Content Unit

[Nano in Electronics]

Name of author: [Christopher Fischer, Andrea Deißenberger, Peter Grambow]
Last processing date: [05.11.2021]

Nano in Electronics has a history of about 56 years.

First introduction

Nano in Electronics has a long history. One of the characteristic milestones about 56 years ago was an important statement from Gordon Moore. He, as one of the cofounders of Intel (one of the most important semiconductor companies worldwide), postulated his idea of industrial growth in this field in a simple rule. He expected that about every two years the storage capacity of an IC (integrated circuit, an important electronic device) will increase by a factor of two. This was an amazing statement at that time, because most of the people had no idea how this should be possible. The boundary conditions for this rule are still driving the electronics industry worldwide. The maximum storage capacity and the velocity of data transmission are still increasing. Without these developments smartphones or computers would not exist.

Practical relevance – This is what you will need the knowledge and skills for

Modern electronic products, like smartphones and tablets would not work without the deep knowledge of nanotechnology. Therefore, it is important to understand the meaning of structure sizes in this area.

Overview of learning objectives and competences

One of the most important factors in electronics nowadays is size. Producing the structures as small as possible is the only way to be able to control high amounts of data like in a smartphone or in a tablet computer. You will learn the key features in the game of electronics. What are the most important developments today and where are the physical limits in the future.

Learning objectives	Fine objectives
LO_Examples in Electronics 01	FO_Storage capacity 20 years ago and now_01_01 FO_Features of computers today and 30 years ago_01_02 FO_Smartphones and phone boxes_01_03 FO_Current sensor examples / phyphox_01_04
LO_Transistor, the magic device 02	FO_What is a computer?_02_01 FO_What is an IC?_02_02 FO_What is a transistor?_02_03
LO_Size matters especially in electronics 03	FO_Smallest visible electronic structures by the naked eye_03_01 FO_Limitation of light microscopes_03_02 FO_One of the challenges in electronics today is to build structures as small as possible. How does that work?_03_03

	FO_Explanation of Moore's Law_03_04 FO_Where are the limits? What can we learn from the Greeks?_03_05

1. Examples in electronics

In this section we want to understand how nanotechnology impacts electronic applications. A good way to approach this topic, is to examine how computers became so powerful over a short period of time. So, the computer basically is built from several "chips", which are also called "integrated circuits". We will discuss them in more detail later. For now, we just need to know that the key elements of those chips are "transistors". These are the building blocks, which allow us to control if an electrical current is flowing through a certain conducting line or not. So, there are two cases for the computer: electrical current or no electrical current. This is interpreted as 1 and 0 – the **bi**nary dig**its**, the so called "bits". Those two numbers encode all the information a computer is working with. Rephrased, a computer breaks every problem down to the degree, where he just must handle those 1s and 0s.

Important
Basic concept of a computer A computer breaks every problem down to the point, where he just must deal with 1s and 0s. This is done by controlling if a current is flowing (1) or not (0).

After that discussion we can understand that an increasing number of transistors on the chip comes along with more computing capacity or in general a faster data processing. Obviously, those requirements became more and more important in our modern world, which led to intense research and development of those chips. This seems very logic now, but back in the days this trend wasn´t obvious. So, it is even more outstanding, that Gordon E. Moore predicted not just the increasing number of transistors on a chip, but he quantified it correctly. He stated that the number of transistors an on integrated circuit doubles approximately every two years. The figure below proofs that his prediction turned out to be correct.

Important
Moore´s Law The number of Transistors gives a computer chip its processing speed. The number of those transistors on a chip is doubling approximately every two years.

