Epilogue and Acknowledgements

First and foremost, I want to thank my parents.

Almost half a century ago, they brought me a book about basic electronics circuits from the local library, with a comment "you might be interested in this". After building the first, simple flip-flop circuit that blinked a tiny incandescent lamp, I was hooked, and that book sparked my lifelong interest in everything electronic.

What followed was a string of implementations, from illegal FM transmitters with three-kilometer range to detonation timers for dynamite-based "experiments" with enthusiastic friends: being a teenager in a small, rural village in the back country of Finland had some interesting side benefits at the time, and surprisingly, everyone got away alive, with eyes and fingers intact…

Naturally this is a credit to the public library system in Finland, which has been and still remains world class, nowadays even maybe a bit over the top, as it has been stretched way beyond the original idea of free access to books, newspapers and magazines. But the bottom line is that the variety of material available in the tiny, municipal library of Rasila, Ruokolahti, already during my youth in the 1970s, was simply amazing.

The library even had a special bus that took a selection of books to people in the more remote villages, and if you did not find the book you were looking for, you could order it to arrive on the following week's bus round.

A lot of discussion lately has been centered around the comparative quality of education systems around the world, and on the attacks on science that are happening due to the growing push of faith-based "facts" into the curriculum. This kind of activity has been prominent especially in parts of the United States, often backed by huge amounts of money from highly religious donors, many of which had actually made their fortunes by applying hard science to some business opportunities in the past, yet now appear to be actively trying to suppress scientific thinking for the future generations.

P. Launiainen, *A Brief History of Everything Wireless*,
https://doi.org/10.1007/978-3-319-78910-1

In stark contrast to this kind of development, the homogeneous, fact-based education that is the norm in the Nordic countries, is currently being slanted towards giving students the necessary base they need for critical thinking later in their lives, in an environment where fact and fiction are becoming increasingly hard to distinguish from each other.

The differences between societies are much deeper than just in education, though: access to good quality libraries that provide ample amounts of both fact and fiction for young minds has certainly in my case been one, very important aspect in the process of shaping my future, and often this goes hand in hand with the appreciation of learning in general.

I see the wireless, global access to the Internet as the next, great equalizer that allows practically anyone anywhere to tap into the existing and ever-expanding information pool at their own pace: the Internet is becoming the new, universal *Library of Alexandria*, for all mankind.

Information is power, and by empowering all citizens of the Earth with access to it, we can achieve great things together. Having the "Internet in your pocket", as enabled by wireless technologies, makes access to this new, ubiquitous virtual library amazingly easy.

As our continued prosperity is so deeply intertwined with the capabilities of these new technologies, it is hard to grasp the growing keenness to willingly throw away all that cold, hard science and start promoting baseless beliefs instead. This kind of development is most apparent in violent, religious groups like *ISIS* and *Boko Haram*, the latter of which loosely translates into "Western education is a sin". These groups despise education and the use of technology, yet, without the results of all the scientific research that has been behind even their most basic weaponry and communications systems, they would be fighting their wars with sticks and stones, and would be wiped out in an instant against a modern force based on advanced technology.

More alarmingly, the new trend of promoting "make believe" over science appears to be a growing rot even in the most advanced societies. It is raising its ugly head on issues like denying global warming and denouncing vaccinations. Both of these can mean fatal consequences to thousands, even millions of people in the longer term.

Another example of incoherent behavior comes from the silly group of people who have decided that the world is flat and spaceflight is just a big lie, propagated by all of the tens of space agencies in the world—according to these *Flat Earthers*, hordes of well-educated, trained professionals waste their lives by meticulously creating fake imagery from "space", to the count of thousands of pictures per day, even creating a non-stop fake video feed from the *International Space Station (ISS)*.

And if you point out the glaring inconsistencies in their Flat Earth view by referencing everyday services that they use, or even the simple fact that ships clearly appear to slide below the horizon as they sail away from their ports, it does not matter: any "scientific proof" is only cherry-picked when it can somehow be shoe-horned to match their world view—everything else is a deception of some kind.

Why is this kind of behavior so prevalent these days?

My assumption is that the vast improvements in our technical capabilities and the growing ease of use of new inventions have separated us too far from the actual roots of all the advancements benefiting our daily lives. The technology behind these new gadgets is superficially incomprehensible, and hence easily ends up lending artificial credibility to alternative explanations that are pure hogwash.

Add to that the fact that consumers are bombarded with movies that make almost everything look capable of breaking the rules of physics, and the boundaries of reality are blurred further: if Spiderman can walk on a wing of a flying jet plane and Wonder Woman does not even get a scratch while charging into relentless machine gun fire, why wouldn't the shape-shifting lizard people who *really* run the world use commercial jets to control us all with those heinous *chemtrails*, too?

The need to be "special" is a common trait for us humans, but unfortunately it often ends up being specialty of the Forrest Gump style, just without any of the kind-heartedness of the character.

Even worse, as peddling these "alternative truths" to a receptive audience can create sources of income, scrupulous players have stepped in en masse to gain financially from this new level of voluntary ignorance: create a wild claim that makes some atrocious event look like part of a conspiracy, and there's apparently enough audience to turn it into a profitable and repeatable enterprise.

In 2018, this has reached a point where even the current President of the United States is re-tweeting provably false material from highly questionable sources, while at the same time calling professional news organizations "fake news". These tweets then get "liked" instantly by millions, as well as parroted as truths by a major television network and a handful of so-called "real news" websites.

As a result, despite our almost unlimited access to knowledge, our collective ability for critical thinking appears to be rapidly diminishing. The ultimate paradox of the newly relinquished power of the Internet is that the people who have abandoned their trust in science are promoting their views by using the very same tools that were made possible by the science they are actively disparaging.

The *Library of Alexandria* was gradually destroyed over a long period of time by mobs commanded by ignorant rulers. Let's not allow an equivalence of that to happen to the Internet in the form of letting ourselves being overwhelmed by intentional falsehoods.

If the approach of basing important decisions on *faith* instead of *fact* becomes the accepted norm, it will eventually cripple the technological competitiveness of the country it is happening in.

In the words of the late Carl Sagan:

> We live in a society exquisitely dependent on science and technology, in which hardly anyone knows anything about science and technology.

With this book, I have tried to add my small share to the understanding of what makes our wireless world tick. It may often look like a cool magic trick, but it is still based on hard, cold physics.

As for digging up material for this book, the amount of information that is currently at everyone's fingertips through the Internet is absolutely mind-boggling: never before has verifying even the most obscure detail been as easy as today.

In my youth, being able to fix problems came only after a lot of experience: you became a *guru* that could solve things because you had spent countless days and nights relentlessly finding the solutions and workarounds to various problems.

Nowadays, the most useful capability is being able to search quickly and effectively, and increasingly, being able to rate the *quality* of the information that you come across.

In terms of most technological and scientific issues, the amount of false data on the Internet is not really significant, at least yet, but the more you shift towards historical and especially political subjects, the more you really need to weigh the *accuracy* of the information source you are using. Even with credible sources, the details sometimes do not match, and the further back in time you go, the larger the discrepancies appear to be.

Successfully evaluating the reputation of your source of information is a mandatory skill on the 21st century, and should be taught in schools all over the world, not only in the Nordic countries: as we have seen in 2016, absolute falsehoods that are spread via the Internet and other media can even turn elections around, with potentially very detrimental results to all of us.

For creating the basic building blocks for the new, virtual *Library of Alexandria*, I have to give credit to *Google*, despite the fact that I am not at all a fan of their corporate approach towards privacy: the way they hook you up into their network of information harvesting the moment you sign up to any of their services is nothing to be proud of.

Although many companies were pioneering the Internet search, it was *Google* that emerged as the best solution for indexing the information of the world, and they at least appear to be actively working on the immense task of ensuring that their search results have credibility.

I just hope that they would not have so obviously given up on their original mantra of *don't be evil*—privacy should be our birthright, not a commodity that is being exorbitantly collected, benefited from and sold off to the highest bidder.

On a personal level, learning to explain complex issues in simple and palatable terms, I remember the thorough discussions about minute technical details that I had in the early days of my computing career with Heimo Kouvo.

Those sessions were great private tutoring from an experienced guru at a time when you just could not find a solution to your problem through a simple web search. Not only did Heimo teach me a great deal, but also he taught me the way to address complex subjects so that they can be understood by a person with much less background in the subject matter at hand. That helped me immensely in many public presentations that I have ended up giving throughout my career.

Closer to the actual effort of putting a book like this together, my first great thank-you goes to my early round editor Grace Ross.

Not having English as a first language unavoidably means plenty of subtle errors and weird sentence structures, and Grace did a great job in cleaning up an early version of this manuscript. Her comments also helped me to improve those parts of the manuscript that were not clear enough for readers with less exposure to the digital world.

Identically, the final copyedit round done by Lyn Imeson was a true eye-opener. Her attention to detail was impeccable.

