1

Controlling Your Volume

Cell phones are now part of our daily lives. Take illustration 1.1, which shows the **mobile penetration** for selected countries in mid-2015. This is the average number of cell phones per person in the country. Notice that the leftmost five are over 100%, meaning there are more cell phones than people in these cases.

Also, the 13 countries in illustration 1.1 were the ones in mid-2015 with over 100 million subscriptions. If we took the worldwide total number of purchased cell phone subscriptions at this time, it was over 6,800,000,000 (that's 6.8 billion!).

With these enormous numbers, you might ask, how is it possible that we can all communicate in the air effectively without disrupting one another's calls, messages, or Internet usage? We'll see a couple of methods for sharing in this part of the book, starting with power (speaking volume) control in this chapter.

The modern mobile **cellular** system is the result of decades of technological innovation. As mobile devices moved from being a luxury item in the 1940s–80s to an utter necessity by the twenty-first century, engineers had to come up with different ways of letting people share the air.

FROM TELEPHONE TO CELL PHONE

Before the advent of wireless networks and cell phones, communication networking was more about **wireline**. Wireline is communication using wires, as opposed to **wireless**. The first phone call was made all the way back on October 9, 1876, by Alexander Graham Bell, over a 2-mile wired stretch from Boston to Cambridge. The following year, Bell founded the Bell Telephone Company, the first company that provided a **public switched telephone network** service (we usually refer to this as "landline").

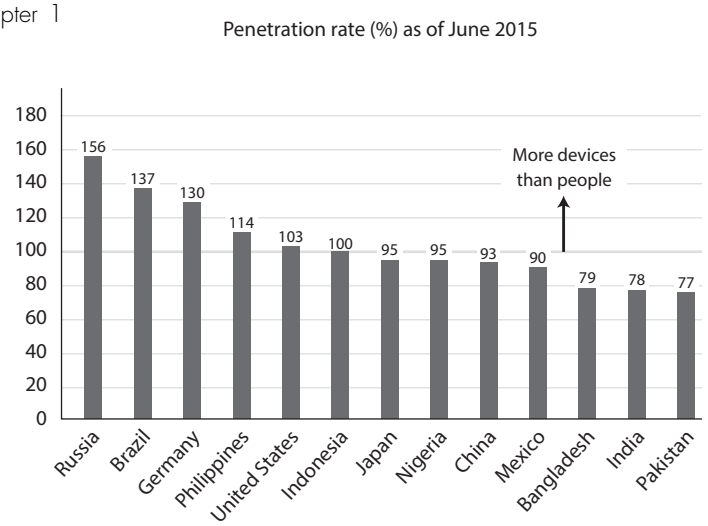

Illustration 1.1 Mobile penetration, or number of cell phones per person, in selected countries as of June 2015. Six countries have a penetration of at least 100%, meaning there are more phones than people.

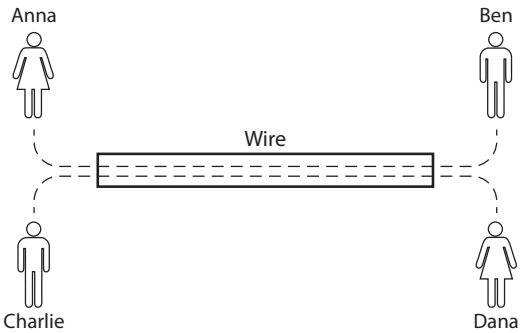

Illustration 1.2 Suppose Anna and Ben are having a phone conversation, as are Charlie and Dana. How is it possible for both pairs to use the same wire without interfering with each other?

Before Bell designed the telephone, he was experimenting with the telegraph, an earlier invention. The "multiple telegraph" was one where multiple **transmitters** (senders of messages) and **receivers** (receivers of messages) could operate over a single wire.

Stop there for a second: how is it possible for us to have many different people sharing the same wire, as in illustration 1.2? If Anna and Ben are trying to talk to each other, as are Charlie and Dana, wouldn't this cause them to disrupt each other's conversations?

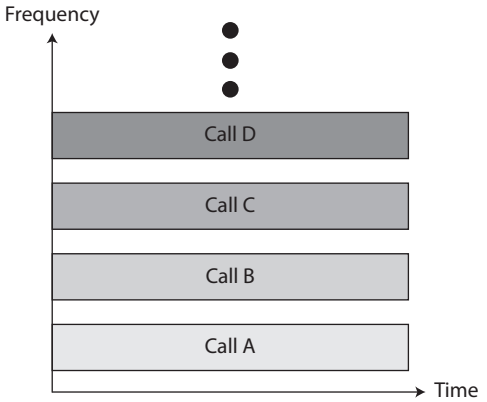

Illustration 1.3 With frequency division multiple access, calls are distinguished by their designated frequency channels: the users on Call A get one channel, those on B get a different one, and so on.

Not necessarily. Even though they are sharing the same *space* (the wire), we can separate them along some other dimension. The most intuitive one is perhaps time: let Anna and Ben use the wire for a bit, then Charlie and Dana, then back to Anna and Ben, and so on. We could also try language: have Anna and Ben talk in English, and Charlie and Dana talk in Spanish. Then they can just listen for their own language while still talking at the same time. In this case, we still have to worry about different voices overpowering each other.

These dimensions—time and language—are simplified examples of different **multiple access** technologies. These techniques allow multiple users to share the same network medium (e.g., wire or air). We take a look at them in more detail throughout this chapter.

Sharing by Frequency

The multiple telegraph separated conversations along different *frequencies*, in what is known as **frequency division multiple access**, or FDMA. FDMA allocates each transmitter-receiver pair, which we call a link, a separate **frequency channel** by which they can communicate. You can see this in illustration 1.3.