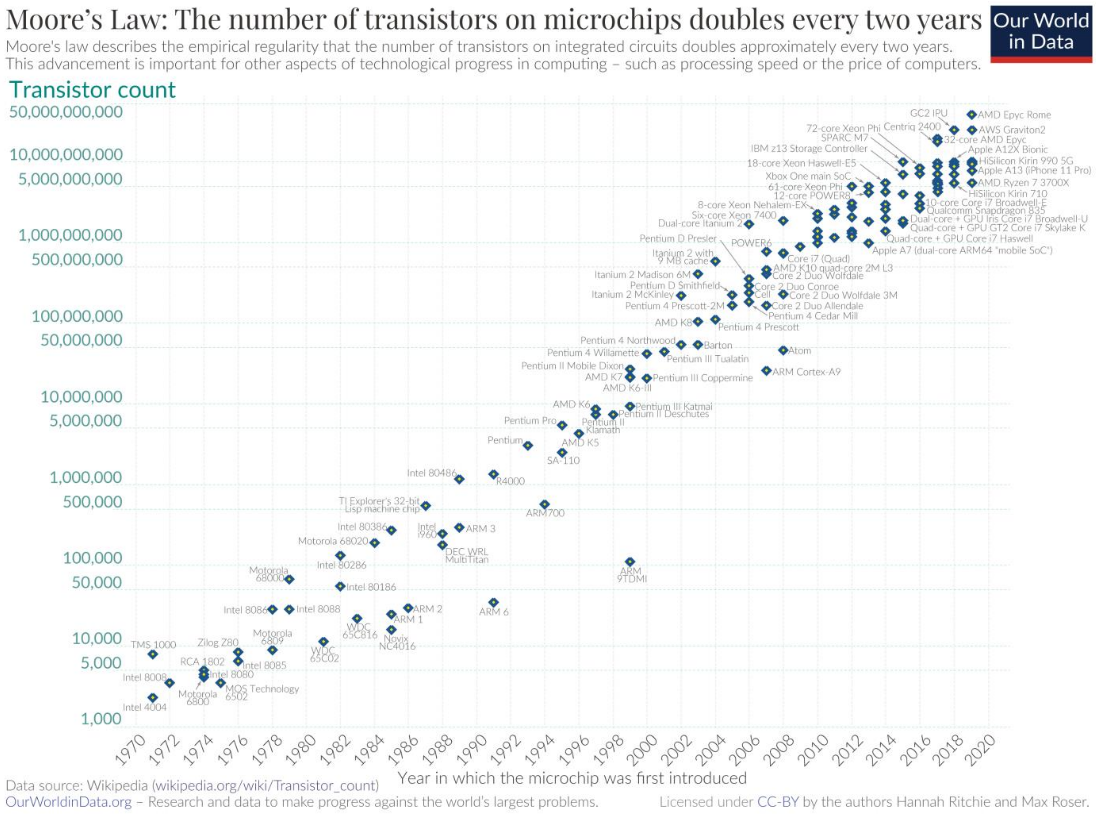


Figure 1: The picture shows the increasing number of transistors during the last fifty years.

Basically, everything you are doing on your computer right now – moving the cursor, scroll through this text – is done by simple summations, comparison of data and short-term-storage of data. The "CPU" (**C**entral **P**rocessing **U**nit) is the building block, which takes care of this. Of course, there are huge numbers of transistors involved too. So, let us also have a look at the impact of Moore´s Law in this area and what kind of computer-features have been implemented in this context over the years. The first chip was developed in 1971 by Intel. The Intel 4004 with 2300 transistors on it and a clock rate of 108kHz was able to take care of all the problems for programmable calculating machine. The clock rate gives an idea of the speed the data is processed with. Later, you will see that the kHz-clock rate is very low in comparison to today´s standards. Yet in the following years the efficiency improved rapidly. In 1972 the Intel 8008 had 3500 transistors and 200 kHz; the Intel 8080 two years later already had 6000 transistors with a rate of 2 MHz. The 5 MHz chip with 6500 Transistors on the Intel 8085 was the most efficient in 1976. Up to this point the application was most likely traffic light switching or steering of production machines. In 1982, the development of computer-chips with a performance of 25 MHz enabled the breakthrough of the desktop-PCs. And when we think about this, the word "desktop" describes the important development very well at that point: Now it was possible to concentrate enough efficiency on such a small space that the "calculating machine" fitted on top of a desk. This was not always customary. The next important period was 1990 to 1999, when Intel developed chips, which reached the 1-GHz-mark. In the years of 2000 until 2008 a new way of increasing the chip performance came up. Additionally, to higher numbers of transistors it is possible to implement more than one "core" on a chip. One core can work on a single task at a time with the specific clock rate. Latter is defined by the number of transistors. So, if you add another core multiple tasks can be processed simultaneously. We can compare this to a situation, where one is waiting for his ordered drink at the bar. If only one bartender is taking care of the whole line, it will take much more time compared to the same situation with two bartenders. However, Intel reached 3.33 GHz with two cores in 2006. The number of cores increased in 2008 to 2013 from two to four. Up to this day the core-number in the high-end-segment of available chips increased to 16. So, for example this explains how all of today's great and realistic videogames run so fluid and without a problem.

