EPILOGUE

Four Laws of Nature and Society: The Governing Principles of Digital Wireless Communication Networks

Andrew J. Viterbi

Four laws, two each from the natural sciences and the social sciences, have formed the basis for the development of digital wireless communication networks. This essay describes their interaction, as well as their logical support for spread-spectrum multiple-access techniques.

E.1 Overview

In this techno-philosophical essay, we attempt to demonstrate that the implementation and success of digital wireless communication networks depends primarily on four basic laws and their underlying theories, which are attributed respectively to:

- Maxwell and Hertz
- Shannon
- Moore
- Metcalfe

The first two laws are laws of nature, while the last two, though often mistakenly thought as such, are in reality laws of behavior. The order is in the sequence of their discovery and their importance; additionally, as the field of wireless communications has matured, the emphasis and immediate relevance has shifted gradually downward in the list. Without an appreciation for Maxwell's and Hertz's theories, there would be no controlled wireless propagation of electromagnetic waves. Without an understanding of Shannon's theories, efficient use of the spectrum through sophisticated signal processing could not be achieved. Without the consequences of Moore's law, these signal processing techniques could not be implemented in a useful and economic fashion. And finally, Metcalfe's law, which we shall explore last, helps to predict the success or failure of large new network deployments and, consequently, the wisdom of business strategies involving proportionally large capital investments.

E.2 Wireless Propagation and Its Anomalies

In a remarkable sequence of achievements in theoretical and experimental physics toward the end of the nineteenth century, the basis for electromagnetic propagation was established and proved both theoretically and experimentally. Though numerous academic researchers, residing in the musty lecture halls and laboratories of that period, shared in the success, the two that stand out are James Clerk Maxwell and Heinrich Hertz. Maxwell's equations, learned by every electrical engineering undergraduate as the elegant synthesis of all the fundamental laws of electricity and magnetism, represents the framework upon which, with the aid of a few unifying steps, the theoretical proof of electromagnetic wave propagation is readily constructed. Hertz was perhaps the first to verify this theory experimentally. Thereafter, just after the turn of the century a succession of pioneering "communication engineers" defined this new profession with gradually more convincing experimental successes, culminating in commercial deployments for which the name of Guglielmo Marconi stands out for his outstanding blend of experimental and business acumen. Our purpose, however, is not to review the scientific and historical record, far better recounted elsewhere, but to note the particular features that impact modern wireless multiple access communication embodied in digital cellular networks.

Thus, we take for granted electromagnetic propagation but, as discussed in several chapters of this volume, note that the direct path from transmitter to receiver may not be the only path of signal propagation and, in some cases, may be blocked and hence attenuated far more than other indirect paths created by reflections off terrain or buildings. Consequently, for a transmitted signal

$$x(t) = A(t)\sin[2\pi f_0 t + \theta(t)], \tag{E.1}$$

where f_0 is the carrier frequency and $A(t)$ and $\theta(t)$ are, respectively, the amplitude and phase modulation which bear the information transmitted, the received signal will be of the form

$$\begin{aligned} y(t) = {} & B_1(t-t_1)\sin[2\pi f_o(t-t_1)+\theta_1(t-t_1)] \\ & + B_2(t-t_2)\sin[2\pi f_o(t-t_2)+\theta_2(t-t_2)] \\ & + \cdots + B_L(t-t_k)\sin[2\pi f_o(t-t_L)+\theta_L(t-t_L)] + \eta(t), \end{aligned} \tag{E.2}$$

where $t_1, t_2, \ldots, t_L$ are the propagation delays of the various propagation paths and $B_k(t-t_k)$ and $\theta_k(t-t_k)$ are, respectively, the received amplitude and phase for the kth path. Finally, $\eta(t)$ represents the additive noise at the receiver, partly of thermal origin, but which may also include interference caused by other emissions and transmissions not under the control of the communicators. The amplitudes $B_k(t-t_k)$ and phases $\theta_k(t-t_k)$ may be distorted versions of the transmitted amplitude and phase, which vary with time. As discussed in several of the chapters of this book, one model for this process is that of a time-varying delay line as shown in Figure E.1, where both the delays between taps, $\Delta t_k \triangleq t_{k+1} - t_k$, and the complex tap multipliers $\alpha_k(t) = B_k(t)e^{j\theta_k(t)}$ are time-varying functions. If these values can be measured exactly, then the optimum receiver, in a minimum-mean-square-error sense and in the presence of Gaussian interference, is the matched filter, which can be implemented as the time reverse[1] of Figure E.1, but with the complex amplitude $\alpha_k(t)$ replaced by its conjugate $\alpha_k^*(t) = B_k(t)e^{-j\theta_k(t)}$.

