# Cellular Phones

## Professor Richard Harvey FBCS

### 8th March 2022

What does the word "invention" suggest? Even I, who am broadminded about such things, have a tendency to let my mind wander to a lone eccentric tinkering a basement – a Wallace or a Gromit. Although such inventions are heroic, I'd like to enthuse about two other types of invention. One type is a way of thinking, physicists would call this a theory, but they also occur in the real world, and several of the previous lectures in this series have considered inventions that are made of pure thought. They are ways of thinking about a problem – a algorithm or a scheme. But there is also a third type of invention, one that gets even less attention, and that type is an agreement that, of the many ways of doing something, we are going to do it this way. The internet, mentioned in previous lectures [1], is just such an invention. And this lecture is about another such invention: cellular phones.

One of the challenges of these *protocol inventions* is trying to keep the standards ahead of the actual technology but not too far ahead otherwise the standard cannot be implemented. And certainly not behind, otherwise you are constraining everyone to an antique system. Instead, the goal is to set standards now for what humans might achieve in the future.

Mobile telephony, as it is called in the UK, is hardly a new idea. The dream, or for some, the nightmare of everyone having a mobile phone was certainly alive in 1906 via a Punch cartoon which showed a couple with mobile phones – one "sexting" as we would call it now and another having a go at online gambling [2] And in 1926 the German satirical magazine Simplicissmus was treating mobile telephony as an example of how sophisticated, and full of themselves, Berliners were compared to the inhabitants of Munich [3].

Mobile telephony, or cellular telephony as it is called in the US, is in principle, remarkably simple – if we can convert the voice into an electrical signal and then amplify and connect it to a long piece of wire, called an antenna or aerial then, bingo, we have a radio transmitter. The recipient amplifies the faint signals on their antenna, connects to a loudspeaker and we have one-way, or simplex communication. Generally speaking, the antenna needs to be noticeably long compared to a wavelength. The formula for wavelength, $\lambda$, is $\lambda = c/f$ where $c$ is the speed of light ($3\times10^8$ ms$^{-1}$) and $f$ is the frequency. The human voice can be understood in a frequency range of 400 Hz to 4 kHz so at 400 Hz, we would have $\lambda = c/400 = 75$ km. Here is our first problem. If the antenna was even one-tenth of a wavelength, then it would be 7500m long – hardly handy (handy is the German word for mobile telephone by the way).

All voice radios therefore use some form of *modulation*. They might alter the amplitude, the frequency or phase of a carrier wave. 4G telephony uses bands up to around 6 GHz for which $\lambda = 3\times10^8/6\times10^9 = 5$ cms. Ah! That's more like it! We could easily conceal an antenna of a couple of centimeters in a mobile handset. Generally, the higher the frequency, the shorter the range, and the more expensive the hardware. All of which tended to encourage lower frequencies. Marine VHF, for example, uses 160MHz (around 1.9 m wavelength). This gives it a range to the horizon. A further issue is that a receiver cannot be listening on the same channel on which it is transmitting. Hence the convention is to use two channels – duplex operation. The selection of channels is a major faff. If you have taken a marine VHF exam, then you will recall that there have to be agreements about which channel is used for what. Thus, communication becomes irritating and laborious. Old military radio sets used to come with a selection of colour-coded crystals and radio operators had to select the right one. Not a pleasant task in a dark ditch under fire from the enemy.

For mobile phone designers, the situation looked discouraging. There was a need to keep the frequency low

to be affordable and to maximise range, but there wasn't much spectrum available at those frequencies, and how on earth can we allocate frequencies to hundreds of users in a manner that doesn't get chaotic and uncontrollable. This is where the idea of a cell arrives.

The initial idea was to model the region around a phone mast with a hexagon – hexagons fit together neatly and allow us to think about the frequency assignment problem. Obviously if adjacent cells use the same frequency, then there is the potential of interference, so a hexagon tessellation implies that we must have no frequency re-use inside seven channels. Some designs are more stringent than that and insist on frequency reuse into the second ring of hexagons, others, near impassable regions, can use fewer frequencies. Deciding this is the topic of network planning – a rarefied art carried out by a few specialists[1]. More recently the hexagon pattern has been put to a different use – positioning directional antennae on the intersections of the hexagons allows several beams into a cell — something we will examine later.

