, as well as a methodology for combining such data if it were available, a numerical value of 5 was chosen to represent the improvement of trunked systems over individual channels, for the calculations in Section 4.

There is a major assumption in the above analysis, since it assumes that all channel usage is independent of other channel usage. This means that there is no statistical correlation between the times that User X wants to use a channel and the times that other users want to use a channel. However, if all of the users sharing the trunked system with User X needed to communicate for the same reason (at the same time) as User X, the statistical analysis of channel availability would not be valid. Only 4 of them will simultaneously be able to use a 5-channel trunked system—no matter what the statistics say. The use of a larger number of channels on a trunked system yields better statistical performance—partly because the raw statistics are better, but also because there is a better chance that the population of users will need to communicate at a more randomly mixed set of times.

The above limitation becomes pertinent to emergency communications when all of the users need to communicate during major catastrophes. Although trunked systems can provide greatly improved channel availability during times of moderate use, they may become overwhelmed by demand during times of large-scale demand—which is exactly the situation in which these systems are needed to perform their most critical functions. Because some situations trigger a simultaneous need for many users to communicate, many trunked systems incorporate the ability to give priority preference to certain sets of users. The priority users would generally be expected to change, depending on the nature of the emergency situation.

The problem of statistical dependence also suggests that (despite common practice) the worst strategy would be for each agency to build its own private trunked system to ensure reliable emergency communications. This practice is poor on two counts. First, being a single-agency system, it will have fewer channels and will give less statistical improvement on that basis. Second, since the members of a given agency tend to need communications for the same set of reasons, a single-agency strategy will tend to decrease the independence between users, ensuring that all of the users will want to use the system at the same time. Providing a mix of users from various agencies tends to increase the number of channels available in an emergency, as well as randomizing the times when users make heavy use of the system.

The advantage of statistical independence does not mean that there are no advantages in selecting the type of user allowed on a trunked system. There are substantial practical advantages to have users requiring a common set of communication system capabilities and who can agree on a common set of administrative priorities. On the other hand, a system sized to handle a large number of mixed high- and low-priority users provides excellent and economical service during normal times. During emergencies, a large number of channels are available for high-priority users, while the low-priority users are temporarily shut down.

2.3.5 Compression Efficiency

This factor applies especially to digital systems, where it refers to the amount of data compression that is performed between the derivation of original digital data and the amount of digital data that is transmitted. If redundant coding is required for efficient operation, it should be included in this factor. This factor is set to 1 for analog systems.

Compression is possible for voice messages because voice messages contain a tremendous amount of redundancy, when viewed as a simple information channel. As mentioned in Section 2.1.2, there is a ratio of more than 1200:1 between the number of bits needed to transmit the voice and what is needed to transmit the information contained in the voice message. Much of this difference represents redundancy in the speech process that can be removed by proper digital compression.

Proper digital compression is not that easy to perform, however, and much research is currently underway in the field of digital signal processing to obtain better algorithms and processing hardware. A general rule of thumb is that more-natural-sounding speech can be obtained by using more bits/second or more computer signal-processing "horsepower." Low data rate vocoders either sound poor or require more expensive signal processing and smarter algorithms. The current state of the art for inexpensive, high-quality vocoders is in the range of 16,000 - 32,000 bits/s. These systems typically require more bandwidth than efficient analog systems.

On the other hand, digital compression techniques should be expected to continually decrease the number of bits needed to transmit a voice message. For example, the new APCO-25 vocoder requires only 4,800 bits/s (plus another 4,800 bits/s for error correction, timing, and other overhead) and takes 12.5 kHz bandwidth. It should be assumed that future research will lead to the development of vocoders that require as little as 1000 bits/s (possibly 5-10 years away). Many digital systems plan to incorporate more efficient vocoders when they become available, giving an immediate increase in the number of voice channels available.

2.3.6 Modulation efficiency

This factor includes all of the factors that determine the amount of bandwidth required for an analog system or the number of bits/Hz for a digital system. For an analog system, this factor is simply the ratio of bandwidths. For a digital system, the efficiency factor is divided in two factors: one associated with data compression and one associated with the number of bits transmitted per Hz. The digital system is split into two factors since these two factors can often be separated and engineered independently.

Most modern digital mobile systems will use low-level modulation like 4-QAM (2 bits/sample). Although fixed microwave links often use 64-QAM (6 bits/sample) and 256-QAM (8 bits/sample), these higher levels of modulation need better signal-to-noise ratios and are less tolerant of multipath distortion. The fixed microwave systems can get by with the use of a high-performance adaptive equalizer, but the rapidly changing mobile radio signal path degrades the performance of some equalizers.

2.3.7 Frequency Reuse (Range) Efficiency.

This factor is related to the distance within which one system could cause interference to another system. A short range is preferable, since it allows a frequency to be reused more often by other systems. Range efficiency is inversely proportional to the square of the frequency reuse distance. This may seem counter-intuitive, since one often considers long range an asset—the longer, the better. From a standpoint of spectrum efficiency, however, long range is associated with the consumption of more spectrum space (it is part of the "S" factor in Equations 1 and 2). A longer range means that the frequency cannot be re-assigned as often, decreasing the amount of service obtainable at that frequency. This does not mean that long-range radio links are necessarily bad, but only that they consume more spectrum space than a short radio link. (In other words, don't use a long-range system if a short-range one will do as well.)

The actual length of a radio link is not what consumes the spectrum space. The spectrum space is consumed wherever the radio transmitter produces a signal strong enough to cause interference to another system. Thus, good spectrum management will use the minimum amount of transmitter power needed to perform the required communications function. Use of more power results mainly in the consumption of additional spectrum space, causing interference to other potential users without adding other benefits. Cellular systems dynamically adjust the power level transmitted by the mobile unit to the minimum required for reliable operation. This not only prolongs battery life in portable radios, but provides better statistical frequency reuse in adjacent cells.

Reducing the power of a transmitter and decreasing antenna height are only two of the possible techniques to improve frequency reuse. Some cellular systems use directional antennas to divide coverage from a site into 3 or 6 sectors. This allows the benefits of up to 6 cells from a single base station. Two orthogonally polarized antennas have been employed to double the use of each frequency at many fixed satellite and terrestrial microwave sites.

It should be noted that merely allowing additional assignments to be made within a shorter distance of an existing assignment does not necessarily give better range efficiency. If the shorter assignment range is less than the actual interference range of the system, decreasing the assignment range does not improve the system efficiency, but merely allows the assignment of more users at the cost of increased interference. A comparison of range efficiency must assume an equal degree of interference (preferably none) in each case.

It can be difficult to obtain realistic information on the typical frequency reuse range of many systems. Unlike many system parameters that are set by hardware, the range of many radio systems can be easily changed over wide ratios. In cellular systems, for example, down-sizing of cells is a common way to increase the traffic capacity of the system.

2.3.8 Assignment Efficiency

The category of assignment efficiency includes those factors in the assignment process that allow a maximum amount of use to be derived from a channel. An assignment process that is too conservative—possibly because of over-simplified EMC models, incomplete or incorrect frequency listings, or lack of detailed user information—will fail to allow the maximum number of users on a channel. This factor could also include bottlenecks that prevent or substantially delay the assignment of frequencies to intended users. We do not have sufficient information to know how to calculate this number in most cases, but it is a factor affecting the efficient use of frequencies.

One example of assignment efficiency is observed when a certain set of frequencies is reserved only for base station transmitters and another set (often paired with channels of the first set) is reserved for mobile transmitters. Since base stations have higher antennas, with higher gain, and more powerful transmitters compared to the corresponding mobile stations, the possibility of base station-to-base station interference would require very long distances—possibly 160 km (100 mi)—between base stations operating on the same frequency. If all base stations use a common set of transmitter frequencies, however, there is no interference between base stations, since no base station receives on the same frequency that another base station transmits. Therefore, one can calculate a separation distance based on the distance a base station can interfere with a mobile unit. This distance is usually a much smaller distance (possibly 40 km), since the mobile unit has a lower antenna, which is often blocked by terrain and buildings.

The mobile units experience a similar improvement of efficiency, since two foreign mobile units operating close together (but at their maximum operating range from their respective base stations) cannot interfere with each other, since neither of them will use a transmitter frequency that is received by the other.

2.3.9 Allocation Efficiency

This category includes those frequency management decisions that allocate frequency bands for particular functions and set the operating rules for those frequency bands. It includes allocating enough bandwidth to meet user demand and reallocating bands (or changing rules) when necessary to meet new requirements. It includes actions like re-channelization and narrowbanding when crowding becomes sufficient to recommend the action. It includes the selection of frequency bands that are most suitable for each service, based on propagation and technology considerations. In most cases, a good allocation efficiency would require that bands be allocated according to real functional needs and that all allocations represent a level of control and efficiency sufficient to meet the crowding in a given band.