The problem, however, is in the feasibility of the measurement and its accuracy when the parameters vary rapidly in time. A fundamental limitation on feasi-

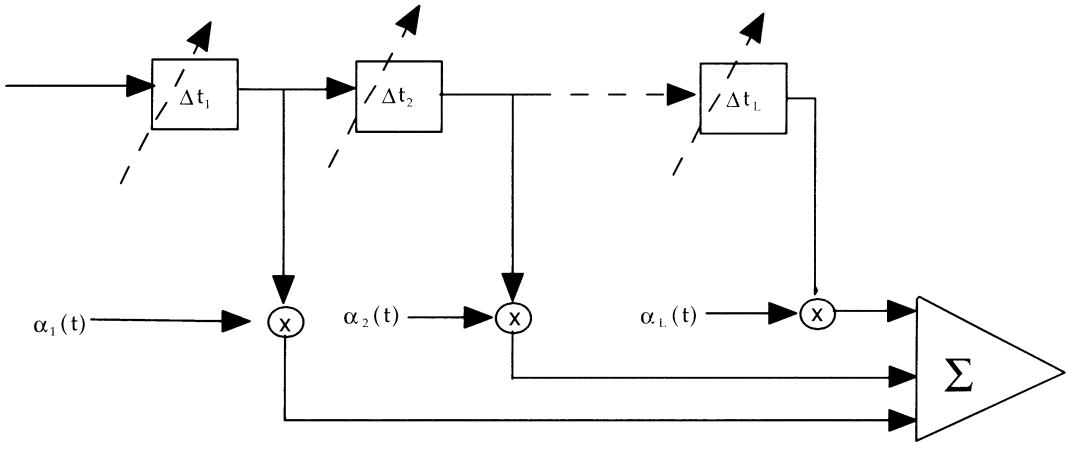

Figure E.1 Multipath propagation channel model: $[\alpha_k(t) = B_k(t)e^{j\theta_k(t)}]$.

[1]Usually referred to as a RAKE receiver.

bility, known as the uncertainty principle, dictates that measurement resolution in time is inversely proportional to the signal (and receiver) bandwidth. Thus, for example, if we wish to resolve two paths separated in time by $\Delta t = 1\ \mu$s, the bandwidth of the complex signal $A(t)e^{j\theta(t)}$ must be at least on the order of 1 MHz.

If the signal and receiver bandwidth W is less than $1/\Delta t$, then the paths will appear smeared together; for paths whose relative delays are separated by less than $1/\Delta t$, their complex amplitudes will occasionally cancel one another, thus causing deep fades in that composite component of the received signal. But measurement accuracy depends on more than bandwidth. Fundamental estimation theory (the Cramèr-Rao bound) leads to the expression for the standard deviation (inaccuracy) of the estimate for the peak-time of the received signal,

$$\sigma_{\Delta t} \geq \frac{k}{W\sqrt{E/N_0}}, \tag{E.3}$$

where W is the signal bandwidth, N_0 is the noise or interference density (assuming wideband noise of uniform density), and E is the energy of the signal component measured; k is a proportionality constant that depends in part on the definition of bandwidth but is not far from unity. Thus, the longer the measurement time, the more accurate the result, *provided* the signal parameters remain nearly constant over this duration. This phenomenon implies that for rapidly varying environments, such as with speeding vehicles, the measurement time must be shortened and the accuracy reduced. Though it would appear that wider bandwidths are always preferable, this too has its limitations. For while the far field propagation effects produce distinct multipath components (such as reflections from different buildings or hills), the near field effects may create many components, closely spaced in time. With a very large bandwidth, we may resolve these close components, but each will produce very little received energy, jeopardizing its measurement accuracy. Thus, keeping the bandwidth large enough to resolve a reasonable number of moderately spaced multipath components is important. Beyond this, it may be better to allow very closely spaced components (closer than the inverse bandwidth) to combine, occasionally causing fading of the combination through cancellation, but on the average with enough energy for accurate measurement. The key is to have a sufficiently wide bandwidth to isolate enough components (or composite components) and thus provide sufficient path diversity to guarantee an overall adequate signal strength for the optimally combined paths at the output of the matched filter. This strategy is usually effective for outdoor propagation, where the propagation time dispersion is on the order of several microseconds. Indoors, where propagation time dispersion is much smaller, other techniques for temporal and spatial diversity must be employed.

