

Lecture 9- Pre-lecture reading

The evolution of cellular

Portability, digital voice and smart phones

In this lecture we will discuss the evolution of cellular technology, services and economics over more than 30 years. The first important evolutionary trend was the shrinking size and resulting portability of the cell phone. It started as a car phone, but got small enough to carry in a pocket, ushering in the world of truly “personal” communications.

Then we will take our first look at digital transmission in cellular systems, which began in the late 1990s. With new research in digital “voice compression” (fewer bits per minute of speaking), three voice calls could be carried by the 30-KHz cellular channels that had carried one analog voice. This tripled the voice capacity of cellular systems and changed the economics of “second generation” (2G) cellular systems, but competition created a big surprise in how the increased capacity was used by competing cellular providers. 2G digital channels also ushered in a profitable new service called text-messaging.

Finally, in the last decade-plus, third and fourth generation (3G and 4G) cellular systems have led to smart phones, largely replacing voice with information and entertainment as the dominant cellular services. The science and economics of all these dramatic changes is a significant strategic issue in cellular and will be discussed in much greater detail in future lectures.

The pocket phone

An important trend in the early cellular years was a dramatic reduction in the size and cost of a cell phone [slide 4]. The early cell phones were powered by the car battery and had a handset mounted on the dashboard of a car, with cables running to a big, heavy radio in the trunk of the car and more cables connecting the radio to an antenna on the roof of the car. The installation took several hours and usage was limited to the time users spent in their cars. As early as the mid-1980s, however, Motorola had introduced its portable “Dyna-TAC” (see below), which included the telephone, radio, antenna and battery in a portable package. It was the size of a brick, weighed about 2 pounds and cost about \$3000, but it could be carried in an attaché case and represented a dramatic change in the nature of cellular communications. Prices and sizes dropped very quickly, and by the early 1990s there were cell phones weighing ounces, that people could comfortably carry in a handbag or shirt pocket. Their cost was below \$100, and most were provided at no extra cost by cellular operators, with a 1–2-year service contract.

The impact of this transition on the use of cell phones has been profound. The early car phone was mobile, but it was still a phone in a place— even if the place was a moving car. If you called a cell phone, you got whoever was in the car at that time, just as when you dialed a home number, you got whoever picked up the phone at that location. When the cell phone got small enough to carry, and cheap enough for families to have several of them, we began to think of a cell phone as the same as a person, and the era of truly “personal” communications was born. Moreover, the user carried a portable

phone all day, not just when in a car, so the caller was “always there” and call minutes per user increased dramatically.

From car trunk to pocket

The dawn of true personal communications



**A radio in the
car trunk- 1983**



**DynaTAC
1983**



**MicroTAC
1989**



**RAZR
2004**

A phone in your pocket is different from a phone in your car

4

Roaming

Now we call people instead of places, and we expect them to answer whether they are in New York or New Delhi. In fact, we are often unaware of where the person we are calling is located at that moment, but this total and effortless “roaming”, which we have now come to expect, was slow to evolve. In Europe, the first generation of cellular systems used different standards and different channels in different countries, so roaming was very limited (the “Nordic Network” in Denmark, Norway and Sweden was an exception). In the US, as we have noted, the FCC tried to achieve nationwide roaming by requiring a “compatibility standard.” There would be two systems in each market, and the goal was to allow any cellular phone to access any system as a “roamer” when away from home. The FCC envisioned that cellular operators would reach mutual agreements to serve each other’s “roamers” and that efficient nationwide service would result.

In practice, such agreements were slow to evolve, in part because of the large number of cellular operators (recall that the A-Band licenses were assigned by lottery, so there were many new operators). A worse problem was fraud, in the form of “cloned” cell phones. When the first systems were planned, the problem of fraud was underestimated. It was thought of in terms of an individual, unsophisticated user, who might try to change the assigned telephone number that was inserted into the cell phone when the subscriber arranged for service. There was provision for a chip to be

buried in the circuitry that held a serial number that could be compared to the telephone number, so that a failed match would represent a fraudulent phone.

