A Brief History of Mobile Communications

Abstract
Sometime during the last decade, the cell phone became ordinary. We now take for granted that we can call anyone, anytime and anywhere, and have moved on to ask whether such a call would be safe while driving, or socially acceptable in public places. The technology arena has also moved on-- from voice to wireless information. Yet the cellular revolution is less than 20 years old, and the entire history of "mobile radio" is barely 100 years old. In this article, we take a brief look at that 100-year history, in part to derive those elusive lessons that historians are fond of, but mostly just for fun.

The Magic of Radio
In the second half of the 19th century, wireline telegraphy and telephony changed the nature of communication. Within a few years, the movement of information had gone from the speed of people to the speed of light. Then, in the 20th century, a second revolution took place, as radio freed those same communications from the constraint of a wire.

Radio waves have almost magical properties that capture the imagination and lead to a variety of different uses. Travelling through our atmosphere or the universe, they connect points that could be connected in no other way. Traveling at a known speed, they allow us to locate flying objects, pinpoint our own location on the surface of the earth, and even chart the universe. Perhaps no other property, however, has been more profound than what we might call their ability to "fill space." While the radio signal may appear to move from point to point in a linear fashion, it is in fact expanding spherically. It passes through many solids, and fills in behind obstructions through the mechanisms of reflection and diffraction, so that the information it carries is available "everywhere" throughout a large area of coverage. As a result, millions of people can access and share the same "broadcast" news and entertainment at very low cost. Notwithstanding that a new generation of Americans believes that entertainment arrives on a cable, it is impossible to overstate the impact that broadcast information and entertainment have had on the world.

This property of radio waves that we have described as "filling space" leads not only to broadcast, but to the last and newest form of radio communications. It allows us to reach a particular person "anytime, anywhere," and since the antenna can move through that space while sending or receiving the signal, it makes "mobile communications" possible.

The Origins of Mobile Communications
In the mid-1860s, the Scottish mathematician James Clerk Maxwell produced a pair of equations whose solution predicted electromagnetic waves propagating at the speed of light. It took 20 years to verify this prediction in the laboratory, and another 20 years for the first "mobile" application to take place. [1] In September of 1899, Guglielmo Marconi ushered in the era of practical mobile radio communication with his historical radio telegraph transmissions from a ship in New York Harbor to the Twin Lights in Highlands New Jersey. His first transmission covered the arrival of Admiral Dewey from Manila, and soon thereafter, coverage of the Americas Cup races was provided as well. This was accomplished with high energy, wideband pulses of radio noise, created using a spark generator coupled to an antenna. The signal was received with a similar antenna attached to a fascinating device known as a "coherer," a tube in which metallic particles were caused to cohere to one another by the presence of the radio signal. The reduced resistance of the coherer in this state allowed the radio signal to be detected, after which a mechanical "tapper" returned the particles to their original state. Each telegraphic symbol was represented by a rapid series of these spark-generated signals, which was detected as an audible tone. [2].

Within a year, Marconi had added filtering to create separable channels, and thus to allow multiple simultaneous transmissions in the same area. Within three years, he was able to cross the Atlantic with a radio transmission, and radio telegraphy was soon used on many ocean-going vessels (most notably, it was used to report the sinking of the Titanic in 1912). Analog (voice) transmission was used as early as 1905, but early work was driven primarily by military applications, and it was not until 1919 that an experimental ship-to-shore radiotelephone service was initiated. Commercial radiotelephony for passengers on ships in the Atlantic was begun in 1929. By this time, radios were small and rugged enough to be installed in automobiles, and the first "land mobile" radio system was put into operation by the Detroit police in 1928.

By 1934, there were 194 municipal police systems and 58 state police radio stations serving more than 5000 radio-equipped police cars [3]. The age of mobile radio had begun.