Problems with allocation efficiency are caused, for example, when a standard set of frequencies is reserved for a particular type of mobile user (e.g. forest products, motion pictures, fisheries, etc) even though some of these users have widely-varying requirements in various geographical areas. Another example is the division of mobile radio bands into a set of Government and non-Government bands, irrespective of the amount of relative demand in various areas.

2.3.10 Resource Efficiency

This function is not—strictly speaking—a factor of spectrum management, but more of a figure-of-merit that shows how well national spectrum management is performing its job. In a market-driven spectrum economy, where spectrum shortages are reflected in higher user prices for spectrum, this factor would show how well the spectrum management process allows users to minimize total system cost (including the cost of the spectrum used by the system). This factor may be more difficult to measure and interpret if spectrum cost cannot be determined, but still deals with achieving the maximum total service for the minimum cost.

Spectrum efficiency remains a goal, to be sure, but only a sub-goal that is subservient to a goal of maximum service for least cost. For example, one could impose expensive and rigid technical standards in mobile bands (e.g., very tight spurious response limits on receivers) that would contribute slightly towards better spectrum management. If the tighter limits would not provide benefits worth the cost, however, such a measure should not be instituted. Each change that is suggested should have its cost measured against the likely benefits, and this difference should be an important part of the decision. The factors will be different for each frequency band, including the degree of need for greater efficiency (the amount of present and future crowding) and the operational and fiscal costs of making the changes. The acceptance of this as a general principle does not suggest that the numbers for costs and benefits are necessarily easy to obtain or not subject to extensive interpretation.

3. DESCRIPTIONS OF MOBILE SERVICES

The category of "mobile services" includes a number of distinctive applications, some of which are inherently able to use the radio spectrum more efficiently than other applications. For some readers, this chapter may be quite elementary and can be skipped. It is included here to provide a common understanding of the services whose spectrum efficiencies will be calculated in the following section. One must note that each of the services described meets distinct functional needs. One should not necessarily assume that these services should be considered interchangeable when applied to the communication needs of a particular user. The services described in this section include the following examples:

a. Single-channel, single-user systems
b. Repeaters
c. Trunked systems (including SMR)
d. Cellular telephone systems
e. Future PCS and other systems

3.1 Single Channel, Single User Systems

These systems are technically the simplest type of system and were historically the first type of system to gain widespread use. In this type of system, each organization wanting communications obtains a frequency (perhaps a pair of frequencies) and builds a system operating on that frequency. If communications are needed over a large area, e.g. a 50-km radius circle, the typical user would build a base station with a tall (15- to 50-m) transmitter tower and a high-gain omnidirectional antenna. The high antenna will provide better coverage than a lower antenna, since a larger geographical area will be line-of-sight or almost-line-of-sight from the base station antenna. Fewer obstructions will stand between the base station antenna and mobile units, and radio waves will travel with lower losses to a receiving antenna. To ensure reliable reception over that distance, the base station transmitter will use a relatively large amount of transmitter power (50-100 W). The vehicle-mounted mobile units will be constrained to use smaller antennas and lower transmitter power (10-50 watts); portable hand-held units (walkie-talkies) will be even more seriously constrained in antenna size and transmitter power (typically 1-3 W). Although the size of batteries in hand-held units has traditionally been the major limitation on portable transmitter power, possible health risks (or the fear of health risks) from higher-power transmitters held close to the user's head may also become a factor.

Because two-way communications are needed in most situations, the performance of the mobile transmitters is the weak link that limits the system's maximum useful range. In fringe locations, the mobile unit may be able to hear the more powerful base station transmitter, even though the base station cannot hear the weaker mobile signal. Therefore, some systems employ multiple receiving sites placed throughout base station transmitter service area to compensate for the shorter effective range of the mobile transmitters. At the base station, the receiver site with the best signal is selected for the mobile-to-base channel.

The useful range between mobile units will be even shorter. In some situations, though, the major role for communications will be between mobile units working close by each other on a common project while outside of the range of the base station. Here, it will be crucial to maintain mobile-to-mobile communications and less important whether they are out of range of the base station.

A single base station might transmit to 25 mobile units. In a simple system, all 25 users will hear the messages broadcast from the base station, even though the message was intended for only one of those users. To eliminate the distraction and inconvenience on the part of the 24 "unintended" listeners, some systems transmit a coded signal containing the ID of the intended listener. Although all 25 mobile units receive the message, the loudspeaker on the intended receiver is the only one turned on and the message is heard by only the intended listener. In some applications, of course, it is desirable for all 25 users to hear all transmissions, so that they all remain aware of the progress of the work that is underway. Or, possibly only a subgroup of 10 users needs to be made aware of the activity on a particular project. The use of IDs for individual listeners or groups of listeners can accomplish this. The ability to simultaneously call a group of listeners is termed "group calling" or "partly-line" (if all listeners are allowed to hear). "Tone-coded squelch" refers to the use of specific audio or sub-audio tones in a coded combination causing the unaddressed loudspeakers to remain turned off (squelched).

Simplex and duplex.[4] If the base stations and mobiles use a single frequency for all transmitting and receiving, the system is called a "simplex system." In a "half-duplex" system (also called a two-channel simplex system), a station uses a pair of frequencies, one for transmitting and one for receiving, however a station cannot simultaneously transmit and receive. Simplex and half-duplex systems both require "push-to-talk" operation, where the station is normally in a listening mode. The operator typically pushes a button on the microphone to turn on the transmitter. If two stations both transmit at the same time, neither one will be able to hear the other's message or be able to determine whether the other party is continuing to speak or trying to interrupt. Thus, certain conventions and radio disciplines have been developed to signal the end of a transmission, "over" being one phrase used to turn the radio channel over to the other transmitter.

Push-to-talk operations allow many users to listen on a channel (or pair of channels), with the only constraint being that only one user can transmit at a time. In simplex operation, mobile users will hear all of messages to and from the base station. In half-duplex operation, mobile users will hear all of the messages from the base station, but they will miss the messages from other mobile users.

In a "full-duplex" system, a pair of frequencies is also used, but each station simultaneously and continuously listens on one frequency and transmits on the other. This allows customary social conventions to determine who continues speaking if both parties attempt to speak at the same time. Although full-duplex systems are easier to use (no push-to-talk action required), they use radio frequencies more of the time and hardware costs may be 20-40 percent more than a comparable half-duplex radio. Less-experienced (or wealthier) users may prefer full-duplex systems because of the greater ease of operation. A more serious limitation on full-duplex radios is that only two users can simultaneous share a conversation. Since each user is continuously transmitting, a third user will necessarily continuously interfere with one of the other users.

In some half-duplex dispatch operations, the base station maintains full-duplex capabilities, but transmits only on a push-to-talk basis. This allows the base station dispatcher to hear incoming mobile messages even when the base station is transmitting.

3.2 Repeaters

Repeaters are used primarily to solve the problem of limited range. The location of the user's office might not permit tall antenna towers, or it might be in a valley that severely limits the radio coverage range.

---

[4]We will follow a set of definitions from Duff [2]. Unfortunately, there is not industry-wide agreements on these terms and alternative definitions have been used.

Talking between two mobile units (each with a low antenna and weak transmitter power) a also has very limited range. A repeater station on a mountainto allows the user (in effect) to put his base station on the mountaintop, while sharing cost with other users.

Communications from the user to a mobile unit proceeds in a two-stage process. The user transmits to the repeater station on frequency A. As soon as the repeater station begins receiving the signal on frequency A, it begins re-transmitting the signal on frequency B. The mobile units receive the signal on a frequency B. When the mobile units want to transmit a reply, they transmit a signal on frequency A, which is be received by the repeater station and re-transmitted on frequency B, where it is received at the user's office.

The repeater station handles signals from the office and mobile units in the same way. It does not know whether the signal comes from the office or from a mobile unit. The repeater station originates no signals; it simply receives any signal on frequency A and retransmits it on frequency B. Functionally, the office station is identical to any of the mobile units, and can be expected to operate with the same type of transmitter power and antennas as the mobile units—assuming that the office is located within the range of the repeater. The range of coverage between any two mobile units is limited only by the requirement that both units are within the repeater site area of coverage.