Remember

Features of computers over the years

The CPU is processing all the data in the background, what makes your computer work. The performance of those chips improved drastically. In the beginning it was enough to control traffic lights, nowadays computers can run video games or movies very fluid without a problem.

So, until now we gained knowledge about the miniaturization and performance-optimization of hardware for computers. Yet this is not a discussion, which is only interesting for specialists. The huge impact on our daily life becomes clear, when we look at the situation 30 years ago and now. In Europe at least most of the people had a telephone in their home. But sometimes making phone calls was a difficult task, because you have to be at home to call someone and you basically needed to be lucky to be home, when you received a call. At this time phone boxes have been a solution to this problem. Consequently, the people were able to call each other when they were not indoors. Except the phone booth was occupied, you didn't have coin money, you forgot the phone number or you simply couldn´t find a phone box. With regards to all of those disadvantages no explanation should be needed for the triumphant advance of smartphones. The important aspect right here is that the usage of phones gone beyond making phone calls. Basically, the development we discussed in the beginning of the topic came so far, that we carry a million times the computing power of earlier PCs in our pockets today. Hence it is safe to say that nanotechnology made it possible to concentrate all of that efficiency in chips with the size of the cent coin, with which you actuated the phone booth back in the days.

Besides the enormous computing power of the modern miniaturized devices, you can find multiple sensors on the same small amount of space. Therefore, let us have a quick look at what kind of sensors there are and what they are doing. Position sensors tell your smartphone its location. In this regard the Magnetometer always points to north, the proximity sensor turns your screen black when you are taking a phone call, so that you don't touch anything accidentally on your display and the GPS (Global Positioning System) uses satellites to locate you. Another type of sensors is analysing your motion by detecting speed and rotation. The gyroscope tells the device where it is pointing in a three-dimensional room. Accelerometers are sensing the vibration and acceleration tilt, what makes it possible to see your speed in navigation apps or switching the phones orientation when it is turned. Moreover, a bunch of environmental sensors are common in modern smart phones. Their function is pretty much self-explanatory: Thermometer, Hygrometer, Barometer, ambient-light sensing. We can imagine that a pretty clear picture of our everyday life is painted, when a bunch of those sensors work together. On the other hand, this comes along with a whole bunch of advantages. For instance, when the Gyroscope and the magnetometer work together, tilts and turns can be added to maps, which is giving us improved navigation in the future. A practical insight in all the sensors you carry with you is given by the app "phyphox". The app allows you to perform physical experiments directly with your smartphone. For example, if you ever asked yourself in an elevator how fast your motion actually is, this application helps you with it. By making use of the barometer in your smartphone, the atmospheric pressure and its variation can be detected, which allows a conclusion regarding the speed. Performing different experiments really gives an idea of how accurate your smartphone is detecting your surroundings and how it was able to unite a lot of things like cards and fitness trackers on a single screen.

Important

Current sensor examples

Today´s smartphones are equipped with a lot of nanosized sensors, which enable the high functionality of the devices.

2. Transistor, the magic device

In the very beginning, we briefly mentioned how a computer works. Let us examine that concept step by step, starting from the whole ensemble down to the transistor.
Obviously, you can write documents, email your friends or browse the world wide web with your computer. But what is actually happening in this machine to make those things work? Expressed in a very simple way, a computer only manipulates data. In this context the information is stored, retrieved and processed. Nonetheless what kind of data or information we are talking about? Back in the days the input as well as the output was limited to digits, which also is the reason why the name of the device is derived from the verb "compute". Even though modern PCs can also handle letters or sounds, those in turn are coded with digits. Consequently, the computer is a "simple" calculating machine up to this day. At this point we should remember that we use the binary system for that coding, which only uses 1s and 0s. For instance, every single letter is coded with a defined number and series of those binary digits. Thus, every movie, picture or song becomes a series of 0s and 1s for your PC.
Now that we know about the basic operating-principle of a computer, a remaining question is: How do we generate those 0s and 1s? This can be imagined by thinking of the computer as a number of switches. Each and every one of those can be turned on or off, which is defined as 0 or 1. Increasing the number of switches means you increase the number of 0 and 1.

Definition
What is a computer A computer is a calculating machine. The device only operates with 0s and 1s. Latter are created with miniaturized switches.