As for my experience in the wireless domain, my special thank-you goes to my former employer, *Nokia*, and especially to *Nokia's* former CTO, Yrjö Neuvo. He was the guiding light behind the technology that enabled the explosive success of *Nokia*, and he gave me the idea of moving to their Brazilian research institute, *Instituto Nokia de Tecnologia (INdT).*

I gained great, global insight into the wireless revolution during my time at *Nokia* and *INdT*, and being located in Brazil instead of the corporate headquarters in Finland also gave me a great, combined outsider/insider view into the decline of *Nokia's* groundbreaking handset division.

I'm very pleased that despite the missteps that doomed their handset division, *Nokia* has been able to continue as the leading wireless network infrastructure provider. Their research unit also seems to be actively looking into new, interesting areas, this time hopefully trying to exploit them better than during the handset days: so many of the great R&D projects inside *Nokia* never saw the light of day.

When this book started getting a more solid structure, the first readthrough of my rough manuscript was done by my friend and former colleague André Erthal.

André eats scientific literature for breakfast, and has the right kind of endless curiosity towards all new things in technology, as well as life in general, and his initial feedback helped me fix the direction of the story.

Early feedback came also from my lifelong friend and fellow computer guru, Yrjö Toiviainen. His "kyllä tästä vielä kirja tulee" (yeah, it will be a book one day) was the kind of traditional low-key Finnish thumbs up that helped me kick the can forward.

Some of the technical aspects regarding the cellular evolution were reviewed by my former colleagues at *Nokia* who prefer to remain anonymous: this does not stop me from expressing my greatest appreciation for their valuable insight. You know who you are…

It took about two years to get this story together, and making a list of every book, article, newsreel, video and website I came across during this time would be impractical. My idea was to make a collection of interesting and entertaining stories, not a scientific study.

To learn more about some of the issues touched in this book, here are some other publications that I would suggest as excellent further reading:

My personal favorite from all the persons mentioned in this book is by far Nikola Tesla.

All in all, Tesla remains one of the most versatile inventors of the 20th century, and his work in wireless technologies is just a small part of his lifetime achievements.

You can entertain yourself for days with the conspiracy stories around Tesla's inventions, or you can get a balanced and comprehensive picture of Tesla's life by reading ***Tesla: Man out of Time*** by Margaret Cheney.

Encryption has become a fundamental requirement for modern communications, and the use of encryption in human history started much earlier than most of us think. This fascinating part of our history is described in ***The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography*** by Simon Singh.

Last but not least, going back to the comments I made above about the alarming increase in nonsensical pseudo-science, I don't know of a better book that approaches this issue than ***The Demon-Haunted World: Science as a Candle in the Dark*** by Carl Sagan.

Denying or artificially diminishing the findings of science for whatever reason is ultimately detrimental to the goal that should be common to us all: making this planet a better place for all of us to live and learn.

I hope that in this book I have done my tiny bit in trying to give insight into these developments we are so utterly dependent on today. They may look simple on the outside, but this is just an illusion, created by the thousands and thousands of great minds that have spent sometimes their entire lives in improving the underlying technologies.

Therefore, my final thanks go to all of the named and unnamed heroes that have made all this possible, making this a better world for all of us.

TechTalk
Sparks and Waves

From a technology point of view, the very first radios were crude electro-mechanical devices, utilizing coils, transformers and the so-called *spark-gap* approach to generate radio waves.

Spark-gap is basically just a method of generating radio frequency noise through an actual stream of high-voltage sparks, and to make matters worse, the resulting generated frequency is not very precise.

Therefore, the transmissions would "bleed" to adjoining channels, making it difficult to use a receiver tuned on a different channel in the vicinity of a spark-gap transmitter. Two transmitters on adjoining channels would easily cause considerable interference to each other, and although adding more transmission power would expand the range, it would also expand the radius of the potential interference.

During the early years, this nasty side-effect was unfortunately often used to deliberately interfere with competitors' transmissions in various contests and demonstrations between rival equipment manufacturers, unnecessarily tarnishing the perceived usefulness of radio.

There was no *modulation* for the generated signal: the transmitter was turned at full power and back to zero power via the *telegraph key*, so effectively the telegraph key was used as a nimble on-off switch for the spark-gap transmitter, producing high-frequency pulses that were interpreted as *Morse code*.

In Morse code, each letter is represented by a set of short and long pulses, *dots* and *dashes*, and these combinations were optimized according to the statistical representation of letters in the English language. In this way, the overall time needed to transmit English text as electric pulses was minimized. For example, the most common letter is *E*, and it is presented as a single short pulse (dot) in Morse code, letter *I* got two short pulses (dots), letter *A* one short (dot) and one long (dash), and so on.

Naturally, text in German or Swahili does not share the same statistical distribution of letters, and hence would not be optimized as Morse code, but the same

P. Launiainen, *A Brief History of Everything Wireless*,
https://doi.org/10.1007/978-3-319-78910-1

letter coding is used globally anyway, with new, lengthy codes added for the local special letters.

This pulse-based communications method was already in use in the telegraph lines crisscrossing the world, so it was natural to use the same approach with radios as well. Finding persons that were fluent in Morse code was easy, thanks to the pool of existing telegraph operators.

As systems were purely electro-mechanical, it was difficult to reduce the amount of interference that was generated by the transmitter. Considerable improvement was only possible when vacuum tube technology become available years later.

Things were not much better on the receiving side:

In the early designs, a component called a *coherer* was used to detect the radio signal. This consisted of a tube with two electrodes, filled with metal filings, the conductivity of which changed in the presence of electromagnetic waves that entered the coherer via the connected antenna.

In this way, the presence of a *carrier wave*, the unmodulated high-frequency transmission, was detected, and the resulting change in electrical conductivity could be used to drive a ticker tape mechanism or feed a sound to headphones.

In order to reset the metal filings inside the coherer for another pulse, a mechanical tap on the coherer was needed. This "knock" returned the filings back to their low-conductance state. Therefore, the coherer was very susceptible to both mechanical and electric disturbances, and a lot of effort was put into the design of the necessary *decoherer* assemblies that were needed to reset the coherer after each pulse.

The first coherer was invented in 1890 by a French physicist Edouard Branly, and the inner workings of it and its successors were a kind of black art of the era: progress was made mainly through an endless loop of tweaking and retesting by the radio pioneers of the time.

In hindsight, having all this hodge-podge circuitry providing communications over thousands of kilometers was another amazing example of how human ingenuity relentlessly pushes the limits of available technology, however primitive it happens to be. Even today, the physics that make the coherer work are not fully understood, yet it did not stop the inventors utilizing it as a key component of a radio receiver, constantly testing new designs in order to improve the selectivity of the receivers.

The next major step was to resolve the problem of interference that was caused by the wide frequency spectrum of the carrier wave generated by the spark-gap transmitter technology. To cure this, the research focus moved to developing highly tuned, *continuous wave* transmissions. These systems would be able to generate a *sinusoidal wave*, a pure signal of a fixed frequency, greatly reducing the broad-spectrum radio noise created by the spark-gap.

Again, the earliest continuous wave attempts were electro-mechanical, using traditional alternating current generators, but with higher rotating speeds and denser coil structures that were able to generate radio frequency oscillations. The most notable of these systems were the *Alexanderson Alternators*, which were powerful enough to enable cross-Atlantic transmissions.

The generational breakthrough away from electro-mechanical systems came after the introduction of the *triode*, which was invented in 1907.

A triode, as the name implies, is a *vacuum tube* that has three embedded metallic terminals sealed in glass tube that has all air removed from it. Current flows through two of the terminals, the *cathode* and the *anode*, and a controlling voltage on the third one, the *grid*, can be used to increase or reduce this main current flow. The fundamental, groundbreaking feature of a triode is that a small variation of the control current can cause a much higher change in the flow of current from cathode to anode, and hence, for the first time, it was possible to amplify the weak signal that was received through the antenna.

The downside of a triode is the fact that in order to maintain a flow of current, you had to provide internal heating of the cathode through a special, glowing filament. This filament, just like in conventional light bulbs, has a limited life span, and when it breaks, the triode is no longer functional. For commercially produced triodes, the expected life span varied between 2,000 and 10,000 hours. Hence, in the worst case you had to replace them after three months of continuous use.

Heating the filament required considerable extra energy that was in no way participating in the current flow of the actual circuitry. This was problematic for any battery-operated devices. Finally, the glass tube that was used to contain the vacuum in a triode was very fragile against physical shocks.

Despite these shortcomings, the invention of the triode revolutionized radio technology by moving it into the *solid-state electronics* era: the equipment no longer needed bulky and complex moving parts, thus becoming more robust, smaller, and superior to the first-generation devices in terms of transmission signal quality and receiver selectivity. Large-scale production of vacuum tubes brought their price down rapidly. All in all, this fundamental change was behind the *broadcast revolution*, as discussed in Chapter 4: ***The Golden Age of Wireless***.

The second revolution in solid-state electronics started in 1947, with the invention of the *transistor*.