The repeater site is not selective about what signals it re-transmits. It repeats any signal it receives on frequency A (including signals not necessarily intended for that site). Thus, several independent users could uze the same repeater site (as long as they transmit to the site on frequency A and receive from it on frequency B) using the ID codes discussed earlier to keep the unwanted signals from distracting the other users. Many repeater systems have been built for Government use, for business use and for amateur operation. Since the repeater site owner may want to prevent unauthorized operation of the side, the repeater may look for a special "password" code, so that only authorized messages are repeated.

It should be noted that, although a repeater uses a pair of frequencies, it does not provide full-duplex operation to the user. Repeater operation requires a pair of frequencies for single-direction half-duplex operation. Every transmission is simultaneously transmitted on two frequencies: frequency A from the transmitting mobile unit to the repeater site and frequency B from the repeater site to the receiving mobile unit. A repeater system providing full-duplex service to the user would require four frequencies. From a frequency usage standpoint, repeater operation uses twice as many frequencies as conventional operation, and is—in that respect—inherently wasteful of frequencies. Moveover, the major objective of using a repeater is to be able to cover a maximum geographic area. This means that the frequency pair is not easily amenable to reuse within a nearby area, further aggravating the spectrum efficiency problem.

If each mobile unit can talk to another mobile unit only through the repeater, all communications between mobile units would cease if the repeater stopped operating properly or if the mobile units moved outside the range of the repeater site. Therefore, many of the mobile units offer the option of "talk-around," which allows mobile units to talk directly with each other, bypassing the repeater. With talk-around, the mobile units switch to simplex operation, transmitting and receiving on the same frequency. Of course, the advantages gained by repeater operation are lost, but mobile units can continue to communicate with each other, independent of the repeater site.

3.3 Special Mobile Radio (SMR) and Trunked Systems

The specialized mobile radio (SMR) service was created to provide repeater services to business and other organizations. Particularly for the small business having only a few mobile units, the cost of setting up a base station giving full geographic coverage could be avoided by paying a small monthly fee (typically $12-$15 per unit per month) to rent some services from a repeater operator. Even in this single-channel mode, repeaters offer a valuable service extending the coverage range of mobile/portable transmitters.

The usefulness of repeaters (and their spectrum efficiency, as well) is multiplied substantially when they are operated in a "trunked" mode. The trunked mode uses several frequency pairs integrated into a single system. For the following discussion, we will name these frequencies A_1 - A_5 and B_1 - B_5. When a user wants to transmit a message, the SMR system selects one of the channel pairs that is currently unused (e.g., A_4 and B_4) and assigns that pair to the user and the intended listener for the duration of the conversation. This means that a user has the immediate use of a channel pair whenever any one or more of the trunked channel pairs is available. The discussion of queuing efficiency in Section 2.3.4 describes the very substantial improvement in channel availability provided by trunked systems.

Non-Federal LMR bands that have been described as extremely crowded have been measured to have overall occupancy during the most crowded hour of the day of somewhere in the 15 percent range. Figure 1 in Section 2.3.4 shows that one could put 5 times as much usage (75 percent utilization) into a 10-channel trunked system before it would experience apparent usage in the 5 percent range. (Note, however, that the use of trunked operation implies the use of a repeater site, which doubles the number of channels in use. Therefore, the repeater band occupancy of 75 percent would correspond to a non-repeater occupancy of 37.5 percent. Thus, this scenario would correspond to an increase in user traffic of 2.5 times, instead of 5 times.)

The efficiency continues to increase with the number of channels trunked together. For a very large number of trunked channels, almost 100 percent usage in each channel would be obtainable before the user experiences crowding. Even greater benefits can be obtained by combining users who have peak usage at different hours of the day. Thus, the ability to serve more users at a lower apparent level of crowding more than compensates for the spectral inefficiency of using two radio frequencies for each signal. The user, of course, usually subscribes to an SMR system because of the benefits related to lower cost or better service—not because of the greater spectrum efficiency. The trunked SMR systems represent the great majority of SMR systems today, and they are expected to continue to make up an increased percentage of trunked systems.

Telephone interconnects are allowed by many SMR systems, giving SMRs a valuable auxiliary dual-mode capability. A special code identifies a telephone connection mode, and the SMR base station then switches to a full-duplex mode with the mobile station. In this case, the SMR base station still uses the same pair of frequencies (receiving on A_n and transmitting on B_n), but now the circuits are to and from the same mobile unit, instead of being used to receive from one mobile unit and transmit to another mobile unit. Full-duplex telephone operation, of course, implies that the mobile unit has to be capable of full-duplex operation, increasing the mobile hardware cost considerably. In half-duplex telephone interconnect mode, the mobile user has to use "push-to-talk" procedures. Although the SMR mobile unit operator might be comfortable with this operation, an inexperienced telephone user on the other end may feel awkward with this constraint. In contrast, the cellular network is a full-duplex network, permitting normal use of a telephone.

It is important to note that telephone users generate a different set of message-length statistics than a dispatch user. Although dispatch messages usually last for only a few seconds (5-20 s), telephone users often have conversations lasting several minutes. A relatively small number of telephone users will tend to occupy all of the available air time on a system. Since the FCC measures channel loading by the number of customers, rather than by the amount of air time used, the use of SMR for telephone interconnect may prevent channels from being used by enough customers. Without at least 70 customers per channels, an SMR system is not considered fully loaded and the SMR operator cannot ask the FCC for additional channels. Therefore, SMR telephone interconnect operation in urban areas is discouraged. In rural areas, there is usually no competition for additional channels and the loading criterion does not need to be met to request additional channels. Therefore, telephone interconnect is more widely used in rural areas.

Recent technical developments and changes in the regulations governing SMRs has encouraged interconnection with the telephone system. Especially in rural areas, where frequency crowding is less of an issue, telephone interconnect has become a major service for many SMRs. Although the SMR operators are not allowed to make a profit on the telephone services, they can charge for air time associated with the telephone services.

Most SMR services are billed on a flat monthly rate according to the number of mobile units. To ensure some degree of fairness to other customers, many SMRs put a several-minute time limit on telephone interconnect messages—disconnecting or billing on a per-minute basis after the time limit has run out.

SMRs currently can be licensed as public carriers (regulated under Part 22) or as a private user (regulated under Part 90). The services provided are essentially identical, though the regulatory environment is somewhat different. These regulations are being revised to reflect the status of many smaller private SMR systems which are providing services to the public at large.

Trunked systems are being strongly recommended to Government agencies, partly because of the spectrum efficiency obtained with trunked systems and partly because of the cost savings obtainable by combining the radio requirements of several Departments or agencies. A number of groups of 5-pair of channels are specifically designated for trunking systems in the 406-420 MHz Government LMR band. The Federal Government can also obtain service on commercial SMR systems, but cannot be the owner of an SMR system. Licenses for Government users are obtained through the NTIA, not the FCC.

3.4 Cellular Telephone

Cellular telephone provides easy (but somewhat expensive) mobile connections to the local switched telephone network. The service to mobile or personal telephones is provided by 30-kHz-wide channels using conventional analog FM transmission and duplex operation. Although the use of two 30-kHz channels for each message requires much spectrum, this system achieves additional spectrum efficiency by using trunked technology and by dividing a metropolitan service area into a number of smaller cells served by low-power transmitters. The total set of frequencies is divided into subsets (typically 7-12 subsets) with each cell being served by one of these subsets. Since the range of coverage is designed to remain deliberately limited, a given subset of frequencies can be reused in several cells within a metropolitan area.

As the demand for additional cellular service grows, the existing cells can be continually reduced in size through the use of lower antennas, directional antennas, and lower transmitter power. This permits more (smaller) cells, permitting still greater frequency reuse. In Toronto, for example, Bell Mobility Cellular has l-km-radius cells with 60-degree sectorization [3]. This is claimed to give 150-200 times greater spectrum efficiency than the earlier 8-km-radius omnidirectional cells. The evolution of smaller microcells has the dual advantage of providing greater frequency reuse and lower transmitter power in the mobile unit. The low transmitter power has made small personal cellular phones quite practical, and the trend toward smaller cells will continue, especially in major metropolitan areas.

Cellular systems continue to expand rapidly. At the beginning of 1992, there were 8 million users, with another 2 million added in 1992. In June 1992, cellular systems had been established in all of the 734 U.S. market areas. In the year preceding June 1992, the number of cellular sites increased from 6,685 to 8,901. Some of these new sites represent service to new geographic areas (especially rural areas); others represent downsizing to allow for more intense frequency reuse in crowded areas. All of the new sites represent increased cellular capacity. Digital data services over the cellular network were added in 1992, partly to provide a wireless communications path to portable laptop and palmtop computers. New nation-wide calling services that use a single telephone number to reach a given cellular telephone anywhere in the U.S. are becoming available and are beginning to provide some of the advanced features that are envisioned as part of a future Personal Communications Service (PCS).