The cellular idea has had a remarkably long lifetime, persevering over all the generations of mobile phones. The first generation were analogue and emanated from the USA where Martin Cooper had built the first ever handheld mobile – the Motorola Dynatac [4] (featured in the hands of Gordon Gecko in the film *Wall Street)*. One irritant was that the analogue signal was easy to eavesdrop, and the systems in a variety of different countries were not compatible. In Europe there was a concerted effort to design a more secure digital system that was interoperable. A relatively little-known international agreement, the Bonn Agreement, was signed in 1987 and that was the start of roaming – mobile phones hooking into foreign networks as visitors with the billing information being passed back to their home country. GSM was born.

This is probably an apposite time to taste the alphabetical spaghetti that is mobile phone technology. GSM originally stood for *Groupe Spécial Mobile*, a sub-committee of the CEPT committee that governed European telephony (*Conférence européenne des administrations des postes et des télécommunications)*. However, as it became clear that GSM was going to cover more than just the EU, the name morphed into Global System for Mobile communication. Mobile communication standards start by defining the required specifications for end-to-end communication. There is then a call for technological solutions. There is some to'ing and fro'ing about which technology is likely to be ready within the next 10 years. The functional interfaces are then defined, and a new standard, or release, is produced. The Gs of 1G, 2G, 3G etc do not exist within the standards which leads to massive confusion.

To return to the generations of mobile phones we can summarise their principal features as follows. 1G was analogue; 2G was the first all-digital standard and introduced the idea of an SMS message or text (there is a whole another lecture to be given on the history of the SMS[2]. 3G was driven by the need to provide mobile data to phones. 4G was all about data (by now the voice capacity of a phone is of little importance) and was the first standard to abandon circuit-switching and move to voice over IP. 5G is yet more data but with a recognition that there are platoons of devices waiting to use mobile data if only it was fast enough, ubiquitous enough, and secure enough.

The development of standards is a tricky business. For mobile telephony, the whole cycle from inception to implementation takes around 10 years. So, the standards bodies need to predict what technology will be capable of and what society will need within that time. And unlike other branches of engineering, the pace of change is very rapid. In 1975 there was a bidding war for the most powerful supercomputer in the world, the Cray 1. They cost millions of dollars and required special adaptations to the building to house the cooling system – 'cool' in multiple senses of the word. In 1997 IBM developed a computer that managed to beat Gary Kasparov, the world Chess Champion. Compared to, say, and iPhone 13, which one of these computers has the most computing capacity? The answer, as I expect you have guessed, is the iPhone 13. I borrowed this example from an online article [5]. The comments protested that measuring Deep Blue by its ability to multiply two floating point numbers together (floating point operations per second are known as *flops*) was unfair as that machine was designed to solve integer problems. Fair enough, but let's not dispute the general point which is that the computer in your pocket is more powerful than all the computers used by NASA for the moon landings (by all I mean all the computers NASA had ever purchased up to that point!) When capabilities are growing so quickly, it's not that easy to predict ahead, so mobile phone standards are enhanced more rapidly than the Generations. This is something of a marketing challenge — how does anyone know what precisely they are buying?

The situation is not helped by the proliferation of naming conventions. 3G for example was proposed as an international standard by the International Telecommunications Union, ITU, who proposed a standard called

---

[1] I imagine a convention of cellular network planners could meet in a small pub!
[2] I heard that SMS was a designer's afterthought – no-one expected it to become so important.

IMT-2000. The standards are developed by a working group known as 3GPP (3rd Generation Partnership Project) comprising seven telecommunications standards' bodies covering the world continents. They called their version of the system UMTS. Thus, UMTS is the system that will meet the specifications of IMT-2000. UMTS comprises several systems: there is the Radio Access Network or RAN, denoted UTRAN; and then there are the core standards and system which for 3G are imaginatively named UMTS-core and UMTS-system. I mention this tedious nomenclature because, without it, much of the literature is incomprehensible.

Let's see how this works in practice with the latest generation of mobile known as 5G. There is a much-copied diagram which illustrates the key parameters for a cellular system [6]. It shows the difference between 4G (labelled IMT-advanced) and 5G (labelled IMT-2020). This diagram was produced long before the 5G standard was defined so this is a comparison of something that exists (4G) and something that the standards body was willing into existence (5G). It's probably worth dwelling on what is meant by 'willing to exist'. The standards body means that when the 5G infrastructure is fully deployed and the user equipment (that's code for the handset) is also up-to-scratch then this is what we expect to achieve.

