Technology Hardware & Equipment

Thank you to Swedbank and Kepler Cheuvreux Swedbank for releasing the material

This text was originally published June 5, 2015 as a thematic piece in Swedbank's quarterly small-and mid-cap product called The Companion. It is, with slight cosmetic alterations, republished by InvestingByTheBooks thanks to Swedbank's and Kepler Cheuvreux Swedbank's kind permission. Note that this altered text is solely the responsibility of the author and not of the above-mentioned firms.

Mats Larsson, May 20, 2018

Contents

Introduction

Information Technology
- Semiconductors & Semiconductor Equipment
- Software & Services
- Technology Hardware & Equipment

Technology Hardware & Equipment is one of three industries that make up the Information Technology sector, the others being Software & Services plus Semiconductors & Semiconductor Equipment. The Technology Hardware & Equipment industry further contains three sub-industries: 1) Communications Equipment, 2) Technology Hardware, Storage & Peripherals and 3) Electronic Equipment, Instruments & Components.

Globally the second sub-industry is the dominating one with the larger market capitalization, but in the Nordics the first sub-industry is clearly the most important with companies like Ericsson and Nokia - but also Swedish Transmode, Doro and Net Insight. The third sub-industry Electronic Equipment, Instruments & Components includes companies making test instruments, barcode products, electronic components like resistors and printed circuit boards plus contract manufacturers and distributors of technology products. In relative terms this is a very small sub-industry and we will not focus further on these companies below.

Ericsson, Nokia, Transmode, Doro, Net Insight etc.

The sub-industry Technology Hardware, Storage & Peripherals, that we will from now refer to as “IT Hardware”, itself includes many different types of businesses. Swedish companies in IT Hardware are for example Axis and Fingerprint Cards.

Axis, Fingerprint Cards etc.

Picture 7.1

Examples of Nordic technology hardware companies

IT Hardware and Communications Equipment are intertwined, interdependent and sometimes companies labeled under one sub-industry produce products that rightly fall under both sub-industries. Despite this we have chosen to discuss the two sub-industries separately below as it adds to the clarity of the text.

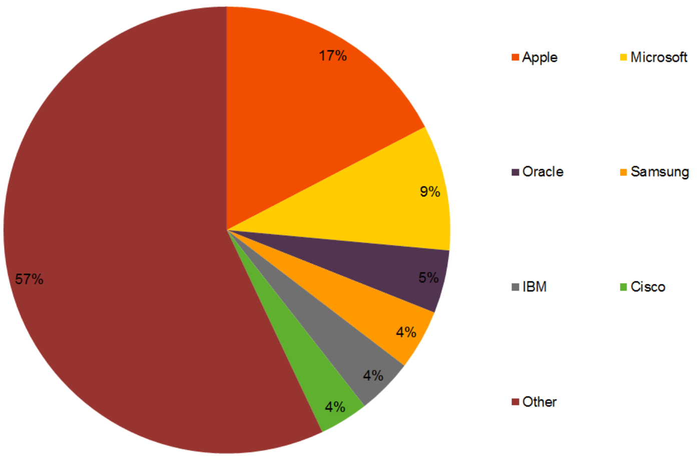
Further, hardware and software are also mutually interdependent and even if they might be logically easy to separate, many of the important technology trends of our time such as The Internet of Things, Cloud Computing, Big Data, Software as a Service, Social Media or even the Internet itself are the result of developments and progress within both hardware and software and could as such be discussed under both headings. As it is then probably beneficial for a reader to know the basics of hardware and software (plus semiconductors) before digging into these hot topics we will describe them in a forthcoming issue of *The Companion* when the industry primer will cover the Software & Services industry. Be patient.

Cool stuff like IoT, cloud, Big Data, SaaS we save for later

We will leave out printers, TVs, set-top boxes, gaming consoles and wearables

The IT Hardware sub-industry is the largest one in hardware so we'll kick off with this and then continue with Communications Equipment. As you will notice both sub-industries include a vast number of products and we will have to concentrate on the more important ones. Due to the space requirements, we will unfortunately leave out a number of products like printers, TVs, set-top boxes, gaming consoles and so called "wearables" like the iWatch.

Picture 7.2



Relative market capitalization of Tech Hardware companies in Bloomberg's BICS classification. Note the doubtful classifications of Microsoft & Oracle as hardware companies. Source: Bloomberg

Technology – a means to fulfil a human purpose

Before digging into the details and getting seriously techy we want to add a more philosophical angle to the topic. The concept of technology is something more than just electronic gadgets or some clever computer coding – it's a means to fulfill a human purpose. The economist and complexity researcher Brian Arthur of The Santa Fe Institute describes technology as capturing natural phenomena and putting them to use. This is always done by combination. New technology is a combination of elements that already exist. Further, new technology is not only something smaller and faster; it's also about interconnection, efficiency and new possibilities – i.e. better ways to fulfill a human purpose.

A) IT Hardware

IT Hardware

- Personal computers
- Tablets
- Handsets
- Servers
- Storage

The Information Technology sector includes firms that make or distribute electronically based products or services. The IT Hardware sub-industry focuses on the most tangible parts of the space. The more important product segments in IT Hardware are personal computers, tablets and handsets, servers and enterprise storage. We will try to give a brief overview of them all and start with the PC as it is the quintessential IT hardware product.

The Products and How They Work

The Personal Computer

PC – Personal Computer
- Desktops
- Laptops
- Notebooks
- Netbooks

OEM – Original Equipment Manufacturer

A personal computer (PC) is defined as a stationary or portable type of computer with a keyboard. The former are called desktops and the later laptops if thicker and more powerful or notebooks if slimmer and with less functionality. A netbook is a further slimmed down version of a notebook. Smaller in size and with fewer components and less software they primarily function as Internet portals and they are as such cheaper and lighter than traditional notebooks. However, apart from in emerging markets they have had a hard time competing with tablets. PCs are often marketed by original equipment manufacturers (OEMs) like Hewlett-Packard, Apple or Dell. The PC market is the largest portion of the IT hardware sub-industry.

