3

The Longest Journey

There are many extraordinary journeys in the natural world. The European eel's long, mysterious migration to the Sargasso Sea; the pole-to-pole journey of the Arctic tern, which can cover a million and a half miles in its lifetime; the rush of king salmon hundreds of miles upstream, up waterfalls, into tearing river torrents.

Journeys through the Material World are no less awe inspiring, especially when viewed through the dimension of time. Consider the humble sand grain, a tiny speck of quartz. Over millions of years – sometimes billions – it has witnessed the passage of deep time. It has been rock and then grain and then rock again. It has been compacted into stone, pressed under the earth for hundreds of millions of years before being freed by the erosion of a river or the wind. It has been swirled downstream into estuaries and beachheads, tossed and turned in the tide, packed and compressed back into rock. Ground down by glaciers, tipped through channels and waterfalls, this grain has lain dormant before being picked up for the cycle to begin all over again.

It is thought that about half of all quartz sand grains have been through six cycles like these, from rock to sand to rock again. The rocks may wear down through the ages but the sand grain remains, on its onward voyage. It is perhaps the second most extraordinary journey in the Material World.

Why only second? Because, remarkable as this journey is, it pales in comparison with another epic odyssey. This journey begins with a piece of rock pulled from the ground; it ends up in your pocket, having been transformed into one of the most technologically advanced pieces of machinery in existence. Along the way it involves two, maybe three or more treks around the planet. This voyage doesn't merely span continents and time. It takes us into the furthest reaches of chemistry, physics and nanotechnology, involving processes so far out, so seemingly implausible, that they sound like science fiction. But don't be fooled; this is happening in the real world every day on a mammoth scale, because this, the longest odyssey of all, is the supply chain that turns silicon into silicon chips.

That these chips make the modern world go round is, by now, quite well-established. Everyone knows they represent the brains of our computers and smartphones; but few realise just how prevalent they are. They are *everywhere* . That main processor in your phone is only one of dozens of different chips, each controlling discrete functions. A modern car might have hundreds of chips: for the entertainment system, the navigation, controlling the engine, the windows and so on. Even in devices that don't have 'smart' in their name, mechanical linkages have long since given way to a network of semiconductors. They are not just the world's brain or even its nervous system but, increasingly, its sinews, veins and receptors. Nearly every economic activity, nearly every dollar of global GDP, relies in one way or another on the microscopic switches of semiconductors.

Glance at one of these wafers of pure silicon circuitry and you might struggle to reconcile this shiny, metallic object with the silicon we are more familiar with: the primary ingredient in sand, stone or concrete. Yet the marvel of silicon is that not only does it have unique properties that enable it to become a glass; not only is it strong enough to hold up buildings in concrete form; it also has electrical attributes that set it apart from most other occupants of the periodic table, for it is a semiconductor.

Semiconductors are anomalous materials that for many decades did little more than perplex scientists. They didn't conduct electricity like copper does, but nor did they insulate its current like, well, glass. No one could quite think what to do with them, but eventually they discovered that they worked brilliantly as a kind of switch. The first such switch, or transistor, as it was named, was made a couple of days before Christmas 1947 by Walter Brattain and John Bardeen, two physicists working under William Shockley at Bell Labs in the US. When you see it today (there is a replica in the Smithsonian Museum in Washington, DC) it looks somewhat ungainly, like a soldering experiment gone wrong. It dangles from a Perspex block: a mess of tangled wires plugged into what looks like a wedge-shaped piece of plastic, all sitting atop a dirty-looking slab of dark metal – in this case germanium.

But this first solid state switch represented a revolution. For, by combining enough of these switches, each one a tiny, physical manifestation of the binary code, zero or one, you could create a computer on a tiny piece of silicon, chipped off a circular wafer (hence 'chips'). These leaps of innovation, from the switch itself to the 'integrated circuit', the first of which was etched on to silicon by Robert Noyce at Fairchild Semiconductor in 1959, represented the physical foundation of the computing age.

That word, physical, matters here. Sometimes innovations are the fruit of a simple brainwave. As historian Anton Howes has pointed out, there was no inherent reason why the flying shuttle – John Kay's 1733 invention which revolutionised the weaving of wool – couldn't have been produced thousands of years earlier. But often it takes decades or centuries of material advances before ideas can manifest as working contraptions. That was the case with Leonardo da Vinci's sketches of a helicopter, which couldn't become a physical reality until the materials caught up, and it was the case with computers, which were conceptualised by Ada Lovelace and Charles Babbage more than a century before the invention of the transistor. There were computers with glass vacuum tubes playing the role of switches from the 1940s, but a solid state switch had long been the goal. Far more efficient, far more reliable, far smaller, with no moving parts, save for the electrons buzzing silently through the switches – the transistor was the material advance that birthed the modern age. The latest smartphones and devices and, these days, the latest cars and fridges are all reliant, in some cases utterly, on these small, thin slices of silicon, doped (a form of chemical infusion) with various other materials and etched with a set of microscopic transistors.[1]

That first contraption from 1947 was about the size of a small child's hand, but the part that really matters, the transistor itself, was perhaps about a centimetre. Now consider what happened in the following 75 years. By the time of the Intel 4004, the first modern computer chip, in 1971 there were just over 2,000 transistors crammed into roughly the same area, each single one about the size of a red blood cell. Roll forward to the early 2020s and smartphone processors could fit around 12 billion transistors into an area slightly *smaller* than a square centimetre.

This race to the bottom of the metric system, most famously formalised by Intel co-founder Gordon Moore, who noted in the 1960s that the number of components inside these integrated circuits was doubling at a constant speed, is one of the wonders of the modern age and it represents a virtuous circle. The smaller the transistors get, the better they get: they can be turned on and off again (remember each one is ultimately just an electrical switch) even faster and with less power.