What is a "frequency"? For the ones we can hear, we distinguish them as different pitches of sound. Frequency is measured in units of **hertz** (Hz), which indicates the number of cycles per second in a wave. So 10 Hz means

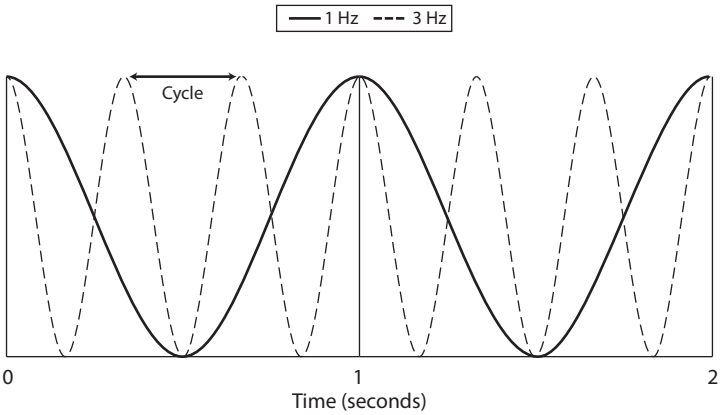

Illustration 1.4 Two waves of different frequencies, which is the number of cycles the wave completes in one second. The solid wave has a frequency of 1 Hz, while the dashed wave has a frequency of 3 Hz.

that a wave completes 10 cycles per second (see illustration 1.4). For more on frequency channels, check out Q1.1 on the book's website.

The unit of frequency will be mentioned many times in this part of the book, but the actual ranges we will deal with are many orders of magnitude higher than a hertz. Wireless frequency bands are typically referenced in millions of hertz, known as megahertz (MHz), or billions of hertz, known as gigahertz (GHz). To give you an idea of how large these are, the highest frequency a human can hear is about 20,000 Hz.

The first mobile phones, dating all the way back to the 1920s–30s, used FDMA. They were **analog** in nature, meaning that their signals traversed the air in their exact electrical forms. In 1946, the first "mobile-phone" network, called the Mobile Telephone Service, was introduced by Bell Telephone. This was an FDMA system, as was its successor in 1964. They are considered zeroth generation, or **0G**, technologies, as opposed to the **4G** technologies that we use today.

The First Handheld

Back in the 1970s, Martin Cooper of Motorola was convinced that handhelds were the wave of the future. In 1973, he and a team spent 90 days making the first-ever mobile handheld phone: the DynaTAC.

The DynaTAC was not what we view today as a handheld. Weighing close to 2 lbs, it cost almost $3,000 (in 1973 dollars!) and offered 30 minutes of call time before charging was necessary. By comparison, an iPhone in 2016 weighed less than one-third of a pound, cost as low as $150 (depending on the model and the wireless contract), and offered many hours of voice and data applications after each recharge.

It wouldn't be until the mid-1990s that the industry would really start switching from car phones. Similar to digital networks, palm-sized phones only became practical once the cost of electronic components began to reduce drastically, which is in turn partially driven by the volume of demand, which is in turn partially driven by the applications enabled by such technologies.

The "Cell" in "Cell Phone"

In 1976, New York City alone had about 500 mobile subscribers, and more than six times this number on the waiting list. There was a dire need to increase network **capacity**. So what could network operators do? There were really only two options: petition the Federal Communications Commission, or FCC, for more spectrum, or figure out a way to fit more users into the same spectrum.

For more information on the FCC licensing process, check out Q1.2 on the book's website. How can we put more users in the same spectrum? Maybe we can reuse the channels? That seems kind of far-fetched: if we have two links right next to each other, and they are using the same channel, certainly they would interfere. But what if they aren't right next to each other? If they are far enough away, then could we reuse the same channel?

The answer is yes. When signals propagate through the air (as well as through wires), their power levels **attenuate**. This means that they diminish with further distance, as in illustration 1.5.

Typically, attenuation is looked on as a negative quality. It causes a signal to weaken, making it harder to transmit over long distances. But that's exactly what we need here: if you and I are far enough apart, we can each make a call without overlapping in space.

The nature of attenuation led engineers to begin dividing mobile regions geographically into **cells**, often represented as "hexagons" (for good reason, beyond our scope here). The idea is that any given cell can be assigned a set of frequencies not being used by a cell adjacent to it. In

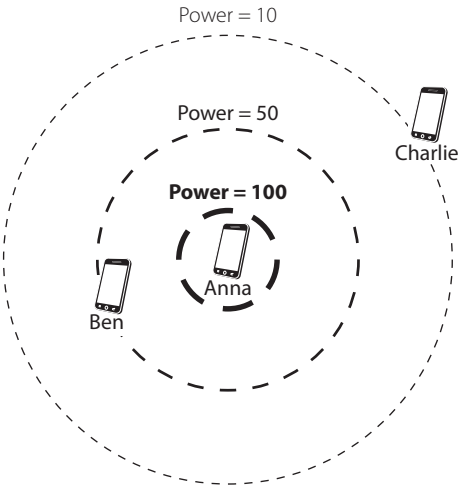

Illustration 1.5 As signals propagate through space, their power levels attenuate. Around Anna's phone, her transmit power level is 100. By the time the signal gets to Ben, it's 50, and at Charlie, it's 10.

this way, cells that are using the same channel will be far enough away from each other that it won't matter, allowing us to be more efficient with the available resources.

You can see an example of a cell diagram in illustration 1.6. Here, any of the hexagons with the same shade will be assigned the same frequencies, since they are not adjacent. Let's say the darkest shade gets channels 1–4, the middle shade gets 5–8, and the lightest shade gets 9–12. Rob is in a dark cell, and is on channel 2. Someone else in his cell may have 1, 3, or 4. Since Rachael is in a different dark cell, she can also be assigned channel 2, because she is far enough away. Ben, in a middle cell, cannot get channel 2, because he is too close to the dark cells. When we assign colors (frequencies) to cells, we often want to use the smallest number of colors possible. Finding that combination can actually be quite difficult, especially as the number of cells in the diagram gets very large.

So, what's in a cell? There are **base stations**, or BSs, and **mobile stations**, or MSs. The BSs in each cell are connected on one side to the wired core network and the Internet, and on the other to each mobile station assigned to it. An MS could be a cell phone, tablet, or any device that can transmit and receive according to a cellular standard.