Picture 7.3

Laptops and desktops from Dell. Source: www.computer-show.com

von Neumann Architecture

ALU – Arithmetic Logic Unit

CPU – Central Processing Unit

A computer's purpose is to make computations. In 1945 the mathematician and physicist John von Neumann defined an architecture to do this that still describes the layout of computers today (se picture 7.4). The input device in the architecture would for example be the computer mouse, the keyboard, a scanner, a touch screen etc. The original architecture showed a separate control unit and an arithmetic logic unit (ALU). The former is the component that directs the operation of the computer and gives instructions in accordance to a software on how the computer's components should act and interact. The ALU is the component that does the computations on the data provided by the input device. Today both these logical components are parts of the computer's central processing unit (CPU). The CPU is the brain of a computer.

HDD – Hard Disk Drive

SSD – Solid State Drive

USB – Universal Serial Bus (a bus is a parallel or serial channel that transfers data between computer components).

RAM – Random Access Memory

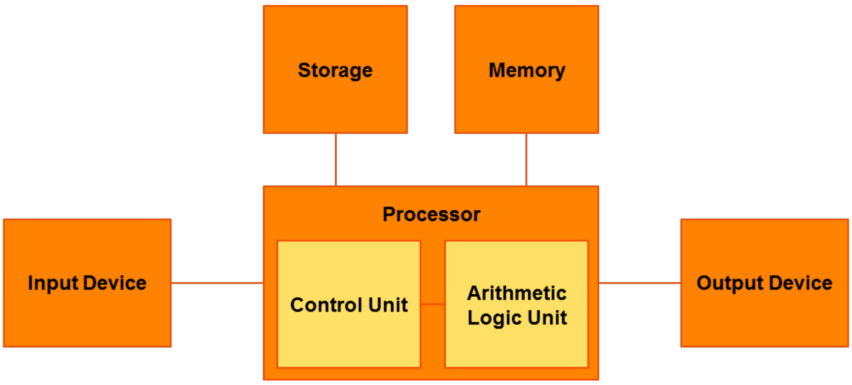
The original architecture further showed one Memory Unit. We have chosen to show Storage and Memory separately. Storage is the permanent memory with products like hard disk drives (HHD), solid state drives (SSD) or even a USB stick, while what's called memory is the non-permanent quick access, working memory of a computer, i.e. the cache memory and the random access memory (RAM). The difference between storage and memory is that non-volatile storage is used to hold programs and data until purposely changed or removed by the user, while volatile memory is a temporary workspace for retrieving programs and processing data, but then loses the content when power is removed.

The cache memory that resides in the CPU is the fastest memory and it's used for the computations that are made at a specific point in time. The RAM is somewhat slower and it's for example used for temporary storage of the operating system while the power is turned on. There is a tradeoff between speed and storage capacity where the storage capacity of the cache memory is minute compared to the RAM that in turn is a midget compared to the HDD. Storage and memory are hardware products and should not be confused with database products that is software and will be covered in a future industry primer.

LCD – Liquid Crystal Display

Further, the output device presenting the computations that have been made could be a printer, a graphics card plus a liquid crystal display (LCD), a loudspeaker etc.

Picture 7.4



The 1945 von Neumann Architecture. Source: Wikipedia

PSU – Power Supply Unit

GPU – Graphics Processing Unit

Actually, there are a few more components to a computer. The power supply unit (PSU) supplies required electricity and also transforms the current of the electricity grid to a current that suits the computer. The graphics card could be described as a separate unit that feeds picture signals to the display or the printer (although the graphics processor – GPU – is sometimes integrated with the CPU) and the motherboard is the traditionally green board that connects all the components to each other – the nerve system of the computer. Finally, the chassis is wrapping everything up in metal and plastics. There's your computer – at least seen from a hardware and semiconductor perspective.

8 bits = 1 Byte
1 kB (Kilobyte) = 1024 B
1 MB (megabyte) = 1024 kB
1 GB (gigabyte) = 1024 MB
1 TB (terabyte) = 1024 GB

Before the modern computer the programming was done on punch cards. A hole in the punch card represented one value and a space without a hole another value. From this the binary system with 1s and 0s developed. A 1 or a 0 is called a bit and 8 bits are called a byte. The specific significance of 8 bites is that it takes a combination of 8 separate 1s and 0s to write a letter of the alphabet. For further information of the binary system see the semiconductor primer in *The Companion 2015:1* from March3, 2015.

The pulse of a computer, i.e. in some respect the speed of the computing performed, is called the clock rate and it's measured in hertz. A computer with a processor with a clock rate of 5,5 GHz has 5,5 billion cycles per second. If a computer can perform one computation per cycle it then can perform 5,5 billion computations per second. However, higher clock rate also consume more energy and thus creates more heat.

The previous historical focus on ever higher clock rates has for the last few years been redirected to developing processors that are able to do more computations per cycle. Further, the ability to add several processors ("dual core", "quad core", "octa core" etc.) working in parallel has made the race to increase clock rates less important. Also, as with all semiconductors the design is vitally important. A badly crafted graphics processor can underperform a well-designed one even with higher clock rate and more cores.

BIOS – Basic Input / Output System

There is one piece of software that is such an integral part of the computer's hardware that it should be covered in a hardware primer. And it's not the operating system. If you think of it, you can strip a computer of its operating system and install a new one, say going from Microsoft Windows to Linux.

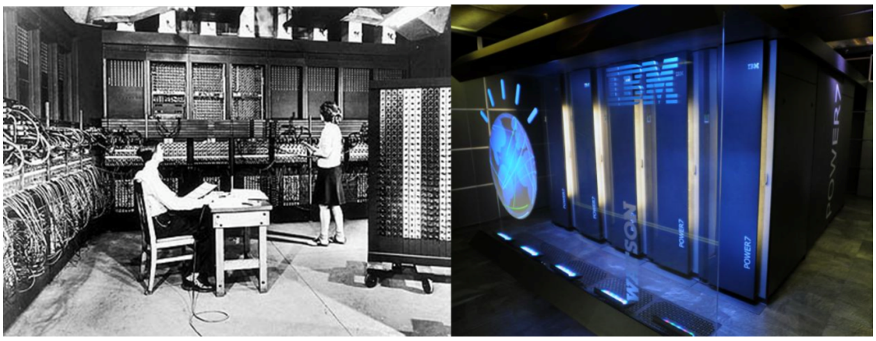
This process runs on the most basic pre-installed software of the computer called basic input / output system (BIOS). It's stored on a separate memory with a battery that will last for years; it identifies hardware, starts the computers components, informs them about the processors clock rate and where in the storage the operating system can be found. BIOS has its roots in the 1980s. Hence it's old, making the booting of the computer slow. The good news is that a successor software is just around the corner.