Some of these parameters seem easy to interpret. 5G must be capable of a peak data rate of 20 Gbits s$^{-1}$. This figure is useful for network planners, since it implies immediately that 5G base stations should be connected via fibre-optic cable into a fast network. Furthermore, users should experience rates of 100 Mbits s$^{-1}$. So, up to 200 users per cell. We will come back to that number as it sounds low — how are we going to serve video of the latest goal to a crowd of 100,000 at the Camp Nou stadium in Barcelona? There is another speed number which is the amount of capacity per square m. If users are to experience 100 Mbits s$^{-1}$ with a maximum data density of 10 Mbits/s/m$^2$ then our imagination has rows of users squeezed into boxes of 3.3m by 3.3m. Even with 5G we cannot cope with a whole aircraft or trains wanting to stream video. That said, not everyone wants to stream data so we might also measure the number of connections which here is desired to be a million devices per square km. One million devices per box measuring 333 by 333m! Is that realistic? Well, 5G covers Internet of Things devices and many of us carry three or four personal devices which potentially could require mobile bandwidth: it all adds up. Spectrum efficiency is a measure of how much data we squeeze into a fixed radio bandwidth. In many countries there are severe constraints on available bandwidth so good spectrum efficiency is helpful. As is a commitment to use less energy. Actually, it is a commitment to be more energy efficient which is not the same as using less energy – generally speaking IT equipment has become more and more energy efficient but there is a constant battle with adoption – the more people use something the more energy is consumed. There are two parameters left – latency and mobility. I'll look at those in a moment when we discuss some of the technical aspects of 4G versus 5G.

Qualcomm are one of the significant companies in mobile telephony developments and in their backgrounder to 5G they identify five innovations that are associated with 5G: a flexible slot-based framework; a scalable OFDM-based air interface; advanced channel coding; massive MIMO and mobile mm-waves [7]. I think this is an interesting set. Are any of them innovative in the sense that we have not seen them before? No. Have any of them been applied to mobile communication before? Yes – in parts. Is 5G innovative? Yes, because the parts have not been put together before. Why have they not been put together before? Mostly the answer is that suitable interfaces have not been defined that allow these subs-systems to work together effectively. That's what I meant by a protocol invention at the start of the lecture.

The traditional band for mobile communication is 0.9GHz ($\lambda = c/f = 3\times10^8/0.9\times10^9 = 0.3$m). This was about the limit for cheap technology in the 1970s. But by 5G we have extended to around 50GHz (or around a wavelength of 3mm). Millimetre-band communication is strictly line of sight and attenuation from buildings is very significant. Also, even the small antennae on the mobile phone (or *user equipment* (*UE*) in the patois of mobile technology) can be highly directional, so 5G phones are proposed with multiple antennae. In the UK spectrum is highly regulated and the government makes a pretty packet selling off parts of the spectrum via spectrum auctions – providers who own the most spectrum can provide the most coverage and so can charge their users a premium. In other countries, not all the spectrum is regulated which can lead to spectrum clashes. The most recent of these is in the USA where one of the new 5G bands overlaps the frequency used by aircraft radar altimeters. How this will play out, I am not sure. For the time being, pilots landing in the USA would be well advised to cross-check their height using the conventional barometric altimeter or GPS!

5G also differs in the way it uses the spectrum. The key issue is that the capacity of a radio tower has to be shared with other users. Various generations of mobile telephony in various countries have tried pretty much every combination of spectrum sharing. The simplest is known as *frequency division multiplexing* or FDM. In this scheme each user is allocated a small band within a band – their voice is modulated onto a *sub-carrier* — and they have exclusive use of that band. Data do not fit so easily into that scheme and early mobile