Early Mobile Telephony

Spectrum that could be exploited for practical systems was always in short supply, and mobile communication services were in competition with military and broadcast services for the available channels. As a result, most mobile radio channels were devoted to emergency and public service uses until the cellular revolution began some fifty years later. Despite these difficulties, an early mobile telephone service was initiated in 1947, using several channels at 35 MHz. Additional channels were soon allocated at 150 MHz, and later at 450 MHz. Improved filtering and frequency stability allowed these channels to be narrowed, eventually creating a total of more than 40 channels for mobile telephony. In an effort to create competition, the FCC divided these channels equally between the local telephone companies and a group of new entities called Radio Common Carriers (RCCs), allowing two competing carriers in each geographic area.

Early mobile telephone systems resembled broadcast systems, in that powerful transmitters were used to cover a distance of 20-30 miles from a high tower or rooftop. The reuse of any channel for a different call required separations of more than 50 miles. The New York channels were reused in Philadelphia, so that each city and its suburbs were limited to about 40 simultaneous calls. The demand for service was great, resulting in severely overloaded channels and long waiting lists for service. As a practical matter, people with an important need for service (e.g., doctors) were given preferential treatment, and the average person might face the discouraging reality that his position on the list was actually becoming worse over time.

The earliest mobile telephone systems were “manual”, in that calls in both directions were placed through a mobile operator (a process still common on marine radio systems). In the 1960s, automatic systems were introduced, with significant advances. An “idle tone” was put on an idle channel, so that a scanning receiver could identify and camp on that channel. The mobile telephone could decode its own dialed number on the idle channel, and send dialed digits for outgoing calls. In combination, these capabilities yielded a mobile telephone service that approximated conventional telephone service of that period (except that due to overcrowding, an idle channel was seldom available). For the several decades of their existence, these manual and automatic systems created an image of mobile telephony as a crowded but “elite” service, and one that was generally unavailable to the public.

Early Cellular Proposals

The origins of cellular lie surprisingly deep in the past. As early as 1947, at essentially the same time that the first 150 MHz system was installed in St Louis, the Bell System (AT&T) proposed a “broadband urban mobile system,” and requested 40 MHz for its implementation somewhere in the region between 100 and 450 MHz. At that time, the idea that a large geographic area could be served using many small coverage areas had already been put forward at Bell Laboratories, primarily to allow low-power radios to be used. The FCC denied this request, citing the unavailability of frequencies in that range. In 1949, the FCC considered allocation of the band from 470 to 890 MHz, but chose to reserve this band for the vast educational and entertainment opportunities thought to be offered by the newly proposed UHF TV band. Once again, in 1958, the Bell System requested an allocation for mobile telephone, this time in the range of 764-840 MHz, and once again the FCC declined to take action. By this time, however, the essential ingredients of the cellular system were under discussion—the use of small cells and the reuse of channels that would increase dramatically the number of simultaneous calls per channel.

Docket 18262

Finally, in 1968, with the UHF TV band continuing to fall far short of its original promise, the FCC opened Docket 18262, which proposed to allocate the upper portion of this band to mobile systems for both private and public uses. In 1971, the Bell System submitted a detailed proposal for a cellular system to be implemented in this band. [4] This began a decade-long battle among a diverse set of “stakeholders,” many of whom had a strong interest in maintaining the *status quo*. Broadcasters did not want the frequencies reassigned. Existing manufacturers were threatened by the prospect of new systems, markets and competitors. Fleet operators wanted the spectrum for conventional uses. RCCs feared domination of a new, capital intensive service by the Bell System. In contrast, there was no lobby for the potential customers, who were generally unaware of the proposed new service. Cellular proponents came to fully appreciate the words written by Machiavelli, 400 years earlier:

"It must be remembered that there is nothing more difficult to plan, more doubtful of success, nor more dangerous to manage, than the creation of a new system. For the initiator has the enmity of all who would profit by the preservation of the old institutions, and merely lukewarm defenders in those who would gain by the new ones."