We move on now to the second set of laws.

E.3 Shannon Theory: Limitations on Signal Processing

Claude Shannon, beginning with a remarkable series of papers in 1948, established the theoretical basis of digital communication with two well-known theorems: for source coding and channel coding. The first established the minimum bit rate required to reproduce a source signal within a given degree of accuracy; the second established the maximum rate at which transmission is achievable with arbitrarily high accuracy, in terms of channel parameters such as bandwidth and signal-to-noise ratio.[2] Underlying the proof of the channel coding theorem was the concept of signal randomness, which is closely related to wide bandwidth spread-spectrum signaling. We shall concentrate on the latter as it applies to multiple-access communication, which we have previously labeled "Shannon's Third Theorem." It is best expressed in terms of a game between a communicator and an interferer (or jammer). If both are restricted to transmit at power levels S and J, respectively, the communicator's channel capacity is bounded according to the Shannon-theoretic limits,

$$\begin{aligned}\frac{1}{2}\log[2\pi e(S+J)] - H(J) &\geq C \\ &\geq \frac{1}{2}\log(1+S/J)\text{ bits/channel symbol,}\end{aligned} \tag{E.4}$$

where C is channel capacity, $H(J)$ is the entropy of the jammer's signal of power J, and all logarithms are to the base 2. (This formula can be converted to bits/s by multiplying all terms by $2W$, where W is the communicator's bandwidth, assuming Nyquist-pulse signal modulation.) The minimax solution to this game, meaning the joint selection of communicator and jammer signals that maximizes capacity for the worst case jamming or that minimizes capacity for the best case communicator signal, is for both the communicator and the jammer to employ random signals whose first-order distributions are Gaussian and for which successive symbols are independent. The per symbol entropy of the jammer signal is thus $H(J) = (1/2)\log(2\pi eJ)$.

Consequently, from (E.4) we obtain

$$C = \begin{cases}\frac{1}{2}\log(1+S/J)\text{ bits/channel symbol} \\ W\log(1+S/J)\text{ bits/s,}\end{cases} \tag{E.5}$$

[2]This maximum limit, known as channel capacity, has occupied legions of channel coding specialists for the intervening half-century. While simple techniques for reaching within about one-third to one-half of channel capacity (for a wideband Gaussian channel) have been known and employed for at least 30 years, only within the last few years has a composite technique involving iterative soft decoding of parallel or serial concatenated codes, known as "turbo" decoding, shown that efficiencies above 80% of channel capacity are practically achievable, provided sufficiently long decoding delays can be tolerated.

which is the usual capacity expression for a Gaussian channel. Gaussian signals can be approximated by spread-spectrum signals that are implemented by modulating the digital information onto a carrier already modulated by a random (or pseudorandom) sequence of symbols generated at a much higher rate than the information, approximately equal to the spreading bandwidth W. The ratio of the random sequence rate to the digital information rate is the spreading factor. Actually, provided the random carrier bandwidth is much larger than the information rate (the condition for spread spectrum), the symbols of the random sequence need not be Gaussian distributed. It suffices that the random sequence consist of independent, equiprobable binary symbols and thus be Bernoulli distributed; given the large spreading factor, the aggregate (sum) of the independent spreading symbols over the duration of one data symbol will approach a Gaussian distribution, according to the central limit theorem.