Computer chips can be used for all sorts of purposes: at their simplest, they can act like a switch, turning on or off the headlights in a car. You are surrounded by more of these ‘power silicon’ chips than you probably realise, since they are in every single modern electronic product, from a hairdryer or vacuum cleaner to a power cable for your phone. There are semiconductors everywhere: in your smartphone there are many different varieties, some serving as the sensors in the camera, others storing photos of your family. Most exciting, however, are those that act as the brains of our devices, where the more transistors you have, the more calculations you can carry out and the more powerful the computer. The latest such chips can fit roughly 15 million transistors into a dot the size of a single full stop on this page. The transistors in today’s smartphones are not just smaller than a red blood cell (about a thousand times smaller, as it happens); they are smaller than the COVID-19 virus. Actually you could fit four of them inside a coronavirus, each transistor having about the same dimensions as one of the virus’s spike proteins, those club-like tendrils radiating out from its centre.

It is tempting to describe these features as microscopic but, at the risk of being pedantic, today’s transistors are even smaller than the wavelength of visible light and are thus totally indiscernible to the naked eye through even the most powerful conventional microscope. They are measured not in micrometres but in nanometres. According to Intel we are actually now only a few chip generations (new sizes tend to be developed every couple of years) from the ‘Angstrom era’, where transistors will be measured in angstroms, which are 0.1 nanometres, or 0.00000001cm. For reference, a silicon atom measures just over five angstroms.

Perhaps the reason so few people are astonished about this – for it truly is astonishing – is that these products are invisible to us, packed deep in the innards of everyday gadgets. Perhaps it’s because, unlike iconic engineering feats of old, like the 1970s jet Concorde, whose tickets were so expensive only the very wealthiest could experience supersonic flight, this marvel is something available to nearly everyone, even in the developing world.

Perhaps the reason Silicon Valley is oblivious to the march of the Material World is that these days no silicon chips are made in Silicon Valley. The last major semiconductor manufacturing plant there shut down a decade and a half ago, and while there are still some firms in the valley focused on the design of semiconductors, these days they are outnumbered by what is sometimes called ‘soft tech’: apps, platforms and services.

So if you want to trace the most extraordinary journey in the world you must look somewhere else entirely. Let us imagine you are lucky enough to have picked up the very latest iPhone (we are using the iPhone as the example here, but in practical terms the journey we’re embarking on is much the same for most new Android devices too – indeed their semiconductors are often made in the very same factories as those of their competitors). Look at the smartphone’s packaging. It says it was designed in California and assembled in China, but this is a vast oversimplification, for this miniature computer is a tapestry of technology that comes from all over the world. The display, the glass covering it, the battery, the cameras, the accelerometer, the modem and transceivers, the storage and power management chips; each comes from a different factory before they are assembled in China and then shipped to you.

Much of this activity does not happen in China or California. Indeed, it’s worth noting at this stage that pretty much all the physical components in the phone are not made by Apple itself, which is not really a manufacturer at all, but a brilliant re-packager of the technology made by other people. Even the chips that bear Apple’s name – the A16 Bionic was the latest iPhone chip at the time of writing – are in fact manufactured by another company altogether, Taiwan Semiconductor Manufacturing Company or, as it’s better known, TSMC. That company in turn was only able to make the chip with the help of machines made by another, even more obscure company, ASML. And at the heart of ASML’s machines are critical components made by other companies, some of which will be familiar (the lenses are made by Zeiss, with glass from Schott) and some less so (the lasers are made by another German company, Trumpf).

All of that covers only a fraction of the journey – the final steps our silicon atom takes before it is admitted into your smartphone. But our journey must begin not at the assembly plant or the silicon foundry where those tiny transistors are etched on to a silicon wafer, but with the very moment the silicon contained in that computer chip was first removed from the ground. It begins not in a manufacturing plant so sterile that there is barely a mite of dust to be found, but amid dirt, smoke and fire.

The Birth of a Silicon Chip

We are on a dusty stone track in a forest about 15 miles south of Santiago de Compostela. This part of Galicia is best known for the pilgrimage that draws more than a quarter of a million people from around the world each year to visit the tomb of St James the Apostle in Santiago's cathedral.

The hills are green and lush, dotted with medieval monasteries and pretty stone villages. Early in the morning the clouds here often sit low in the valleys, with the hilltops floating above them, seemingly suspended in the sky. The most distinctive of those hills is Pico Sacro, an outcrop you can see from miles around, a pyramid jutting over the fertile fields below.

Locals tell a fair few stories about the hill. One is that the two disciples carrying St James's body to Santiago appealed to Queen Lupa, the devious mythical ruler of this region, for a cart to help them. She sent them up to Pico Sacro, knowing there was a dragon there and hoping it would devour them. When the dragon appeared, the disciples made the sign of the cross, defeating it instantly. A different story goes that the original plan was to bury St James's body here at the summit until the dragon was discovered, forcing them into an abrupt change of plan.

The geological explanation for how this hill came to be is only slightly more prosaic. Around 350 million years ago this was where two supercontinents, Laurasia and Gondwana, collided, thrusting up the land and forcing a quartz outcrop hundreds of metres into the sky. Walk up the hill and you can see the white quartz stone alongside granite rocks. Look down towards the valley and there, alongside the high-speed railway, you see something else: a long, deep gash in the green fields and woods of the hillside, gleaming so bright it looks almost like a snowfield. Drive down the dusty road through the forest and eventually you come across a sign marked 'Serrabal'.

You hear the trucks and the clank of rocks long before you enter the premises. But once inside you realise what that blinding white expanse really is: an enormous rock quarry. Serrabal is a quartz mine. The vein of rock that lifts the Pico Sacro and its adjoining hillocks heavenwards is one of the purest quartz deposits to be found anywhere in the world, a rock so white that it is sought after far and wide.

The quartz pulled out of the ground here is sometimes used to make kitchen work surfaces. It is ground down into gravel for ornamental gardens and pure white sand for golf bunkers. But the real reason we have come here is for the bigger lumps of quartz that come from this hillside. These white, dusty chunks of stone are the raw materials that will eventually – months or more likely years down the line – become the next generation of silicon chips. This is the place we have been looking for, the beginning of our journey.

The company that owns the mine is Ferroglobe, a Spanish business which is the world's biggest silicon metal producer outside of China. This final distinction is all-important, for the vast majority of these raw ingredients for the technological revolution are mined and refined in China these days. Ferroglobe, which also has quartz mines in the US, Canada and South Africa, is one of the rare exceptions.