As Docket 18262 evolved, other technologies were emerging that would be critical to cellular design. The first "stored program controlled" central office switching machines were being introduced, providing a powerful central controller capable of the new functions, such as vehicle locating and call handoff, that would be required to allow calls in small cells. Integrated circuits offered the opportunity to create complex logic in a mobile radio at small size and low cost. Synthesizers were designed that would allow the mobile units to access hundreds of channels. Minicomputers offered a powerful option for control of the complex base stations, which were called "cell sites." The very long period of controversy even allowed for new technologies, like microprocessors, to emerge, evolve, and influence design choices. Given the many years that passed in litigation and confrontation, it is perhaps remarkable that the final design remained quite similar to the initial proposal.

Ultimately, in 1981, the FCC crafted a Final Order that paralleled earlier decisions on spectrum allocation for mobile systems. About half the spectrum was allocated for "trunked" systems, in which groups of 20 channels were used to provide conventional wide-area services such as fleet dispatch. The other half was allocated for cellular systems, once again to be divided equally between the local "wireline" telephone companies (primarily operated by the Bell System) and competing Radio Common Carriers. Rules were written to assure that cellular telephones could access either system, creating an open, compatible national service.

Early Cellular Deployment

The first cellular system in the U.S. was put into operation by the Bell System in Chicago, in 1983, as part of a rapid deployment plan. Systems were also being installed in Japan and Europe, the beginning of a worldwide deployment that has now touched virtually every corner of the earth (in fact, the first Japanese system was installed in Tokyo in 1979, while the U.S. was still mired in regulatory and political delays). Ironically, at its moment of triumph after almost 40 years of dockets, litigation and development, the Bell System was split up in 1984, and the "telephone company systems" were deployed and operated by regional telephone companies. And in a further irony, when AT&T (parent of the Bell System) decided to enter the cellular market a decade later, it was forced to do so by buying some of the RCC systems that had been created by Craig McCaw and others. Repeated mergers and sales have rendered the original distinctions created by the FCC meaningless.

The intent of the FCC to assure seamless national service was dealt a severe blow by the very competition they sought. While the capabilities needed for an efficient "roaming" service had been provided with the standards that were adopted, the many system operators were slow to seek inter-carrier billing agreements, and fraud involving "cloned" cell phones actually caused many operators to deny service to customers from other systems. Only in the late 1990s, with the advent of additional standards, improved fraud prevention and inter-carrier billing agreements, have cell phone users in the U.S. been provided with true nationwide service.

Pocket Phones and the age of "Personal" Communication

Cellular began as an "automobile" system, with relatively large trunk-mounted radios that were connected by cables to dashboard-mounted "control units." Even as service began, however. "satchel" units were offered that provided a "portable" option. More significantly, Motorola soon introduced the "DynaTAC," a 2-pound hand-held unit that was about the size of a brick, and could be carried in an attaché case. The evolution toward the pocket phone had begun.

The evolution to the tiny pocket phones of today has had tremendous significance. A telephone attached to an automobile may provide great utility (and in fact was an extremely popular service), but was still a location-based phone rather than a "personal" device. In the 1980s, when we called a car phone, it was because we expected someone to be in a particular vehicle. We might have called a home or office with the same expectation. Today, when we dial a cell phone, we are calling a particular person, and may not know or care whether that person is home or across the country. This is the essence of personal communications.

The Transition to Digital Cellular Systems

The cellular systems deployed during the first decade of service used digital signals for control, but the voice signal was carried as an "analog" waveform. Even in the 1970s, however, there was an early debate as to the potential advantages of a fully digital system. At that time, it was concluded the technology was not ready for a fully digital cell phone that would achieve the same spectrum efficiency, voice quality and cost as the analog design, but by the late 1980s this was quickly changing. This primary reason was that voice processing technology, using digital signal processing (DSP) chips, had made dramatic advances. Good quality voice, once requiring 30-60,000 bits per second, was becoming possible at rates approaching 10,000 bits per second. This allowed three times as many voice channels to be created from the same spectrum, which had important economic implications.