Severe crowding in the major metropolitan areas has pushed the development of several new technologies to increase the traffic capacity of cellular systems, including 1) a digital system TDMA system (IS-54), 2) a narrowband analog system (narrowband advanced mobile phone system, NAMPS), 3) a digital CDMA system (IS-95), and 4) continued downsizing of cells. The TDMA system and NAMPS will each produce 3 times as many channels with the present spectrum. The TDMA system uses 9.6 kbits/s vocoder and has the advantage of allowing complete privacy through digital encryption, as well as the extension to advanced digital data services. Future versions of the TDMA system will use 4.8 kbits/s vocoders, providing 6 channels/30 kHz. NAMPS is mainly a narrowing of the analog FM channel. Although the tripling of capacity can be obtained fairly easily, no further extension of capabilities is likely. All of these improvements to AMPS have the major disadvantage of being incompatible with AMPS and each other.

The ability of cellular systems to incorporate very small cells allows very high spectrum efficiency in dense traffic areas. To efficiently utilize these small cells, however, the mobile cellular phone must adjust its transmitter power to match the cell size that it is currently using. In a large rural cell, the mobile unit may need a full 3 watts to reach the relatively distant cellular base station site; far less power will be needed to reach the base station in a microcell. The transmitter power of the mobile transmitter is adjusted, based on commands received from the base station.

3.5 Comparison Between Cellular Telephone and SMR

Cellular telephone service is fundamentally different from SMR/trunked, and it reaches different customer needs, as these services are presently constituted. However, this difference is based more on the selection of operational parameters designed to service different types of customers, rather than being an inherent and inflexible division. SMRs primarily provide mobile-to-mobile simplex repeater service, between a few specifically-designated private customers, using maximum-sized coverage areas. SMRs allow convenient dispatch operation, where it is only necessary for a user to pick up a microphone and call your intended listener (who can hear the message without any action on his part). SMRs allow selective calling, group-calling, and partly-line operation. These are important features for many operations, where several mobile and/or fixed units need to be continually kept up-to-date on the progress of the many independent parts of an operation. Law enforcement operation, fighting a forest fire, coordinating the parts of a large construction operation all can require this function.

On the other hand, cellular telephone systems are intended to add a mobile full-duplex connection to the telephone network. A mobile-to-mobile connection with cellular telephones is clumsy and expensive. Cellular phones are not configured (nor permitted) to enable group call or party-line operations. A mobile-to-mobile connection can be made only through the base station and requires 2 independent mobile channel pairs.

Cellular systems are designed for small cells, which allows frequencies to be reused several times in a metropolitan area. The shorter frequency reuse distance gives current cellular systems a larger traffic capacity than an SMR system (which has a larger coverage area and a larger reuse distance). Cellular systems pay for that extra capacity, however, with the added expense of constructing many cell sites, as well as the extra complexity needed to route signals to the appropriate cell and to allow a mobile unit a smooth transfer from one cell to another. The spectrum efficiency that cellular gets from frequency reuse is dependent on a given conversation using only one cell at a time. However, mobile-to-mobile, group-calling and party-line features would typically require the simultaneous use of frequencies in multiple cells, increasing the cost of these features in a cellular environment. Full-duplex and group call are fundamentally incompatible capabilities. If two users in a group were located in a single cell, they would both be simultaneously calling back to the base station repeater on the same frequency. These two simultaneous mobile-to-base signals would interfere with each other. Therefore, SMR systems provide group call services only in a half-duplex mode.

Nextel and other SMR providers are building an enhanced SMR (ESMR) service, combining some of the features of SMR and cellular telephone. The cellular-like features will include full-duplex telephone interconnect, as well as small cells using directional antennas for cell segmentation. The SMR-like features include dispatch mode operation with a full range of group-call features. Advanced technical features include a TDMA/digital compression architecture that will provide 6 voice channels in 25 kHz of bandwidth, and a full set of digital message and paging services. Nextel will first provide this service with a 125-site system in Los Angeles in the first quarter of 1994, with other systems in San Francisco, Chicago, and New York opening later in the year. Other companies are building similar systems in other cities.

Though we have no pricing information at present, we have been told that Group Call features will be priced according to how many cells must be used to reach all members of the group.

3.6 Future Personal Communications Services (PCS).

The next major evolution of the telecommunications industry will probably be a set of services known as PCS. Although the exact nature of this service remain subject to change, it is likely that it will include many of the following features. It will be a short-range digital wireless service that will provide 2-way voice and data to a variety of portable terminals or personal communicators. It will probably operate in a number of modes, including an in-building private branch exchange (PBX) (to provide cordless telephone service within a building) and wireless local area network (to provide wideband data services to portable computers within a building). Outside a building, it would operate like a short-range cellular telephone, accessing signals radiated by the neighborhood PCS port located on a lamppost at each city corner. Outside the city, it would access a cellular medium-range service or possibly a trunked LMR system. In remote areas, it might access a mobile satellite. The personal terminal for PCS would operate in all of the above modes, seeking out the best (and cheapest) communication pathway in every location.

An intelligent switched network tied together by optical fiber and microwave would feed all of the wireless ports and would keep track of the location of all users, so that the nearest PCS radio port could be selected for communications with a particular user. Billing, call priorities, mode of communications (audio, data, and/or video) would be sorted out and tracked by the intelligent network. The large number of unknown factors make specific predictions about the use of PCS systems difficult. However, such a system might provide many of the services currently provided by LMR and cellular systems. The Federal Government has stated that it expects to make major use of PCS and it has requested that it be able to purchase and operate PCS systems, as well as purchasing commercial services.

Depending on the nature of each personal terminal/communicator and the services selected, a range of services would be available. A minimum level of services would include at least narrowband data, speech, and paging services. Other alternatives might provide wider bandwidth services, including one- or two-way video, updated geographical location, health or security alarms, etc. Many of these services are less than 10 years away. It is anticipated that these services will be priced for the mass market, and they would surely be a promising alternative to the piecemeal way that many mobile communication services are provided today.

4. RELATIVE TECHNICAL SPECTRUM EFFICIENCIES OF MOBILE SERVICES

A variety of mobile services are used in the United States today, and additional ones will soon be available. We have selected a representative variety of these systems to provide an approximate estimate of the relative technical spectrum efficiency of a number of mobile systems. It was convenient to ignore some of the finer points of comparison, chiefly because it would complicate the calculations beyond the intent of this paper. Although this is a potentially inaccurate approach, it allows systems to be compared even when detailed information is not available. Because of the possible inaccuracies, the reader should pay more attention to the larger differences in computed efficiency.

It is important to recognize that the systems that are compared here do not necessarily perform the same functions, nor do they operate within the same constraints, nor have cost or other system characteristics been considered. Therefore, a low efficiency number does not necessarily represent a poor choice of system or even a poor choice of engineering parameters. Instead, low efficiency factors may reflect an operating environment or a service that is inherently difficult. Neither has there been any effort to compare the suitability of a particular system for a particular function. Therefore, the comparisons made within this chapter should not be used to infer that any particular system is necessarily a better choice for any particular function. The focus of this chapter is spectrum efficiency, as defined in earlier chapters.

4.1 Comparison Methodology

In comparing the efficiency of mobile services, we will calculate an overall relative spectrum efficiency by comparing the individual spectrum efficiency of several factors listed in Section 2.3 as "technical spectrum efficiency factors." The omission of the other five factors should not be construed to suggest that they are not important, but rather that they are less directly applicable to the engineering/technical parameters of a communication system. The five technical spectrum efficiency factors that are being considered in this section are repeated here:

Configuration efficiency - how many channels/message?
Queuing efficiency - trunking, waiting times, blockages
Compression efficiency - voice and video compression algorithms
Modulation efficiency - binary FSK, 16-QAM, etc.
Range efficiency - Frequency reuse power control

We will compare all of the selected systems to a commercial dispatch system (whose characteristics will be described below) on a factor-by-factor basis. The overall relative efficiency of the system is the product of the individual factors. Each of these five factors has been described in more detail in Section 2.3. Here, we will repeat a short description and list some of the assumptions employed for these calculations. For each factor, a numerical rating greater than "1" indicates a higher spectrum efficiency than the reference system.

Configuration efficiency. This factor describes how many radio frequencies are required to send a message. The commercial dispatch reference system uses a pair of frequencies in a half-duplex mode. It could also use a single frequency for send and receive (simplex mode). These are considered approximately equivalent, since the half-duplex configurations uses two channels—each of them for about half the message time. The simplex system uses a single radio channel for the duration of a message. A full-duplex channel or a half-duplex repeater uses two channels for the duration of message, giving a relative efficiency of 0.5 for a message. A full-duplex channel with a repeater uses four channels during a message, yielding a relative efficiency of 0.25.