Unfortunately, the criminals were more sophisticated than expected, and they learned to modify phones to charge calls to fraudulent numbers. These “cloned” phones were used in other cities, as “roamers”, and the carriers weren’t able to share information well enough to verify legitimate customers while the call was in progress. This was also a period when international calls were quite expensive, so there was a real incentive to “clone” a cell phone for international calls. Scammers would buy a “cloned” cell phone and run up thousands of dollars in international charges, and when the cellular operator tried to collect, there was either no subscriber or a subscriber who hadn’t made the calls. The international charges had to be paid by the cellular operator, and many refused to accept calls from roamers unless prior arrangements had been made. Roaming service became almost non-existent in the US.

Eventually, in the 1990s, the cellular industry consolidated to a more manageable number of operators, and new technology and network standards were developed to prevent fraud and coordinate “roamer” information between carriers. Carriers negotiated inter-carrier agreements and “seamless” roaming in the US finally became possible.

In Europe, cellular operators also focused on “seamless” and innovative services. A new European standard called GSM was adopted by the European Economic Union (EEU). This level of cooperation among the Europeans was a new and somewhat surprising phenomenon, and had important results. It allowed seamless roaming across Europe, and in other countries that adopted the GSM standard, and it enabled other personalized services. For example, by plugging a “SIM card” into any phone, users could transfer their service (with all personal information) to the new phone, and since there was a single system standard in all GSM countries, international roaming was achieved. There were even some GSM systems installed in the US. This new focus on personal services led to the new systems being called “Personal Communication Systems” (PCS), but they were just cellular systems with better features.

Second generation (2G) digital voice systems and the standards wars

A new technology called “voice compression” emerged in the 1980s, allowing a high-quality digital voice signal to be transmitted with fewer bits. Then, in the 1990s, an inexpensive chip called a Digital Signal Processor (DSP) came into use. It could handle voice compression in a cell phone and thus it dramatically increased the voice capacity of a digital cellular system.

In with the environmental radio noise and interference of a mobile radio system, the radio channel could only transmit about 1.5 bits per second per Hz. Thus, the 30 KHz (1-way) cellular channel could only transmit 45,000 bits per second. This was less than the 64,000 bits per second that was needed for “telephone-quality” voice in the telephone network, so there was no point using digital voice in the early cellular systems. With voice compression, however, that same 30 KHz channel, transmitting 45,000 bits per second, could carry three calls. By going to digital voice and using voice compression, service providers could triple the voice capacity of their existing base stations. [For future purposes, we must understand that this breakthrough did not increase the total “bits” that the system could carry. It simply reduced the number of

bits needed to transmit a voice.] Nevertheless, the result was to triple the “voice-capacity” of the system” without expensive cell splitting.

The most obvious way for service providers to benefit from this “technological windfall” would be to serve three times as many users with no new base stations. It would require new base-station radios and new cell phones, but those are modest costs compared to cell-splitting. No new base-stations would be needed until the number of users (and therefore the revenues) had tripled. Their profits would increase dramatically.

That would be the strategy a monopoly would choose-- triple the capacity, reduce the monthly charges a bit (to keep the users and regulators happy) and rake in the profits. As we know, of course, cellular by the late 1990s was a highly competitive world, with four competing systems in most locations. In a competitive world you do things to get an advantage over your competitors and the customer wins. In a superb irony, the company that upset the 2G financials was AT&T, the ex-monopoly. Before we discuss that, there was also the political matter of the “standards war” between two competing 2G digital cellular standards in the US.

In addition to new networking standards, many new “digital cellular channel standards” emerged in the 1990s, replacing the analog channels that were used in the first systems. This led to new “second generation” (2G) cellular standards and the replacement of all cell phones and base-station radios in the world, but that was a small price to pay to triple the capacity of all those million-dollar cells. GSM was a 2G digital cellular standard created in Europe, and two 2G digital radio standards emerged in the US. They were called CDMA (Code Division Multiple Access) and TDMA (Time Division Multiple Access). We will discuss these technologies in more detail in Lecture 16 but a high level description that follows will suffice for now.