The primary cost of a cellular system is in its cell sites, and the radios themselves are but a small part of that cost. The costs of the land, tower, building, antennas and cables, the control and maintenance equipment, the transmission facilities and backup power—all these lead to cell sites that can cost as much as a million dollars. This cost is obviously supported by the traffic that the cell site can carry. For this reason, a reduction in the spectrum allocated to a system would have significant impact on the per-user (or per-minute) cost of cellular service. The number of channels in a cell would obviously be reduced, and the traffic-handling (or "trunking") efficiency of these smaller groups would be reduced as well. This explains why the FCC decreed two systems per coverage area, rather than a larger number. In fact, during the long debate on this subject in the 1970s, the Bell System argued (correctly but unsuccessfully) that allocation of all the available spectrum to a single entity would reduce per-user costs by more than a factor of two.

The same logic applies to the creation of more channels though digital voice processing. By allowing a threefold increase in the number of channels, digital voice processing allowed the system operator to spread the cost of the cell site over more than three times as many customers, reducing the overall cost per user by a factor of about three. (Initially, of course, there was a cost to upgrade the systems to these digital channels, and the cost savings was only achieved over time, as more customers were added without requiring more cells.)

This economic motivation for digital cellular was compelling, and a standard was proposed using Time Division Multiple Access (TDMA) to achieve this 3-to-1 advantage. It might be expected that this standard would have been quickly adopted, but that was not to be the case. A second standard, using Code Division Multiple Access (CDMA) was proposed as being still more efficient, and many years were spent in an energetic debate that caused significant delay. The details and merits of this dramatic but time-consuming battle cannot, alas, be explained within the bounds of this humble history. Ultimately, both standards were adopted and are in current use in the United States.

Cellular Standards in the U.S. and Europe—Contrasting Examples

The evolution of cellular standards in the U.S. and Europe provide an interesting contrast. In its initial standards for cellular, the FCC was determined to create a single specification for compatibility between bases stations and mobile units, so that wide-area roaming would be possible. At that time, European access standards and even frequency bands varied from country to country, so that wide area (i.e., pan-European) roaming was not possible.

In approaching "second generation" (2G) digital cellular systems, motivations in the U.S. and Europe were once again different, but in new ways. In the U.S., these systems were to occupy the same frequency bands as the existing systems, and the primary motivation was to increase capacity and reduce cost. Both the TDMA and CDMA proposals had this objective, but engineers seemed unable to agree on which standard would best accomplish this goal. The FCC ultimately required a "dual-mode" approach. All new systems were required to provide service using the old standard, and could combine either TDMA or CDMA with that standard. While logical, this approach resulted in added complexity, market confusion and delay.

In Europe, the primary focus of second generation planning was to provide a richer set of features (such as digital messaging and improved security) and set a single standard that would allow seamless roaming service. New bands were set aside, so the issue of increased capacity was less central than in the U.S.. To the surprise of many in the U.S. (including this author) the Europeans agreed to a common standard, and GSM (a form of TDMA) was born. GSM used more bits per second for the voice signal than the U.S. form of TDMA, and additional coding bits doubled that rate, so it provided about the same number of voice channels per megahertz of allocated spectrum as the older systems. The additional coding provides a more robust signal in the presence of interference, however, which can theoretically be used to reduce the "reuse distance" between cells having the same channels, and thus increase the "spectrum efficiency" of the system. It is difficult to ascertain the degree to which this increased efficiency has been realized in practice.

Although initial plans for digital cellular in the U.S. assumed the same UHF spectrum as for analog, the FCC relented in the early 1990's and allocated 120 MHz of additional spectrum near 1.9 GHz for what came to be called PCS (Personal Communications Services). Plans were made to clear this spectrum of point-to-point microwave services and to auction it off to as many as six operators in each area. Thus, cellular service in the U.S. today operates in two bands (800 MHz and 1.9 GHz), two modes (analog and digital) and, in the digital mode, two radio technologies (TDMA and CDMA). GSM has also migrated to 2 GHz, while preserving the radio parameters and protocols of the earlier GSM systems.

The continuing TDMA/CDMA battle in the U.S. created a certain degree of uncertainty among service providers throughout the world, and undoubtedly contributed to the success of the less controversial GSM standard. It may also have resulted in a strong U.S. position in the new technology of CDMA, which could play an important role in the next generation of cellular systems. It is likely that the many participants in this drama draw differing conclusions as to the benefits and drawbacks of their competitive strategies. It is clear, however, that for thirty years the politics of competition and regulation have played as great a role as technology in shaping the cellular market in the U.S.