phones used modulation schemes that made digital data look like an analogue signal so that it could be transmitted. Furthermore, accurate generation of sub-carriers is not so easy, and filters are imperfect so there have to be *guard bands* between the channels which is wasteful. Another alternative is *time-division multiplexing* or TDM. In TDM we establish a synchronous clock between the user equipment (UE) and the radio mast, and each user takes their turn in a slot. It's a hassle establishing the slots, and because voice is a real time signal, we cannot wait too long between slots. In practice TDM was often used in sub-bands leading to guard-band wastage again. From a commercial point of view, both FDM and TDM have a *hard* failure when there are no more bands or slots which is still commonplace in countries where users exceed installed capacity. An ingenious attempt to avoid hard failure was to use *code-division multiplexing* or CDM. If you have been following these lectures you will have come across CDM in my lecture on GPS [8]. It's a favourite of professors because it is very elegant. Each user is assigned a unique code that is very long, looks random and repeats. The digital data are exored with the code. Other users' codes are chosen such that they have zero correlation with the other codes thus, to other users, look like noise. As the cell becomes busier and busier, the signal to noise decreases and so there is notionally a soft failure as the eventually high error rate causes users to dial off. In practice, most digital systems exhibit a *threshold effect* so that in high noise conditions they fail catastrophically thus losing the advantage of soft failure. Plus, distributing and decoding non-correlating codes is a pain. 5G has reverted to FDM but it multiplexes all the signals together digitally using a mathematical algorithm called the discrete Fourier transform or DFT (there are speedy versions of this algorithm known as Fast Fourier Transform or FFT algorithms). If we are using the DFT then we create the sub carriers in exact sync with each other (because they are in sync, they do not overlap so they have a property known as orthogonality) so there is no need for guard bands hence Orthogonal Frequency Division Multiplexing or OFDM. In 5G there is a further possibility of time slotting (TDM). Both the channels and the timeslots can be assigned dynamically by the network depending on load and need and because handsets can roam there has to be some flexibility on both sides about what channels and slots are used when – there is a flexible *numerology*.

Qualcomm also list three other innovations. The first we have covered in a previous lecture on Error Control Coding [9]. In that lecture I noted that error control coding had been very much improved in the previous 20 years by the discovery, or rediscovery, of Gallager (Low-Density Parity Check or LDPC) codes and Polar codes. These codes allow us to get much closer to the Shannon limit (see previous lecture) which means more data and fewer error control (or parity) bits. Every digital communication system has been revised or will be revised to use these codes — it's a very low computational cost upgrade that provides free bandwidth. So, it would have been surprising if 5G did *not* use those codes.

As frequencies increase, aerials become more directional which in turn leads to an intriguing possibility of arrays of antennae and *beamforming*. In sonar it has long been practice to replace long continuous sensors with an array of separate (or discrete) sensors. The array performs just as well as a continuous sensor and its convenient having gaps between sensors. A line array of sensors is most sensitive when the wave arrives normal to the receiving elements (known as the *broadside* or *broadfire* configuration). If we apply linearly increasing delays, then we can make the array sensitive to arrivals from other directions. We form a *beam* that points in a certain direction. *Beamforming* has some fascinating aspects (one little known one is that it is possible to oversteer an array beyond endfire and hence correct for the natural beam broadening that happens when an array is receiving waves from along its length) but at such high frequencies and data rates, 5G appears to stick to forming simple far-field beams where, instead of delays, the system approximates these with phase adjustments. In principle both your phone and the radio mast can beamform. Trials have shown beamforming to be especially effective in crowded urban environments — one segment of a football stadium connects via a different beam to another part. At the start of the communication, there is some negotiation between the UE and the mast on the correct beam to use.

Now I would like to talk about the final property which is latency. If we look at the architecture of 3G (otherwise known as UMTS by the cognoscenti) then it looks a bit of a mess. 4G is worse largely because it had to interoperate with 3G. In 3G there are two channels, one for voice and one for data. There are interfaces all over the place and signalling data flying backwards and forwards between the various boxes. This leads to two issues, the first is a simple one which is that it takes time for calls to get set-up, time to transfer data and that waiting time is variable. This time is called *latency*. Latency is a real bugbear of digital systems – many efficient systems like to work on large blocks of data. It takes time to fill-up those blocks, and that time can be damaging, so virtual reality, which is one of the use cases of 5G, can become implausible. If you think virtual reality is frivolous, then consider *platooning* which is another use case. In this scenario, platoons are trucks are connected together by 5G – the trucks whiz along at high speed in a convoy separated by only a meter or two. If your control system has a few milliseconds of variable delay then it will take too long to make

the adjustments and the M25 is closed in a very sudden and undesirable manner. Latency can also be costly in the signalling and control aspects of cellular telephony. As you move from one cell to another there is a complex process known as *handover*. Your mobile device is constantly signalling back its signal strength and when your current cell dips in strength, a new channel is established with a neighbouring cell, or with a different beam. This soft handover process takes time and if you are moving too quickly it fails (it also fails in other conditions in my experience). If we want 5G to work on drones or aircraft, then handover needs to be more robust. To achieve this 5G has proposed a new configuration, NR or New Radio, that simplifies the data flows around the system and hence improves latency. However, no-one can afford to tear-up all their existing 4G and 3G infrastructure and replace it with 5G, so 5G provides a number of bodge configurations that allow one, for example, to use NR with EPC and EPS, the 4G backend.