To get a step closer to the understanding of how a computer works, we should ask the question: How do those circuits, containing the switches, look like? What is the difference to the ones we are familiar with from the physics lessons in school or the fuse box at home? Latter are made of discrete building blocks such as transistors, diodes, capacitors, and resistors. The circuit in your PC on the other hand takes all of the discrete simple functions of all of the devices and integrates them on one single chip. This is the reason why they are called integrated circuits. If you ever opened such an electric device, you could have seen them as the little black squares or rectangles as in the picture below.

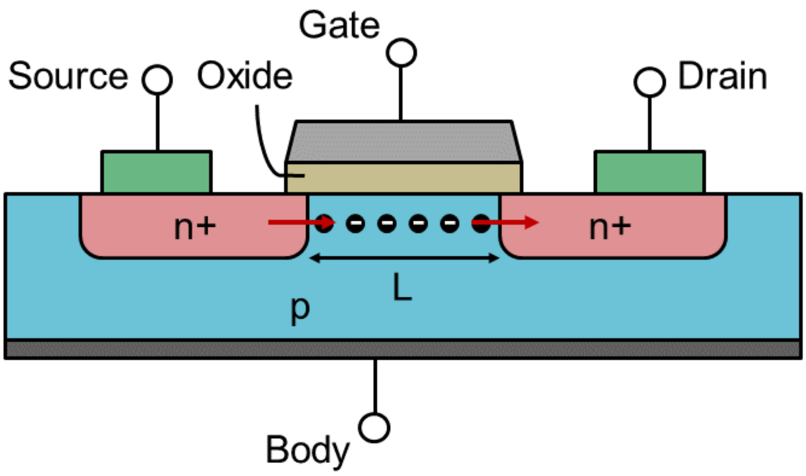
One of those black boxes, basically contains a single silicon chip on the inside, on which all of the building blocks mentioned before are printed on. The fact that nowadays it´s possible to bring billions of electronic components on those semiconductors, paved the way for microprocessors not being bigger than a bunch of square millimetres. You can imagine that with the miniaturization of ICs also the development of innovative physical and chemical processing technologies became more and more interesting. Just to have that mentioned alongside: This is the point where nanotechnology meets the different disciplines of science again, to produce such great modern devices. And we already know that the efforts of many developers paid off, by recalling the graphic of Moore´s Law to our mind again. There we saw that the number of transistors on a single chip increased tremendously and is still rising until today. Right now, we should understand that with more of these structures on the chip, microprocessors can reach better performances, because they can handle more 0s and 1s at the same time. And to implement an increasing number of components on the same sized chip, all of those single building blocks have to shrink in their size. For this reason, it is a good point in time, to have a closer look at the transistor, which can be understood as our miniaturized switch.

Definition
What is an IC An integrated circuit consists of discrete electronic components of a single silicon chip.

There are different kinds of transistors. We want to understand the structure and operating principle of the "MOSFET". The name MOSFET stands for Metal Oxide Semiconductor Field Effect Transistor and already tells us the structure of the device. The basis of the device is a flat silicon bulk material, which contains intended impurities. Latter is called "doping". This bulk material can be p-doped (NMOS) or n-doped (PMOS). Those types of material differ in their charge carriers, which are electrons for NMOSFETs and holes for PMOSFETs. Our discussion will be about NMOSFETs. In this case the p-doped silicon substrate (you can imagine this as a slightly positive charged region) is drawn in blue in the picture below. In this material are two red areas, which are doped with a different type of ion. This generates regions of a n-type (you can imagine this as a slightly negative charged region) material, the so called "source" and "drain". The region between those two is called the "channel" and is covered with an insulating oxide layer. On top of this oxide layer is a conducting metal material, what is referred to as the "gate".

The very basic concept of a MOSFET is to control the electron flow between the source and the drain, which is driven by a voltage between source and drain. The electron flow is then controlled by applying a voltage to the gate, what induces an electric field and influences the charge carriers in our doped material. Latter effects how good or bad electrons can be moved from source to drain. Nevertheless, let us dive a little bit deeper into those processes step by step.