Quartz rocks like the ones you find here are not common, but nor are they incredibly rare. There are veins of quartzite in Norway, Russia, China, Turkey and Egypt. Surprising as this might sound, the rock that comes out of the ground here in Serrabal, snow pure as it is, has a slightly lower silica content than the sands at Lochaline, Fontainebleau and many other sand mines. But silica content isn't everything; if you want to make silicon metal, what matters far more is shape. Actually, what we're looking for here isn't exactly sand – at least as the Udden–Wentworth scale would have it – but more a sort of large gravel, with chunks the size of a cricket ball.

After they are blasted out of the ground, cleaned and emptied into lorries, the chunks are driven an hour or so north towards an industrial park just outside the port of A Coruña. This is the Sabón plant, a set of hulking blue corrugated iron sheds and warehouses. On the other side of the inlet is an enormous chimney belonging to a gas-fired power station. The fact that the Ferroglobe processing plant is next door to a power station is not a coincidence; you need a lot of power to turn quartz into silicon metal.

The rocks from Serrabal are emptied on to the floor outside the warehouses, a pile of white stone on the grey concrete. After a while they are mixed with coking coal (a baked form of coal) and woodchips and tipped into a furnace, heated up above 1,800°C. What happens in that furnace, where an electrical current is run into the mixture of quartz and coal, remains something of a mystery.

‘Even after more than a hundred years of production, there are still things people don’t understand about what’s happening in this reaction,’ says Håvard Moe, one of the directors of Elkem, a Norwegian company, which is another one of Europe’s biggest silicon producers. ‘It’s just too complex. There are lots of chemical reactions; it’s all happening under a very strong electrical field, which also affects the reaction. Making a mathematical model of this is quite difficult.’

The upshot, however, is that the now-molten silicon from those quartz stones parts with its oxygen and sinks to the bottom of the furnace, where it is released through a tap. For every 6 tonnes of raw materials thrown into the melt – quartz, coal and woodchips – about a tonne of silicon metal comes out. The mechanics of this furnace, by the way, are what explains why granular sand won’t work as the raw material for silicon chips. While there’s nothing wrong with its chemical composition, sand grains are simply the wrong size.

‘These are huge furnaces and there are convection flows of carbon dioxide bubbling through them,’ says Reiner Haus, a German scientist who is one of the few people in the world who understand the complexities of this supply chain. ‘If you were using sand it would be blown up into the filters and wouldn’t melt. So you need large, fist-sized lumps of quartz.’

If at this stage you’re starting to wonder whether this doesn’t all sound a little, well, industrial, like the production of steel or aluminium, then you have a point. These furnaces are roaring, smoking, red-hot cauldrons of molten metal and coal. You can’t get anywhere close without wearing a heatproof suit, a little like the kind you sometimes see volcanologists wearing near the crater of a spewing volcano.

In fact, the terminology used to describe processing quartz does evoke volcanoes. The process of getting silicon from quartz rocks is called smelting. The heart of the furnace is called the crater. Watching these rocks being melted into their metallic form is a little like witnessing the industrial revolution in all its glory and brutality. It is, as one industry analyst told me, ‘like the Middle Ages. There are guys heaping coal. It’s like the mines of Moria from *The Lord of the Rings* .’

Nor are the consequences of all this smoke and heat trivial. It takes about 45 megawatts of electricity to power one of these furnaces – enough to power a small town. It is, says Moe, frankly impossible to turn quartz into silicon at scale without emitting carbon dioxide, which means that even if the electricity powering the furnace is generated from hydropower, these silicon smelters would still be contributing to global carbon emissions. But few pay much heed to the carbon emissions produced during the creation of a silicon chip. And anyway, we are still only at the beginning of this journey.

After the silicon metal produced by the furnaces here at Sabón is poured out of the furnace and solidifies, it is smashed up into a kind of granulated metal. At this stage the silicon is around 98 to 99 per cent pure, which sounds pretty good to most of us, but is a long way from the purity you need for a silicon chip or a solar panel.

Purer Than Pure

On to the next stage of our journey, with the metallurgical-grade silicon whisked away from Ferroglobe to another company most people have never heard of, where it will be turned into an even purer version of silicon known as polysilicon. The firm in question is Wacker, a German business which produces more polysilicon than anyone else outside of China. Its main plant is in Burghausen, an hour and a half east of Munich, just on the Austrian border. This town prides itself on having the world's longest castle, a handsome fortress that snakes along a ridge overlooking a bend in the River Salzach. Burghausen has always been two things: small and rich, for this was one of the main hubs on the ancient salt route between Austria and Germany. Having been ferried down the Salzach ('salt river') on barges, salt was unloaded here, taxed and then sent on its way. The old salt trade may be long gone, but Burghausen is still one of Germany's richer towns, because, like so many former salt-trading communities, this is now a hub for the chemical industry.

Actually that somewhat understates it, for the chemical plant at Burghausen is so vast it takes up about the same footprint as the town it sits alongside. There is a power station, an oil refinery, there are more chimneys than you can count and neighbourhood after neighbourhood of chemical silos and sheds. There is even a football stadium for the plant's club (SV Wacker Burghausen), which plays in the Bavarian League. And this – the chemicals plant, not the football pitch – is the next stop for the silicon metal from Spain.

What happens next is known as the Siemens process and it involves breaking down that pure silicon metal into its elemental pieces and reforming them all over again. The metal is ground into a powder, mixed with pure hydrogen chloride, distilled and then heated up in a bell jar to 1,150°C. At the end you are left with long rods a little like the heating elements in an old kettle, except that the material here is not furred-up limescale but ultra-pure silicon.

Tearing apart atoms and reconstituting them, which is essentially what happens here, is another energy-intensive exercise. According to scientist Vaclav Smil, the energy cost of ultra-pure silicon such as this is more than 3,000 times that of cement and 1,000 times that of turning iron into steel. It's true that the sheer quantities here may be smaller, but this is a demanding, costly and frequently dirty process. The end product, however, is silicon of astonishing purity – purer than nearly anything else on the planet – which can now be called polysilicon. Depending on how thoroughly the solution is distilled, you get a variety of grades, each of which is named on the basis of how many nines there are in the purity number.[2]

There is solar grade polysilicon for multicrystalline cells, with up to eight nines (99.999999 per cent pure silicon). There is monocrystalline solar grade polysilicon, which has up to nine nines (99.9999999 per cent). Indeed, the vast, vast majority of polysilicon goes towards solar panels, and the vast, vast majority of that is made in China. What is striking, however, is that China has yet to master the manufacture of the *pièce de résistance* of the silicon world: semiconductor grade polysilicon. This can have as many as ten nines (99.99999999 per cent purity), where for every impure atom there are essentially 10 billion pure silicon atoms.