The First Computer

1. Z1
2. Colossus
3. ENIAC

To get a wider perspective on the computer we thought a walk down memory lane could be worthwhile. The early history of the computer is tightly linked to military applications. ENIAC, often dubbed the first computer, contained almost 18 000 vacuum tubes, weighed 30 tons and filled an entire room. It was built during the Second World War by the US military to compute artillery-firing tables. In reality ENIAC was only the third computer in the world with the Z1 built in 1938 in Germany by Konrad Zeuse claiming the number one spot. The original to this computer – and Zeuse's parent's apartment - was destroyed in a bomb raid 1944. The second spot goes to the British computer Colossus (probably an apt name). Colossus 2.0 was used by amongst others Alan Turing to crack the encryption of Nazi Germany's communication.

Picture 7.5



To the left: ENIAC, the "first" computer. Source: University of Dublin, USD School of Mathematical Sciences. To the right: Supercomputer Watson – the reigning champion in Jeopardy! Source: IBM.

IBM – International Business Machines

In 1959 International Business Machines (IBM) lead the way to the future when they released the first transistorized computers. In 1964 IBM developed System/360 which was the first computer with interchangeable hardware and software. Beginning in the 1970s smaller and cheaper personal computer companies started to emerge with the likes of Northgate, Atari and Commodore. Forced to adapt IBM 1981 made a comeback with the "personal computer", the PC, targeted to the enterprise market and for the first time with third party off-the-shelf components like an Intel microprocessor and Microsoft's operating system.

GUI – Graphical User Interface

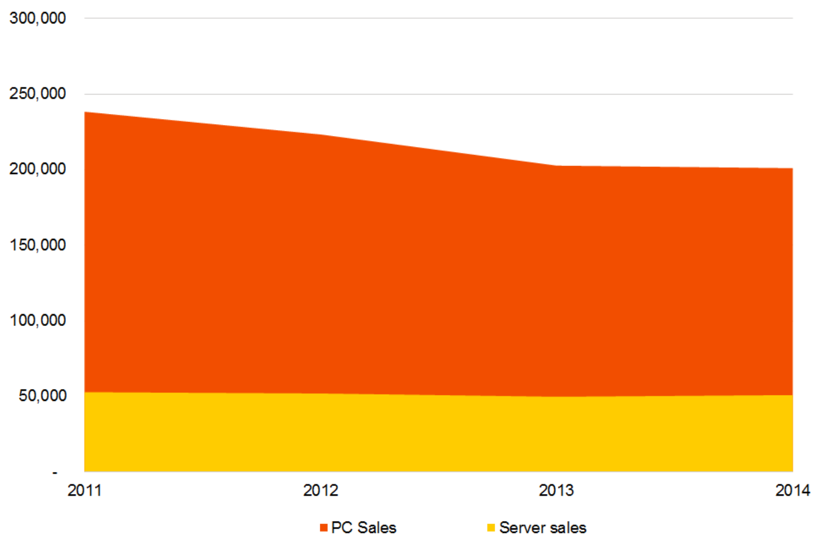
The PC became the benchmark but due to the set up with standard components available for everybody the product over time has become increasingly commoditized. Apple released its Macintosh in 1984. This product for the first time incorporated a graphic user interface (GUI) instead of coded instructions. Charged with a premium and with proprietary software and hardware design it however failed to become an industry standard. In the 1990s Dell entered the market with a unique strategy. The firm built customized computers to order and shipped them directly to customers cutting out the middlemen. With fewer inventories and less distribution cost, prices could be kept very competitive and Dell for a while dominated the market. In time however competitors caught up, evening out the advantage.

The current fate of the different PC-types varies. Desktops have steadily been giving up market share to various mobile computing devices over the past decade. Netbooks never really took off due to the competition with tablets. Thanks to declining prices and the usability of Wi-Fi connections lap-tops have gained market shares over the last decade even though this growth trend got dented by the meteoric rise of the number of tablets post 2010. Before the arrival of the tablet the laptop satisfied the consumer's need as both a media consumption device and a production-oriented device. The tablet separated those two customer needs into two devices and represented a truly disruptive force for laptops. Post the initial media consumption cannibalization, laptops have stopped declining while the growth rates of tablets has come down considerably – in fact Digitimes Research estimated that global tablet shipments in the first quarter of 2015 fell 15.3 percent y-o-y.

In the PC-market there has been a continued consolidation during the previous decade. Lower margins drive a need for scale to reap scale benefits and with declining prices plus the arrival of notebooks many of the previous successful white-box vendors have been outcompeted in Emerging Markets.

Picture 7.6

The PC and server sales growth is pedestrian at best



Global PC and server sales in USD mn. Source: IDC, Bloomberg

Servers

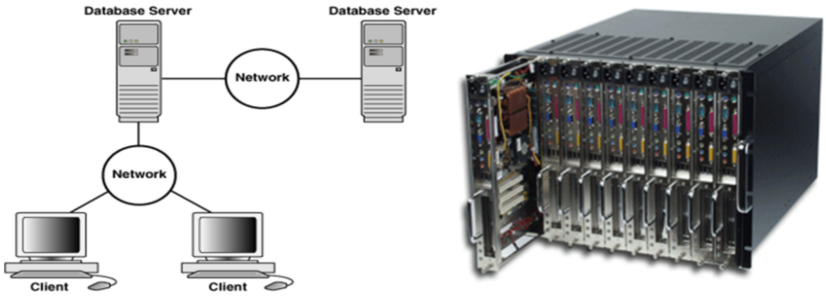
Servers are of many different types but they are essentially powerful computers with pretty much the same components as a PC but with a lot more muscles and without monitors. A server runs application-specific services for multiple computer clients. Some function as application servers, meaning that they serve software applications to clients connected to the network. Others function as web servers and connect the clients to the Internet. Servers do much of the heavy lifting in the enterprise environment running networked software, serving web pages, allowing users to access data etc.