The original transistor worked on the same principle as a triode, offering a controlling electrode, the *base*, which could be used to govern a much stronger current flow between the *emitter* and the *collector*. But as transistors are based on materials called *semiconductors*, they do not require a glass tube with an internal vacuum, or a special, power-hungry heating filament, and are therefore much smaller and considerably more energy-efficient than triodes. Without a filament, the lifespan of transistors is in all practical terms unlimited, as long as their operating conditions stay within specifications.

The most common material to provide the semiconducting material for transistors is *silicon*, which is abundant and cheap, and unlike the case with vacuum tubes, the manufacturing of transistors can be easily scaled up to very large-scale production: as a result, the cost of an individual general-purpose transistor drops to a level of less than one cent when purchased in bulk.

The most fundamental benefit that arose from the improving manufacturing processes for transistors was the possibility to create complete *integrated circuits*, consisting first of thousands, and currently billions of interconnected transistors.

These *microchips* are behind our current, computer-driven world, and they come in all shapes, sizes and functionalities: if you go to the website of the microchip reseller company *Digikey,Inc.* and search for "integrated circuits", you get over 600,000 hits. Many of them perform just one specific electronic task, but one important subset of microchips, *microprocessors*, has made it possible to add generic computational functionality to almost every device today.

Transistor technology has gone through several generational improvements over the years, resulting in a stream of constantly cheaper and more energy-efficient components. Thanks to large-volume production methods, the price of the simplest microprocessors, consisting of hundreds of thousands of internal transistors, is below one dollar.

By altering the internal organization and wiring of transistors inside a microchip it is possible to provide almost limitless variation of highly specific functionality in a very small space. Good examples of this kind of specially designed component are the *graphics processors* that are able to provide us with an immersive, real-world-mimicking 3D-experience in the latest *Virtual Reality (VR)* systems.

At the other end of the scale are *microcontroller* chips, which combine the logic functionality of microprocessors with suitable interfacing circuitry that makes it possible to create versatile devices with minimum external components. These can be programmed to perform whatever functionality the user desires.

With microchips, the human imagination is the final limit, and our current, transistor-based technology appears to have lots of room for feature expansion in the coming decades.

Any kind of electronic component that can amplify signals can also be made to oscillate on a highly precise frequency through suitable feedback circuitry. Therefore, the introduction of triodes was a death knell for all those bulky electro-mechanical components that had earlier been used for radio frequency signal generation.

It was not only possible to produce pure, continuous sinusoidal waves, but the triode also enabled the generation of much higher frequencies than had been possible with its electro-mechanical predecessors. Hence, more channels opened up for communications.

Pure sinusoidal waves, as well as precise control of the transmission frequency, resulted in a huge reduction of interference between channels. These were both superior features compared with the earlier spark-gap transmitters, and this development only got more energy-efficient and reliable with the introduction of transistors.

On the receiving side, a major breakthrough came in the form of *super-heterodyne* technology, an implementation by Edwin Armstrong in 1918, which was based on the theoretical *Heterodyne Principle*, patented by Reginald Fessenden thirteen years earlier.

Directly amplifying the high-frequency signals that are received by an antenna is difficult because the higher the frequency, the lower the gain that can be achieved, even by active electronic components.

In a super-heterodyne receiver, the weak high-frequency signal from the antenna is mixed with a stable, low-power frequency from a local oscillator, and this mix then creates an intermediate, much lower frequency that can be amplified easily. This ingenious way of amplifying the intermediate, low-frequency signal instead of the directly received high-frequency signal results in a superior reception sensitivity, which in turn means that radio stations that reside much farther away can now be received, without the need of increasing their transmission power.

Along with the possibility of selecting the frequency with precision, the new transmitters enabled the use of *Amplitude Modulation (AM)*: instead of having the transmitter just on or off, as controlled by the telegraph key, it was possible to adjust the transmission power dynamically between minimum and maximum. Thus, by using an audio signal as the driver for the Amplitude Modulated transmitter, the spoken word could be used for communications instead of the relatively slow Morse code.

Although voice soon took over the majority of radio communications, Morse code remained relevant all the way to 1999, when it ceased to be used as an international standard for long-distance maritime communications.

The remaining notable uses for Morse code are *Amateur Radio* communications, and the identification of *VHF Omni Range (VOR)* radio navigation aids in aviation, which were discussed in Chapter 6: ***Highways in the Sky***.

Modern aviation equipment automatically turns the received three-letter Morse code into text on the navigation equipment's screen of the airplane, relieving the pilot from doing the station's Morse code identifier verification by ear, but the corresponding dots and dashes for VOR station identifiers are still printed on aviation charts.

Despite this automation, the continuing use of Morse code in VOR stations makes it the oldest electronic encoding system still in use in the world.

Amplitude Modulation has an inherent limitation: when the modulating signal level is low, the resulting transmission power is low as well, and the lower the received signal, the more any interference on the same channel reduces the quality of the received signal.

As Amplitude Modulation was also used for analog television transmissions, the interference would be most visible when no picture was transmitted and the television showed just a black screen—a passing motorbike with a poorly shielded ignition circuit would cause random white dots to be visible on the black image.

The logical step to circumvent this problem came by inverting the modulation signal: when the screen was all-black, the transmission was at its strongest level, thus masking any low-level interference that would otherwise be very visible on the black background.

For audio transmissions, a far better way was devised to circumvent this—*Frequency Modulation (FM)*, which was patented in 1933, allows the transmitter to work with continuous maximum power, while the modulating signal is used to cause slight variation to the actual transmitted frequency instead of its amplitude. Hence the received signal is always at maximum level, whatever the momentary modulation level may be, and thus the reception is much less susceptible to spurious interference. Also, the potential transmitting range is at its practical maximum.

Frequency-modulated signal therefore has a center frequency around which the frequency keeps on shifting, and the width of this shift depends on the depth of the modulation. This constant shifting around the center frequency is then detected in the receiver and turned into the original modulating signal.

As explained in TechTalk ***There is No Free Lunch***, the wider the frequency spectrum is in the modulating signal, the wider will be the required bandwidth for the transmission channel. Therefore, Frequency Modulation is only practical if the transmission frequency is at least several tens of megahertz (MHz).

But the improved sound quality of FM, in combination with the use of the *Very High Frequency (VHF)* band, was the final improvement that elevated broadcast radio to its current, almost universally available status in the familiar 87.5–108 MHz broadcast band. Only a handful of exceptions exists to this universal standard, most notably Japan with its 76–95 MHz allocation. Some of the former Soviet Bloc countries also used to have a different allocation of 65.9–74 MHz, but most of them have now shifted to the 87.5–108 MHz model.

AM transmissions remain standard for lower frequency bands, on which the signal can traverse long, even intercontinental distances. This is due to the fact that the lower-frequency signals follow the curvature of the Earth as a *groundwave* and also via reflection back from the *ionosphere*, an electrically charged region of the upper atmosphere.

This effect is reduced when higher frequencies are used, and for FM transmissions in normal conditions it has entirely vanished: the transmitted signal moves directly away from the transmitter antenna in a straight line, without reflecting back from the ionosphere and without having a groundwave effect. If the FM receiver is far enough from the FM transmitter so that there is no longer a line-of-sight access to the antenna due to the curvature of the Earth, the reception quality quickly deteriorates, because the signal just passes overhead into space. The relative altitude of the receiver has a bigger effect on the receiving distance than the huge transmitting power that is invariably in use in most broadcasting FM transmitters: if you check your FM radio on a plane flying at 10,000 feet or higher over a densely populated area, your dial is suddenly full of faraway stations—another simple proof that the Earth is not flat.

This direct-line propagation of FM signals is the reason why the antenna towers are usually very tall, or built on hilltops. Depending of the height of the transmission antenna, you can normally only hear FM broadcasts to about 200 kilometer distance, whereas due to the groundwave and ionospheric reflection, AM transmissions can be received from stations that are thousands of kilometers away.

Long-distance high-frequency reception of both FM and television signals can happen only in rare, reflective atmospheric conditions, often related to large high-pressure areas, during which the *troposphere*, the part of Earth's atmosphere where all weather occurs, reflects the signals over longer distances, or when the ionosphere is agitated by higher levels of *solar wind*. Both of these create short-to-medium term reflective situations, during which even the higher frequencies can get mirrored back, reaching receivers that are hundreds or even thousands of kilometers beyond the horizon. As the FM radio network's channel allocation was planned with the line-of-sight limitation in mind, this kind of "radio weather" may result in severe interference in situations where the receiver resides far from the desired transmitter—the reflected signal from another transmitter on the same channel may be strong enough even to mask the other, desired transmission completely.

As this kind of special reflectivity situation varies continuously, and due to the way FM receivers try to lock onto a single signal when two signals appear on the same channel, you may experience a situation where your FM receiver seems to have a mind of its own, flipping continuously between two transmissions.