First of all, we want to describe the "off"-state of our switch. When no voltage is applied at the gate, no free charge carriers (electrons in our case) are present in the channel. Hence no electrons are moving from source to drain and there is no current. With this simple case we already defined the 0 for the computer. But how do we get the 1s? What we have to do is to apply a voltage at the gate, the so-called gate-voltage. This voltage in the case of NMOSFETs is always positive. For modern processors a gate-voltage around 0.2 V can be enough, which is very little. To get a feeling: A normal household battery operates at 1,5 V, a doorbell at 8 V, and a car battery at 12 V. So, we really do not need much energy. But why do we need this gate-voltage? What comes along with this voltage is the electric field that effects the p-doped bulk material even through the oxide layer. This induces holes in the bulk p-type substrate near the insulating layer to be repelled. Hence, there are less and less holes at this surface with an increasing gate-voltage. It makes sense that we call that area the "depletion-zone". After that, electrons from the n-type areas (the source and the drain) are attracted in this former depletion-zone and we generated a channel with a magnitude of negative charge carriers. We call this one the "inversion-layer". To not lose the insight, this situation is shown in the figure below.



Now we end up with a n-type channel in-between the also n-type source and drain. This enables electrons to flow through the device, which is shown with the red arrows. Thus, we reached our target and can define the digit 1 for our computer. By the way: We can imagine that higher gate-voltages attract more charge-carriers, the inversion-layer gets wider, and we create a channel, which potentially can transport more electrons.

Definition
What is a transistor The most common type on ICs is the MOSFET (metal-oxide-semiconductor field effect transistor). One of the applications of these transistors is to produce 0s and 1s in computers, as we described earlier.

3. Size matters especially in electronics

Last but not least we want to get an impression of how small those ICs and transistors actually are and what challenges come along with it. After the schematic observation of the MOSFET the desire to have a look at one on a chip alleviated. For better understanding: In the very first text we learned how small a human hair (50 – 80 µm) is, to understand the nanoscale. When we lay two hairs on top of each other it gets very difficult to even see the cross section. In fact, behind that cross section fit thousands of MOSFETs.

Let us examine other possibilities for making very small structures visible e.g. by a light microscope. This discussion will also lead us to the problems we encounter by the manufacturing of MOSFETs. Fundamentally, in light microscopy a specific organized ensemble of lenses is creating the magnified picture of a smaller object. In this context a light beam shines through the object and the lenses into the eye. The question you may ask now is: How small could an object possibly be in this regard? For the answer to that question, you have to know that light microscopes are operating in the visible wavelength-area. This means that there is only a defined range and the half of the shortest wavelength is the smallest object you can see with such a microscope. Or expressed in other words: Two lines have to be half of the smallest wavelength distanced to each other, so they can be distinguished. This law is known as the Abbe limit and is described by the following equation:

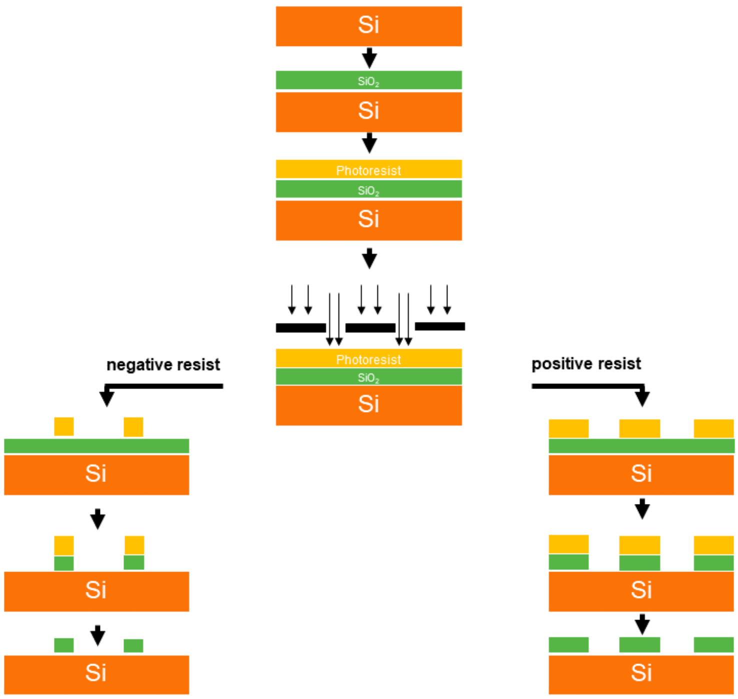
$$d = \frac{1\lambda}{2NA}$$

In this case λ is the wavelength, NA the numeric aperture (containing the refraction index and the angle of the light). Thus, you can come to the resolution limit d, which tells you how far the lines have to be apart, to be distinguishable. The equation also shows how the performance of a microscope can be improved: Either by increasing the numerical aperture or by using shorter wavelengths.
The wavelength of visible light ranges between 380 and 780 nm, therefore, by making use of the Abbe equation, we can estimate the smallest possible resolution of a light microscope to 200 nm. For this estimation we assume that the smallest wavelength is about 400 nm and the numerical aperture (NA) of air is about 1.