But even after being exploded out of a rockface, melted down in a roaring furnace, smashed into oblivion, ground down and dissolved into solution, distilled at extraordinary temperatures and snapped into pieces, our silicon is still not quite ready to become a silicon chip. We have barely passed the midway point of our journey.

The Purest Substance in the World

At this stage you could be forgiven for asking: why all the fuss over a few decimal points? Is it honestly worth the effort? Would anyone actually notice if we skipped a step? Short answer: yes, absolutely. A lone, rogue atom in an otherwise pure silicon matrix is enough to disrupt the flow of electrons in a transistor. If one of the secret weapons of cement is that it is a surprisingly forgiving recipe, where you can usually get away with a slight deviation from the strict instructions, the very opposite is the case here. And purity isn't the only thing that matters: so too does structure. The more perfect the atomic structure in your silicon, the more easily and freely electrons can flow around. The more defects – so-called grain boundaries – the greater the chance of that flow being disrupted and the semiconductor conking out. Picture eggs packed neatly in a box, as opposed to lying around higgledy-piggledy.

This is another one of those recurrent reminders of how much the material bit of the Material World really matters. Most of the initial efforts to create transistors were hamstrung not by a shortage of brainpower or a lack of imagination but by the absence of truly reliable materials. That very first transistor made by Brattain and Bardeen at Bell Labs in 1947 was made not of silicon but germanium. However, germanium was ill-suited for use as a transistor. It didn't function very well at high temperatures, something that is deeply inconvenient given semiconductors can, as you'll know if you've worked your laptop hard while perching it on your knees, get very hot. Silicon, with its high melting point, was a far more attractive material, at least in theory. But this was back before the Siemens process, so, frankly, those first semiconductors were impure and hence a bit rubbish.[3] [fn1]

That brings us back to our silicon, which began its life in a quarry in Spain but has now been blasted and reconstituted into ultra-pure polysilicon. The next task (and the next leg of the journey) involves rearranging its pure but higgledy-piggledy atoms into a perfect matrix. That means flying the polysilicon to the other side of the planet, to a plant on the north-west coast of the United States, just outside Portland, Oregon. We are just on the other side of the mighty Columbia River, in a complex of grey buildings that sit amid the suburban sprawl. The name on the outside is another one of those companies few have heard of but which are titans of the Material World: Shin-Etsu.

This Japanese company is, among other things, one of the world's leading producers of silicon wafers. Indeed, it is arguably here, on the banks of the Columbia River, that you will find the epicentre of the twenty-first-century US silicon industry. That we are by this river, which rolls down from the Rocky Mountains towards the Pacific, powering 14 hydroelectric dams along the way, is no coincidence. For turning polysilicon into wafers

– the pure, crystalline structures, which can then be sent to semiconductor foundries – is, like most other links in this extraordinary supply chain, another energy-hungry process.

The air is fresh here but the silicon has already bidden its farewell to the outside world with all its imperceptible microbes and dust. It is now so pure that it needs to be handled with extreme care; from hereon it will be kept in sterile conditions, in ultra-clean factories and beneath protective seals, until the moment it reaches your front door.

The engineers at Shin-Etsu are some of the world's finest practitioners in something called the Czochralski technique, or CZ as they prefer to call it. 'Everyone here just calls it CZ,' says Neil Weaver of SEH, the US division of Shin-Etsu. 'I actually just had to look up how to spell it.'

There are many weird and wonderful manufacturing techniques in the Material World but the Czochralski process is among the most captivating. The polysilicon is tipped into a quartz crucible (the crucible must be incredibly pure, or else it may introduce impurities back into the silicon) and heated up to just under 1,500°C. A seed crystal, a pencil-sized rod of silicon, is dipped into the melt and is then slowly pulled upwards, rotating slightly. Gradually, a perfect, solid ingot, a boule, begins to form out of the melt.[fn2]

Perhaps the best analogy is someone pulling candyfloss on to a stick, except that this is the very opposite of a fairground filled with screaming children; the Czochralski process occurs inside a chamber filled only with argon gas. The boule slowly turns and is gradually lifted, until eventually above the crucible there is an extraordinary-looking, long, shiny, dark metallic cylinder hanging by a thread only a few millimetres thick. But only when you use X-ray diffraction to examine this torpedo, this silicon sausage, do you see the most extraordinary thing: the atoms are arranged into a quite perfect crystal.

This silicon boule is about 2 or 3 metres tall by the time it is finished, but soon enough a silicon carbide wire-saw carves it into incredibly thin slices, each less than a millimetre thick. These circles, each about the circumference of a small pizza, are buffed and cleaned with chemicals until they are absolutely flat, and there you have it: what left the ground as a chunk of quartz is now a silicon wafer. This is, as you might imagine, an enormous oversimplification. In fact, the silicon will spend months inside

Shin-Etsu's laboratories, being shuttled from one machine to another, being pulled, sliced, smoothed, cleaned and tested. But you get the idea.

For much of the past 70 years or so, these processes were as much an art as a science. In the early days of Silicon Valley, back when most chip manufacturers made their own silicon wafers, crystals were mostly pulled by hand, with machine operators (almost all of whom were women) eyeballing the molten crucibles through black glasses in an effort to judge whether they were lifting and turning the crystals at the right speed. The most skilled crystal pullers swiftly gained a reputation. These artisans became prized employees, fought over by the emerging tech firms as they sought to improve the quality of their wafers.[4]

That, of course, was a long time ago. These days none of the companies whose names appear on silicon chips has much to do with the manufacture of the silicon, leaving it instead to Shin-Etsu or a handful of other companies from Germany, Singapore or Korea. The process occurs in sterile rooms where few human beings – especially those from outside these companies – are ever admitted.

When I ask Neil whether I could have a brief peep into the room where they do the Czochralski process, he simply laughs. 'It won't happen. We're really paranoid about these secrets getting out. There are all these trade secrets and we really want to protect them.'