Queuing efficiency. The number of users that can obtain a specified level of service on a trunked system depends on many factors, including the number of trunked channels, the amount of time the average user needs the channel, the maximum amount of blockage that can be tolerated, etc. For simplicity, we will assume that any trunked system serves five times more users than a non-trunked system. We realize that the efficiency of a 10-channel system is greater than a 5-channel system (especially if one channel is dedicated for a control channel), but a given trunked system technology will have a wide mixture of systems with different numbers of trunked channels and user requirements. It does not seem possible to calculate a more precise number to represent a large collection of systems. Note that many trunked systems operate in a "repeater" mode, so that some of the trunked efficiency is cancelled by the 0.5 configuration efficiency of a repeater operation.

Compression efficiency. This factor is used to show extra channel capacity gained by using digital compression techniques. It is set equal to 1 for all analog systems. In a practical sense, it is closely tied to modulation and bandwidth requirements, and there may be a somewhat arbitrary distribution of system efficiency between the two factors.

Modulation efficiency. This factor considers the efficiency of modulation and/or bandwidth compression employed by a system. It is closely tied to compression efficiency factors in a digital system. For most of the systems reviewed here, this factor is the ratio of system channelization bandwidths.

Range efficiency. This factor describes the amount of frequency reuse possible within a system. It includes the frequency reuse made possible because of reduced transmitter power, sectorized antenna coverage, reduced antenna height, terrain and building shielding, etc. When a frequency reuse distance is given as an average distance, the efficiency factor increases as the inverse square of the distance.

Pinning down range efficiency is somewhat elusive for some systems, since it is not a fixed engineering quantity, but rather a condition of implementation. Depending on how crowded a situation is, systems may be installed close together or far apart. The downsizing of cellular telephone cells in crowded cities is a good example of this; the answer keeps changing. Therefore, we have tried to estimate the normal "state-of-the-art" situation, or we have included a range of values.

Overall score. This factor is the numerical product of the individual factors for configuration, queuing, etc. It represents the ratio of communications capacity achievable with the specified system compared to the reference system. A ratio of 150, for example, would imply that the specified system would be able to provide service to 150 times as many users as the reference system with no increase in the amount of spectrum used. Under more ideal circumstances, it would also imply that the specified system provided equal services to those 150 times as many users, but we have not constrained our comparison to systems that provide equivalent services.

4.2 Relative Spectrum Efficiencies of Various Mobile Systems

This section provides calculations of relative technical spectrum efficiencies for selected mobile systems. Some of these mobile systems provide much different services from others, so it is not proper to judge the relative merit of these systems by a simple comparison of overall spectrum efficiency.

Commercial dispatch. This system is the reference system against which all other systems will be compared. The pertinent characteristics are shown in Table 1, including a frequency reuse range of 80 km (50 mi). The selection of an 80 km reuse range is somewhat arbitrary, since smaller ranges are used in crowded vicinities and larger ranges are used in less-crowded areas. The same 80 km range was used wherever the same basic equipment was employed. A paired-channel, half-duplex system is used as the basis for comparison, though a single-channel simplex system is numerically equivalent.

Commercial LMR bands use a variety of channelizations, including 20 kHz (in the 30-50 MHz band), 15/30 kHz (in the 150-162 MHz band), and 25 kHz (in the 450-470 MHz band). The same F3E modulation with a 20 kHz bandwidth is used in all of these bands, as well as in the Federal LMR bands (162-174 MHz and 406-420 MHz). Therefore, the equipment is identical in all of the Federal and commercial bands. In the commercial bands, various techniques have been employed to squeeze more channels in. The use of 20 kHz wide channels with 15 kHz spacing, was judged to be the most efficient of the techniques, though it requires a minimum 16 km separation between base stations on adjacent frequencies. Strictly speaking, this technique should be counted as an assignment efficiency factor (factor #8 in section 2.3). Since we had included no place for assignment efficiency factors in our tables, we included the 15 kHz channelization in the factor for modulation efficiency.

Table 1. Commercial Dispatch (15 kHz BW)

Factor name	Comments	score
Configuration	paired-channel, half-duplex system	1
Queuing	paired-channel shared with other users	1
Compression	analog FM	1
Modulation	15 kHz channel bandwidth (assignment spacing)	1
Range	power 70 W, ant. gain 6 dB, ant. height 30 m, frequency reuse range 80 km	1
Overall score		1

Federal Government LMR dispatch (25 kHz BW/12.5 kHz BW). This type of system (Table 2) is widespread throughout the Federal Government LMR service, often with an LMR site serving only one agency with several paired-channel, half-duplex systems. The 162 MHz band has been recently re-channelized to use 12.5 kHz channels; the 138 MHz and 406 MHz bands will also be re-channelized soon to 12.5 kHz channels. Specifications remain the same, except for the re-channelization bandwidth.

Table 2. Federal Government Dispatch (25 kHz BW/12.5 kHz BW)

Factor name	Comments	score
Configuration	paired-channel, half-duplex system	1
Queuing	paired-channel shared with other users	1
Compression	analog FM	1
Modulation	25-kHz channel bandwidth/ 12.5 kHz channel bandwidth	0.6/1.2
Range	power 70 W, ant. Gain 6 dBi, ant. height 30 m, frequency reuse range 80 km	1
Overall score	analog FM	0.6/1.2

Cellular telephone. The use of cellular telephone, advanced mobile phone system (AMPS), has grown rapidly in the Government and the commercial markets. Cellular telephone is a full-duplex system, using low base station transmitted power, as well as carefully sited (and sometimes directional) antennas to control the base station coverage area. Cellular mobile systems adjust the reverse path (mobile-to-base) transmitter power to the lowest level needed for effective communications. In rural areas, the cells are spaced up to 30 km apart, with frequency reuse permitted at 55-km increments. In some very dense urban areas, microcells have been developed with coverage areas less than 2 km across and frequency reuse distances of 8 km. Typical urban cell sizes are 8 km across, with a frequency reuse distance of 25 km. The current analog cellular telephone uses 30-kHz analog FM channels in a full-duplex, trunked system, using 15 channel pairs at each site.

Improvements to cellular telephone systems include a digital TDMA-30 system that uses 9.6 kb/s vocoders, giving 3 digital channels per 30 kHz for the first stage of implementation and 6 channels in future versions [3]. An improved narrowband AMPS (NAMPS) is based on narrowing each voice channel to 10 kHz. TDMA-30 and NAMPS will provide an additional factor of 3, in addition to gains from downsizing cells. AMPS spectrum efficiency factors are summarized in Table 3.

Table 3. Cellular Telephone (AMPS)

Factor name	Comments	score
Configuration	full-duplex system	0.5
Queuing	trunked, 15 channel pairs	5
Compression	analog	1
Modulation	30-kHz FM bandwidth	0.5
Range	frequency reuse range: 8 km minimum power 3 W (mobile), 1 W (portable) 25 km typical 55 km maximum	100 10 2
Overall score	minimum distance typical distance maximum distance	125 12 2.5

Trunked systems (SMR and Government). Trunked systems (Table 4) provide substantial user convenience at a moderate level of improved spectrum efficiency. The two systems described here include the commercial special mobile radio (SMR) systems available in the 800 MHz and 900 MHz frequency range for private operations and the Federal trunked radio systems in use in the 406 MHz band. These systems are approximately equivalent except for the frequency band, and they can be operated in a half-duplex repeater mode or a full-duplex connection to the telephone system. In most urban systems, telephone interconnection is discouraged because users tend to spend too much time in that mode, denying the system to other users. The use of trunked systems is expected to grow rapidly, a result of the FCC Refarming studies and similar NTIA studies of Government systems.

Table 4. Trunked Systems (SMR and Government)

Factor name	Comments	score
Configuration	repeater half-duplex system, or telephone interconnect full-duplex	0.5
Queuing	5-channel or 10-channel trunked system	5
Compression	analog FM	1
Modulation	25 kHz channel bandwidth	0.6
Range	power 70 W, ant. height 30 m, frequency reuse range 55 km	2
Overall score		3

Enhanced SMR systems (Nextel ESMR). The Nextel ESMR system (Table 5) provides substantially improved spectrum efficiency, using cellular-sized cells and digital voice-compression. This system is starting commercial service with a 125-site system in Los Angeles in early 1994. Additional systems will be coming on-line in several other large cities before the end of the year. The Nextel system has features that bridge the gap between cellular telephone and conventional SMR, including convenient mobile-to-mobile calling (via the base station), multi-cell selective group call, convenient telephone interconnect, and a variety of enhanced digital data/paging functions.