While TDMA and CDMA were both digital radio standards, they were very different. TDMA was more like the old cellular standard. It continued to use the same 30 KHz channels that had carried analog voice, with the same 7-cell reuse pattern, but instead of the analog voice signal they carried a digitally modulated voice signal. The 30 KHz channel could carry 3 10 KHz digital voice signals. The system was organized by slotting time (hence the name time division multiple access (TDMA)). CDMA was a much more dramatic change in technology, which made it more difficult to determine its capacity. It used a “spread spectrum” channel that was more than 125x wider than TDMA (1.25 Megahertz to transmit a voice call in one direction, compared to 10 Kilohertz per voice for TDMA), so the CDMA channel could carry a lot more information. The compressed voice information was still just 15,000 bits/second so all that extra information capacity was used for a special type of coding that could reject very high levels of interference. As a result, the CDMA voice signal could work in the presence of interference that was much greater than the desired voice signal. With TDMA the desired voice signal had to be much greater than the interference. As a result, the TDMA channel could be used once in 7 cells, while the CDMA channel could be used many times in every cell.

Whether CDMA was really a capacity increase over TDMA or not would depend on how many calls it would allow in every cell. With 1.25 MHz of spectrum in one direction, TDMA could support 18 calls per cell. The proponents of CDMA would initially argue that it had as much as 30x the voice capacity of TDMA, but the argument was complex, highly mathematical and based on many design modifications. Some of those design modification were never implemented and others had nothing to do with CDMA. At the

time there was a lot at stake for Qualcomm (Qualcomm was the source for CDMA technology). In 1999, at the height of the dot-com bubble, Qualcomm's stock rose more than 1000%, before falling back in 2000. It was a bitter battle, much like the battle between AT&T and Motorola two decades earlier.

Important lessons may be learned from this. Depending on the assumptions made, wildly different conclusions can be reached. CDMA was a difficult and exacting technology that presented practical challenges in implementation, so Qualcomm was a year or two behind TDMA in delivering real equipment. Since CDMA proponents couldn't prove their claims without operating real systems, many cellular operators delayed committing to either standard, or chose the less controversial European GSM standard. As it turned out, CDMA probably had a small advantage over TDMA in terms of capacity, but no uncontested data-driven results were ever presented (even when systems began operating). Such situations are a dilemma for strategists. Decisions must be made, sometimes based on who you trust or the costs of being wrong. Sometime waiting is the prudent choice, but that isn't always possible. In this case, the long and bitter battle between CDMA and TDMA delayed the benefits of 2G in the US and gave an advantage to the European GSM standard. GSM was reasonably "spectrum-efficient", but chose to focus on better service with less controversy.

From a political perspective, it's interesting to compare the standards philosophies adopted by governments in the US, Europe and Japan, both for the first generation and second generation of cellular. In the US, the first generation focus was on spectrum efficiency and nationwide compatibility, using the best possible technology. Given the conflicting claims of AT&T and Motorola, determining the "best" standard caused years of delay, and since the problems with fraud and the lack of carrier agreements delayed true nationwide service, the benefits of nationwide compatibility were not realized for another decade. In Europe, there was no attempt to have first-generation roaming service between countries, so compatibility between systems wasn't needed (different countries used different spectrum, so compatibility wasn't possible in any case. As we have discussed, standards were quickly adopted and systems were in operation before those in the US. Japan was always interested in selling equipment around the world, and was willing to build to any standard.

For the second generation (2G) standards, the FCC now changed strategies. To avoid being blamed for delay again, they took a hands-off approach, leaving manufacturers to fight it out. They claimed that this would "allow the best technology to win", but ironically, this once again caused years of delay. In the US, "best" came to mean "most spectrum-efficient"—which standard could serve the most people per cell on a given amount of spectrum? This is an important concern, but the standards war between TDMA and CDMA caused uncertainty and delay. The arguments were difficult to understand and impossible to verify, and they went on endlessly.