Perhaps this is to be expected: so much of what goes on inside there is enormously valuable intellectual property. Everything from the speed at which that growing boule of silicon is rotated and lifted and the way the temperature of the bubbling crucible is managed to the methods used to avoid any defects turn out to be trade secrets, and some companies – some countries, even – will go to extreme lengths to uncover them.

It doesn't take much imagination to realise which country Neil is most concerned about. In the past couple of decades, China has come to dominate much of the silicon business. About 90 per cent of silicon production these days is not for computer chips but for solar panels, and nearly all of that takes place not on the east coast of the US but in China. That has a couple of important implications. First, while much of the silicon in Europe is produced using alternative energy sources, especially hydropower, the Chinese silicon industry is far more reliant on coal for the enormous amounts of power needed to turn quartz into polysilicon. Producing silicon is a dirtier business than you might have thought, but especially so in China. Second, there are concerns that some of China's silicon producers, especially those in the Xinjiang Uyghur Autonomous Region, have inhumane working conditions.

Indeed, as this book was being written the US imposed a ban on imports of silicon from Hoshine, one of China's, and the world's, biggest producers. The White House alleges that the company has intimidated and threatened its workers. The Chinese producers, and Beijing itself, have in turn accused the US of imposing these sanctions in an effort to protect its own economy and stifle China's.

But here's the thing: while China controls much of the global supply chain of metallic silicon and solar polysilicon, it has yet to crack the manufacturing techniques needed to create wafers for the most advanced silicon chips. In much the same way as it has yet to master the processes Wacker uses to turn out polysilicon with less than one in a billion atoms of impurity, it has not yet refined the Czochralski process enough to produce wafers as perfect as those pulled out of the crucibles at Shin-Etsu. Which is why you or I won't ever be allowed into this holy of manufacturing holies: for fear of industrial espionage – that these methods and secrets could be stolen and replicated in a factory in Xinjiang.

There is another thing preventing China from dominating this end of the market, and it comes back, once again, to sand – a very particular type of sand. For those crucibles in which Shin-Etsu melts the hyper-pure silicon before pulling it into that perfect boule and slicing it into wafers are all made – every single one of them – out of a very particular type of quartz, one you can only get in a single place in the world.

It is rare, unheard of almost, for a single site to control the global supply of a crucial material. Yet if you want to get high-purity quartz – the kind you need to make those crucibles without which you can't make silicon wafers – it has to come from Spruce Pine, a small town on the Blue Ridge escarpment in North Carolina. For a long time, this mine – and by extension the entire global supply of high-purity quartz – was operated by a single company, a secretive Belgian business called Sibelco.

You don't have to go far around Spruce Pine to hear rumours about the lengths to which Sibelco will go to protect its privacy. Few people are admitted to the mine, still fewer to the facilities where the quartz is processed and ground into the high-purity product out of which those crucibles are made.

According to one person who has done business with Sibelco, going to their headquarters was ‘a bit like getting into Fort Knox’. There are 25-foot-high fences surrounding the complex, ringed with barbed wire, and there are security cameras and frequent security patrols.

‘When contractors from other companies are brought in for repairs [at the plant] they are literally blindfolded and marched into the factory up to the machine they need to fix,’ says another insider. ‘It’s like something out of Willy Wonka.’

Why this cloak-and-dagger behaviour? Why the secrecy? According to Reiner Haus, the silicon analyst: ‘If you have a monopoly, why would you want to talk to anyone? There’s no need to market your product. There’s no need to talk to anyone.’

These days there are two mines at Spruce Pine. Alongside Sibelco is another, smaller operation: the Quartz Corp, which ships the stones it mines to Norway for processing. The rocks, which look a little unremarkable from the outside – more like granite than the bright white quartz of Serrabal – are washed, crushed, ground up, magnetically separated and rinsed in chemicals. Eventually they become a special type of sand – one pure enough to hold that molten silicon as it is pulled up into a silicon wafer. There are five or ten companies scouring the geological records of other countries around the world to find alternative seams of quartz that might match the quality of these rocks. China has tried for decades to produce quartz of a similar standard, to no avail. It turns out that while it is pretty hard to find sources of quartz as pure as the snow-white rocks of Serrabal, it is nigh on impossible to find quartz as pure as that of Spruce Pine.

In the Material World, there are few such cases where we are so utterly reliant on a single place. There are a few micro producers in India and Siberia, but nothing to rival the consistency and quality of the two mines in Spruce Pine, which raises some unsettling questions. What if something happened to those mines? What if, say, the single road that winds down from them to the rest of the world was destroyed in a landslide? Short answer: it would not be pretty.

‘Here’s something scary,’ says one veteran of the sector. ‘If you flew over the two mines in Spruce Pine with a crop duster loaded with a very particular powder, you could end the world’s production of semiconductors and solar panels within six months.’ No high-purity quartz means no Czochralski crucibles, which means no monocrystalline silicon wafers, which means, well, the end of computer chip manufacture as we know it. We would adapt; find a new process or an alternative substance. But it would be a grisly few years. Perhaps this is why those who work in high-purity quartz are so jumpy. Perhaps that's why the man who passed on that scary thought exercise insisted that I didn't print the type of powder that would play such havoc with the processing of those mines in North Carolina, which quietly serve this tiny but pivotal role in the functioning of the modern world.

Crunching Through Wafers

Our silicon, which began its life (or at least this chapter of its life) on the side of a Galician mountain and has been through numerous transformations – from solid to liquid to solid to vapour to solid to liquid to solid again – now finds itself inside a sealed canister on the other side of the world. We are on the outskirts of Tainan, the old capital of Taiwan. Head north from the city centre and the office blocks and houses soon give way to fields of sugarcane and cabbage. The air is sticky and for a moment it feels almost as if you are heading for the middle of nowhere. But no: for anyone who works in computing, this is the centre of the universe.

Rising from the fields ahead of you there is a gatepost and, some way behind it, a complex of shiny, silvery buildings. The full name of the site is Southern Taiwan Science and Technology Park, but it is better known as the main production hub for a company whose name is emblazoned in red on the buildings: TSMC. This is Fab 18 – the most advanced factory in the world.