LAN – Local Area Network

Client-server architecture

The enterprise IT environment requires networking between users. In the early days there was the mainframe, a powerful and versatile central computer to which a number of dumb terminals were connected. With the arrival of cheap PCs in the 1980s the preferred model for the enterprise's local area network (LAN) shifted to the client-server architecture. The client-server architecture is a networked environment where individual PCs – i.e. the clients – are connected to one or more servers. This allows for the sharing of information among all clients, easier maintenance and better security.

Picture 7.7



Client-server architecture to the left. Source: Oracle. Blade server to the right. Source: maxi-pedia.com

Servers

- Big iron servers
- Blade servers
- ISS

Servers are sometimes divided into 1) so called "big iron" servers (originally meaning a heavy handgun), 2) blade servers and 3) industry standard servers (ISS or x86 servers). The first are expensive high-end servers with proprietary technology providing powerful and fast computing with minimal downtime. They are used in the back end of corporate networks or in datacenters. Examples are UNIX servers and mainframes and IBM is a dominant supplier.

These high end products are based on old technologies with roots back in the 1960s and the combination of server virtualization and cheaper multi-core ISS has caused big iron servers to come under secular decline the last decade. Still, mainframes will exist for a long time and competition that drop off also means huge pricing power and nice margins for those who remain in the market. Mainframes have the ability to replace hundreds of smaller servers by running multiple virtual machines and as long as new functionality and applications are added they are hard to replace as the core of many enterprise IT-systems.

Blade servers are rack-mounted servers. The removable “blades” shown in picture 7.7 are the computers, with each blade operating independently. The servers are often made of standard components while the backplane in the chassis connecting the servers is proprietary providing bandwidth, power and cooling technology and blade management software. Blade servers have gained popularity thanks to low space requirements, consolidated cabling and cooling, reduced power consumption and easy, flexible deployment as new blades can be added to the existing chassis.

There are no real standards to the size of the chassis of blade servers so the blade inserts of one vendor’s solutions are often not compatible with the chassis of a competitor’s solution and this segment thus offers higher margins than ISSs. With a razor-razorblade-model blade server manufacturers offer their differentiated and often valuable chassis technology for a cheap price trying to lock in the customers and then drive future sales of higher priced server blade inserts and software/services. The size of the installed base becomes hugely important for establishing a competitive advantage.

x86 – An Intel semiconductor architecture suite

The ISS is the dominating part of the server market with HP, Dell and IBM as large suppliers. These are lower-priced servers made by commonly available components like Intel’s x86 processors and that use Windows or Linux operating systems. Differentiation is relatively low. These servers often reside at the front end of a corporate network and handle applications like business software, file serving or email. The last few years the dominant suppliers have begun to lose market share to so-called “white-box” commodity server vendors (and thus have started to sell cheaper white box servers themselves).

In the server market competitors quickly mirror cost measures. A way to increase the otherwise rather slim profit margins has been to expand the offering to bundling services such as application consulting or installation. Remotely administered services can provide stable revenue streams with good margins. Direct distribution has the upper hand when it comes to supplying large customers with ISSs. Selling through distributors is often preferable when selling to smaller customers as the distribution partner can provide valuable add-on services.

Enterprise Storage

A server supplies centralized computation ability to the clients of a LAN. A centralized storage capacity can also boost a LANs total storage capacity, increase utilization, limit duplication and simplify maintenance. Enterprise storage is a product and a market segment that combines hardware, storage software and a sizable amount of service revenues. The main suppliers of enterprise storage systems are EMC, IBM and NetApp.

DAS – Direct-Attached Storage

NAS – Network-Attached Storage

SAN – Storage Area Networks

External enterprise storage hardware can be set up in multiple ways depending on the computer environment. Direct-attached storage (DAS) was the first to be established, i.e. when a storage device is connected to a server to boost that individual server's storage capacity. This is now a product in steady decline. In networked environments network-attached storage (NAS) and storage area networks (SANs) are the primary employed technologies. Networked storage arrays are not attached to specific servers but instead serve the entire LAN.

The core layer of software for any so called storage array (pictured below) is the file system. This software determines how data is laid out on disk or flash and how data is retrieved for use in applications. NAS incorporates a file system in the storage device itself while in SAN the file system is separate. The benefits of networked storage are boosted total storage capacity through higher utilization of the equipment, allowing the sharing of information and spreading resources over multiple end-users increasing productivity. Compared to DAS networked storage units are individual network components with IP-addresses and they fit well into an environment of virtualization and next generation datacenters.

Picture 7.8

Storage arrays with funky lightning from Dell. Source: storagereview.com

Saving space

- Compression
- File de-duplication
- Data de-duplication

The amount of digital data handled in the world is increasing rapidly. To avoid storage overload there are a number of technologies such as compression, file de-duplication and data de-duplication. Compression removes redundant bits of data to shrink a file's size. File de-duplication removes redundant files. Say that 100 persons in an office get an email containing a file that 20 of them save in their LANs storage. The file de-duplications make sure that only one copy is stored and that this is accessible for all 20 persons. Data de-duplication goes further and goes through files to determine which data is the same as in other files and then only saves the parts that differ. While compression normally reduces needed storage capacity in a network by 2:1 and file de-duplication with about 3:1, data de-duplication technology can reduce storage needs with 20:1.

The enterprise storage market is dominated by proprietary file systems and management software leading to high switching costs, good pricing power and high return on capital. These characteristics also have downsides as it is difficult for the vendors to change initial architectural decisions if the market changes which has been an issue with the entrance of Flash replacing HDDs. It further attracts competition trying to replace the incumbents with next generation products. Thus, it is vital for the incumbents to spend a healthy part of sales on R&D and to bundle their products with additional software and services to stay in the saddle.

The hard disk drive has long been the undisputed king of storage hardware technology and the main input material to an enterprise storage system. IBM designed the first disk storage systems in the

1950s. Seagate and Western Digital are the dominating suppliers today. The HDD is constructed by a stack of rotating magnetic discs with mechanical arms that read the data – pretty much like the vinyl record player of your youth. The construction is sensitive to physical stress if for example the HDD is dropped to the floor. Usually measured in GB/in^2 (billions of bits of data per square inch), areal density describes how much data that fit on the disc's surface. Improving areal density improves performance and allows fewer components per given storage capacity and by this drives cost down.