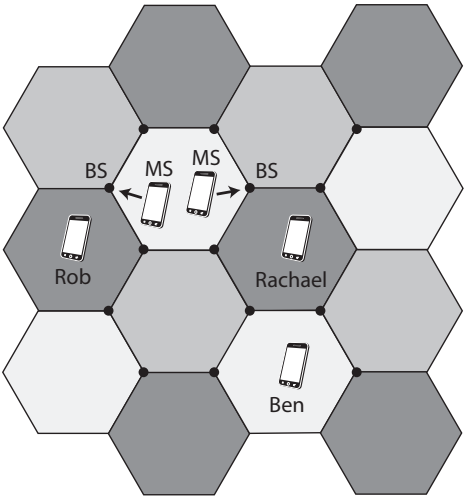

Illustration 1.6 This is a diagram of a cellular network. Each cell is a hexagon, with multiple mobile stations (MSs) and base stations (BSs). The shading of a cell indicates the frequency band that the cell is using. No two neighboring cells have the same shading, and so use different frequency bands to prevent interference.

Cells were first used in the advanced mobile phone system, which marked the beginning of **1G** technology in the United States. Under this system, the number of mobile subscribers skyrocketed. By the 1990s, there were 25 million cellular subscribers in the United States alone. This also meant that, due to the high usage rate and low capacity, analog just could not cut it anymore.

ENTER DIGITAL

With analog networks once again overcrowded, the United States and other countries began experimenting with an alternative: **digital** systems. An analog signal will be "digitized" by converting it into a sequence of **bits**, or 1s and 0s (see illustration 1.7).

Digital systems can offer enormous advantages in terms of capacity, as they enable use of two other multiple-access technologies that we discuss next. Prior to the late 1980s, the small-scale electronics necessary to realize these networks were just not yet available at a low enough cost.

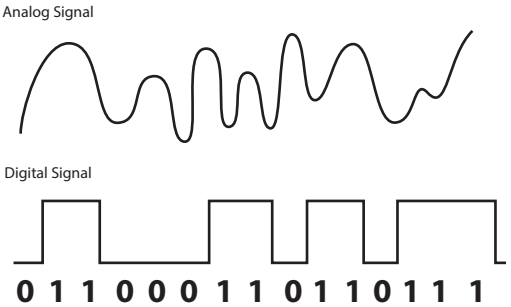

Illustration 1.7 An analog signal is one that changes continuously in time. In contrast, a digital signal is a series of bits, or 1s and 0s.

Sharing by Time (and Frequency)

The transfer from analog to digital cellular marked the migration from 1G to **2G**. The first 2G standard was the global system for mobile communications, or GSM, which began in 1982. By 1987, it was able to achieve three times the capacity of analog.

Digital coding allows us to compress multiple conversations into one band. So, even within one cell, we can have a number of people sharing the same frequency channel. We just have to add another one of our dimensions into the mix. The most obvious choice for the extra dimension is time.

In other words, multiple users can share the same frequency channel, but they all have to take turns. Each is allocated a different timeslot, in a scheme known as **time division multiple access**, or TDMA. You can see an example of TDMA in illustration 1.8.

GSM was adopted in much of Europe rather quickly, as the European Union favored the development of one common standard. GSM is still used in parts of the world today, operating mainly in the 900 and 1,800 MHz frequency bands. It drove down the cost of phones and marked the transition to handhelds that offer texting, gaming, and other entertainment.

Sharing by Codes

The story of 2G standard adoption in the United States is even more interesting. With the knowledge of increasing capacity demands, the US Cellular Telecommunications Industry Association posted a set of performance

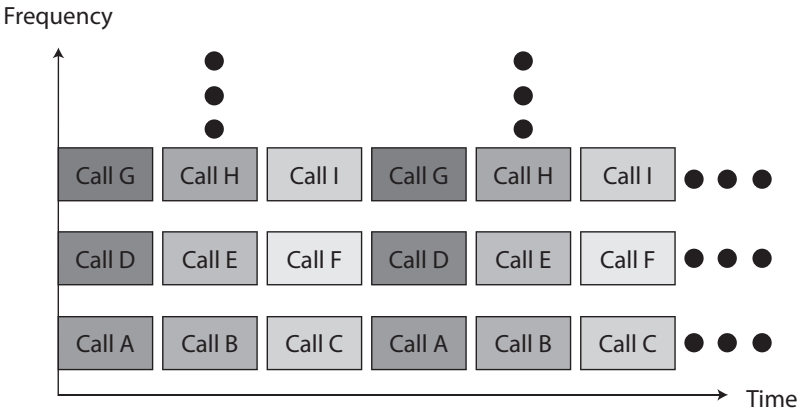

Illustration 1.8 With time division multiple access, a certain number of calls (in this case three) can share the same frequency channel. For instance, calls A, B, and C are assigned the same channel but are separated in time.

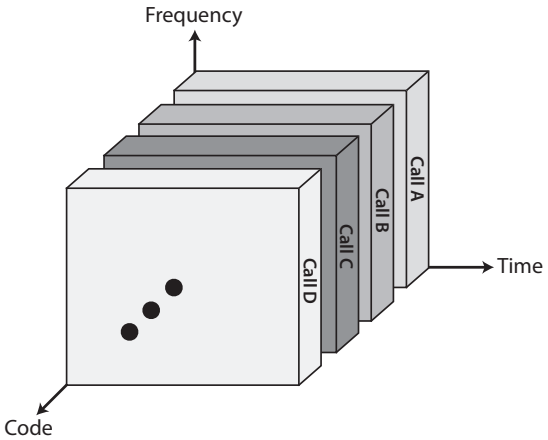

Illustration 1.9 With code division multiple access, calls are distinguished along the ``code'' dimension. All calls may operate over the same frequencies and at the same times, because each transmission in the network is assigned a unique code.

requirements in 1988 that the industry should aim to meet with the first digital cellular standard. The main one was a ten-fold improvement in capacity over legacy analog networks.

By this time, virtually every network operator and device manufacturer in the United States felt TMDA was the best way to go. But not the company

Qualcomm. They championed another technology, **code division multiple access**, or CDMA. With CDMA, users are separated over the "code" dimension and are undistinguished in both time and frequency, as shown in illustration 1.9. The best analogy to a code is perhaps language: it's as if you gave each link a different one to speak.