Table 5. Enhanced SMR (Nextel)

Factor name	Comments	score
Configuration	repeater half-duplex or telephone interconnect full-duplex system	0.5
Queuing	6-to 24-channel trunked system	5
Compression	digital compression - 6 channels, 4.8 kb/s V-SELP vocoder.	6
Modulation	25 kHz channel bandwidth (per 6-channel TDMA block)	0.6
Range	frequency reuse range: 8 km min power 3 W (mobile), 1 W (portable) 25 km ave 55 km max	100 10 2
Overall score	minimum distance average distance maximum distance	900 90 18

Amplitude-companded single-sideband (ACSB) systems. ACSB systems use a single-sideband, AM signal transmission technique to achieve narrow bandwidth, and a pilot tone with non-linear compression to provide a solid frequency lock and improved dynamic range and noise reduction. Modern versions of ACSB place the pilot tone in the middle of the bandpass (where it is relatively free from adjacent band interference) and use digital signal-processing techniques to notch out the pilot from the audio channel. Compressing the amplitude-modulated signal before it is transmitted produces more uniform transmitter output power, and de-compressing of the signal in the receiver provides better dynamic range and lower background noise. The pilot tone is used for establishing a reference amplitude for de-compression, as well as an accurate reference frequency for the single-sideband detector. This combination provides a usable signal in a 5-kHz-wide channel. The use of a digital signal processing chip for most detection/decompression functions helps to decrease hardware misalignment, providing improved operational characteristics.

The 220-222 MHz band was recently re-allocated with this type of system as the intended occupant. Other systems are also permitted, subject to fitting within the 5-kHz channelization. ACSB systems do not have scrambling, though the peculiar modulation will discourage eavesdropping by conventional scanners. Table 6 summarizes the spectrum efficiency factors of ACSB systems.

Table 6. Amplitude Companded Single-Sideband Systems)

Factor name	Comments	score
Configuration	Single-channel, half-duplex system Repeater/trunked system	1 0.5
Queuing	Single channel shared with other users 5-channel trunked systems	1 5
Compression	analog, ACSB	1
Modulation	5 kHz channel bandwidth	3
Range	power 25 W, ant. gain 8 dB, ant. height 30 m, frequency reuse range 80 km	1
Overall score	single-channel system repeater/trunked system	3 7.5

Proposed APCO-25 law enforcement standard. The systems built using the APCO-25 standard will be similar in range to the existing mobile radio systems. They are designed to meet a common need for a large number of law enforcement agencies. They will use digitally compressed voice (4,800 b/s for the vocoder, plus another 4,800 b/s for system overhead and error correction). The development of this system meets several needs, including

1. Full encryption of voice or digital messages (allowing secure transfer of information).
2. Compatibility with the new Government band 12.5-kHz channelization plans.
3. Interoperability between agencies for joint operations and coordinated emergency response.
4. Priority access during system crowding.

The details of this standard have not yet been determined, but they are expected to be available soon. A large number of these radios are expected to be purchased by Federal, state, and local governments for use in a number of frequency bands. This standard will be available in single and trunked systems; a trunked system is evaluated here, based on LMR RF technology. Table 7 summarizes the spectrum efficiency factors of the APCO-25 system.

Table 7. Trunked APCO-25 Law Enforcement standard

Factor name	Comments	score
Configuration	half-duplex repeater, full-duplex telephone interconnect	0.5
Queuing	trunked system	5
Compression	4,800 b/s vocoder (plus 4,800 b/s overhead, error cor)	1
Modulation	12.5-kHz channel bandwidth, QPSK	1.2
Range	frequency reuse range 80 km	1
Overall score		3

Future PCS systems (for example, Bellcore Wireless Access system). (The inclusion of this particular TDMA system as an example does not imply any endorsement or particular suitability of the system; some CDMA PCS candidate systems claim even higher spectrum efficiencies.) It should be noted that many details of this system are not yet fixed, but the stated intention of such PCS systems is to provide essentially ubiquitous coverage over most of the United States for voice and data, at a cost considerably below that of cellular telephone. This system will use encrypted, digitally compressed voice with 9 or 19 voice channels (depending on whether 16-kb/s or 32-kb/s vocoders are used) being developed from a pair of 400-kHz-bandwidth channels [4]. Other types of digital message service can be provided to digital paging systems, laptop computers, etc., using data rates between 4,000 and 320,000 b/s. A typical deployment of a PCS system would use base station antennas hung from 10-m telephone poles spaced at 600 m intervals. A frequency reuse plan breaks the available channels into about 25 frequency groups, giving a 3 km frequency reuse distance. Power control will allow future downsizing of cells when needed to handle more traffic. Table 8 summarizes spectrum efficiency factors of a TDMA PCS system.

Table 8. Personal Communication Services (Bellcore proposed TDMA standard)

Factor name	Comments	score
Configuration	telephone system interconnect, full-duplex, TDMA	0.5
Queuing	9- or 19-channel trunked system	5
Compression	Digital compression - 9 channels (32 kb/s vocoder) 19 channels (16 kb/s vocoder)	9 19
Modulation	400 kHz channel bandwidth	0.0375
Range	power 0.8 W, ant. gain 8 dB, 10-m antenna height 600-m-diameter cells, typical frequency reuse 3 km 300-m-diameter cells, typical frequency reuse 1.5 km	 710 2,850
Overall score	3 km reuse, 32 kb/s vocoder - initial deployment 1.5 km reuse, 16 kb/s vocoder - follow-on	600 5,075

VHF air traffic control (ATC) channel. The VHF ATC system uses 25-kHz channels with AM voice. The aircraft carrying these systems can fly at altitudes above 50,000 ft. At this altitude, the horizon is more than 500 km away from the aircraft, and the FAA requires a 2080-km separation distance for frequency reuse [5]. Lower altitudes permit shorter reuse distances. We have selected 1280 km as a compromise between separation distances required at lower and higher altitudes. The use of a simplex channel configuration allows all users to hear both ground-air and air-ground conversations, but increases the reuse distance. This service is a "safety-of-life" service, requiring excellent reliability, with technical characteristics set by international agreement. The very low relative spectrum efficiency is caused mainly by the very long frequency reuse distances. Fortunately, the users of ATC systems are highly trained to limit the length and content of messages, decreasing the amount of traffic that the system is required to carry.

Table 9. VHF Air Traffic Control System

Factor name	Comments	score
Configuration	simplex AM	1
Queuing	well-disciplined user community dispatch system	1
Compression	AM analog	1
Modulation	25 kHz channel bandwidth	0.6
Range	up to 50,000 ft altitude, 100 W, 1280-km reuse distance	0.008
Overall score		0.005

Aeronautical public telephone service. This system provides air-to-ground full duplex voice communications for airline passengers, typically via phones located in seatbacks throughout the aircraft. Although they operate from the same aircraft that use ATC systems, they are not a "safety-of-life" service, and they operate with less-conservative frequency reuse distances. The use of trunked channels allows efficient use of a channel, without depending on tight user-community discipline. In addition, this system uses much more efficient amplitude companded single-sideband (ACSB) modulation using only 6 kHz bandwidth. Newer versions of this service will be implemented with compressed digital voice modulation, permitting digital services (e.g. fax and computer messages). This service incorporates recent technological innovations, and it is not subject to control by international standards. Table 10 summarizes the spectrum efficiency factors for ATC systems.

Table 10. Aeronautical Public Telephone Service

Factor name	Comments	score
Configuration	full-duplex, ACSB	0.5
Queuing	29-channel trunked system	5
Compression	analog	1
Modulation	ACSB modulation with 6-kHz channel bandwidth	2.5
Range	50,000 ft altitude, power 30 W, 640-km reuse distance	0.03
Overall score		0.19

Other advanced systems, including CDMA PCS and LEO MSS. There are several other systems whose spectrum efficiency would be particularly interesting to analyze in this fashion. They include several spread spectrum (frequency-hopping and CDMA) systems, including terrestrial PCS, cellular, and low earth orbit mobile service satellite (LEO MSS) systems. Unfortunately, for this greatly simplified process to work, it is necessary for the systems to be sufficiently similar to the reference system (analog FM dispatch) that the systems can be broken apart and the factors compared on a one-to-one basis. These systems are sufficiently different in their mode of operation that it was impossible to discover credible ways to compare the factors. Therefore, such systems will have to be compared on a more fundamental basis, which is outside the scope of this survey paper.