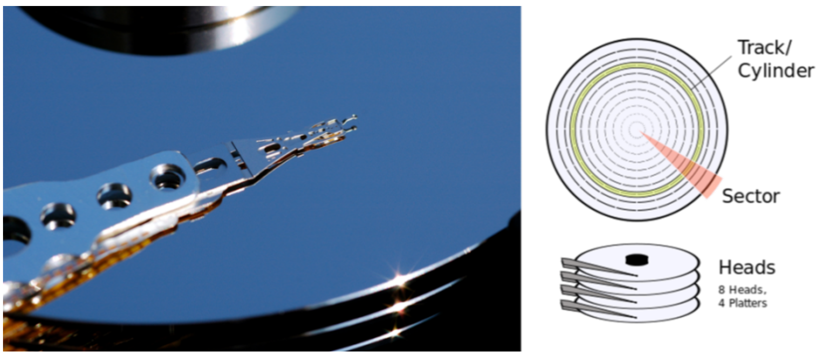
Areal density – GB/in^2

The speed of HDDs is often measured in the number of revolutions per minute (rpm). The higher the number the faster the record is spinning. However, it's not only the rpm that matters. The technology to read and write the discs also matters. Through better reading technology and more intelligent print-heads speeds increase. When data is to be found the HDD must have instructions on the cylinder-head-sector address (CHS) of the information. The cylinder is equivalent to the track of the vinyl disc; the sector is like a slice of a cake cut through all the disc layers and the head is simply which arm that is to be used. The right print-head moves itself to the right cylinder and then waits while the record spins to the required sector. The time it takes the arm to move from where it is when receiving the CHS address to where the data it's supposed to read is stored is called access time.

rpm – revolutions per minute

CHS – Cylinder-Head-Sector

Picture 7.9



The functioning and parts of a hard disk drive. Source: Wikipedia.

Due to Moore's Law and the escalating capacity of microprocessors the bottleneck in today's enterprise IT-environment isn't computing power. It is instead the input-output capacity of stored data (measured in so-called IOPS, input-output operations per second, pronounced "eye-ops"), i.e. the problem is getting the processor something to work with fast enough and having time to store the result. Due to this HDDs are to some extent being replaced with faster solid state drives (SSDs) like NAND Flash as the computer's primary form of storage. The SSD does not have to first move an arm and then wait while a disk is spinning like

Moore's Law – Gordon Moore's prediction that the number of transistors on in integrated circuits would double every 18 months.

IOPS – Input-Output Operations per Second the HDD; it has no mechanic delays. With a switch from HDDs to SSDs comes quicker booting of the operating system and other software plus shorter response times during software operations. The use of Flash storage arrays also leave a data center's architecture unchanged, it's just that the hardware is based on flash rather than disk drives.

SSDs are faster and lighter plus more durable and energy efficient as they lack movable parts, but also more expensive and have more limited storage capacity. However, over time NAND Flash has seen a fall in prices and increasing capacity making SSDs more competitive. Further, the increasing amounts of unstructured and scattered data of today's IT environment suit storage with NAND Flash relatively better than it does HDDs. By mixing disk and Flash, producers of hybrid arrays attempt to mix the best of the two technologies and offer an inexpensive way of deploying enterprise Flash.

ASP – Average Selling Price

The amounts of data that need to be computed and stored are constantly rising and should form a solid foundation for growth for servers and enterprise storage. The combatting forces are a constant technological development that adds to the efficiency of the equipment and also to declining ASPs. Despite a meteoric rise in stored data volumes the price per stored GB has declined at almost the same speed making the sales growth for enterprise storage equipment rather pedestrian. HDDs are slowly losing their position as the main storage medium but since the two dominant players Seagate and Western Digital have bought up all competitors their returns have actually been better the last few years than when the market was growing and the competition was fierce.

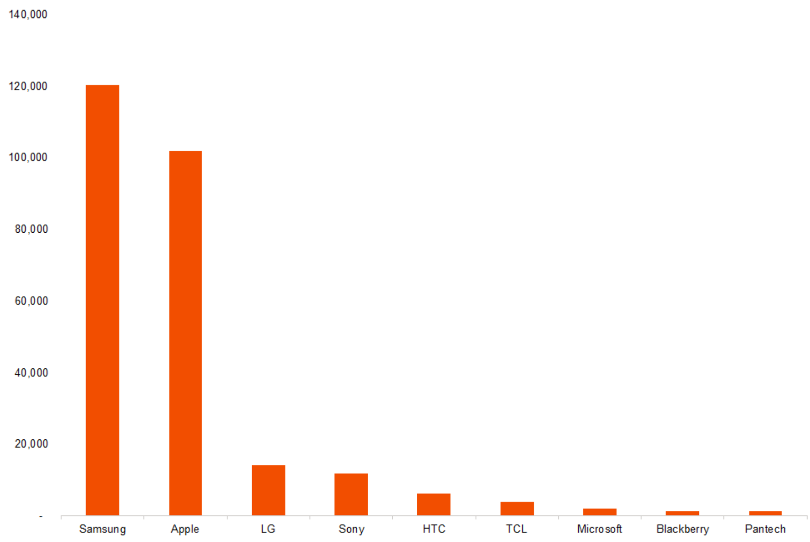
Handsets

Handsets

- Mobile Phones
- Feature Phones
- Smartphones

A handset (or cell phone) is a device that can communicate over a radio link while the user is moving around. Originally the handset handled voice telephony. In addition to voice calls, modern day handsets can also handle various types of messaging, gaming, photography and handle more general computing allowing downloadable apps (short for applications) to be run on the phone. The market is categorized into a) mobile phones, b) feature phones and c) smartphones. Handsets that offer a highest level of general computing capabilities are called smartphones. It's no exaggeration to say that with the entrance of smartphones the value of the market has consolidated to two large players. Picture 7.10 below shows 2014 handset revenues for the market leaders.