That the building is here at all is testament to a man called Morris Chang, who founded the Taiwan Semiconductor Manufacturing Company in 1987. Having moved to the US from China in 1949, Chang climbed the rungs of the American silicon industry, eventually becoming the head of the semiconductor division at Texas Instruments. But when he was overlooked as chief executive, suddenly, at the age of 51, he found himself at a career dead-end. Then the Taiwanese premier called and asked whether he could help set up a semiconductor industry in his country.

Given computing was then so entirely dominated by American firms, the proposal was somewhat bold. Taiwan had little experience of this kind of engineering; nor did it have much of a skills base. But over the following years, in much the same way as the Prussian state helped the glass industry of Jena, the Taiwanese government steadily supported TSMC. What really set it apart from established rivals like Texas Instruments or Intel was its business model: while those companies designed and manufactured the chips themselves, TSMC would make chips for other people – it would be a ‘foundry’.

If you are looking for a company that embodies the Material World, you could hardly do better than TSMC. Here is a business whose sole purpose is to manufacture the processors dreamed up by Apple or Tesla, or ‘fabless’ chip companies like Nvidia and Qualcomm (‘fab’ being short for fabrication plant). Obscure enough that few outside the computing sector have heard of it, it has nonetheless pushed the boundaries of physics, in the process becoming one of the world’s most valuable, and most important, companies. But this dominance does not come cheap. Over a three-year period from 2021, TSMC was budgeting to invest $135 billion, more than many developed countries would outlay over that period, and the equivalent of ten US Gerald R. Ford-class aircraft carriers.

How on earth does a single company, based mostly in a few sites in a single country, spend these kinds of sums in such a short space of time? The answer is here before you in the sweaty fields of southern Taiwan, for Fab 18 is very, very expensive. Actually Fab 18 isn’t a single building but a whole complex of six interconnected units, some of which were still under construction at the time of writing. The budgeted cost is $17 billion, which makes this factory a touch more expensive than the Channel Tunnel (adjusted for inflation). The likelihood is that another factory will overtake Fab 18 in cost terms in the next few years and the likelihood is it will be another semiconductor plant, possibly another TSMC plant. Such is the logic of Moore’s law, where every two years the transistors must get smaller and the factories more expensive.[fn3]

The buildings themselves are each about the size of a multi-storey carpark, but were you to peel away that silver exterior you would discover that a large proportion of the space inside the shell is given over not to working space, offices or, for that matter, cars but to enormous filtration and air-conditioning systems, designed to keep the most critical part of the building utterly, utterly clean. An office building might have its air circulated and filtered five or six times an hour; a class 1 cleanroom, which is what this is, has the air changed 600 times every hour. The cleanroom at Fab 18, which, when you combine all the different units, will be about the same area as 25 football pitches, is one of the cleanest places on earth – provided you ignore the toxic solvents about to be doused on the wafers. That brings us to what you'll find beneath the cleanroom: a whole other floor known as the sub-fab, where the cocktails of chemicals used to wash and treat the wafers are constantly churning and sloshing their way up to the waiting machines above. A semiconductor plant without chemicals is essentially useless: without them there would be no transistors whatsoever.

Beneath the sub-fab is one of the most sophisticated set of dampers on the planet, meaning the edifice is almost entirely disconnected from *terra firma* , which, given this is one of the more seismically active parts of the world, is no bad thing. Any movement, however indiscernible, can disturb the workings of the machinery here, which is why you don't tend to find fabs very close to airports or motorways.

Those machines are comfortably the most expensive thing in the fab. There are whole rows of them, most of them white, self-contained units about the size of a minibus. Some of them implant chemicals on to the wafers – doping, as it's called – others deposit nanolayers of material and some use lasers to etch circuits on to the silicon. Back in the 1950s, semiconductor assembly involved lines of women working at desks, using tweezers to manipulate transistor wires into place. In today's fabs you might see one or two human beings, kitted up in white body suits so as not to introduce impurities to the cleanroom, but almost everything here is done by robots. These plants are so automated they are known as 'lights-out' fabs – where they can operate almost entirely without workers. It is a surreal, somewhat dystopian vision: a world without humans. Yet from a silicon wafer's perspective, human beings with their grimy nails, their flaky skin and tainted breath are walking vectors for impurities. All you need in one of these fabs is for a single stray atom to float into the machinery and thousands of dollars' worth of transistors will be instantly ruined. If all goes to plan, our silicon wafer will not be touched by a single human hand until its surface is sealed up and it is ready for dispatch.[5]

That moment won't come for some time, however, for a single wafer such as ours can spend three or four months whizzing around a fab like this, being carried from one machine to another in a sterile box called a FOUP (a front-opening universal pod). Arguably the most important of these machines carry out what is known as photolithography. For some decades, transistors have been etched on to silicon wafers not by hand or by physical machinery, but by light. The principle is a little like a movie projector, except in reverse. While a projector takes a small image and uses lenses to blow it up to a cinema-sized screen, photolithography begins with a big blueprint of a silicon chip, with all its transistors and features, and uses lenses to project that image down into mind-bogglingly small dimensions. When that computer-chip-sized micro-projection hits the silicon, with the help of some laser light and chemicals coated on to the surface, it fries away tiny channels and grooves, effectively engraving the circuitry on to the wafer.

This might just sound like a convenient analogy, but in the 1950s at Fairchild Semiconductor, Gordon Moore and his colleague Robert Noyce would buy up old 16mm movie camera lenses to make some of their earliest chips. Early Silicon Valley chip designers quite literally drew their initial designs, transistor by transistor, on to a blackboard-sized piece of film, and then allowed the lenses to do their magic. Today those designs are so intricate that were you to do the same exercise with, say, one of Intel's computer chips, you would need a blackboard about the height of the world's tallest building, the Burj Khalifa, and more than a kilometre wide. These so-called photomasks – the modern-day equivalent of those blackboard-sized films chip designers used to draw on – are made out of fused silica, itself made from sand. Sand upon sand upon sand.[6]

How on earth can one physically beam that kind of detail down on to a chip the size of a few centimetres? With one of the most expensive machines in the world. The TWINSCAN NXE:3600D – made by ASML, a Dutch business which used to be part of Philips – costs hundreds of millions of dollars. That might sound excessive, given it is ultimately just bouncing light around a box, but this is no ordinary light and no ordinary box. After all, recall that the transistors TSMC wants to make are so small they are quite literally invisible, so a conventional wavelength laser and a series of lenses will no longer do. In order to get the finest of all resolutions you need light with the smallest of all wavelengths, which in this case means extreme ultraviolet (EUV) light.