4.3 Summary of Mobile System Efficiencies

This section contains a summary of the relative efficiencies of the previously described systems (Table 11). It should be noted that relative efficiency refers to a definition of spectrum efficiency compared to a conventional 15 kHz mobile radio system. No claim is made that these systems are functionally equivalent; a higher relative efficiency factor does not necessarily suggest that a given system is a better choice for a particular function.

Table 11. Relative Spectrum Efficiencies of Mobile Systems

Type of System	Table	Conf.	Que.	comp	Modul.	Range	Total
Commercial dispatch: 15-kHz bandwidth	1	1	1	1	1.0	1	1
Government dispatch: 25-kHz bandwidth	2	1	1	1	0.6	1	0.6
Government dispatch: 12.5-kHz	2	1	1	1	1.2	1	1.2
Cellular telephone (AMPS): 55-km reuse	3	0.5	5	1	0.5	2	2.5
Cellular telephone (AMPS): 25-km reuse	3	0.5	5	1	0.5	10	12
Cellular telephone (AMPS): 8-km reuse	3	0.5	5	1	0.5	100	125
Trunked systems	4	0.5	5	1	0.6	2	3
Enhanced SMR (Nextel): 55-km reuse	5	0.5	5	6	0.6	2	18
Enhanced SMR (Nextel): 25-km reuse	5	0.5	5	6	0.6	10	90
Enhanced SMR (Nextel): 8-km reuse	5	0.5	5	6	0.6	100	900
Amplitude-companded sideband (ACSB)	6	1	1	1	3.0	3	3
Trunked ACSB	6	0.5	5	1	3.0	7.5	7.5
Trunked APCO-25 law enforcement	7	0.5	5	1	1.2	1	6
PCS (TDMA, 3-km reuse, 32 kb/s)	8	0.5	5	9	0.0375	710	600
PCS (TDMA, 1.5-km reuse, 16 kb/s)	8	0.5	5	19	0.0375	2850	5075
VHF air traffic control channels	9	1	1	1	0.6	0.008	0.005
Aeronautical public telephone	10	0.5	5	1	2.5	0.03	0.19

5. SUMMARY

Table 11 shows a remarkably wide range of relative spectrum efficiencies, varying over a value of 1.0 for current typical commercial mobile radio systems, to a value of 5075 for a proposed PCS system, to a value of 0.005 for a typical VHF ATC channel. This 1 million-to-one ratio suggests that spectrum efficiency is a topic that should be considered seriously when systems are selected for deployment. (Apparently, "a radio is not a radio is not a radio.") Although the most inefficient of the systems studied (ATC) owes much of its inefficiency to the airborne environment in which it is used, it must be noted that recent commercial systems (e.g., Airfone) provide full-duplex service in the same environment with 40 times greater efficiency.

This chapter examines the various systems to see what technologies might provide the greatest improvement in spectrum efficiency. This can be approached partly by regrouping the data in Table 11 to show the best (most efficient) and worst (least efficient) values for the respective efficiency factors (Table 12). Because of the manner in which values were assigned to "compression" and "modulation" in the original tables, TDMA systems appeared to have very high compression efficiency and very low modulation efficiency.

Since the actual factor of interest is the bandwidth per voice channel, an additional row was added to Table 12. This row is called "bandwidth/channel" and was calculated as the product of modulation factor times the compression factor for each system times the 15 kHz bandwidth of the reference system. The resulting worst and best systems were the 9-channel PCS (requiring 44 kHz per voice channel) and the Nextel system (requiring only 4.2 kHz per voice channel).

Table 12. Minimum and Maximum Efficiency Factors

Factor	Worst	Best	Ratio
Configuration	.5	1	2
Queuing	1	5	5
Compression	1	19	19
Modulation	.0375	2	67
Bandwidth/channel	44	4.2	10.5
Range	.008	2850	350,000
Total Efficiency	.005	5075	1,010,000

The "Ratio" column shows the ratio between the worst (least efficient) and best (most efficient) values for a particular efficiency factor. The ratio can be taken as an indicator of the amount of effect that each factor can have on overall communication system efficiency. We will discuss the implications of these numbers in the remainder of this section.

Configuration efficiency. In the case of configuration efficiency, a ratio of 2 shows that configuration is not a particularly important factor. Presumably, it would not be worth spending a lot of effort trying to change the configuration efficiency number. Of course, the configuration efficiency of a particular system could have a major effect on the functional uses of a system, which might be much more important than the relatively small change in spectrum efficiency.

Queuing efficiency. The ratio of queuing efficiency is 5. This ratio was determined when we assigned an efficiency factor of 5 for all trunked systems and a factor of 1 for non-trunked systems. We made a reasonable case for that number, and noted that it could change considerably depending on the particular system configuration and user requirements.

The use of trunked systems makes only moderate improvements in spectrum efficiency, especially when the trunked system is operated in a "repeater" mode (which has a configuration efficiency of 0.5) and when the user can tolerate a moderate blocking probability. The net improvement for a trunked repeater system is only 2.5. If the Federal Government chooses to operate their trunked repeater systems with a 25-kHz bandwidth, while the rest of their systems use a 12.5-kHz bandwidth, the net spectrum efficiency for these systems would be only 1.25.

Although the queued access of trunked repeater systems may provide only a modest improvement in spectrum efficiency, it is often a basis for other important features. These features include greater coverage areas (especially from mobile/portable units), the ability to share costs, access to the telephone network, access to a large variety of digital features, etc. One of the greatest advantages of trunked systems, however, is that multi-site trunking technology is the basis for small-cell multi-site systems like cellular telephone and ESMR, allowing these systems to benefit from short-range configurations having very great spectrum efficiency.

For users who need high channel availability (e.g. greater than 99 percent), trunked systems provide efficiency gains much greater than five-fold. For radio system users who have only an occasional need for radio communications, but who need a high assurance of channel availability, a trunked radio system with priority access features offers an excellent solution that has very high relative efficiency.

Bandwidth per channel. In spite of the great theoretical promise of highly efficient digital vocoders, it should be noted that the most efficient and the least efficient systems on a channel/bandwidth basis were both TDMA digital vocoder systems. The fact that the best vocoder was only about as efficient as an ACSB analog system (and the worst one was ten times less efficient) suggests that highly efficient digital vocoders are still hard to build. There is continued rapid progress in digital signal processing (DSP) chips and better algorithms, so there is hope. But considerable effort is needed to merely match an efficient analog circuit. The relatively poor showing of digital vocoders probably reflects the relatively undeveloped state of digital vocoders and modems. At some future time, there may be considerably more advantage to be gained by using state-of-the-art bandwidth compression techniques than there is at present.

Voice channel bandwidth could be reduced by using a lower-bit-rate vocoder or by using a modulation technique that provides more bits/Hz. Although there is probably more improvement possible in vocoder technology than in modulation technology, an improved vocoder will only improve the channel performance when voice is being used. Improved digital modulation technology will also improve the performance of all pure-digital functions such as paging, digital messages, fax, e-mail, etc. Since it is expected that purely digital functions will grow in importance, it may be advantageous to pay more attention to efficient modulators.

Although reducing the bandwidth may provide less improvement than using smaller cells, it has several advantages. It allows piece-by-piece replacement of older radios, and often allows the user control over part of the additional created channels. Although there may be interoperability problems during the changeover, it will probably not require major changes in operating procedures or in the communications organization. Piecemeal changes to narrowband systems may be the better alternative in areas where spectrum crowding suggests that improvements are needed, but where the market is too fragmented to pay for a fundamentally different infrastructure (such as a dense infrastructure of short-range cells).

Frequency reuse (range) efficiency. Table 12 shows an extraordinary variation of range efficiencies (350,000:1). The theoretical possibility of using the "wrong" communications range for a job and thereby consuming 350,000 times more spectrum than was necessary is a real and (hopefully) sobering fact. Although the mobile radio community has acknowledged that one should not over-design the usable range of a communications link, the consequences never seemed this serious.

The selection of communications range, however, is not a simple choice that can be arbitrarily made and isolated from other aspects of system design. The selected range has to meet the operational requirements of the mission. For example, an air traffic control (ATC) system will necessarily involve large operational ranges because of the need to communicate with aircraft over a large range and the very slow rate at which free-space propagation attenuates a signal.

On the other hand, a wireless PBX system is inherently short-range. Short-range systems usually have much higher infrastructure costs (because of the need for many radio sites, as well as an elaborate network to connect and coordinate the sites), and they will usually be practical only when costs of the required infrastructure can be shared with many other customers (as in a common carrier situation). In addition, such a change may require substantial modifications in procedures and organization to efficiently utilize the new communications architecture.