Picture 7.10



2014 handset revenues in USD mn. Source: IDC, Bloomberg

GUI – Graphical User Interface

Our way of using handsets has changed radically the last decade in the developed world as the screen as a graphical user interface (GUI) for software applications (apps) has become more important than the microphone and loudspeaker used for voice telephony. We will not discuss the inside of the mobile phone. The March 3, 2015 industry primer on semiconductors in *The Companion 2015:1* covered most topics around the inside of a mobile phone so we refer the reader to this.

Picture 7.11

The screen size of a smartphone ranges from 3.5” to 6”. The stated size is measured diagonally over the screen. Apart from the screen size the proportions of the screen’s width versus its height also matters. Modern widescreen movies are filmed with the proportion 16:9 making this the ideal proportions for a smartphone screen used for watching movies. The resolution of a smartphone’s screen (or any screen for that matter) is measured in the number of pixels of the screens width multiplied with those of its height. With the proportion 16:9 the number of pixels can for example be 1280 * 720. When the number of pixels along the width and along the height doubles the total number of pixels on the grid quadruples.

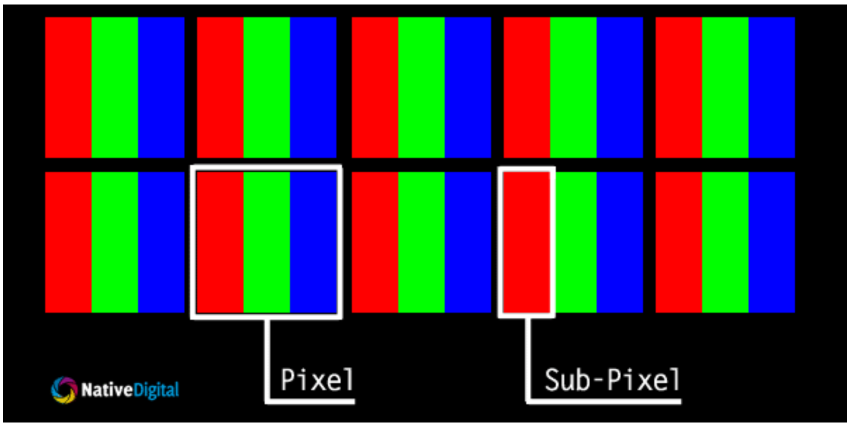
LCD – Liquid Crystal Display

Each pixel in turn consists of a red, a green and a blue subpixel that can shine with varying strength and in combination build up millions of color nuances. The most common screen technology is LCD (Liquid Crystal Display) as it is relatively cheap and also gives a good picture irrespective of the users viewing angle.

OLED – Organic Light Emitting Diode

The downside of the LCD-technology is its poorer resolution compared to for example the more expensive OLED (Organic Light Emitting Diode). This poorer quality is due to that LCD-screens consist of a background light and of a number of filters letting the light through to the subpixels; while each subpixel in the OLED-technology shines of itself. This makes the OLED-screen thinner and when an area of the screen picture is supposed to be black the LCD filters will try to block out the light as best they can. The OLED-screen simply turns off the light.

Picture 7.12



LCD type of pixels and subpixels. Source: navigatedigital.co.uk

Resistive touch screen

Most smartphones today have a touch screen. The early touch screens used a so called resistive technology where two separated electrical layers are pressed together by someone pointing at the screen with a sharp object like a stylus pen and through coordinates the phone knows where the pressure is applied.

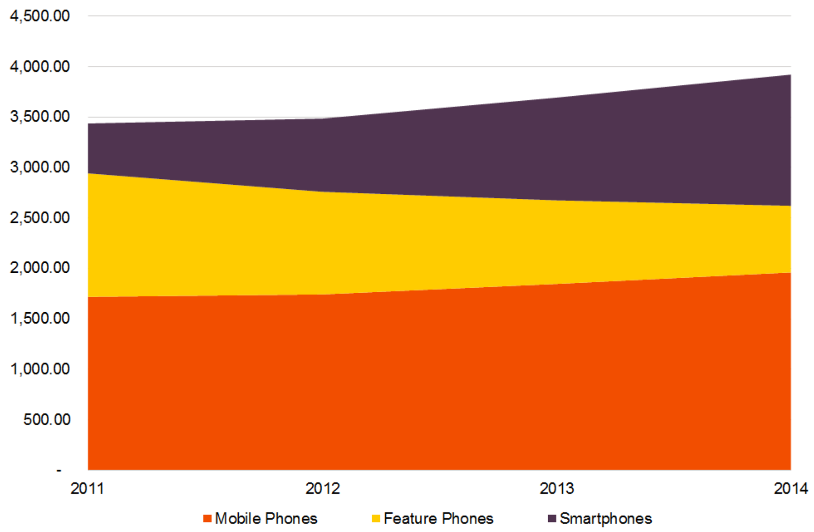
Capacitive touch screen

Today almost all smartphones use a so called capacitive touch screen. The surface layer of a capacitive screen contains an electrostatic field that gets affected by the electrostatic energy stored by the human body. The upside of this technology is that it is possible to construct harder and more scratch resistant screens than before. With fewer layers the screen also becomes thinner. The downside is that you cannot point to the screen with other objects than with naked skin and without the ability to use a stylus pen, some of the ability to point to the screen with precision is lost.

The smartphone has in only a few years' time managed to capture a large part of the value of the mobile terminal market. Looking at units, smartphones has gone from 14 percent 2011 of the market to over 33 percent in 2014. The average unit growth has been 38 percent per year the last three years. Given the much higher ASP of the smartphone the value share is naturally also much higher. The segment giving up share is the stuck-in-the-middle feature phone while lower end mobile phones are growing in mid-single digits, much thanks to increasing penetration in emerging markets. The total number of mobile terminals has been growing with 4,5 percent per year the last 3 years.

IDC – International Data Corporation

Picture 7.13



Yearly mobile phone unit sales 2011 – 2014. Source: IDC, Bloomberg

<u>Tablets</u>

IDC defines tablets as "portable, battery-powered computing devices in a slate form factor that do not have a permanently attached keyboard". Like a smartphone it has wireless connectivity but unlike a smartphone it is relatively big (even though the line between the two is becoming more and more blurred). Further, tablets have a touch screen and tend to be used to browse the Internet and to run small size software called apps that are generally downloaded from the Internet and that communicate over the Internet when being used. The latter is a huge shift from traditional PCs that stored larger size software applications offline.