This leads to a slight curiosity. I'm giving this lecture from the Barbican Lecture Theatre in what is called the square mile — the inner sanctum of London known as 'The City'. The UK is a world leader in deploying 5G and the City of London is the area where you want to deploy it first – my phone declares confidently there is 5G available here. In what sense is it 5G though? Does my phone have a beamformed 5G aerial in it? Probably not. It's probably got a 4G antenna modified to cope with 5G bands. Is the 5G RAN connected in standalone mode? Probably not? It's probably bodged onto an existing 4G infrastructure. Does it meet any of the ITU requirements for 5G? Probably not. Does it outperform my old 4G phone? Actually, I think not, 5G seems to have introduced some strange latency effects and its common for me to get data blockages in 5G. It is classic misrepresentation and I do not find it helpful. The idea, at least in the UK, is that there should be an efficient market for cellular telephony — providers who invest in proper 5G ought to be rewarded with premium customers. If all that is required is to slap on a different antenna, then it's called 5G, then consumers are not getting accurate information. Since your phone knows what service it is getting, it is a bit mysterious that your phone does not tell what you are getting from your service provider but, as we saw, it seems 5G phones might not actually be 5G too – everyone's at it.

Looking at it from the other side, in many countries the cellular infrastructure has to be provided by private investment, so there is a need to encourage early adopters otherwise investment slows. Ideally, I'd like to take you round the country and show the scale of that investment, but fortunately that is has been done for us by the energetic 'Peter C' who runs a YouTube Channel and website that gets excited about mobile phone infrastructure, particularly the RAN. I'll play you a clip from his channel and invite you to find out more if you wish.

In conclusion, the most important thing we can say about cellular telephony is that it is not telephony. When was the last time you used your phone to call someone? I think I make a call about once or twice a month, but I use my phone every day. Cellular telephony is a massively important driver of technology. Many important ideas were either created or developed for the mobile phone: M-pesa the digital money system which originated in Kenya; end-to-end encryption brought to prominence by WhatsApp; the touchscreen; face identification; fingerprint sensors and so. Without the driver of mobile phones all those things would have remained as rarefied and expensive curiosities. This series of lectures has been about inventions – cellular telephony is one of the most important inventions of the last century – it's not an invention in the common sense of the word but inventive it is and long may it continue.

© Professor Harvey, 2022

## References & Further Reading

55
1. *Networks: The Internet and Beyond*, Gresham College Lecture, April 2021, https://www.gresham.ac.uk/lectures-and-events/networks-beyond

2. *Cartoon captioned 'Forecasts for 1907. IV. - The Development of Wireless Telegraphy. Scene in Hyde Park. [These two figures are not communicating with one another. The lady is receiving an amatory message, and the gentleman com racing results.],'* Punch Magazine 1906, https://punch.photoshelter.com/image/I00006GHuH4c0Ojo

3. Simplicissimus magazine, 20th Dec 1928, 31, (38), http://www.simplicissimus.info/uploads/tx_lombkswjournaldb/pdf/1/31/31_38.pdf#page=2

4. *Motorola DynaTAC phone*, https://en.wikipedia.org/wiki/Motorola_DynaTAC

5. *A modern smartphone or a vintage supercomputer: which is more powerful?*, Nick T, June 14th 2014, phonearena.com, https://www.phonearena.com/news/A-modern-smartphone-or-a-vintage-supercomputer-which-is-more-powerful_id57149

6. IMT VISION – "FRAMEWORK AND OVERALL OBJECTIVES OF THE FUTURE DEVELOPMENT OF IMT FOR 2020 AND BEYOND, WORKING PARTY 5D, ITU, Recommendation M.5/BL/22 (09/2015)

7. HTTPS://WWW.ITU.INT/DMS_PUBREC/ITU-R/REC/M/R-REC-M.2083-0-201509-P!!MSW-E.DOCX

8. Making 5G a commercial reality," Qualcomm Feb 2020 https://www.qualcomm.com/media/documents/files/making-5g-nr-a-commercial-reality.pdf

9. *GPS*, Gresham College Lecture 12th October 2021. https://www.gresham.ac.uk/lectures-and-events/gps

10. *Error Control Coding*, Gresham College Lecture 1st Feb 2022, https://www.gresham.ac.uk/lectures-and-events/error-control