Each code is like a key. The sender locks the message, sends it out, and only gives the key to the receiver. The difficulty in designing such codes is that only one key should be able to "unlock" any given signal. If another receiver tries to descramble this message using its own key, it should appear as noise. A collection of codes possessing this property, where each one "cancels" the other, is referred to as a family of **orthogonal codes**. For more information on how CDMA works, check out Q1.3 on the book's website.

An early prediction claimed that CDMA could provide a capacity improvement of 40 times over that of legacy analog networks. Despite this claim, most of the engineers, manufacturers, and operators at the time resisted CDMA. For one, there had not yet been a demonstration of CDMA in a prototype cellular network.

In 1989, the Cellular Telecommunications Industry Association voted and approved TDMA as the first 2G digital standard in the United States. It would take more proofs-of-concept over the next 4 years before CDMA would be approved.

THE COCKTAIL PARTY ANALOGY

Here's a useful analogy that illustrates some of the technologies that we have introduced up to now. Imagine a cocktail party taking place in a large mansion with many rooms, in which there are many conversations occurring. Given that there are a lot of people at this party, if everyone was crammed into the same room and talking simultaneously, it would be difficult to hear what was going on in your own conversation. We leave it up to the host to determine the best way to manage this.

The host first decides that there can be one couple having a conversation in each room. Each couple has their room until their conversation is finished, so each person can speak at a comfortable volume, because voices will have attenuated by the time they reach the other rooms. But if we think of rooms as cells, this would be like only allowing one link per cell at a time.

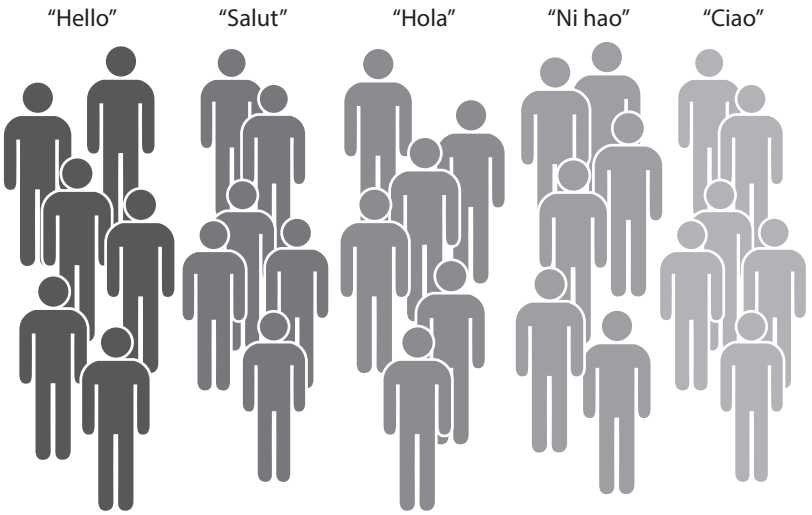

Illustration 1.10 With CDMA, each code is like a separate language. In the cocktail party analogy, multiple conversations can occur in a room if they use different languages. The issue then becomes controlling speaking volume levels.

Given that there are probably many more guests than rooms, this would not be desirable to many of the couples who were not assigned one.

To deal with this capacity issue, the host decides to allow many couples to share each room (i.e., many people per cell), telling each pair to speak at separate times. So, in any given room, the first pair has maybe 30 seconds, while everyone else remains silent, followed by the next group, and so on. Again, each person can speak as loud as she wants, because conversations will not overpower one another. This is an example of TDMA, where in each room every conversation is assigned a separate timeslot.

Rather than assigning timeslots, suppose the host asks each pair in a room to use a separate language. Then, everyone can speak simultaneously, since each pair is listening for one language in particular. This is an example of a CDMA system, where each language represents a different code (see illustration 1.10). But human languages were not designed as perfect codes. Further, volume control becomes an issue here, because everyone in a room can hear all the other conversations, regardless of what language is spoken. We would require some type of coordination, where individuals would adjust their volumes based on their distances from one another.

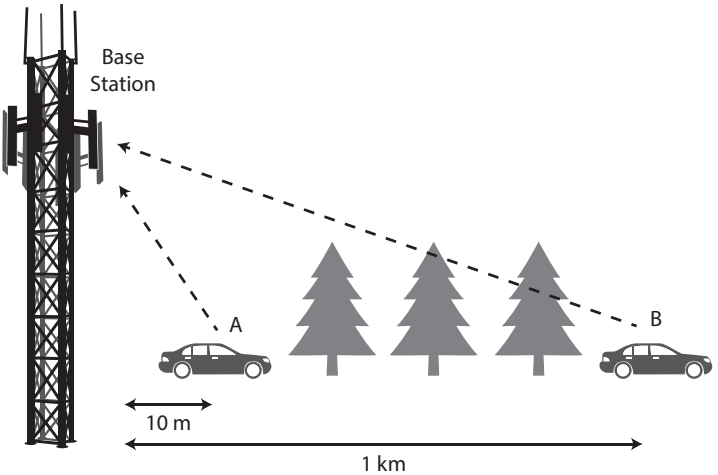

Illustration 1.11 The farther a transmitter is from its receiver, the higher the attenuation is, and the more objects there are to obstruct the path. Here, A has a short, clear path to the tower, while B has a long path that is obstructed by objects (e.g., trees).

CONTROLLING POWER LEVELS

CDMA had its share of complications. We now explore some of the major problems its advocates had to overcome in the early 1990s.

Near-Far Problem

Whenever signals are traveling at the same time, there is inevitable **interference**. The problem becomes worse when you take distance from the BS into consideration. How can someone who is a mile from the BS make a call without it being ruined by someone only a few meters from it? Not only does this person face more attenuation, but also there are likely more objects (like trees) to obstruct the signal's path. This leads to differing levels of **channel quality**, as in illustration 1.11.