A very rare occurrence of reflection may also happen if a *meteor* enters the atmosphere in just the right location between you and a faraway transmitter: the *ionized air* caused by the burning meteor may reflect a strong–enough signal that masks a weak FM transmission for just a couple of seconds, and again, it feels like your radio just flipped channels for a brief moment on its own. I personally experienced this kind of *meteor scatter* phenomenon while "hunting" for remote television broadcasts as a kid—I had the television on an unused channel in Finland, hoping to catch some long-range reflections caused by a hot, summery high pressure, and suddenly got a picture-perfect television test picture from Denmark, only for it to vanish after 2–3 seconds, never to be seen again.

As I explain in ***Epilogue and Acknowledgements***, growing up in rural Finland made you quite inventive in finding ways to spend your spare time…

For radio transmissions over very long distances, another effective variation of Amplitude Modulation, *Single-sideband Modulation (SSB)*, can be applied. It was patented in 1915 by John Renshaw Carson, but only saw its first commercial use twelve years later as the modulation method for the transatlantic public radiotelephone circuit between New York and London.

SSB has better transmission efficiency than normal AM transmission, thus enabling a longer reception range with equal transmission power.

Amplitude and Frequency Modulation are both well suited to be modulated by analog signals, but when your signal is digital and consists of just a stream of ones and zeros, as discussed in TechTalk ***Size Matters***, these original modulation modes are not optimal in terms of both spectral efficiency and susceptibility against interference.

To squeeze the very maximum bandwidth from the available channel, not only the signal frequency and amplitude, but also the phase of the signal is modulated.

The most commonly used digital modulation scheme is called *Orthogonal Frequency Division Multiplexing (OFDM)*. Going into details of OFDM is beyond the scope of this book, so let's just say that it is very efficient in terms of its use of the available spectrum, as well as relatively immune against various common forms of radio interference, like *multipath propagation interference* that is caused by reflections from large objects or mountains.

Hence OFDM is the modulation used in many of the wireless solutions discussed in this book, including Wi-Fi, WiMAX and 4G LTE.

The slice of the *electromagnetic spectrum* that in practice is exploited for radio, the *radio spectrum*, covers frequencies from *Very Low Frequencies* (*VLF*, down to about 3 kHz) to *Extremely High Frequencies* (*EHF*, about 300 GHz).

As the frequency change is a linear progression, the boundaries of the various frequency bands are not absolute: the change from one group to another is gradual, but the commonly mentioned boundaries are as follows:

Very Low Frequency (VLF), starting from 3 kHz
Low Frequency (LF), starting from 30 kHz
Medium Frequency (MF), starting from 300 kHz
High Frequency (HF), starting from 3 MHz
Very High Frequency (VHF), starting from 30 MHz
Ultra High Frequency (UHF), starting from 300 MHz
Super High Frequency (SHF), starting from 3 GHz
Extremely High Frequency (EHF), starting from 30 GHz

If we go below 3 kHz, there is very little room for meaningful modulation, for reasons discussed in TechTalk ***There is No Free Lunch***.

The low end of the spectrum has some very special uses, like communicating with submarines, as the VLF frequencies can be received at depths of hundreds of meters, requiring an antenna that can be kilometers long, towed by the submarine.

At the top end of the range, at *Super High Frequencies (SHF)*, we first hit microwaves, which are utilized by a lot of our current technologies, from radar to Wi-Fi to microwave ovens. Thereafter, when passing beyond the upper microwave limit of 300 GHz, we reach the very top portion of the electromagnetic spectrum, where the propagation properties of the electromagnetic waves change radically: the first part is infrared radiation, followed by visible and ultraviolet light, and finally reaching deeply penetrating X-rays and extremely powerful gamma rays.

All in all, this offers us a wide selection of useful frequencies to exploit, with wildly differing properties: lower frequencies can be used for communication between continents and even underwater, whereas high frequencies can pack enormous amount of information, thanks to the wide modulation they can support.

Gamma rays are the result of nuclear reactions, and the most energetic gamma rays can reach us from the other side of the known Universe, created by *supernova* explosions or colliding *black holes*, and they will pass through the Earth more easily than a ray of light passes through clear glass.

The high end of ultraviolet radiation, together with X-rays and gamma rays, present the part of the electromagnetic spectrum that contains *ionizing radiation*, and hence these high-energy rays are harmful to any living tissue. As discussed in Chapter 11: ***Home Sweet Home***, microwaves are **not** ionizing radiation, they only generate highly localized heat.

As a summary, the solid-state electronics revolution, with its precisely tuned, high-frequency transmitters and selective super-heterodyne receivers, was the most fundamental breakthrough on the path towards our modern wireless society.

Thanks to this improved technology, legislation was drawn up in 1927 that considerably restricted the use of spark-gap transmitters: due to the advances in technology, the pioneering equipment had suddenly turned into a major source of interference for the receivers of the triode era.

After the invention of the transistor, the fundamentals of radio receivers and transmitters remained the same, but the advances in energy efficiency, reliability and size have kept progress in high gear. And the recent microchip revolution is finally bringing entirely new approaches to this technology, as explained in TechTalk ***The Holy Grail***.

Radio waves, by their very nature, are a shared and limited resource, and as they do not adhere to national borders, close collaboration is needed to avoid interference between different users. The *International Telecommunication Union (ITU)* acts as the top-level allocation entity for resolving any potential issues in a global scale, as in the discussion about the allocation of geostationary satellite frequencies in Chapter 7: ***Traffic Jam Over the Equator***.

On top of this international collaboration, practically every country has their own regulatory entities that govern the allocation and use of the various frequency bands within the borders of the country, and due to historical reasons, these allocations often differ between geographical areas, practical examples of which were discussed in Chapter 9: ***The American Way***.

As the technology developed, the usage of some bands became obsolete and these bands were reallocated, like what happened with the release of some of the former television frequencies, as described in Chapter 5: ***Mesmerized by the Moving Image***.

TechTalk
Size Matters

All of us are familiar with the concept of analog and digital wristwatches: one has linearly sliding pointers for time, another has distinct numbers that flip in succession.

While these two concepts essentially present exactly the same thing, they offer great examples of what *analog* and *digital* actually mean.

In the analog world, things happen in linear fashion, whereas in the digital world, things happen in fixed steps, and unless we are observing the world around us on an *atomic* or *quantum* level, everything around us seems to be analog: however precisely we measure some natural event, it seems to move from A to B so that every possible point between A and B is touched along the way—the sunlight dies away without any noticeable steps at nightfall and the stars appear linearly brighter and brighter as the sky around them darkens. Even the lightning bolt, when recorded with a high enough frame rate, has a clear start, growth, decay and end.

But when it comes to computers, the situation is very different:

Computers store information internally in *bits*, which only have two states—either zero or one. This is like a light switch: it is either on or off; there is no concept of *half* in computers.

If we were limited to choose something like the volume level for our music system only as *on* or *off*, that would be very annoying indeed, as anyone with teenagers at home has surely experienced. To circumvent this, computers handle any arbitrary data as a collection of bits, grouped into larger units for the convenience of transferring and processing them. What such a collection of bits actually means is totally up to the implementation: together they can present a digitized picture, the event history of a home alarm system, your favorite piece of music, a copy of this book. Anything.

This description pretty much sets the scene for all *digital data*: everything is just a smaller or larger collection of bits, and what these bits represent is totally up to the implementation. And the implementation is just a gentlemen's agreement between

P. Launiainen, *A Brief History of Everything Wireless*,
https://doi.org/10.1007/978-3-319-78910-1

man and a machine: we have simply defined that data in a certain format means a certain thing.

In order to bridge our analog world to computers, we need to convert the linearly varying values into a stream of fixed numbers, and as long as we do it with small enough steps and repeat this process quickly enough, our limited senses can't tell the difference.

For example, our standard *Compact Disc Digital Audio (CDDA)* has a resolution of 65,536 steps for the possible volume level on each stereo channel at any moment in time of a piece of music. A value of zero means no sound, whereas the value of 65,535 indicates the maximum theoretical level, which with CDDA is about 96 dB.

Without going into too much detail, let's just note that the 96 dB range, if fully utilized, is good enough for any projected use case for CDDA, considering that a quiet room has an ambient noise level of about 30 dB, and the human pain threshold is around 130–140 dB. Hence if you set up your audio equipment amplification so that even the very lowest sounds levels on your *Compact Disc (CD)* are audible over the ambient noise, the highest levels will just about start hurting your ears.

To convert the original audio into digital audio in CDDA format, the signal on both stereo channels is sliced, or *sampled*, into these 65,536 steps 44,100 times per second. This is called the *sampling frequency*, and the magic number of 44,100 was chosen due to a *Nyquist Theorem*, which states that in order to be able to reproduce the original waveform, a sampling frequency that is at least twice the maximum frequency in the source signal is needed.

And as the top frequency range of human ear is around 20,000 Hz, sampling music with 44,100 samples per second was deemed sufficient by the engineers of *Philips* and *Sony*, the inventors of Compact Disc.

Conveniently, two bytes, or 16 bits, is enough to represent 65,536 distinct values, and hence each sample of a stereophonic signal consumes four bytes. With 44,100 samples per second, storing one second of stereophonic CDDA audio requires therefore 176,400 bytes.