Returning to the communicator-jammer scenario, since the spread spectrum jamming is approximately uniformly distributed over the bandwidth W, we can define its spectral density as

$$N_0 \triangleq J/W. \tag{E.6}$$

Let the communicator's signaling bit rate be R_b bits/s, which is bounded by the capacity formula (E.5). Its bit energy is the ratio of power S to bit rate R_b:

$$E_b = S/R_b. \tag{E.7}$$

Combining (E.6) and (E.7), we have the ratio of tolerable jamming-to-signal powers,

$$\frac{J}{S} = \frac{W/R_b}{E_b/N_0}, \tag{E.8}$$

where E_b/N_0 in the denominator is the minimum value required at the receiver for tolerably low error probability. A lower bound on E_b/N_0 can be obtained from the capacity formula, for since $R_b \leq C$, it follows from (E.5) and (E.8) that

$$\frac{R_b}{W} \leq \log\left(1 + \frac{E_b/N_0}{W/R_b}\right), \tag{E.9}$$

whence,

$$\frac{E_b}{N_0} \geq \frac{W}{R_b}\left[\exp\left(\frac{R_b}{W}\ln 2\right) - 1\right] > \ln 2. \tag{E.10}$$

The lower bound, $\ln 2$, is approached for R_b at capacity and as the spreading factor W/R_b approaches infinity.

We turn now to a nonhostile and reasonably cooperative set of communicators in multiple access to a common receiver, such as a base station of a cellular mobile telephony system. If M communicators are all spread over the same bandwidth by

independent random (or pseudorandom) carrier sequences and each has its power controlled so that they all arrive at the common receiving station with equal powers, S, then the demodulator for each user will be faced effectively with jamming power equal to the sum of the powers of all other users, $(M-1)S$. It follows from (E.8) and (E.10) that the tolerable jamming-to-signal power and hence the tolerable number of other users

$$M-1=\frac{W/R_b}{E_b/N_0}<\frac{W/R_b}{\ln 2}. \tag{E.11}$$

Thus, the overall throughput aggregated over all users, normalized by the total (common) bandwidth occupied, is upper bounded by

$$\frac{MR_b}{W}<\frac{1}{\ln 2}+\frac{R_b}{W}\approx\frac{1}{\ln 2}=1.4\,\text{bits/s/Hz}. \tag{E.12}$$

To approach this bound requires a very large spreading factor and error-correcting coding powerful enough to approach channel capacity. It also assumes that all the interference is caused by other users in the band, ignoring background noise of thermal or other origin. Including background noise and practically implementable coding techniques, throughputs of one-quarter to one-half of this value can be achieved, depending on the time-variability of the physical channel.

It can be shown that, with powerful enough error correction, the overall throughput MR_b can approach the classical channel capacity formula (E.5), with J equal to just the background noise *not* including any of the interference from the other cooperative users. Approaching this capacity requires that all users cooperate further in transmitting at specified but unequal powers, and that the common receiver optimally demodulates and decodes each user successively, subtracting off its effect from the common, overall received signal prior to decoding the next user. This latter procedure is an idealized form of the successive cancellation discussed in Chapter 3. At present, however, this optimal successive cancellation procedure remains a theoretical possibility only. Less ambitious cancellation or cooperative demodulation techniques abound, but they seem to yield only modest improvements.

We proceed now to review the practice of spread-spectrum techniques over the past half century.

E.4 Half a Century of Wireless Spread Spectrum: From Military to Commercial Applications

Spread-spectrum techniques for thwarting hostile interference or jamming date back to World War II. The sophisticated approach employing carriers whose spectrum is spread by a pseudorandom sequence generated by a maximum-length

shift register (with linear feedback) date back to the fifties. In their simplest conceptual implementation, the binary sequences, which appear random and repeat only after $2^N - 1$ symbol times, where N is the length of the shift register, modulate the carrier by shifting its phase by $+\pi/2$ or $-\pi/2$ radians, corresponding to whether the symbol is a "0" or a "1," respectively.[3] The intended receiver knows the parameters to generate the same sequence and thus can demodulate by performing the inverse operation, shifting phase by $-\pi/2$ or $\pi/2$ corresponding to a "0" or a "1," respectively. Incidentally, by performing this inverse operation, the jammer spectrum at the demodulator will appear to be spread, even if the jammer's transmitted signal was originally narrowband.[4]