What we're describing here is known as the **near-far problem**. To deal with it, our phones need some mechanism by which we can adjust our transmission powers to compensate for differences in channel quality, so that we can share the air effectively.

The initially proposed solution to mitigate this was the **transmission power control** (TPC) algorithm. It was an attempt to equalize received signal powers. The BS would measure what it was receiving from each

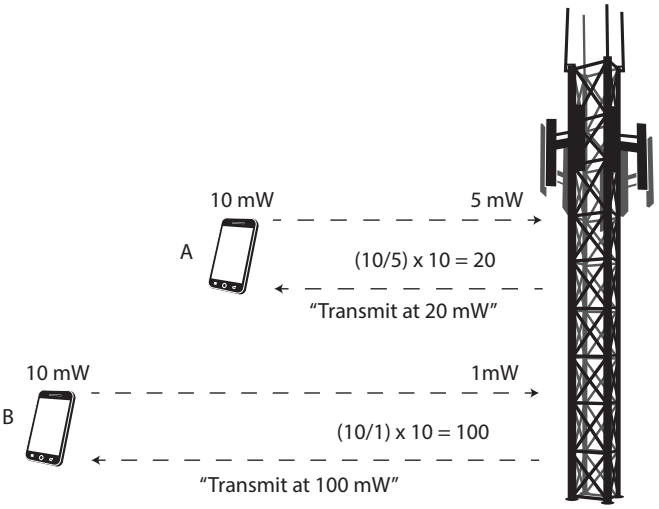

Illustration 1.12 Example of the transmission power control algorithm.

transmitter, compare this with the desired power, and send a feedback message to each device, telling it to adjust accordingly.

How is power measured at a receiver? The standard unit is the **watt** (W), which is the amount of energy transmitted per second. So, 5 W means that five units of energy are transferred per second. In this chapter, we will typically be dealing with power levels that are many fractions of a watt, either milliwatts (mW), which are thousandths of watts, or microwatts (μW), which are millionths.

Back to the TPC algorithm. Let's say the desired power level at the tower is 10 mW. Two cell phones, A and B, start sending at this power, and the BS receives them at 5 mW and 1 mW, respectively, as you can see in illustration 1.12. Channel degradation caused A's power to be halved and B's power to be reduced by a factor of 10. To "reverse" this, TPC asks the transmitters to send at twice and ten times their current transmission powers, respectively. That means A should send at 2×10 mW $=$ 20 mW, and B should send at 10×10 mW $=$ 100 mW.

More generally, the TPC algorithm is based on this equation:

$$\text{Next power} = \text{“The ratio”} \times \text{Current power}$$

where "The ratio" is the desired power (10 mW in this example) divided by the received power (5 and 1 mW).

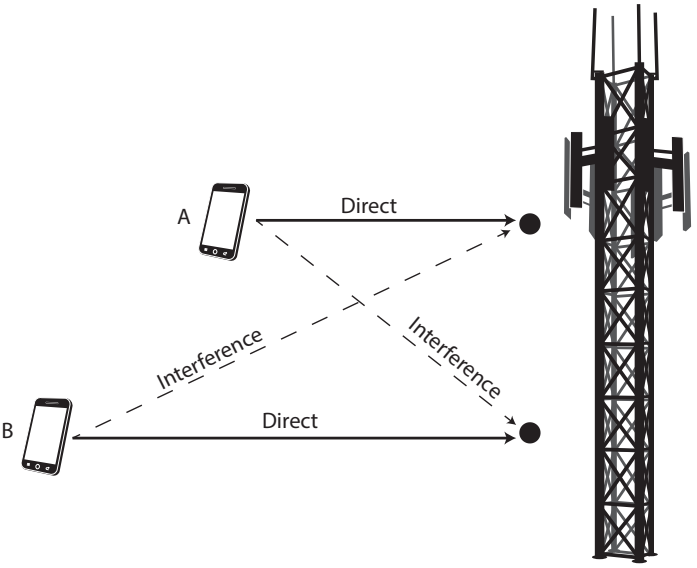

Illustration 1.13 Ideally, only the power from the transmitter of a link would be present at its receiver. But this is not the reality: here, some of A's transmission will be coupled into B's receiver, and vice versa.

Quality Is More Than Power

With TPC, the objective is to equalize the received signal power by boosting the transmissions accordingly. Is this enough to ensure "good reception," though? Not necessarily. The received signal is also going to be impacted by interference from other phones. You can see this in illustration 1.13. Even if the transmission power on link A is high, if the interference coming from other transmitters (link B) is also high, then A's received signals may still have low quality. This is our first glimpse of the impact of the network; in this case, the impact is on multiple users accessing the same communication medium.

For mobile communication, it is typically quality, rather than power, that we need to equalize. So, how do we determine the received signal quality on a link? We can view it as a combination of three factors:

1. The received signal power from the desired transmitter. This is the transmitter the receiver is trying to listen to.

2. The received signal power from the undesired (i.e., interfering) transmitters. These are the transmitters the receiver hopes to avoid.
3. The receiver noise, which is something inherent in every receiver.

Quality is measured as the good (item 1) divided by the bad (items 2 and 3). It's called the **signal-to-interference ratio** (SIR).

Preventing an Arms Race

Achieving target SIRs is more complicated than achieving target powers, because we cannot simply increase transmission powers to achieve the desired SIRs simultaneously. Increasing A's transmission power will increase A's SIR, but it will also cause the SIRs of other links to decrease. We would then have to increase the other transmitter powers to raise their SIRs, which would affect everyone else, causing them to increase, and so on. The result would be an inevitable "arms race," ending in whoever could transmit at the highest power winning. That wouldn't be a very effective way to share.

If each mobile device fixes a desired SIR, is it possible to find a set of transmission powers that will meet all of the targets simultaneously? The answer is yes, as long as the desired SIRs are **feasible**, or mutually compatible with one another. In other words, there cannot be a collective desire for unrealistically high SIRs.

The solution is known as **distributed power control**, or DPC. It works like this:

1. The devices each start with some initial transmission power.
2. The receiver measures the SIR for each transmitter.
3. Based upon the ratio between the target and the measured SIRs, each transmitter adjusts its power level.
4. Repeat steps 2 and 3 as needed.