Working with EUV light is so challenging that, in much the same way as some people assumed in the 1950s that we would never master the manufacture of perfect silicon wafers, up until recently some people assumed we would never have machines capable of doing what ASML's bus-sized unit does. The first and most obvious challenge is creating the light itself, which cannot simply be generated by a laser unit. Instead, we need to enter a kind of parallel universe, which sounds far more like something from an Arthur C. Clarke novel than a manufacturing production line in the real world.

In a vacuum chamber inside this machine, tin is melted until it becomes a liquid. That molten tin is then dropped down into the chamber in a continuous stream. In the midway point of their cascade, each of these tiny droplets is zapped twice by pulse lasers, provided by German company Trumpf, which are powerful enough to cut through metal. These bursts heat the tin up to a million degrees, transforming it into a kind of plasma that simultaneously creates a burst of EUV light. This pinpoint smashing of molecules happens 50,000 times per second, so fast that the stream of tin droplets and the laser explosions are totally indiscernible. All of this to generate a stream of EUV light whose real task is yet to come, for only then is it bounced out towards our waiting wafer.

Actually, calling it light is a bit misleading for it is more like a kind of radiation, a little bit like an X-ray. And like an X-ray, EUV light has a habit of being absorbed by most solid materials, including most lenses. This is where sand makes yet another cameo appearance, for in order to bounce this EUV light down on to the wafer, ASML has contracted Zeiss to produce a set of special mirrors called Bragg reflectors, made from layers of silicon and molybdenum.

Quite how these mirrors are made is yet another closely guarded trade secret, but according to Zeiss, they are ground down from blocks weighing 50 kilograms, and robots are used to polish and correct the outer layer with ion beams. Suffice it to say, they are, according to one ASML engineer, ‘probably the smoothest man-made structures in the universe’. If you blew one of them up to the size of the United States, the biggest bump would be less than half a millimetre high. Having bounced off a staircase of these mirrors, the EUV light in all its 13.5 nanometre glory hits the wafer and etches that intricate design on to it. There is more than a whiff of science fiction about all of this – astoundingly perfect silicon wafers being manipulated by mind-bogglingly flat pieces of glass – yet there is nothing fantastical about it at all. And here at the heart of the silicon supply chain is that very same company that produced the glass in those binoculars the British government secretly traded for rubber during the First World War.

As of this moment, ASML is the only company in the world capable of making these machines, and TSMC is, alongside Samsung, the only company capable of putting such technology into mass production. Intel, which long dominated the industry, is often regarded as lagging at least one generation behind and, despite having been involved in research on EUV from the very earliest days, it is struggling to use those machines to mass produce chips. And then there's China, whose leading chipmaker, SMIC, is prohibited by a suite of US restrictions from buying any of the machines at all.

While in theory the export ban will likely hamper China's efforts to catch up with the Taiwanese and South Korean industries, according to some senior figures in the business, the gap had already been widening rather than narrowing for some time. A decade ago SMIC's technology was four years behind TSMC's. Today SMIC is thought to be 10 or 12 years behind TSMC, despite a flood of government money being unleashed on the sector. In much the same way as they vied with each other to build the biggest bridge or the most high-speed railway lines a few years ago, today Chinese provinces are all building new fabs. The problem, however, is that they are struggling to find the expertise to run them.

A senior executive from one of the world's leading semiconductor companies says: 'If I quit my job tomorrow I would go straight to China and convince any and every province to give me $10 billion to build a fab. Do I know how to build a fab? No way. But that doesn't stop the rest of the semiconductor con artists over there.

'They're building state-of-the-art clean rooms but no one knows how to use them. Most of these new fabs are just big buildings filled with unopened boxes. They have no idea what they're doing.'

One theory is that part of the reason Taiwan has succeeded where China has failed is that it simply struck it lucky with its timing; back in the 1960s and 1970s when it sent many of its graduates to university in the US, they ended up studying engineering and working at companies like Intel and Texas Instruments. They brought that technical knowledge back to Taiwan.

But by the time China opened up and began sending students to the US in the 1990s and 2000s, the American tech industry had changed. The companies in ascendancy were software firms like Microsoft, Amazon and Google. That generation of Chinese students came home and, rather than establishing a hardware industry, they used what they had learned in the US to set up internet services firms. Rather than building the new titans of the Material World, this generation of Chinese entrepreneurs built retail giant Alibaba, TenCent (which owns WeChat) and ByteDance (which owns TikTok).

In steel production, cement, manufacturing, distribution and even social media, China has managed to catch up and even overtake the rest. But not, crucially, in semiconductors, for while it has begun to dominate in less complex, lower value silicon chip manufacture, Chinese fabs still trail those leading-edge designs, however much money and effort they expend. What separates Taiwan and China is not merely an ocean strait but a gulf of technology, which further intensifies the tension in this part of the world. As Morris Chang put it in 2019, 'As the world is no longer peaceful, TSMC is gaining vital importance in geostrategic terms.'[7]

And the scale of this reliance is greater than you might imagine. China spends more money on importing computer chips these days than it does importing oil. Indeed, according to Chris Miller, the author of a history of silicon chips, China's semiconductor import costs as of 2017 were greater than Saudi Arabia's total revenue from oil exports, or for that matter the entire *global* trade in aircraft. 'No product,' he says, 'is more central to international trade than semiconductors.'[8]

Back in Fab 18, our wafer has left the EUV machine but its journey is still not over, for after this first pass it is washed and dried, another layer of chemicals is deposited and then the process repeats itself all over again. Layer upon layer of detail is added, sometimes with EUV, sometimes with other lasers. Sometimes detail is deposited, atom by atom, in other bus-sized machines made by companies like Applied Materials from California. Weeks pass, then months, and still our silicon is clocking up miles within the foundry, being shuttled from one machine to another. The process to create the world's fastest technology turns out to be surprisingly slow and laborious; by one calculation there are more than 10,000 different steps over these months. By the time the work at the fab is finished, months after it began, there may be as many as a hundred different layers of transistors on our silicon wafer, though each is so ineffably thin that, to the human eye, the whole thing looks little bigger than when it entered.