6. CONCLUSIONS AND RECOMMENDATIONS

The preceding summary chapter describes a variety of technologies that can be used to improve the spectrum efficiency of mobile systems. The list of various mobile systems and technologies reviewed is intended to provide examples, not to be exhaustive. In particular, the simplified definition of relative spectrum efficiency is limited in its application to systems that are sufficiently similar to the selected 15 kHz analog FM mobile radio "reference" system. It would have been inappropriate to use this definition to evaluate the efficiency of several important system technologies, including code division multiple access (CDMA), HF radio, and mobile satellite systems.

Nevertheless, even using the incomplete data developed in this paper, some important conclusions can be reached. The wide range of possible spectrum efficiencies should provide some motivation for frequency managers to carefully encourage the use of spectrally efficient technologies whenever it is functionally and economically appropriate. The potential gains in spectrum capacity from using some of the more efficient technologies could provide enormous increases in service, much larger increases than would be available by the reallocation of additional frequency bands for mobile services.

Our conclusions are based on the potential applicability of spectrally-efficient technologies to several types of mobile systems. We will comment on

a. Large- and small-cell, multi-site trunked systems (cellular phone, PCS, ESMR)
b. Single-cell trunked systems (SMR, campus, wireless PBX, wireless LAN)
c. Single user systems.

Large- and small-cell, multi-site trunked systems. The greatest increase in spectrum efficiency is possible using multi-site trunked systems like E-SMR, cellular, or PCS. In dense urban areas, these systems can use many spectrally efficient technologies including extensive frequency reuse, queuing, low-rate vocoders, etc., resulting in relative spectrum efficiencies in the 100-10,000 range. In addition, these systems can offer a wide variety of advanced digital services, like data, paging, stored digital messages, fax, encryption, priority access, talk groups, full-duplex voice, etc. Because of the very high spectrum efficiency of these systems, it is reasonable for spectrum managers to try to place as large a percentage of total traffic as possible on this type of system in geographic areas where spectrum crowding might be a problem.

This type of system is quite "scalable" and can be engineered to meet a wide range of traffic needs, mainly by adjusting the size and number of cells and the number of frequencies available at each cell. In suburban and rural areas, the coverage area for each cell can be greatly expanded by using more transmitter power and higher antennas. Cellular systems operate over at least a 10:1 range of cell diameters, adjusting the power output of the mobile (personal) radio to match the power needed in the local cell. Although the use of larger cells decreases the efficiency gain from frequency reuse, larger cells are generally used only when there is less demand for services (with the corollary greater availability of frequencies).

These systems are complex and expensive, requiring a large initial investment to construct the system and a substantial expense to manage it thereafter. It is not as clear whether there are cost advantages to building a denser system. If one assumes that the number of sites in a system are proportional to the number of customers per acre (this would be the case if one kept the same number of trunked channels at each site), the coverage area of the average site would be inversely proportional to the density of customers. It is probably cheaper to build a site with a smaller coverage area; low-power transmitters, lower antenna towers, possibly fewer site restrictions and zoning problems, etc. On the other hand, there will be more sites, meaning that the network connecting the sites will be more complex (but the lines connecting the sites will be shorter). Altogether, we believe it likely that the cost per customer will decrease somewhat as the density of customers increases.

Single-cell trunked system. A conventional, single-cell trunked system or SMR is considerably less complex than a multi-cell system. Such systems are particularly applicable to serving multiple users in a single building or campus (a short-range or medium-range system) or throughout a metropolitan area or military base (a long-range system). Since there is only a single cell, the entire coverage area must be serviced by that cell. If a large area of coverage is required, the opportunity to achieve spectrum efficiency through frequency reuse will be considerably diminished. Nevertheless, queuing efficiency and narrow bandwidth efficiencies can still be obtainable. With a short-range coverage requirement (e.g., a wireless PBX), there will be high spectrum efficiency through frequency reuse, and total system efficiency can be very high.

Although a long-range, single-cell trunked system is probably less efficient than a short-range multi-cell trunked system providing the same coverage, it will be more efficient than a several single-channel systems providing the same service.

Single-channel system. A single-channel voice system is unable to use queuing and is probably less able to take advantage of frequency reuse. Although there are fewer users to share system costs, the system is relatively inexpensive to construct or operate. The traditional wisdom applies here, including the use of no more transmitter power or antenna height than is needed to provide the required coverage area. Such a system is often an older 15-kHz- or 25-kHz-bandwidth analog FM system that will be eventually replaced with more efficient narrowband technology. Changing regulations for Federal and non-Federal mobile radios will increasingly discourage the use of single-channel systems, partly through efforts to make more efficient systems (such as trunked systems, SMRs, and PCSs) easily available as alternatives.

The preceding paragraphs describe a hierarchy of spectrally efficient systems, with the highest efficiency ascribed to the small-cell, multi-site, trunked system. This list should not be construed to suggest that lower-efficiency systems are not recommended or have no place in a modern telecommunications system. Instead, this list should alert the spectrum manager to the large difference in spectrum efficiencies between the various systems and allow a choice based on the radio environment and the user needs. Short-range, highly efficient, multi-site systems may be available in dense urban environments, and they may meet user needs in an economical manner. Such systems are complex and expensive, however, and they require a large volume of traffic to be economically feasible; they will probably not be available in rural areas.

Lower-efficiency systems can be used in rural areas. This is potentially a good fit, since there is less need for efficiency in the rural areas, and low-cost, low-efficiency systems can be tolerated. In the urban areas, the high efficiency of short-range systems is required to carry the total traffic. Thus, there is a natural division into urban short-range systems and rural long-range systems (and possibly super-long-range systems in remote areas, supported with low VHF or mobile satellite technology).

There may be substantial advantages in combining the communications requirements of smaller users, so that there is enough combined traffic to make a higher-efficiency system feasible. The advantages of shared multi-user systems may not be realized if the users have sufficiently different needs that the common set of user requirements dictates equipment that is much more expensive than needed by a simpler system meeting the requirements of a large subset of users. Problems may also arise if user "cultures" are sufficiently different that there are irreconcilable differences regarding system operational priorities. Other structural barriers that prevent users from sharing efficient systems include lack of common frequencies or frequency bands, conflicting administration and control issues, and conflicting license and procurement regulations. The spectrum management process should be careful to follow policies that effectively encourage agencies to share multi-user systems, where appropriate.

The various spectrum efficiency factors discussed in this paper do not only produce spectrum efficiency. They also produce added spectrum capacity, which is the ability to provide service to many additional customers. The evolutionary development of more efficient cellular systems, for example, could result in a system that provides enormous amounts of service. Stating that the Toronto cellular system is (in certain locations) as much as 100-150 times as efficient as the original cellular configuration means that the present 50 MHz of cellular spectrum is now providing services that the original cellular configuration would have required 5 GHz of bandwidth to provide. This magnitude of spectrum capacity increase can make major shifts in the amount of spectrum required for various services.

Finally, the pursuit of increased spectrum efficiency is an empty goal, unless it results in better service at a lower cost to customers. Any system that is merely spectrally efficient, but not responsive to customer needs and marketplace competition, is a failure. Therefore, the material within this paper must be continually evaluated in the context of customer needs, including some of the less-than-obvious factors of security, robustness during emergencies, and user priorities. In the end, the goal is to provide the best service for the least cost, though we believe it reasonable to include the value of the spectrum used as part of the cost of the system. A market-based spectrum cost may make it easier to understand the trade-offs involved, and spectrally-efficient technologies will help to minimize the cost of spectrum used by the system.

7. REFERENCES

[1]. CCIR Report 662-3, "Definition of Spectrum Use and Efficiency, Volume 1, Spectrum Utilization and Monitoring." Geneva, 1990.

[2]. W. G. Duff. A Handbook on Mobile Communications, Don White Consultants, Inc. 1976.

[3]. J. Sarnecki, C. Vinodrai, A. Javed, P. O'Kelley, and K. Dick. "Microcell design principles," *IEEE Communications Magazine*, p.78, April 1993.

[4]. Bellcore Generic Criteria for version 0.1. Wireless Access Communications System (WACS). Technical Advisory TA-NWT-001313, Issue 1, July 1992.

[5]. Aeronautical Industry VHF Policy. Administrative and Radio Frequency Management Manual, ARINC.

[6]. D. J. Cohen, P. C. Roosa, R. J. Matheson, D. E. Kitzmiller, and G. M. Patrick. "Land mobile spectrum efficiency - A plan for Federal Agencies to use more spectrum-efficient technologies," NTIA Report 93-300, October 1993.