Although the digital audio volume always changes in steps, the steps are so small and happen so quickly that our ears can't tell the steps apart, despite the claims of "golden-eared" audio purists that try to belittle the CDDA quality.

In real life, our analog ears can be fooled with much less quality, as the proliferation of *streaming music services* that rely on *lossy compression algorithms* has proved. For example, both *Spotify* and *Apple Music* rely on lossy compression, meaning that all parts of the sound that are considered non-essential to the perception capabilities of the human ear have been removed. Therefore, if the streamed digital data is compared with the originally sampled signal on a purely numerical level, they have very little in common. But thanks to the *psychoacoustic* limitations of our brains, the more than one hundred million satisfied users of these streaming services could not care less about this very real inherent discrepancy.

These lossy audio compression algorithms first became in wider use through the *MPEG-1 Audio Layer III (MP3)* standard, mainly developed by the Fraunhofer Institute in Germany. Other, more recent standards are *Advanced Audio Coding*

(AAC) and *Ogg Vorbis*. The theoretical work in the area of psychoacoustics is nothing new: it dates back all the way to the end of the 19th century.

The process of taking an analog signal, like our voice while we speak into our smartphone or video from a television camera, and turning it into a collection of bits that computers can handle, is called *digitizing*, and the computers around us do this all the time: they can't handle analog signals, so they have no other choice than to employ electronic circuits called *analog-to-digital converters* to do the job.

On the other hand, when computers are required to restore the digitized data back to analog form for human consumption, they use *digital-to-analog converters* to reverse the process.

The additional benefit of swapping analog data into numbers comes from the fact that these sets of numbers can be further processed in a way that reduces the size needed to save and transmit them. As mentioned above, you can greatly reduce the size of digitized information by the use of lossy compression algorithms, but in cases where you have to be able to *exactly* reproduce the original information, several *lossless compression algorithms* can also be used.

If, for example, the digitized data happens to have 5,000 zeros in a row, instead of saving a zero into 5,000 consecutive memory locations, we can define a scheme that allows us to say in shorthand format that "the next 5,000 numbers will be zeros". Depending on the way we define such a lossless compression scheme, the space needed to save these 5,000 numbers can be cut to under 10% of the original size.

We therefore save a lot of storage space or transmission bandwidth, but the payback comes when we access the compressed data, and we have to use computing power to decompress it.

Digital Signal Processors (DSPs) can do such compression on the fly, and they are the essential ingredients for all digital communications these days. The control programs for these DSPs are called *codecs*, and they were the behind the technological leap that made it possible to jump from analog first-generation cellular networks to digital second-generation networks.

As discussed in Chapter 8: ***The Hockey Stick Years***, during the transition from first-generation to second-generation cellular networks, we were able to squeeze three digital voice channels into the same amount of precious radio spectrum that formerly could handle just one analog one, thanks to the real-time conversion capability of codecs. Therefore, with all other aspects being the same, three times more users could be handled in the same amount of bandwidth.

There are many other cases where digitalization and subsequent compression create huge savings in terms of required memory space, but they are beyond the scope of this book.

But let's go back to the size issue:

As mentioned, a single bit is the smallest piece of information a computer handles, but for practical purposes, computer memory is designed to handle data in bigger chunks, most notably as *bytes*, which consist of eight bits.

As each of the eight bits in a byte can be either one or zero, a little experimentation with paper and pen would reveal that one byte can have 256 distinct values as the individual bits switch to ones and zeros in all possible combinations over the eight available bits. 256 variations are enough to cover the *American Standard Code for Information Interchange (ASCII)* encoding of all letters, numbers and special characters in text written in English, actually leaving 128 possible bit combinations in a byte still unused.

The problem comes with special letters in other alphabets:

Thanks to all the different ways us humans have invented for representing written text in different languages around the world, the number of possible combinations does not fit in the available extra 128 variations inside a single byte.

Hence, the Internet as we know it most commonly uses a text encoding standard called *UTF-8* for the presentation of the pages that we browse daily. Just to prove that real engineers named this format, UTF-8 is shorthand for *Universal Coded Character Set Transformation Format–8-bit*, and it is a dynamic-length character representation: the lowest 128 variations match the original ASCII table, but on top of that, a single character can require anything from one to five bytes to be identified.

For example, the euro currency character, "€", requires three bytes in UTF-8, with values 226, 130 and 172, whereas the letter "a" needs only one byte with a value of 97, matching its value in the ASCII standard.

Therefore, in order to represent letters "a€" in the UTF-8 format, we need space for four bytes, the numerical values of which are 97, 226, 130 and 172.

This combination of consecutive byte values only means "a€" when the content is handled as UTF-8 text, so the computer has to know somehow that we are dealing with text, and that the text should be interpreted in UTF-8 format.

Like this example, any data that a computer is using is eventually broken into chunks of bytes, and when these chunks are stored, there is always some agreed-upon mechanism that tells the computer what the current bunch of bytes actually represent: a text file, a picture, a video, a spreadsheet etc. The way the computer is made aware of this is totally implementation-dependent, and can be as simple as a certain suffix in the file name, or a set of specific bytes in the beginning of the file itself.

Therefore, in essence, all data in computers is a bunch of zeros and ones, usually grouped into sets of eight bits stored in a byte, which is the smallest individual unit that the computer accesses internally.

This book, with all the formatting and other information needed to represent a text file for the *LibreOffice* program that was mainly used to write the original manuscript, fits in about 900,000 bytes.

Playing with such big numbers is hard, so we use suitable prefixes to ease the number game:

900,000 bytes can be presented as 900 kilobytes, "kilo" meaning a thousand of something, as in *kilogram*, which is thousand *grams*.

900 kilobytes is usually written as 900 kB.

The next common multiplier is the *megabyte (MB)*, which is one million bytes. Hence this book with its 900,000 bytes can be said to contain 0.9 MB of data.

When you buy a common *USB Memory Card*, or a *Micro SD card*, which are both pretty much the same thing in terms of their internal storage technology, they are currently sized in *gigabytes.*

One gigabyte is one thousand megabytes, or if represented in short form, 1 GB = 1,000 MB.

For example, at the time of writing this, you could buy a 16 GB memory card for under five dollars, and fit roughly 16,000 copies of this book on it.

So, 16 GB looks big enough to last a lifetime?

Unfortunately, no.

Text is very compact in terms of its memory requirements, but things look different with other types of data: snap a picture with your reasonably good smartphone, and depending on the complexity of the picture, you need 2–5 MB to represent just one image. The actual size varies due to the lossy compression that is used to minimize the amount of space required, and the level of compression that can be applied depends on the structure of the picture.

Hence you can save about 4,000 pictures on a 16 GB memory card.

That still sounds like a big capacity for an average user, but a quick count in my own image repository shows that I have about 16,000 pictures archived so far.

All of them would still fit on four 16 GB devices, or a single 64 GB one, which is a size that is also readily available today for roughly 20 dollars.

But if you turn on the video feature of your smartphone, the requirements for data storage jump again: for one minute of *High Definition (HD)* quality video, you need roughly 100 MB of memory.

Suddenly your 16 GB memory card can only hold about 15 minutes of video.

Not quite enough for a lifetime of video memories.

Add to that new, ultra-high resolution *4K* video and whatever 360-degree *Virtual Reality (VR)* video formats will be used in the future, and it is obvious that storage requirements for data never end.

Luckily enough, the price of memory is going down all the time, while the available capacity is going up: it was not so long ago that a 64 GB memory card cost about 100 dollars, and if you go back fifteen years, having 64 GB on a memory card was pure Science Fiction.

For larger capacities, you need to shift from memory cards to other types of memory, like *hard drives*, and at the time of writing, it is possible to buy hard drives with 10 terabyte (TB) capacity, where 1 TB = 1,000 GB.

And both the available size of these various memory devices, along with the new ways to fill up that space keeps on progressing. Relentlessly.

When it comes to the memory that is in use by the computer or smartphone to accommodate the actual processing of data by the applications used, we are dealing with smaller numbers.

Although the requirements for smooth operation tend to go up continuously as our expectations for our portable devices grow, we can currently manage by having between 1 and 16 gigabytes of device memory to run our applications.

More is always better here, but the reason why we can't have as much as we like is because this memory is different and more expensive than what we discussed earlier.

The type of memory in devices like USB Memory Cards and the so-called *Flash Memory* in smartphones, as well as the *Magnetic Memory* in traditional hard drives, is *non-volatile*: whatever was written in that kind of memory is retained when power is turned off.

In contrast, the memory used by the operating system of a smartphone while running applications is *volatile*, meaning that when you turn the smartphone off, whatever was in the volatile operating memory is wiped out. The benefit of being volatile is that both reading and writing access of this kind of memory is very fast compared with most non-volatile memory access times, allowing the processor to run applications with maximum speed.

The slower, non-volatile memory is only accessed when data or applications need to be loaded or stored: all other processing is done with maximum speed inside the volatile memory.