Remember
Limitation of resolution The Abbe Limit is the formula for the resolution in dependence of the wavelength: In this case smaller wavelength enable higher resolution.

This is an appropriate time for an examination of the manufacturing of chips. How can we develop something that is too small to be seen? The modern era is using a procedure called "photolithography". In this procedure a pattern of a mask is transferred onto the surface of a silicon wafer. Latter is the substrate we already saw in the scheme earlier. The concept is very similar to the operating principle of your printer at home. The difference is that the inkjet is replaced with a light beam. This turns out to be very helpful when we create our extremely small structures, because the dye molecules exceed the wavelength of the light multiple times by their size.

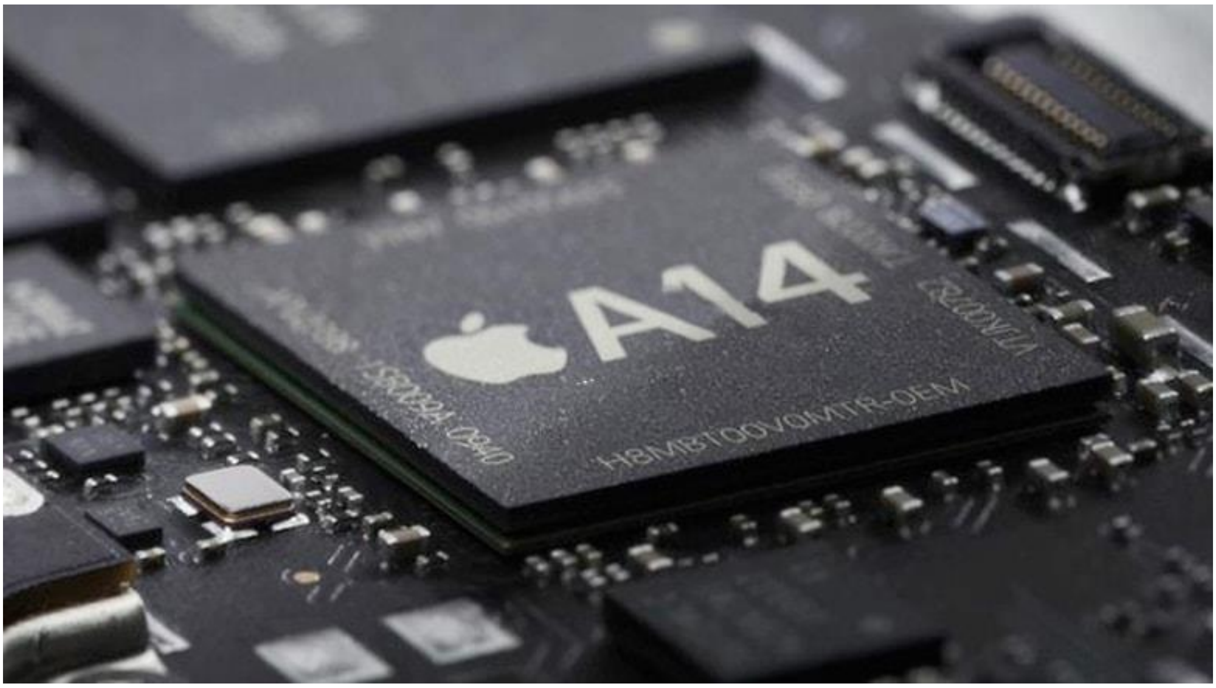
Definition
Photolithography The process of printing very small structures on a silicon surface. The wavelength of has to be short to obtain higher resolution.