The Dynabook

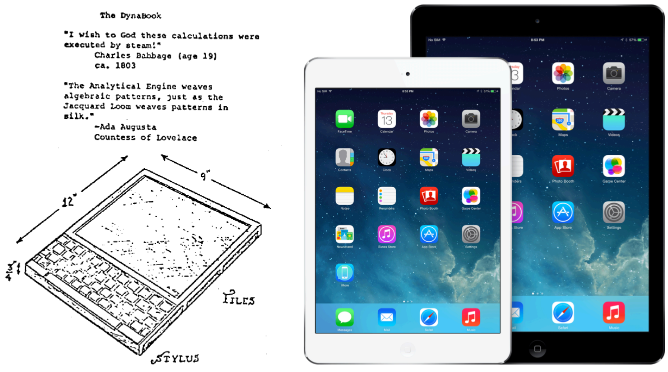
GRiDPad

iPad

Tablets have been around for a long while. Computer scientist Alan Kay proposed the basic idea in 1968 and he called it the Dynabook – a product that had both a touch screen and a keypad plus a wireless connection. In 1989 Jeff Hawkins who was to become the founder of Palm introduced the first commercial product called GRiDPad, but it wasn't until the launch of Apple's iPad in 2010 that this type of product had any real mass market impact.

Apple is still the market leader by far in value but Samsung using Google's Android operating system has taken a relatively large unit market share. Other vendors have found it hard to break into the market in any big way. Amazon's kindle is an interesting niche product as it focuses on certain types of usage, they price the hardware low and Amazon instead tries to profit from accompanying media and content sales. This is the complete opposite to Apple that in general combines low priced software and more expensive hardware.

Picture 7.14

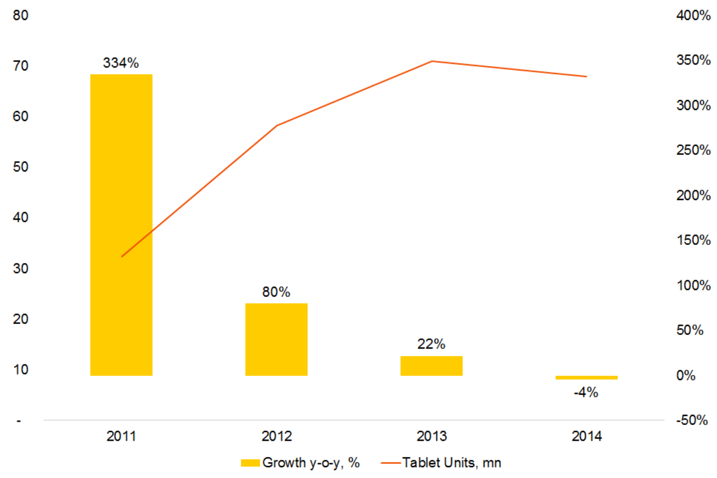


The Dynabook and an iPad and an iPad mini. Source: Wikipedia

ARM Holdings [ARM-GB]

The screen of a tablet is of any of the same two main types that are found in a smartphone so we refer the reader to this section. The inside of a tablet is a PC-light with much of the same components but of less powerful types. Since programs for tablets – the apps - are less complex and robust the higher processing power and storage capacity isn't needed. With portability follows a focus on energy efficiency when choosing components. This has allowed ARM-based microprocessors with their roots in mobile phones to dominate the space over the traditional more powerful but less energy efficient x86 microprocessors with their roots in the PC and server world. Since the usage of tablets is tilted toward media consumption and gaming the graphics processor is important and so is semiconductor content like accelerometers and gyroscopes that help the tablet determine its orientation. Unlike a PC you won't find a fan in a tablet. There isn't enough space so it could become quite warm after heavy usage.

Picture 7.15



Tablet unit sales and y-o-y growth rates. Re-run of the familiar boom-bust-cycle of IT Hardware? Source: IDC, Bloomberg

There have been a number of attempts of reintegrating the tablet mostly used for media consumption with the notebook more often used for work-related production - or at least to develop combined devices with touch screens and (removable) keyboards. So far the attempts hasn't made any material market impact proving Apple right in their comments that the usages of a tablet and a notebook are so different that it would be like combining a toaster with a refrigerator.

The Market and the Business Models

IT Hardware stocks track with the direction of GDP growth and are more likely to outperform when GDP growth surprise on the upside. Hence, IT Hardware stocks tend to move in the same direction as the broader market, though with bigger swings. Naturally, different technology segments can perform differently depending on which stage of the business cycle they are leveraged to and how cyclical they are.

Another important driver is the IT-spending on computers, storage equipment, servers etc. of the enterprise market. Enterprise IT spending tends to be early cyclical as it is a fixed investment made to take advantage of coming business opportunities. IT Hardware stocks often move on the expectation of future changes in IT spending. For an investor it can also be worth tracking for example Taiwanese and Philippine export numbers as the numbers are dominated by technology components.

As a result of this economic sensitivity, technology companies have had a tradition of keeping net cash instead of a net debt. Also, many

TMT – Technology, Media & Telecommunication

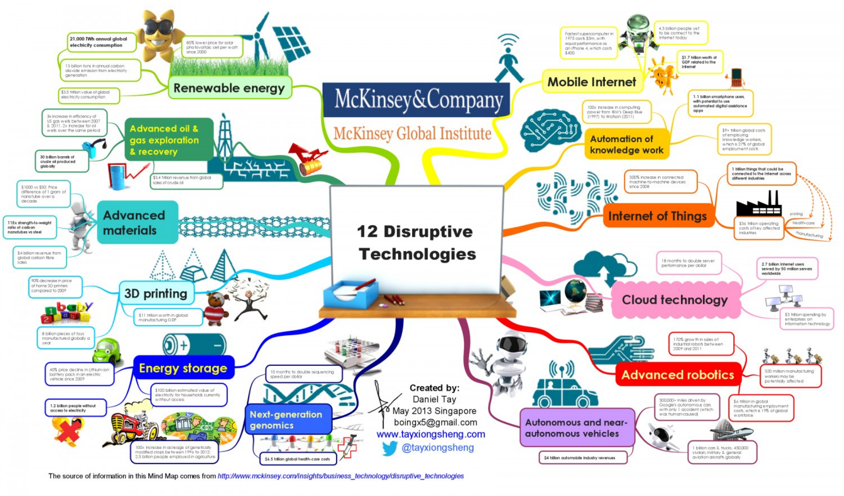
technology companies go public as relatively young companies and young companies are relatively more often financed by equity than by debt. In the TMT-boom of the late 1990s many technology companies issued shares and filled their chests. When the bust came, the value of the staying power that cash brings was evident to everybody and since then few technology stocks carry large amounts of interest paying debt.