With the launch of Sputnik in 1957, the era of satellite communication began. Many of the early launches were of military satellites, which are "sitting ducks" to hostile interferers. Spread-spectrum techniques with very large spreading factors provided for a wide margin of advantage over a jammer, according to (E.8). Commercial satellites began operations in the mid-sixties and the first digital communication satellites were launched in the seventies. Spread-spectrum techniques were not employed commercially until the eighties, when they found their way into mobile terminals operating with very small and hence wide-aperture antennas in the presence of much stronger interference from stationary terminals transmitting through near-orbit satellites. With 250,000 mobile satellite terminals now installed in trucks worldwide, providing two-way communication and position location to their home bases, this spread-spectrum system has dramatically impacted the long-haul transportation industry. Finally, in the nineties, spread spectrum has had an even greater impact on the digital cellular communication industry. Here, it is usually referred to as *code-division multiple-access* (CDMA), to distinguish this access technology from frequency division (FDMA) and time division (TDMA) techniques.

The common thread through all these applications is tolerance to interference through digital signal processing. As discussed elsewhere in this volume, for cellular applications, the interference comes not only from the other users communicating through a given cell's base station but also from the transmissions of users in other cells, which contribute strongly to the background noise in the given base station. For FDMA and TDMA systems, it is generally necessary to allocate different bands or time slots to contiguous cells, thus reducing spectrum efficiency by an order of magnitude. With CDMA, all cells can be allocated the same common spectrum, a feature called *universal frequency reuse.* This of course increases the interference in each cell, reducing the number of users per cell, but only by a factor

[3]As noted elsewhere in this volume, this approach is generally called *direct-sequence spread spectrum.* An alternate approach, known as *frequency hopping,* uses multiple symbols of the pseudorandom sequence to select one of numerous carrier frequencies among which to hop.

[4]This case again implies a minimax solution to the game between communicator and jammer.

of about 1.6, a net gain over the other access techniques. As previously noted, to establish (E.11), each CDMA user must be power controlled to guarantee the maximum number of users per cell. But with tight power control, each transmitter emits the least amount of power needed to achieve reliable transmission, thus avoiding the usual power margin allocation and also reducing the interference to users in the same and other cells. Furthermore, the spreading feature of CDMA also guarantees a bandwidth wide enough to isolate multipath, using the RAKE receiver technique described in Section E.2. Also, with universal frequency reuse, transition between cells can be eased by performing "soft handoff"; as the mobile user approaches the edge of a first cell, it can begin communication with the second cell's base station without dropping the link to the first cell's base station. This artificially creates a dual diversity multipath condition. The same multipath RAKE receiver can thus be employed to handle soft handoff just as it does the natural multipath condition previously described in Section E.2.

Employing all the above techniques in implementing a digital multiple-access system, as well as error-correcting coding to reduce E_b/N_0, and variable rate transmission of digitally compressed voice, results in a spectral efficiency over ten times that of a conventional analog system. Hence, employing techniques based on Shannon laws improves the efficiency of a system based only on Maxwell-Hertz laws by more than one order of magnitude.

Their implementation, however, depends critically on digital signal processing technology whose practical and economic embodiment would be impossible without highly integrated solid state processors and memories. We discuss this technology and its underlying law in the next section.

E.5 Moore's Law: The Socio-Economic Basis for Digital Wireless

Gordon Moore, a founder of Intel Corporation, observed in the seventies that the number of devices per unit area that can be incorporated in a silicon integrated circuit (IC) doubles approximately every year and a half. We may state this fact as the formula

$$v(T) = v(T_0)2^{(T-T_0)/1.5}, \tag{E.13}$$

where $v(T)$ is the device density at time $T > T_0$, where time is stated in years. Although this may appear to be a physical law like those previously discussed (and in some minds it is considered such), it is in fact an empirical observation with no direct physical basis. Rather, it is explained by the fact that human ingenuity, coupled with market forces, produces exponential growth of technological capability, and eighteen months appears to be the time between product cycles in the