There is much debate these days about whether these continual improvements in miniaturisation will eventually cease, bringing Moore's law to a shattering end. Some wonder whether we will soon reach silicon's atomic limits and have to seek out an alternative material. Some talk about 'compound semiconductors' such as gallium arsenide. Others suggest that the future may lie with the material in that first, primitive transistor: germanium. Then again, scientists have been saying such things for decades, and for decades silicon has been the main game in town. Though it is far from the only game.

As transistors have multiplied, so too has the number of chemicals deposited during the average chip's journey through a fab. These days your average chip may be made up of 60 elements, compared with around 15 in the 1990s and perhaps 11 in the 1980s. The typical smartphone, with its display and battery, might contain as many as 70, making it one of the most advanced manifestations of chemistry in history. But for the time being, we invariably end up back with silicon. Even quantum processors, the next generation of computers which function not in a binary world like most transistors but somewhere in-between, are still reliant on silicon wafers. These extraordinary machines, which have to be supercooled down to just above absolute zero (-273°C), may have circuitry of aluminium and niobium, but it all still sits on silicon.[9]

With the currently available technology there are still a few generations' – maybe a decade or so – worth of miniaturisation left. What then? One option is to do what cities do when they run out of landmass, and to build up. To see how this might be achieved, consider the glimpse of future chips served up by IBM, which used to make computers but these days does lots of the semiconductor research that companies like Intel then adopt. It has built prototype chips whose transistor gate length is 12 nanometres, though due to the somewhat illogical naming convention given by chipmakers, this would qualify them as 2-nanometre chips. The upshot is it can fit 50 billion transistors into a space roughly the size of a fingernail.

When you look at these transistors in an electron microscope (and that is the only way to see them) they look a little like a triple cheeseburger without a bun – three slices of silicon stacked on top of each other, with a chemical coating in-between each slice. Of course, these are no ordinary burgers: each slice is about as thick as two strands of DNA. The advantage of stacking the transistors in this way is that you can go on stacking them higher and higher, creating an ever denser computer chip, much faster and more efficient than today's equivalents.

Layer upon layer, material upon material, the onward march continues, as does the journey of our silicon wafer, now etched with all its transistors and ready to be sliced and diced into one or two hundred chips, each of which is placed into a protective shield. The transformation from lump of quartz into a computer processor is now complete. Its journey still isn't quite over, however. For now this chip is shipped to another factory altogether, this one in Malaysia, where it undergoes further inspection. The fragile silicon surface is covered with a protective layer and wiring is added so it can be attached to a circuit board. This fingernail-sized marvel is finally more or less finished, ready to become part of a smartphone. From Malaysia it is then shipped to the assembly plant in China. In these enormous, town-sized factories, most of them run by a company called Hon Hai Technology Group, better known as Foxconn, they are attached to logic boards alongside a dozen or more other chips from companies like Qualcomm and Texas Instruments.

Few, even in the industry itself, understand the length and complexity of this journey, the number of processes involved, the quantity of companies playing a part. The media frequently writes about Apple, sometimes about Foxconn. Occasionally specialist outlets write about TSMC and maybe even ASML. They cite the centrality of Taiwan and the Netherlands to the semiconductor supply chain. But this is only the tip of the iceberg, for there are hundreds of other companies without whom these somewhat more prominent parts of the supply chain would be unable to function. What about Linton Crystal, which makes the furnaces for the Czochralski process and the diamond saws that slice up the boules? What about JSR, one of the world leaders in photoresist technology? What about EV Group and IMS Nanofabrication, which dominate wafer bonding and mask production and are both based in Austria? What about all the other firms providing critical machinery for the fabs, which read like a list of mysterious names and acronyms: Veeco, Tokyo Electron, Lam Research, ASM Pacific, Applied Materials and Edwards …? Remove one or two of these companies and, well: no more computers or smartphones.

This is worth pondering, for the world's two leading superpowers both talk with increasing volume about attempting to 're-shore' that supply chain. Joe Biden wants to bring semiconductor manufacture back to the US, introducing legislation in 2022 to encourage more investment in the business. Xi Jinping has a 'Made in China 2025' policy, which promises to increase China’s dominance and self-sufficiency in the manufacture of everything from complex machinery to semiconductors. Yet it is hard, mind-boggling, even, to imagine compressing the journey we have just experienced into a single country, without relying on companies or imports from other parts of the world.

Even if China invaded Taiwan and even if TSMC’s fabs survived the assault (some have suggested that the company incorporates explosives into the foundations, to be detonated upon invasion much as armies destroy bridges before retreating), that would not resolve its issue. Fab 18 might be where the world’s most advanced chips are made, but they are mostly designed elsewhere, primarily in the US, with intellectual property that derives from a company based in Cambridge, England: ARM. TSMC’s fabs would not function without machine tools from the Netherlands and Japan, or chemicals from Germany and bits and pieces from a range of other nations. There is only a handful of companies capable of making perfect silicon wafers, and none is headquartered in either the US or China. And there is only one site in the world capable of making the quartz sand for the crucibles where those wafers are crystallised. When politicians talk lazily about re-shoring, it often betrays a deep ignorance of what is happening out there in the Material World.

Tempting as it is to conclude that the journey we have just been on is unique – an extreme example of economic intricacy in the twenty-first century – it is anything but. Given how ubiquitous semiconductors are, it turns out that most of the gadgets and tools in your life, not to mention the solar panels bedecking many of our buildings and fields, have been on a similar ride before ending up in your hands. Arduous and wondrous though it might have been, the longest journey of all is not at all unusual. This is just how our world fits together.

And the deeper one delves, the clearer it is that each of these supply chains is interwoven with another. We are in a web, not a chain. There would be no silicon chips without the roaring coal-fuelled furnaces turning quartzite into metallurgical silicon. There would be no polysilicon without the hydrogen chloride we use to dissolve it and initiate the Siemens process. There would be no semiconductors without the chemicals and gases being pumped up into the cleanroom from the sub-fab level below it. And where, pray, do many of those chemicals come from? The answer might just be sitting on your kitchen table.