Some types of non-volatile memory also have limitations on how many times a single storage position can be written over before it wears off and fails, but for most practical uses, this number is so large that it can be ignored.

Last but not least, a practical note: whatever capacity or type of long-term storage you have, **always have at least one backup**, more preferably two, and to be on the safe side in case of fire or theft, **do not keep your backups in the same place as your computer.**

Swap one copy every week with another memory card that you keep in your work drawer. Or give it to a friend, or save your precious data to a *cloud backup service*.

Things break, are stolen or get accidentally deleted, usually just when you least expect them to. And remember, just having a *mirrored copy* somewhere in the cloud is not enough: if you accidentally delete your local copy, the deletion may be mirrored in the cloud the moment you synchronize your device again. There goes your copy...

Save your work and save it often—computers have a tendency of breaking one way or another, just when you need them most.

Also make multiple versions of your work: don't just save with the same name over and over again. There is no limit to version numbers you can use, and this approach allows you to go "back in time" if you inadvertently break something along the way: I ended up saving almost 400 intermediate versions while writing this book, and due to an error I made during one late-night session, one of them was worth all the extra hassle of versioning.

The discussion about mirroring and copying our digital data brings us neatly back to the subject of wireless communications: for *uploading* or *downloading* all these books and pictures and videos and whatnot across any data transmission connection, we use *bits per second (bps)* to present the available *data transmission speed*.

Without taking into account any overheads needed for ensuring successful data transmission, you can deduct from the numbers above that sending one *megabyte* of data over a channel that provides 1 megabit per second speed takes approximately eight seconds, as every byte has eight bits to send.

With all of the overheads needed for data encoding, splitting the data into individual packets and various *error detection and correction* schemes, a multiplier of 10 is quite OK to use for general calculations.

In an *error detection and correction* scheme, additional information is embedded to the transmitted signal, which can be used to mathematically verify that the data was received correctly, and in some cases even restore the original signal if the detected error was not large enough.

The speed that is available for wireless communications depends on various aspects like the frequency band in use, available bandwidth per channel, number of parallel users on the same channel, modulation type and distance between transmitter and receiver. All these cause physical limitations that affect the potential maximum speed of a transmission channel.

As an extreme example, when *NASA's New Horizons* spacecraft passed by Pluto in 2015, it gathered over 5 GB of images and other data on Pluto and its companion Charon, storing it all in its non-volatile on-board memory.

All this data collection and picture grabbing during the fly-by happened automatically in just a matter of hours, but as *New Horizons* was only able to transmit the collected data back to Earth with an average speed of 2,000 bps, it took almost 16 months until the last bit of data was safely received by *NASA*.

Had something gone wrong with the probe during this 16-month period, all the remaining fly-by images would have been lost forever.

And to put the distance into perspective, when a single data packet was sent from Pluto, it took over five hours for it to arrive to Earth, even though the transmission was occurring at the speed of light, 300,000 kilometers per second.

In another example literally closer to home, the common, low-end home Wi-Fi network has a theoretical maximum speed of 54 Mbps. Hence, in optimal conditions, the amount of data sent by *New Horizons* would pass through this kind of Wi-Fi connection in roughly 16 minutes instead of 16 months.

Regarding *cellular data networks*, the current top of the range is called *LTE Advanced* (where LTE is shorthand for *Long Term Evolution*), offering a theoretical speed of 300 Mbps. This would be able to deliver the *New Horizons* images in just under three minutes.

Depending on the number of simultaneous users that are accessing the same cellular base station at the same time, the real data speed will be downgraded. Similarly, your distance from the base station and any obstacles between your device and the base station play a big role here, but in practice it is possible to get several tens of Mbps of average download speed, which is very good for most purposes.

For city dwellers, having a fixed-line Internet connection via *fiber-optic data cable* is by far the best connection you can get, as commonly available urban connectivity speed is 100 Mbps, often both for uploading and downloading of data. And the available offerings are getting faster and cheaper all the time.

But whatever top-of-the line high-speed technology you have today will be old news in a couple of years, as new, more complex and hopefully more entertaining use cases will be invented. Our data communications solutions will keep on playing catch up against the ever-increasing needs: currently, *fifth-generation (5G)* networks are being rolled out, with a promise of a tenfold improvement over the prevalent 4G speeds, and hence, over a 5G channel, all of the *New Horizons* fly-by data would be transmitted in less than a second.

The other side of the speed coin is the fact that higher speeds mean higher power consumption: every data packet has to be processed, the potential encryption and compression need to be opened, and the data has to be stored or displayed somewhere in real-time.

With the increase in connection speeds, all this requires ever-increasing amounts of computing power, which in its turn has a direct relationship with the amount of energy needed for these operations. The advances in microchip technology do counteract this trend, as new *microprocessors* with smaller internal transistors can perform the same amount of processing with less power, but we humans have so far been extremely proficient in using all the available capacity as soon as it becomes available. Therefore, unless there is some fundamental breakthrough in battery technology, the available battery capacity will remain as the ultimate limiting factor to our wireless device usage.

TechTalk
There is No Free Lunch

You can't extract information from nothing.

In the world of radio, this rule of thumb means that even though your transmission is a pure *sine wave*, strictly on a specified *carrier frequency*, modulating the transmission with another signal in order to actually embed some information into your transmission will cause the transmitted signal to occupy a slice of adjoining frequencies instead of just a single, sharply specified one.

Your *channel frequency* will still be the one you are transmitting on, but your signal will "spill over" to adjoining frequencies on both sides of this frequency.

The resulting small block of adjacent frequencies is called the *channel*, and the *bandwidth* describes the size of the channel.

The simple consequence of this limitation is that the higher the maximum frequency of the modulating signal, the more bandwidth you need, and thus the wider your channel will be.

A simplified rule for analog transmissions is that your channel will be twice as wide as the highest frequency of your modulating signal.

Therefore, if you are transmitting on a frequency of 600 kHz and modulate your transmission with a telephone-quality voice signal that is limited to a maximum frequency of 4 kHz, your channel width will be 8 kHz, occupying frequencies from 596 to 604 kHz.

In order to avoid interference, adjacent transmissions must be separated at least by the bandwidth amount from each other. Therefore, if you have a fixed block of frequencies in use, the number of channels you can have in this block depends on how many times you can fit bandwidth-sized blocks, side by side, into the total available frequency block.

The frequency spectrum is divided into bands, and most of us are familiar with at least two common identifiers for this, as every radio tends to have selectors for AM and FM.

P. Launiainen, *A Brief History of Everything Wireless*,
https://doi.org/10.1007/978-3-319-78910-1

Although these abbreviations actually refer to types of modulation, as explained in Tech Talk ***Sparks and Waves***, they also switch your receiver between two distinct frequency bands:

The AM band usually covers frequencies from about 500 to 1,700 kHz, which is officially known as the *Medium Frequency* part of the radio spectrum, and has been defined to contain channels of 10 kHz width in the Americas region. This means that the AM band can theoretically accommodate 120 individual channels of this size.

10 kHz bandwidth is good enough for talk radio but miserable for music, as the maximum modulating frequency has to be below 5 kHz. The frequency range for our ears is at best from 20 Hz to 20 kHz, so it is no wonder that the transmissions of any musical content on the AM band sound very low quality to us. But for talk radio use that's totally fine.

The benefit of using this low-frequency band for transmissions is that, thanks to the way these lower frequencies follow the curvature of the Earth, stations can be heard from great distances, as was discussed in TechTalk ***Sparks and Waves***.

As the name says, all transmissions in this band are amplitude modulated.

Selecting "FM" on your radio usually offers frequencies from 87.5 to 108 MHz, which means that even the low range of this part of the radio spectrum has a frequency almost 50 times higher than the top part of the AM spectrum.

If you kept using the same channel bandwidth as for the AM band, this FM band could accommodate over 2,000 channels, but instead of maximizing the number of channels, the higher frequency is used to widen the individual channel to 100 kHz, thus allowing better quality audio transmissions: FM audio modulation bandwidth is 15 kHz, which is high enough for most of us with conventional hearing abilities, as the upper frequency range that our ears can hear tends to decay with age, and conventional music content usually has very little spectral information between 15 and 20 kHz.

The 100 kHz channel width theoretically allows just over 200 separate channels to be crammed between 87.5 and 108 MHz, and as the audio modulation does not use up all of the available bandwidth, there's ample room in an FM channel for other purposes than just sound. This extra bandwidth is often used to carry information regarding the station name and the song being played, so that this information can be displayed on the receiver, and as explained in Chapter 4: ***The Golden Age of Wireless***, the wider channel also makes it possible to transmit stereophonic sound in a way that is compatible with monophonic FM receivers.

There are some practical limitations that halve the actual channel capacity of the FM band: one less understood one is related to the *Heterodyne Principle*, which is described in TechTalk ***Sparks and Waves***, and which is used by all current FM receivers.

Due to the fact that the standard FM receivers use the super-heterodyne model, an intermediate frequency of 10.7 MHz is used for the local amplification of the received signal. This effectively makes every FM receiver act simultaneously as a low-power transmitter on a frequency that is 10.7 MHz higher than the current channel frequency.