As we have seen, choosing a less efficient standard than one's competitor would be a significant competitive disadvantage (it would lead to higher costs per user). Cellular operators in the US were unwilling to risk choosing the wrong standard, so they delayed purchases of any new technology. In Europe, to everyone's surprise, the European Union was able to agree to the GSM standard. GSM was arguably less spectrum-efficient than TDMA or CDMA, but because everyone would use the same GSM standard in Europe, operators didn't have to worry about a more-efficient competitor. As a result, they cared much less about spectrum efficiency and more about seamless customer services like international roaming. Their systems may have served fewer

users per cell, but since all operators used the same standard there was no competitive disadvantage. Operators were willing to move forward with purchases of systems, and GSM sales grew more quickly than the more efficient but controversial TDMA and CDMA standards proposed in the US. One could make a good case for the FCC having had the wrong strategy both times. They delayed cellular in the US both times, while the Europeans moved forward more quickly, and sold more systems.

A lesson in the economics of competition

In the early 1990s, AT&T was still operating independently and trying to grow. It had not yet lost all its wealth and was spending money getting into new businesses. With its rich cellular history and its old cellular dreams, AT&T decided to re-enter the cellular business by buying the biggest cellular network on the competitive RCC side of the business. They bought that network from Craig McCaw, who we discussed earlier. McCaw had bought licenses from the lucky lottery winners and was creating a nationwide network. His story is worth discussing because it leads to some interesting lessons in finance and competition.

McCaw's network was doing well, but he had taken on a huge debt (about \$4B in loans) and because his business was risky (with all that debt) he was paying high interest rates on those loans. What would have been a decent profit for AT&T was eroded by those interest payments and he was showing a very small profit. He was unable to borrow any more, and he was at risk of defaulting on his interest payments. If that happened, he would lose the company and all the family wealth he had invested (we will discuss bankruptcy later in the course). He had an incentive to sell the business and AT&T had an incentive to buy.

AT&T agreed to buy McCaw Cellular for \$16B. If this seems like a high price (much more than McCaw had spent to create the systems) we must consider that the first spectrum auction had just taken place and McCaw's spectrum was suddenly worth much more than he had paid for it. Perhaps more importantly, AT&T was impatient to operate a significant nationwide system and did not want to spend years creating it (as McCaw had done).

At the time, users were paying a fixed monthly service charge plus a per-minute charge (about 25 cents/minute) that discouraged calling. Many cellular customers bought service for emergencies and made few calls. AT&T was eager to increase its customer base (to steal customers from its competitors). The added system capacity from the 2G conversion was starting to become significant, and AT&T made a dramatic "package" offer. It offered users a large number of "minutes" of talking time for an attractive fixed price. This attracted a lot of customers from AT&T's competitors (in fact they had trouble meeting the demand and had to speed up their 2G conversion). At the same time, their competitors were forced to fight back with similar (and larger) package offers. This created a competitive war that ultimately used up the 2G advantage (and more).

In the years before 1998, minutes of usage had been pretty constant at about 1300 minutes/user/year (see the CTIA data). Under the influence of the package deals, the minutes per user tripled in only 4 years, while the voice revenue per customer remained essentially the same at about \$500 per user per year (less than \$50/month). Thus, under the influence of competition, all the benefits of that new 2G capacity (and more) went to the users. By 2008, in fact, users were using more than 8000 minutes per year, and were paying less than \$40/month for those minutes. This is rather amazing.

Before 1996, the cost of talking was fairly constant at about 50 cents/minute (counting the monthly charge and the per-minute charge). By 2008, it had fallen to 5 cents/minute, but if 2G provided a 3x increase in capacity how did the price per minute drop by a factor of 10?

One reason was that the car phone had become a pocket phone. Instead of only being available a few hours a day when the user was in a car, the pocket phone was always available. A system that had only been full during rush hours was now loaded most of the day. The cost of a system was based on providing maximum capacity during the busiest hour, so all that “off peak” traffic was essentially free. Voice compression and all-day usage allowed that huge increase in minutes per user at little cost to the service providers. They seemed to have passed up a huge opportunity for profits, but the six-fold increase in the number of users reduced SG&A expense per user by about \$20/month, and the digital radios allowed a very profitable new service called text messaging. In the case of 2G, competition worked for both users and providers.