semiconductor industry. This leads us to categorize Moore's Law as a socio-economic principle. Applying the formula and taking the initial year to be 1965, when one IC contained only one device, we find that by 1995 one IC of the same area could contain one million devices and by the year 2000 about ten million devices, both estimates being reasonably accurate. Within a few more decades, the atomic limit will be reached, but already experiments in subatomic storage would lead us to believe that the ultimate limit may be even further out. In any case, the millionfold growth in the last three decades has turned many a system theorist's dream into reality. The early military spread-spectrum systems consisting of multiple racks costing tens of millions of dollars are now implemented on a single chip (embedded in a palm-sized cellular telephone), costing only tens of dollars. In short, the progress described by Moore's Law was indispensable for the realization of the benefits of all three laws of Shannon (source, channel, and multiple-access coding). Yet, there remain the skeptics who do not understand, or merely ignore, the significance of Shannon's laws made practically implementable by Moore's Law and continue to design and attempt to justify systems that do not profit from these guiding principles.

E.6 Metcalfe's Law: Implications for Wireless Networks

Another socio-economic law of more recent vintage, due to Robert Metcalfe, states that the value of any communication network grows as the square of the number of users of the network. More precisely, if the number of users is N, then

$$\text{Network Value} \sim N(N-1)/2 \tag{E.14}$$

since this is the number of connections[5] possible among N users. This law places an inordinate burden on the initiation of a network service that requires a significant capital cost to both the network provider and the consumer. Numerous failures attest to this fact. The most notable was "PicturePhone," an initiative of AT&T in the early seventies, which was discontinued after an initial trial period. As telephone data modems became ever more capable, offering progressively higher data rates at ever lower costs, a few manufacturers offered updated versions of telephone video terminals that could operate over the ordinary public switched network. Even these were not particularly successful because of low consumer demand both caused by and resulting from prices being high. The most common sale involved the purchase of a pair of terminals by grandparents who wished to view distant grandchildren. Turning to successes, we cite the Minitel data terminal operating over the French public switched network. This, in large part, was the

[5]Actually, since each link is bidirectional, one could argue that it should be $N(N-1)$.

result of heavy subsidization by the French government and thus removed the burden from the consumer. Another is the facsimile (fax) machine, which is practically ubiquitous in businesses and is becoming common in consumer households. For this success to occur, first the prices had to fall significantly and then the considerable saving and convenience became clear: a fax message consumes a fraction of the time of a telephone conversation, and more importantly, it gets through with greater accuracy than voice mail when the recipient is not present. This feature is particularly important for transoceanic calls, where the volume of fax messages now exceeds that of voice calls.

The most dramatic current network success is the Internet. Its usage has been growing precipitously in just the last couple of years, after a gestation period of over twenty years, during which the U.S. government through DARPA and, later, NSF, financed an ever-higher-speed data network to interconnect computer centers of universities and government facilities. Turned over to the private sector, its usage and required capacity have grown exponentially with time. The explanation is the growing presence of the personal computer at almost every desk and workstation of businesses and in the majority of homes, coupled with the rapidly falling costs of the embedded data modem. With tens of millions of users comes the opportunity to capture their attention as consumers, resulting in the creation of a multitude of information and other services through the World Wide Web.

Focusing, however, on our theme of digital wireless networks, the question arises as to whether Metcalfe's law applies in a strict sense. Even a single mobile phone has accessibility, through a base station, to all the fixed phones of the public switched wired network. Yet the mobile user wants to connect to the universe of users wherever he or she may roam. Thus, the operator must provide to this "first" user, connectivity through a multitude of base stations throughout a metropolitan area as well as in multiple areas and even in multiple countries. And this works in reverse as well. All the user's wired friends and associates want to reach her or him wherever she or he may be. Hence, an operator's capital expenditures dominate the network economics, at least initially. Thus, effectively, the value should grow as the square of the area covered (assuming, simplistically, a uniform density of users). Each unit area covered captures users linearly, for whom value increases also linearly with the size of the total area covered. Clearly, the network provider's economics are improved, the fewer the base stations required per unit area. A base station's coverage area and capacity, in numbers of users that it can serve, are the key economic parameters. With light usage, coverage is the principal factor. As usage grows, the capacity constraint dictates that more base stations must be provided. Both coverage and capacity are increased from two- to four-fold by the use of spread-spectrum techniques, enhancing economics and consequently network growth.

Case histories of the growth of two very different digital wireless networks are summarized in Table E.1.