Therefore, it would be very unwise to have two stations in the same coverage area with frequencies differing by 10.7 MHz: for example with one station at 88.0 MHz and another at 98.7 MHz, a nearby listener of the 88.0 MHz station would cause interference to another listener who has tuned her radio on 98.7 MHz, as the *local oscillator* of the receiver tuned in the 88.0 MHz channel would generate a low-power 98.7 MHz signal that might be strong enough to block out the reception of the 98.7 MHz transmission. The distance in which this might be a problem is at best calculated in tens of meters, but it could be a problem for car radio listeners in a traffic jam, or between neighbors in a block of flats.

This effect is easy to detect with two analog FM radios: tune one on some frequency and you will find a "quiet spot" on the dial of the other one at a location that is exactly 10.7 MHz higher.

Due to the use of this intermediate 10.7 MHz frequency, digitally tuned FM receivers in some countries only allow the selection of odd decimal frequencies, from 87.5 to 107.9 MHz, as the frequency of the local oscillator in this case will always end up on an even frequency decimal, on which there are no other receivers or transmitters. Every FM transmitter in the country will always transmit on a frequency that ends with an odd number. As a result, you cut the number of your potential channels by half, but all of these channels can be taken into interference-free use in high-density urban areas.

When your modulation signal becomes more complex, you end up generating a wider channel, and to accommodate multiple channels with highly complex modulation signals like television transmissions, you have to use even higher frequencies.

As an example, terrestrial television transmissions use the *Ultra High Frequency (UHF)* part of the spectrum, starting from 300 MHz.

The top of the range has microwave frequencies in the GHz range, where it is possible to cram several digital television transmissions into a single but very wide transmission channel, and as discussed in Chapter 7: ***Traffic Jam over the Equator***, these high-capacity channels are used for satellite communications, as well as for a multitude of point-to-point terrestrial microwave links.

As we enter the microwave portion of the electromagnetic spectrum, the absorption caused by water molecules starts affecting the reception of the signal. This phenomenon and its side effects were discussed in Chapter 11: ***Home Sweet Home***.

With microwaves, heavy rainfall or a snowstorm can momentarily reduce the signal strength of a satellite television broadcast or a low-power terrestrial microwave link to a point that reception is no longer possible.

Further up, when the frequency reaches a couple of hundred terahertz (THz) range, we reach the visible light part of the electromagnetic spectrum, in which we need only a thin physical obstruction to completely block out any transmission.

Increasing the frequency also increases the energy of the generated waves, in direct relation to the frequency. Therefore, when the frequency of the radiation goes past visible light into the ultraviolet part of the spectrum, the waves have enough energy to pass through obstructions again. Hard ultraviolet radiation, X-rays and especially gamma rays can penetrate deep into matter, and together they form the *ionizing radiation* part of the spectrum.

All three have enough energy to knock off electrons from the atoms that they hit along the way, and thus are harmful to living organisms. Ultraviolet radiation is strong enough to be used to kill bacteria, and it can also cause skin cancer as a result of excessive exposure to the incoming ultraviolet radiation from the Sun.

Detecting X-rays and gamma rays forms a major part of *Radio Astronomy*, allowing us to understand complex things like the formation of black holes and the processes that happen inside our own Sun. With gamma rays, we have reached the top end of the electromagnetic spectrum, with the strongest known type of radiation.

Another fundamental aspect of electromagnetic radiation is the concept of *wavelength*.

An electromagnetic signal travels through space as a *sinusoidal wave*, and the distance needed to complete one oscillation defines the wavelength of the signal.

Wavelength can be calculated by dividing the speed of light by the frequency, and hence the wavelength of a signal gets proportionally shorter as the frequency gets higher.

The wavelength varies from tens of thousands of kilometers for the *Extremely Low Frequency (ELF)* band (3–30 Hz) to meters in the television and FM radio bands (30–300 MHz), and down to nanometers (billionths of a meter) for the visible light spectrum (430–750 THz).

The wavelength of a signal affects the antenna design, as an antenna for a certain frequency works best if its length matches the wavelength of the signal in a fixed ratio. For example, a common *quarter-wave monopole antenna* design consists of a metallic rod that is one fourth of the wavelength of the desired frequency, combined with a suitable grounding surface.

Hence, as discussed in Chapter 3: ***Radio at War***, the shift to higher frequencies made truly portable battlefield radios feasible, thanks to the shorter antennas they required, and our current gigahertz-band mobile phones can get away with fully internal antenna designs.

To summarize, the properties of electromagnetic waves differ wildly depending on the frequency band, and each frequency band has its optimal usage scenarios. The higher the frequency, the more modulation it can accommodate and thus more information can be transmitted per second, but the way different frequencies propagate in the Earth's ionosphere and the way they pass through intermediate matter like walls, rain and snow also varies hugely. Therefore, the selection of which frequency band to use depends totally on its planned application and the context it will be used for.

TechTalk
Making a Mesh

Digging cables into the ground or installing them on poles is very expensive, and hence has been a limiting factor for the improvement of communications infrastructure in many less-developed countries.

The value of the copper needed for the cables themselves has caused problems, as thieves have cut out the cabling and sold it onwards as scrap metal. Any copper cables are prone to failures due to thunderstorms, especially in rural areas where the cable runs are long and often exposed to weather: the longer the cable, the more subject it is to induction-induced over-voltages due to nearby thunderstorm activity: such induced peak currents can easily fry electronic equipment at the end of the connection.

Fiber-optic data cables do not have the same susceptibility to electric storms, but at the moment they require even more expensive infrastructure components, which often becomes a limiting factor, especially in rural settings.

Therefore, the possibility of switching directly to *cellular networks* has been a boon for many countries: building cellular towers at the heart of existing cities, towns and villages makes connectivity instantly available to all, without the cost and labor of connecting each individual household to a wired network.

You still need a complex and high-bandwidth wired or dedicated microwave-based connection between the *base stations* and the rest of the communications infrastructure, but you are only dealing with thousands, not millions, of individual connections.

I remember a visit to Kenya over ten years ago, during which I purchased a local SIM card for data usage and was astonished at the speed and quality that it provided in the outskirts of Mombasa. The cellular revolution has given many countries the ability to jump from zero to a totally connected society, thus giving a huge boost to local commerce. The positive economic effect of such cellular installations can be enormous.

And when the cellular phone is the only means of communicating, it is no surprise that cellular phone-based cashless payment systems, like the *M-Pesa* of

P. Launiainen, *A Brief History of Everything Wireless*,
https://doi.org/10.1007/978-3-319-78910-1

Kenya, had been around for almost a decade before starting to appear in the smartphones of the developed world.

The trend of "bypassing the cable" that has been so groundbreaking in the emerging market economies is making its way to the developed world, too: in the United States, the number of households with only cellular connectivity surpassed the number of households with fixed phone lines in 2017. These customers are not bypassing the cable due to the non-existent availability of such a connection, they are cutting themselves off of an existing one.

With "fast enough" fourth-generation networks, even the connection to the Internet is now becoming a wireless one, often subsidized by the telephone companies that are happy to see their costly wired infrastructure become obsolete.

But there are still numerous places where this kind of cellular technology based wireless revolution is not yet feasible: base stations are still relatively expensive to install and maintain, requiring a certain minimum return of investment to become feasible, and in some less-tranquil parts of the world, a continuous armed security presence must be deployed to ensure that the equipment at the base station, or the fuel stored for the resident backup power generator, is not stolen.

And what if your "customer base" is just a bunch of wild animals in the middle of a rain forest, hundreds of kilometers away from the nearest civilization, and your need for connectivity only occurs when one of your "customers" wanders in front of a hidden camera trap?

Or what if you want to provide a wireless Internet service that is going to be used by hundreds of occasional users in a tiny, far-flung village that has only one physical connection to the Internet?

You should build a *mesh network*.

The basic concept of a *mesh node* is radio-enabled equipment that dynamically adapts to its surroundings: it finds other compatible radios within range and maintains a connectivity and routing map of all other nodes in the network.

A good example of such a node would be a solar-powered wildlife camera with a radio transmitter/receiver that can cover a distance of a couple of kilometers.

You spread nodes like these in a suitable pattern, so that each one can connect to at least one other node. As each node further away connects to an additional node or nodes near to its own position, you can cover several hundreds of square kilometers with only a couple of tens of nodes and still be able to transmit data freely from one edge of the mesh to another: the message is delivered, hop-by-hop, by the adjoining nodes.

Connect one node of this mesh to the true Internet, and you can have connection from each node to anywhere in the world, and vice versa.

In populated areas you can use cheap Wi-Fi hardware with improved external antennas to have a mesh grid size of 200–500 meters, benefiting from the high bandwidth of Wi-Fi.