DPC is an **iterative** algorithm that repeats over and over, unlike the single step, near-far TPC algorithm discussed earlier. Given that the target SIRs are feasible, it turns out that the DPC algorithm will **converge**, meaning that the SIRs will be met by power levels that will stop updating. In fact, these convergent power levels will also be **optimal** ones, in the sense that they will use the least amount of energy.

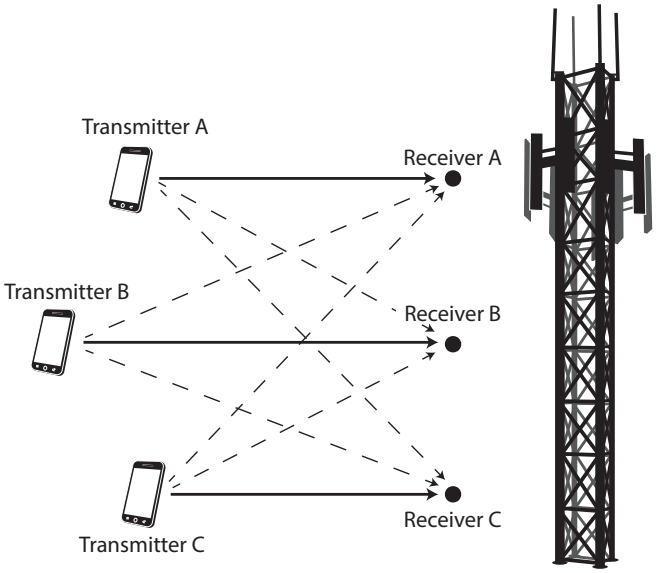

Illustration 1.14 Channels between three mobile stations and the base station. The solid lines denote direct channel gains, while the dashed lines represent the interference channel gains.

The Workings of Distributed Power Control

Think of three MSs, A, B, and C, in a single cell, as in illustration 1.14. Here, the solid lines represent the direct **channel gains** uplink for each of the transmitter-receiver pairs. Channel gain is a measure of how much the power is **amplified**, or, for all practical purposes, multiplied, from source to destination (as these are fractional quantities, the multiplication is really attenuation). Direct channel gain should be as high as possible, because it represents an intended conversation. In contrast, the dashed lines represent the interference channel gains, or by what factor the unintended signals will couple to each receiver. You can see the channel gains, the target SIRs, and receiver noise for our example in illustrations 1.15 and 1.16. (These numbers are meant for illustrative purposes, and don't represent actual numbers typically observed in cellular networks.)

Let's illustrate the DPC algorithm's computation for its first few steps. To update the transmission powers, the DPC algorithm employs an intuitive equation similar to the near-far TPC equation four pages earlier, but

Transmitter of link	Receiver of link A	B	C
A	0.9	0.1	0.2
B	0.1	0.8	0.2
C	0.2	0.1	0.9

Illustration 1.15 Channel gains for our distributed power control example.

Link	Target SIR	Noise (mW)
A	1.8	0.1
B	2.0	0.2
C	2.2	0.3

Illustration 1.16 Desired signal-to-interference ratio (SIR) and noise parameters in our distributed power control example.

involving SIRs rather than channel quality. For each transmitter, the update is

$$\text{Next power} = \text{“The ratio”} \times \text{Current power}$$

where “The ratio” is now the desired divided by the measured SIR.

This update is quite logical. If the measured SIR is lower than the desired SIR, then “The ratio” will be larger than 1. The transmission power will then increase, in an attempt to equalize them. On the contrary, if the measured SIR is higher than the desired one, “The ratio” will be lower than 1, and the transmission power will decrease. This transmitter can afford to use less power, and this action will help improve the SIRs for the other transmitters, too. Finally, if the measured and desired SIRs are the same, “The ratio” will be 1, and the transmission power will stay the same. There’s no need to change if the target is already being met.

Why is such an update necessary? In a cell, each device imposes **negative externality** on the others, by interfering with them. In other words, while achieving its own benefit, a device does some “damage” to the rest of the network as well. This update keeps the devices in check: whenever the SIR is higher than it needs to be, the power level is dropped, and whenever it’s too low, the power level is raised.

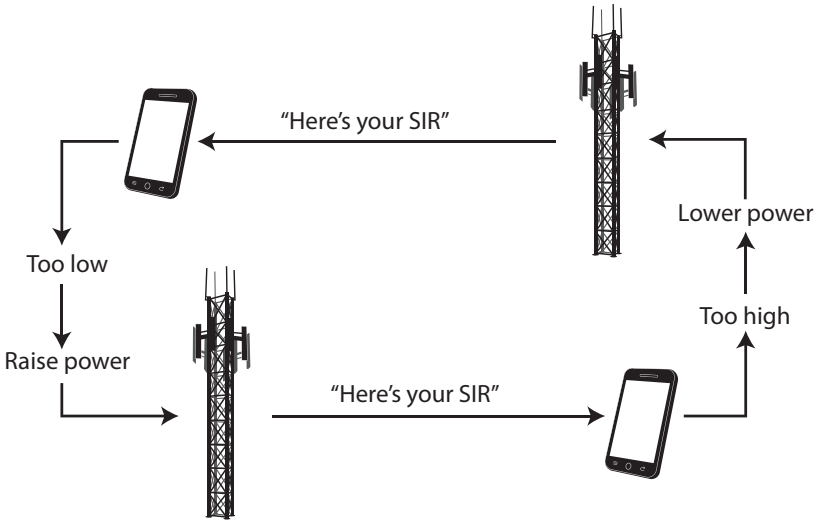

Illustration 1.17 The tower tells each device its current received signal-to-interference ratio (SIR), which serves as a negative feedback signal. Using this, each device can update its transmission power independently.

The process of "message passing" between the BS and the devices to correct for such deviations is an example of **negative feedback** (see illustration 1.17). It forces the transmitters to internalize their negative externalities (i.e., to pay for the interference they cause) by following the rules to make up for their added interference to the system.