The Cost of Spectrum

While spectrum at 800 and 900 MHz has proved very desirable for cellular purposes, and spectrum at 2-3 GHz has now been used as well, higher frequencies pose significant problems in terms of propagation and path loss. At higher frequencies, antennas are smaller and capture less received signal, and the ability of the signal to bend (diffract) around corners and over hills is reduced. As wavelengths are reduced to inches and less, scattering by foliage and even raindrops takes place. Even at 2-3 GHz, systems must begin with smaller cells, raising significantly their initial costs.

Since the spectrum that is suitable for cellular systems is limited, it will inevitably have tremendous value. Initially, governments viewed such spectrum as a national resource, to be allocated to the most socially valuable services based on careful (and time-consuming) deliberations. More recently, governments have discovered that they can sell spectrum for large amounts of money, thus filling their coffers while avoiding the aggravation of protracted allocation proceedings. Thus, the spectrum auction was born, and has led to some remarkable prices for this ephemeral resource.

Today, the cost of spectrum for a new system may equal the cost of the system itself. This reinforces the value of achieving high spectrum efficiency, but makes it difficult for system operators to achieve profitability, particularly in the early years of system operation. As a result, it has become increasingly important to build a large customer base in the presence of multiple new competitors, and the customer has benefited in terms of very attractive rate structures. It is less clear that the investors will benefit, but a euphoric optimism about the future of mobile telephony has kept spectrum prices at very high levels.

Creating the Wireless Internet

Given the popularity of mobile voice communications and the Internet, it seems inevitable that they will soon merge in what has been called the "wireless Internet." Initial email and Internet-based services have been offered in current second-generation cellular systems, with market results that have often fallen short of expectations. This may be because the pocket phones that are most popular have small screens, limited keypads and small batteries. It may also be that these initial service offerings have not been sufficiently exciting to the mobile customer. It is certainly true that the bit-rates offered have been rather low, and the cost of delivering those bits has been rather high, when compared to the fast modems and flat-rate services of the "wireline Internet."

Despite this somewhat disappointing start, the "wireless Internet" appears to be the driving force behind proposals for the third generation (3G) cellular systems that are expected during the current decade. These systems will offer bit rates that are significantly higher than those provided by current systems, but it is not yet clear how the issues of cost and physical configuration are being addressed, or what exciting new services will emerge to capture the hearts of the mobile public.

Designers and investors face difficult issues in shaping the next generation of cellular systems, but they can draw comfort from the realization that few technologies have demonstrated the benefits, popularity and rapid evolution that have characterized mobile communications for more than a century.

REFERENCES

[1] Larry Greenstein, *100 Years of Radio,* Speech at WINLAB Marconi Day Commemoration, Red Bank, NJ, September 30, 1999
[2] Bob Buus, *Marconi, A Man of Vision*, CQ Amateur Radio Magazine, September 1999, pp. 16-23
[3] George Calhoun, *Digital Cellular Radio,* Artech House, Inc., 1988, pp. 25-26
[4] High-Capacity Mobile Telephone System Technical Report, Prepared by Bell Laboratories, December 1971, included within FCC Docket 18262

AUTHOR'S BIOGRAPHY

Richard Frenkiel spent 30 years at Bell Laboratories, where he was responsible for systems engineering for AMPS, the first cellular system in the U.S. He served on the EIA committee that adopted the rules for nationwide cellular compatibility, and invented a method for cell splitting that has greatly simplified the process of growth in cellular systems. He is currently Visiting Professor and Director for Strategic Planning at WINLAB, the Wireless Information Networks Laboratory at Rutgers University.

For his work in Wireless, he has received the National Medal of Technology, the Alexander Graham Bell Medal of the IEEE, and the Achievement Award of the Industrial Research Institute. He is a Fellow of Bell Labs and the IEEE, and a member of the National Academy of Engineering.