The drawback of a deep mesh network is the fact that the same data has to be re-transmitted several times until it reaches its destination, and therefore if the amount of traffic is high, your closest node is constantly using bandwidth and energy to relay messages from other nodes. Also, for cost and power management reasons, most solutions rely on cheap, standard hardware and have only one set of radio circuitry. Therefore, they can't send and receive at the same time, which effectively cuts the available bandwidth in half.

On purpose-built mesh hardware, it is possible to create separate radio channels for the core routing traffic and the network endpoint traffic, which would speed up the overall network throughput performance, but also increase the power consumption of nodes.

When a mesh like this is connected to the Internet, the bandwidth available for an individual user goes down rapidly if many others are trying to access the Internet at the same time, but in many cases, even a slow connection is much better than no connection at all. Normal browsing of web pages on the Internet also does not require a constant stream of data, so multiple users can easily be served, as their requests are interleaved in time. Just hope that not too many users are trying to watch YouTube videos or other streaming content at the same time.

In the optimal case, the network topology is planned to be dense enough so that the overall mesh is resilient against changes—if one node disappears for one reason or another, the routing can work around this and use other available nodes instead.

Therefore, the common denominator for all mesh networks is the automatic adaptability to the potentially hostile and constantly changing environment. In this sense, they mimic the functionality of the existing *Internet Protocol (IP)* networks, which also route each individual data packet according to the constantly changing routing situation of the Internet.

Apart from this basic requirement, mesh systems can be designed freely, depending on their expected usage model and available power limitations.

A good example of real-life implementation is the *FabFi* mesh, which is used to cover towns in Afghanistan and Kenya. The achievable throughput is over 10 Mbps, which is very respectable for occasional connectivity needs.

Mesh networks are not only a viable solution for developed countries: the *Detroit Community Technology Project* is maintaining a set of interconnected mesh networks as part of their *Equitable Internet Initiative* to alleviate the fact that Detroit is one of the least connected cities in the United States: 40% of the residents of Detroit have no Internet access. This kind of digital divide is becoming more and more of a cause for further alienation from society, as most services are increasingly only accessible via the Internet. Providing connectivity through a shared setup enables lowering the cost to a level that can support a large number of non-paying participants and in Detroit's case, help in the recovery of a city that has been hit by a major economic downturn.

The nodes in a mesh can even be portable handsets.

Many countries have deployed emergency radio networks based on *Terrestrial Trunked Radio (TETRA)*, in which the network is supported by a relatively low number of traditional base stations, but if some or all of those get knocked out due to a hurricane or an earthquake, the handsets can configure themselves to act as relaying nodes.

Naturally, if the relay usage is very high, the handsets are on all the time, so their battery life is greatly reduced, but in the case of TETRA, the slimness of the handset is not a driving design parameter, so they can be provided with much beefier batteries than your usual smartphones.

TETRA has its origins in the 1990s, and thus is voice-oriented and very poor in terms of data connectivity, but with its robust handsets and inbuilt mesh feature, it is hard to replace until handset-to-handset mesh becomes a viable feature in revised 4G or 5G specifications.

Mesh technology is one of the potential ways to support the new concept of the *Internet of Things (IoT)*, that was briefly discussed in Chapter 11: ***Home Sweet Home***.

With IoT, your environment is filled with simple sensors and other useful devices that are intermittently generating data and relay their information to some centralized processing unit when the need arises.

This “master node” then offers external Internet connectivity for the control and data access of the entire IoT mesh.

There are also proprietary military versions of mesh networking, which is understandable, as the battlefield environment is probably the best example of a highly dynamic situation in which reliable and resilient communications is of vital importance.

Mesh networks fulfill a “sweet spot” in cases where the expected traffic density and the distance between nodes happens to be well suited to the power capabilities of the nodes. With their dynamic adaptivity, they are very robust against changes in network topology, can be set up and expanded easily and offer yet another alternative for harnessing the power of electromagnetic waves.

TechTalk The Holy Grail

For the actual radio spectrum, the different behaviors of these highly varying frequencies have been discussed in the other TechTalks above.

Similarly, problems encountered in creating these waves have been addressed along the way, from spark-gaps to generators to solid-state electronics, first with vacuum tubes and now with transistors and microchips.

The third important aspect of signal generation is the modulation that is either suitable for the frequency band in use, or optimal to the kind of information that needs to be transmitted, whether it is digital television or narrowband audio for mobile conversation. Again, this was discussed in detail in TechTalk ***There is No Free Lunch***.

Even the best solid-state electronics have physical limitations in terms of amplifying low-level, high-frequency signals that are extracted from the receiving antenna. Therefore, it has been mandatory to turn this weak signal first into a lower-frequency one before further amplification and demodulation in the receiver. This application of the *Heterodyne Principle* was described in TechTalk ***Sparks and Waves***.

Another limitation is the required antenna length, which usually can only be optimally tuned on just a small portion of the frequency spectrum.

All these constraints have, until very recently, forced the use of 100% purpose-built electronics, in which the utilized frequency band and supported modulation scheme are hardwired into the actual circuitry, and thus are practically impossible to change after the circuitry has been assembled.

Four recent technological advances are now challenging this mainstream approach:

First, it is now possible to create special transistors that have very low internal *stray capacitance*, allowing amplification of frequencies in the gigahertz range. This makes it possible to capture the faint signal from the receiving antenna and amplify it directly as-is, without any intermediate frequency conversion steps.

P. Launiainen, *A Brief History of Everything Wireless*,
https://doi.org/10.1007/978-3-319-78910-1

Second, digital signal processing is getting cheaper and faster, often thanks to the ever-increasing demands of our beloved digital “timewasters”, like game consoles and High Definition (HD) digital television receivers.

Third, basic computing capabilities are still expanding along the original *Moore’s Law*, which states that the number of transistors packed into the same physical space in microchips doubles approximately every two years. More transistors per area means that they become smaller, which in most cases reduces the internal stray capacitances and thus improves the maximum achievable switching speed and lowers the overall power consumption. This increase in processing speed can be directly utilized without any modifications to existing software.

We are also in the verge of moving from two-dimensional chip designs to three-dimensional ones, which allows a yet larger number of transistors to be crammed into our silicon chips: the more transistors at your disposal, the more complicated processing methods, like parallel processing, can be utilized.

Last but not least, smart, adaptive antennas are being developed, making it possible to transmit and receive on a much wider set of frequencies than with the traditional, physically fixed antennas.

Put all these together, and we are getting closer to the ultimate Holy Grail of wireless: *Software Defined Radio (SDR).*

In an SDR receiver, the signal from the antenna is directly amplified, fed into a high-speed analog-to-digital converter, and then handed over to the signal processor circuitry, guided by traditional computer logic. The received signal is therefore turned directly into a stream of bits, the manipulation of which from that point forward is only limited by the available computing capacity.

Want to use the circuitry to listen to FM radio?

Just load the appropriate software.

Want to add a new modulation type to your existing mobile handset?

Just load new software that contains the necessary logic for handling the new modulation.

Want to send a search-and-rescue signal to a satellite from your mobile phone when you get stranded on a remote island without cellular coverage?

Just select the appropriate item from your menu, and the corresponding software module that emulates a standard *406 MHz Personal Locator Beacon (PLB)* is activated.

With fast enough SDR technology, combined with smart antenna solutions, your device would not become obsolete as new modulation technologies and frequency bands are taken into use. When a novel modulation method that improves the wireless data connection speed is invented, or a new frequency band that adds channels to your cellular connection is released for use, all you need is a software upgrade.

And instead of having different circuitry to handle your various connectivity modes from Bluetooth to Wi-Fi to cellular to GPS, you could run them all in parallel on a single, fast enough circuitry. The microchips that form the basis of the radio circuitry of today’s devices already handle many of these different types of radios in parallel, thanks to the hardwired logic that is built in, but SDR would add an unprecedented potential for flexibility and future extensions.

Having fully programmable hardware would also speed up the development of new protocols, as anything could be tested in real life by only writing new controlling software: there would be no need to wait for the new circuitry to be first defined, then made into astronomically expensive first-generation microchips, only to be scrapped due to some minor problem that could not be foreseen in the process.

And when everything is based on software and there are clever programmers around, there's always the potential for misuse—an open SDR environment could for example be made to run rogue code that takes down entire cellular networks: the existing cellular networks are based on commonly agreed upon standards and hence expect correctly behaving equipment at both ends of the connection. Sending malformed data at the right time could easily take such a network down, and devising a program for an SDR for this purpose could be fairly easy.

Despite potential problems, we are rapidly heading in this direction. Hobbyist SDR boards can be bought for a couple of hundred dollars, and 100% software-based demonstrators of various protocols have been created with these devices, including the functionality of a GSM base station.

These do still have many practical limitations, especially in the width of the supported radio spectrum and the smart antenna technology, but the direction is clear—the moment SDR circuitry becomes cheap enough, it will become the heart of any wireless device.

This will be the Holy Grail that adds unprecedented flexibility to our exploitation of the electromagnetic spectrum.

Index

P. Launiainen, *A Brief History of Everything Wireless*,
https://doi.org/10.1007/978-3-319-78910-1