TABLE E.1 Case histories of two digital technologies

	GSM (TDMA)	IS-95 (CDMA)
Technology Proposed	1982	1989
Technology Standardized	1988	1993
First Commercial Launch	1991	1995 (late)
1 Million Subscribers	1994	1996
5 Million Subscribers	1995	1997
50 Million Subscribers	1997	1999 (Estimated)
Spectral Limit*	Sooner	Later

*Limit set by Maxwell and Shannon, not Moore or Metcalfe

The first, GSM, is the Pan-European TDMA standard, now available also in the Americas and Asia. It was launched as a "Greenfield" service, meaning that new spectral allocations were provided for it. Also, most countries in which it was launched had either no previous service or a very inadequate and sparse analog service, and GSM was the only digital technology licensed in the countries of the European Union. Probably its most significant innovation was the introduction of "global roaming," whereby a phone purchased in any covered nation can be used equally in any other covered nation, wisely adhering to the just-stated wireless version of Metcalfe's law.

The second technology is the spread-spectrum, or CDMA, technology which has been the major theme of this essay. It was standardized in North America as the Telecommunication Industry Association's IS-95. Its purpose was to provide much higher spectral efficiency than the analog networks, already well established in North America. Thus, it was planned to ultimately serve an order of magnitude more users within the same spectral allocation, over the same base stations, which were nearing saturation with the inefficient analog access technology. A principal requirement, both economic and regulatory, was to not displace the existing analog users, unless they willingly chose digital service. The introduction strategy, therefore, contained three components:

- Provide superior service, including improved voice quality
- Convert only as much of the spectrum over to digital service as the user demand warranted, thus not disadvantaging the remaining analog users
- Provide for staged conversion of base stations so that initially not all stations needed to be converted to digital

CDMA was able to fulfill all three requirements. By its better than tenfold increase in user capacity, it required converting only one-tenth of the bandwidth to digital service to provide enough capacity to potentially serve all existing analog users on the network. By providing improved service, CDMA could attract the

heaviest users; these represent typically 3% of users, consuming 30% of service. Thus, a relatively small percentage of conversions to the 10% digital bandwidth could alleviate congestion on the 90% of the band still serving analog users. Finally, by providing all converted users with a dual-mode analog/digital CDMA phone, the operator did not need to provide digital service on all base stations at once; when digital users roamed away from digitally equipped base stations, their phone was transferred automatically to analog without service interruptions. The statistics in the second column of Table E.1 attest to the success of this approach.

Standardized seven years later than GSM and launched four years later, CDMA now lags in customer adoption by two years. Significantly also, in almost all countries where it was introduced, the consumer had a choice between analog, CDMA, and one or two digital TDMA technologies. The most important conclusion, then, is that where there already exists an installed base, as well as considerable competition, it is essential, if not also mandatory, to provide backward compatibility to the predominant, previously existing technology.

The last entry in both columns deals qualitatively with the time at which spectral allocations will reach saturation. More users are supported by CDMA because of its greater efficiency. One approach for GSM operators to expand their capacity would be to gradually equip their base stations with CDMA technology compatible with their GSM network signaling. The high-usage customers would be provided with a dual-mode GSM/CDMA subscriber unit capable of communication with both types of base stations, in the same manner in which North American and Asian analog users were converted to digital.

We conclude with a look into the future. Unlike the wired network that already serves a substantial and ever-growing percentage of data users at varying speeds, the wireless network still offers primarily voice telephony, with only a minuscule percentage of low-speed data. This situation is certain to change over the next few years. The nomadic nature of business users, particularly for electronic mail and continuous connectivity to a home base, mandates the availability of wireless high-speed data (above 64K bits/s) beyond what is currently provided by ordinary wired phone connections. Several standards organizations worldwide are currently deliberating the merits of a variety of proposals for this service. Given the large number of current subscribers to digital cellular services, adherence to the principles of Metcalfe's law seems logical. In order not to start at the bottom of the quadratic curve, service providers would be well advised to choose a technology that is backward compatible to one of the existing services. The flexibility of spread-spectrum signaling facilitates such compatibility.

The complex interaction of technology and economics, involving the interplay of the four very diverse laws described in this essay, has created a vibrant wireless telecommunication industry, likely to serve an ever-widening percentage of the world's population, with both mobile and fixed service, for many decades to come.