Text Messaging: a popular and profitable new service

A text message was priced about the same as a minute of talking (about 5 cents), but it caused almost no load on the system. Even with voice compression, it takes about a million bits to transmit a minute of voice, but less than a thousand bits to transmit a hundred characters of text, so text messages generated significant new revenues without requiring new investments in capacity. By 2008, text messaging made up for almost 20% of revenues for the service providers, at almost no cost.

The era of 2G is a powerful testament for the benefits of competition. As we have noted, the old AT&T monopoly would probably have used the new capacity to raise profits, by adding users with little new investment. Many users would have continued to buy cellphones for emergency use, and not used them very much. Instead, the call packages changed everything. People started to talk—a lot. The cellphone became an important part of life. By 2008, almost everyone had one, and a trillion text messages were sent. Remarkably, it was a great period of profitability for the service providers as well. They were adding cells for the new customers, but the increased usage per customer was probably free (voice compression and all-day usage). Text messaging was bringing in almost \$10/month/user at no cost and the growth probably reduced SG&A/user by \$20/month.

We have also seen (in Lecture 8) that the new competitors created by the auctions of 1994 would have a hard time making a profit. We will see in Lecture 13 that the happy and profitable era of 2G was not equally shared by the new competitors. Moreover, as strategists with an increasing understanding of the world of cellular, we will see that we can predict what will happen with increasing confidence.

In the 2G era, we see the powerful benefits of competition for both users and service providers, but the era of 3G/4G/LTE systems and smart phones that would follow was quite different. Technology and competition would once again provide great benefits to the users, but huge investments for capacity have created some significant challenges for the service providers' profitably.

3G/4G/LTE: Cellular meets the Internet

The positive experience of the service providers with 2G digital systems (and the growing popularity of the Internet) created an excitement about the possibility of cellular access to the Internet. At the end of the millennium the dot-com bubble added to this excitement. Demonstrations of high-speed download with new “3G” technology seemed very promising. An idea called “convergence” proposed that communication, information, and entertainment could be provided by a single device— the cellphone. This led to very high prices for spectrum in 1999 and 2000, but a variety of problems delayed 3G service for a number of years. We can think of 2008 as the beginning of a transition from 2G to 3G. It was not the carefree transition the service providers might have envisioned.

The fiber-based internet was doubling in capacity every year, but those “high-speed” 3G radios didn’t compare to fiber. High-speed demos were misleading. They used a lot of spectrum to download big files like movies, but the real issue wasn’t giving high-speed downloads to a few people. It was giving a lot of download capacity to everyone, and the 3G/4G/LTE technology fell far short of fiber. In decades that saw million-fold increases in fiber capacity, the capacity of a radio channel (measured in bits/second/Hz) didn’t increase at all. Even 2G didn’t increase the capacity of cellular channels in terms of bits/second/Hz—it only reduced the number of bits that were needed for voice. 3G/4G/LTE technology did finally increase the data capacity of radio channels, but only about by a factor of about 5. The new technology, which was expensive and impressive for a radio system, was only a fraction of what would be needed.

Over the next decade the service providers made heroic efforts to meet the smart phone demands for ever-increasing capacity. They invested almost **half a trillion dollars** in radios, base stations, and spectrum (more than they had invested for voice services from 1985 to 2008) and they were able to achieve an impressive increase in cellular capacity (about 100x). . Once again, competition resulted in wonderful new services for the users, but revenues didn’t increase at all. In fact, revenues have been falling. Spending hundreds of billions of dollars with falling revenues has been a serious strategic problem for the service providers.

Now there is discussion of 5G systems, using huge amounts of inexpensive millimeter wave spectrum near 30 GHz to vastly improve capacity and download speed, but those frequencies have severe range problems (a few hundred feet and no building penetration). We will discuss 3G/4G/LTE in detail, and begin a discussion of 5G, in Lecture 16. Before that, however, we will conclude the first segment of this course with a movie followed by the first Midterm. After the Midterm, we will begin a series of “technology” lectures as well as discuss “unlicensed” spectrum that gave us cordless phones, WiFi, and Bluetooth. As always, we are interested in what things cost as well as how they work.