The concepts of negative feedback and negative externality will come up repeatedly as we look at different networks in this book. More generally, negative feedback is a way of maintaining **equilibrium** in a system by checking for and counterbalancing fluctuations in the output. Positive feedback, where we move away from equilibria, will come up later, too.

Back to our example. Of all the quantities needed in the DPC equation, we know the desired SIR values (from illustration 1.16), and we can start off the current transmission powers at 2 mW. That leaves the measured SIRs. For each link, we can calculate this as

$$\text{Measured SIR} = \frac{\text{Signal}}{\text{Interference} + \text{Noise}}$$

How do we get "Signal," "Interference," and "Noise"? Let's start with Link A:

- *Signal*: This is the direct gain from Transmitter A to Receiver A, multiplied by the transmission power. From illustration 1.15, it is 0.9×2 mW = 1.8 mW.
- *Interference*: This is the sum of the indirect gains from other transmitters to Receiver A, multiplied by their transmission powers. In illustration 1.15, we look at the first column: from B to A, the gain is 0.1, and from C to A it is 0.2. That gives 0.1×2 mW + 0.2×2 mW = 0.6 mW.
- *Noise*: This is the receiver noise, given in illustration 1.16 to be 0.1 mW for Link A.

In reality, the receiver does not even need to do this multiplication and addition, because it can physically measure the SIR.

So, the measured SIR for Link A is

$$\frac{1.8}{0.6 + 0.1} = \frac{1.8}{0.7} = 2.57$$

What about Link B? From Transmitter B to Receiver B, the direct gain is 0.8, so the signal power is 0.8×2 mW = 1.6 mW. The indirect gains come from Transmitters A and C: from A to B, the gain is 0.1, and from C to B it is also 0.1. That means the interfering power is 0.1×2 mW+0.1×2 mW = 0.4 mW. Finally, the receiver noise for Link B is 0.2 mW. So, the measured SIR for Link B is 1.6/0.6 = 2.67.

Using the same procedure, you can find the SIR for Link C to be

$$\frac{0.9 \times 2}{0.2 \times 2 + 0.2 \times 2 + 0.3} = \frac{1.8}{1.1} = 1.64$$

Let's compare these values to the desired SIRs. For Link A, the desired value is 1.8, and so the measured SIR is too high by 2.57 − 1.8 = 0.77. Similarly, the measured SIR for Link B is too high by 2.67 − 2.0 = 0.67, while that of Link C is too low by 2.2 − 1.64 = 0.56.

We can now calculate the new power levels using the DPC equation. What are the ratios? We divide the target by the measured SIRs for each link to get these: 1.8/2.57 = 0.70, 2.0/2.67 = 0.75, and 2.2/1.64 = 1.34. As expected, the ratio is less than 1 for Links A and B (the measured value is too high), and greater than 1 for Link C (measured value is too low). With these values, the next power levels for A, B, and C are

$$0.70 \times 2 \text{ mW} = 1.40 \text{ mW}$$

$$0.75 \times 2 \text{ mW} = 1.50 \text{ mW}$$

$$1.34 \times 2 \text{ mW} = 2.68 \text{ mW}$$

respectively. Negative feedback has caused the power levels of A and B to decrease, and that of C to increase, as we expected.

What is the next step? To calculate the SIRs at these new power levels, we use the equation on page 20. And after that? We adjust the power levels based on the update equation on page 19. The calculations for future steps are carried out in the same manner. If you'd like to see more done by hand, check out Q1.4 on the book's website.

The transmission powers and SIR levels for 30 iterations of the DPC algorithm are graphed in illustration 1.18. After roughly 10 iterations, we can no longer see any noticeable changes in either quantity, indicating that the DPC algorithm has converged to an equilibrium. The measured SIRs have reached their target values of 1.8, 2.0, and 2.2, with power levels of 1.26, 1.31, and 1.99 mW, respectively.

Why does Link C have a much higher power level than the other two? It has the highest noise component of any receiver (0.3 mW), the highest interference gains from other links (both 0.2), and the highest target SIR value of any link (2.2). It needs a higher transmission power to overcome these disadvantages.

Why does the algorithm converge? Since the measured SIRs are the same as the target SIRs, "The ratio" in the update equation will be unity, so the power will not change any more. Negative feedback has brought the network to an equilibrium at which the devices are sharing effectively. It will stay this way until the network changes, like when a device's interference conditions change, a new device enters the cell, or an existing device leaves the cell.

In a real cell, there can be hundreds of phones, and as conversations are started and stopped and people move from location to location, you can imagine that the channel conditions and SIR values from link to link will change quite rapidly. As a result, it is required to have power control implemented up to 1,500 times each second. One benefit of the DPC algorithm is that each device really doesn't need any knowledge of how the other links are operating. To calculate its next power level, all it needs is its current transmission power, target SIR, and current measured SIR. These are all its own parameters, and it makes its current decision independently (e.g., there's no need for it to know the SIRs of any other links). This allows each device to perform its computations internally, without the need to share information with the others. In other words, the DPC is a completely