The steps of photolithography are shown in the scheme above. In the beginning the silicon substrate is covered with a silicon dioxide layer. A Photoresist is applied on top of this layer afterwards by a process called "spin-coating". Now the actual printing-process can take place: A mask with a certain pattern is placed above the system and is exposed to light beams. Hence, only some spots of the photoresist are meeting light. At this point there two different types of photoresist: The first option (left side of the scheme) is "negative resist". In this case the area, which is exposed to light, changes its structure in a way that it becomes stable towards a solvent. So, the masked areas can be washed away. Now the unprotected areas are etched by chemicals, and we already transferred our pattern onto the wafer. The final step is a washing-procedure, where the remaining photoresist is removed. The "positive resist" can be seen on the right side of the scheme. Here, the areas, which are exposed to light, become soluble and are washed away, which tells both procedures apart. With this example we can understand how we bring the insulating SiO_2 in the MOSFET structure, which we saw earlier.
Now we can put the puzzle pieces together, so a bigger picture emerges. We have seen that the length of the insulating layer is approximately the gate length. We also have seen the manufacturing of this component by photolithography. This makes us understand that the performance of the photo-processing is essential for building very small surface structures, respectively our tiny MOSFETs. In conclusion: We need to improve photolithography to still fulfil Moore´s Law with more transistors on less space. The limitation of photolithography in general is also given by the Abbe limit.

The most important factor is therefore the wavelength. Repeatedly we notice that a smaller wavelength is a main factor in the optimization of building nanostructures. Thus, the state of the art is the use of "deep ultraviolet" light, which is about 193 nm. So, in this case we notice that it is possible to leave the visible range in comparison to the light microscope. The aim for the future is at 13.5 nm and is called "extreme ultraviolet" light. Nevertheless, Moore´s Law is present up to this day. Basically, in all applications of integrated circuits the number of transistors on a chip has increased rapidly.
Last but not least we can immerse ourselves into the subject of the "technology nodes". This term defines the state-of-the-art in the manufacturing processes. For example, the "45 nm technology node" describes a certain procedure. In this context it includes a specific way the photolithograph is built with all its components, the wavelength used in the system, etc. The number 45 nm describes the "half pitch". Latter is half of the distance between two single structures like contacts or conducting lines. So, it is not the gatelength (the gatelength can be even smaller). When we explore the history of this process development briefly, we get an idea of how important the optimized manufacturing is up to this day: 1971 to 1998 was the time of 10 – 0,25 µm nodes. 1999 was the year, where we came from micro to

nanometres with the 180 nm node. The following decades showed huge improvements up to the 5 nm node in 2020, which lead to the Apple A14 Bionic chip (shown below). In this context only TSMC was able to place 11,8 billion transistors on 88,45 mm^2. By the way: If you are reading this text on an iPhone 12 or iPad of the 4th generation, one of those chips is operating in your hand right now.



And to keep pace with the prediction of Moore, TSMC is willing to come up with the 4 nm node in 2022. IBM on the other hand already introduced a prototype of 2 nm node chip in May of 2021, which is the first of its kind.

Definition
Technology nodes A description of the manufacturing process of modern chips. New generations of technology nodes are essential for keeping up with the pace of Moore´s Law.

But can this miniaturization be infinite? It is interesting that the answer to this extremely modern topic can be found in an era, which was a very long time ago. The Greek philosopher Democritus already claimed that nothing can be downsized without a limit. At some point you will end up at an undividable unit – the atom. If we think of the 2 nm node chip, there aren´t more than 20 silicon atoms across a single transistor. So, one could think, that the end of Moore´s Law cannot be avoided.

Important
Where are the Limits of Moore´s Law? A gatelength, which comes close to the size of a single silicon atom cannot be downsized any further.

1.Save knowledge

Summary

You reached the end of the content unit about **nano in electronics**. As there was a lot to learn, please receive a quick repetition of the most important things you learnt about this topic:

Computers are basically built from several “chips”, which are also called “integrated circuits”. The key elements of those chips are “transistors”. A computer breaks every problem down to the degree, where he just must handle 1s and 0s. Those 1s and 0s are expressed by flowing current (1) or not (0) in the transistors.

In the recent history, the storage capacity and speed of computers has evolved extremely fast. Gordon Moore predicted this phenomenon already in the 1965. He even quantified the increase by saying, that every two years the number of transistors on a chip and therefore the capacity will double. Till today he was right with his predictions.

The development of smaller structures is highly connected to nanotechnology. The fact that Moore's Law is still valid today is due to great technological efforts, especially in the field of nanotechnology.

Nanotechnology is used to keep the computer systems very small even with more and more functionalities. With your smartphone, you can have a very fast and highly functional computer in your pocket, enabling you to watch high resolution movies, communicate anytime and anywhere, play games, and other things that former generations maybe would not have imagined.