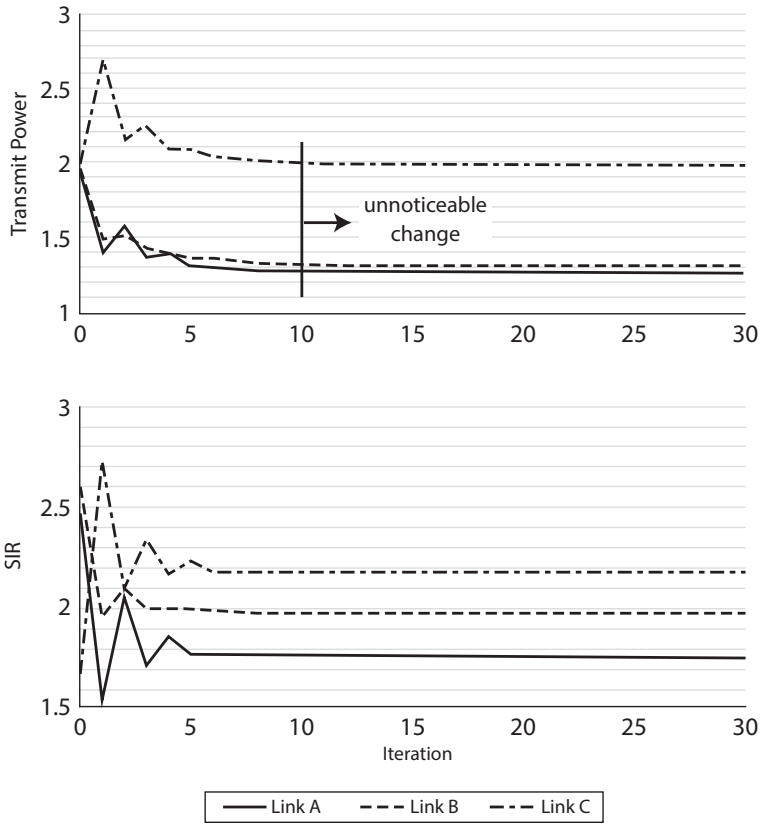

Illustration 1.18 Graphs of the transmit powers (top) and the signal-to-interference ratio (SIR) levels (bottom) for 30 iterations of the algorithm.

distributed algorithm (see illustration 1.19), as opposed to other, more **centralized** ones we will encounter (like Google's PageRank in chapter 5).

CDMA As a Standard

The DPC algorithm was proposed to deal with interference issues for CDMA. Even with this development, it took several large-scale demonstrations throughout the United States under realistic network conditions before major network operators would endorse it.

Finally, in 1993, CDMA was approved as a 2G cellular standard under IS-95, with the brand name cdmaOne. Three years later, the first large-scale commercial deployment of CDMA in the United States was made by

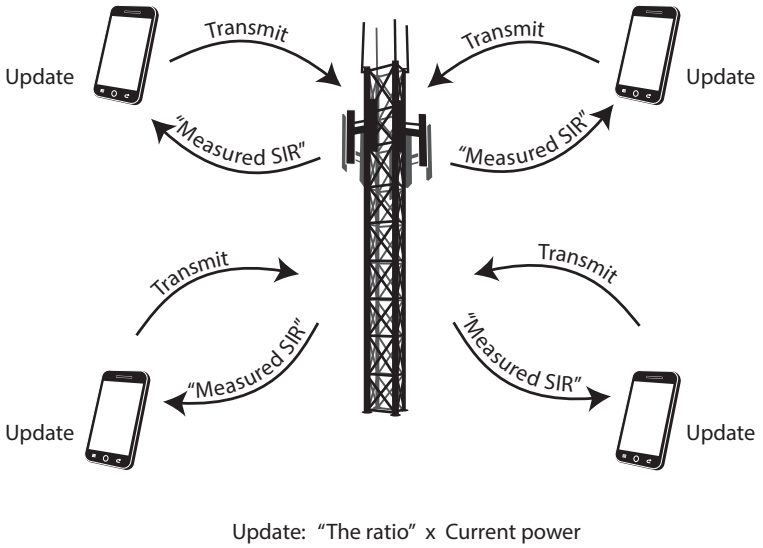

Illustration 1.19 Distributed power control is a fully distributed algorithm.

Sprint PCS. Though it has largely been upgraded to 3G standards, IS-95 and its immediate revisions are still used in parts of the world.

UPWARD AND ONWARD: 3G, 4G, AND BEYOND

The growth in the number of mobile subscriptions over the past few decades has been tremendous. The United States alone went from roughly 340,000 subscriptions in 1985 to 327,000,000 subscriptions 30 years later in 2015, an almost 1,000-fold increase. The mobile penetration rate in the United States has been larger than 100% since 2011.

Since the turn of the twenty-first century, 3G phones have gained momentum worldwide. The International Telecommunication Union's 3G specifications released in 2000 essentially called for cell phones to function as handheld computers: in addition to phone calls and texting, our phones are now capable of Internet access, video calls, and mobile TV. The two main 3G standards are UMTS, used primarily in Europe, Japan, and China, and CDMA2000, used especially in the United States and South Korea. Both technologies are based on CDMA and are typically deployed in the frequency range of 1.9–2.1 GHz.

Roughly 70% of the world's population was covered by at least one 3G network as of the beginning of 2015, up from 50% at the start of 2012. It is forecasted that more than four out of every five people (i.e., more than 80%) in the world will have access to 3G by 2020, making it almost ubiquitous. For information on the emergence of smartphones, check out Q1.5 on the book's website.

Since 1G networks were commercialized in the 1980s, a new generation of cellular network has emerged roughly every 10 years. Keeping on this track, the performance requirements for 4G were released in 2008. They specified higher speed requirements and capacities than the previous 3G specifications. Since then, the major standard that has emerged is long-term evolution, or **LTE**. Rather than using CDMA, LTE is based on a technology called **orthogonal frequency division multiplexing**, or OFDM.

The first LTE smartphones in the United States appeared in late 2011. At the beginning of 2015, roughly 25% of the world was covered by a 4G network, with this expected to increase to more than 60% by 2020. The predicted performance improvements over 3G are expected to attract 1 billion users by 2017. Though 4G network coverage worldwide was less than 3G as of 2016, it is also being deployed at a much more rapid pace.

The story of cellular evolution is a perfect example of how networks have struggled throughout the years to meet capacity demands of consumers. Different methods of sharing—frequency, time, and code-based wireless—have all been developed to do so. Even if we aren't aware of the processes involved, real-time updating and management of the power at which our calls are operating is essential to a cellular network's operation. Coming up with the right methods for sharing is difficult, yet very important.

Distributed power control illustrates several themes that are recurring in networking and in this book: negative feedback, system equilibrium, and distributed coordination. It also illustrates the following, major idea that we will see time and time again: allowing each user to make independent decisions driven by self-interest can be aggregated into a fair and efficient state across all users.

In the next chapter, we turn to WiFi, another type of wireless network. With WiFi comes a different flavor of sharing than cellular: rather than having stringent power control algorithms, WiFi relies on random access to manage interference among users in the same location.