Due to the net cash position some technology stocks actually offered a relatively safe haven in the downturn of the 2007/08 financial crisis. This net cash is also very often used to buy up smaller companies with promising technology. This strategy of using acquisitions to supplement internal R&D has long been championed by companies like Cisco but recently the Internet and social media giants have taken it to a different level further feeding the venture capitalists of Silicon Valley.

The technology sector is typically one of constant innovation and changing fashions. All new hot technologies eventually face obsolescence. A smash hit product matures and then continues to the graveyard of technological wonders. The driver of this is competition. Witnessing a new successful product or technology, competitors develop variations and bring to market. Over time the market becomes saturated leading to loss of profitability and consolidation or a new innovation with better features and customer proposition appears and takes the market with storm.

Some companies like Cisco, Oracle and others have, as mentioned above, for years had a strategy of acquiring and absorbing companies with new promising technologies (or threatening technologies depending on which side of the fence you stand). The acquisitions often look – and are – expensive but once in a while they produce a hit product and one should remember that the strategy is as much defensive as anything else.

Picture 7.16



Mind map of disruptive technologies – not only IT-Hardware ones though. Source: tayxiongsheng.com, McKinsey Global Institute

In part due to Moore's Law, most IT Hardware products are in a constant state of deflation where prices fall over time. This means producers must constantly improve productivity and establish economies of scale at a similar pace to maintain a healthy profit margin.

Most IT Hardware companies operate quite similarly as they sell products that they design and assemble but where the components regularly are sourced from third-party manufacturers – with Intel's x86 chip architecture and Microsoft's Window's operating systems as building blocks for PCs and ISS as the most well-known example. Each component in for example laptops is usually to some degree interchangeable with little differentiation in product functionality among brands. In most cases the assembly is outsourced to primarily Southeast Asia, where companies like Foxconn build the final products on the OEMs design specifications. Notebooks are for example almost exclusively made in Taiwan. IT Hardware companies are because of this left to compete mostly on price, distribution methods and in some cases brand recognition.

Distribution models

- Direct selling
- Distributors
- Consumer retail

Once assembled the products are sold through different channels: direct, distributors and consumer retail. The direct model pioneered by Dell in 1996 is simply selling directly to the end customers without third-party intermediaries. Distributors are one-stop shops for a large variety of technology products. With direct distribution OEMs can hold a lower amount of inventory as each computer is "built to order" rather than to forecasted demand. In a business where the prices of input components are in steady decline lowering held inventory has positive gross margin effects. Additionally, it allows OEMs to keep parts of the profit margin that would otherwise have gone to the third-party.

A distributor sales model sacrifices some of the flexible customization that the direct model offers. This and other disadvantages can at times be outweighed if OEMs through distributors gain a more extensive exposure to the right users and are able to share risks with their distribution partners. Also, the relative margin advantages of direct distribution versus indirect distribution have narrowed over the past decade.

One reason for the previous success of Dell's direct sales model was the lower inventory held in the sales channel. In the 1990s the combination of rapidly declining PC-prices and a system of distributor price protection made selling through distributors a costly business for OEMs. The price protection agreements stipulated that if the OEM lowered the price of a product – which they did constantly – they had to give the distributor a refund on their unsold products purchased at a higher price. The OEM stood all the price risk for the large amounts of inventory in the entire sales channel.

With the rise of direct distribution, shifting the bargaining power towards the OEMs and with less severe ASP-declines, the difference in margins between direct selling and selling through distributors has narrowed. Finally, with the rise to prominence by notebooks and tablets that offers no customization this additional advantage has declined for the direct sales model. In the new world of apps and cloud, users are less interested in paying for specific hardware specifications.

The enterprise market usually uses a direct sales force for more complex and high-end products and distributors for relatively simpler products. In addition to having the most streamlined cost structure, finding the optimal mix of indirect and direct distribution is a main source of competitive advantage to for example PC-manufacturers. Dustin is a Swedish example of a distributor. The retail channel with stores like MediaMarkt or BestBuy is the main purchase venue for consumers, even though the direct channel is growing as a consumer channel as well.

Dustin

Similar to the PC, the tablet OEMs rely on sufficient scale and favorable sourcing agreements to be competitive. However, for tablet OEMs the strength of the tablet's ecosystem is the key competitive advantage. A large supply of high quality apps that are accessible and easy to use for consumers is a formidable barrier to entry for the supplier that lack the same supply; this favors OEMs using Google's Android operating system and Apple. Being more of a portable consumer product, design and branding probably also constitute more of a competitive factor for tablets than for PCs.

The Fight for Digital World Domination

Wintel – Windows and Intel

We cannot write a piece on IT Hardware without commenting on the world's by far largest company measured in market capitalization and the world's most widely held stock among retail investors, i.e. Apple. To some extent much of the above discussion on hardware OEMs business model does not apply to Apple. Apple is a phenomenon. Similarly to the "Wintel"-camp Apple produces almost none of its products and the company is in a way a combination of an R&D-unit, a design agency, a marketing unit and a law firm.

Unlike the other PC-manufacturers however Apple never disintegrated its hardware from its software. In fact Apple designs some of its own semiconductor content, designs its own operating systems and designs all of its hardware. With this tight control the company can better manage the technological roadmap of its products. Additionally, a product where one party controls all design aspects should be better integrated and offer a higher quality. Add to this a laser sharp focus on usability, great innovative capabilities and a captivating corporate history and it all sums up to a truly powerful brand value.

Picture 7.17

The strongest brand and logotype in corporate history? Source: thenextweb.com

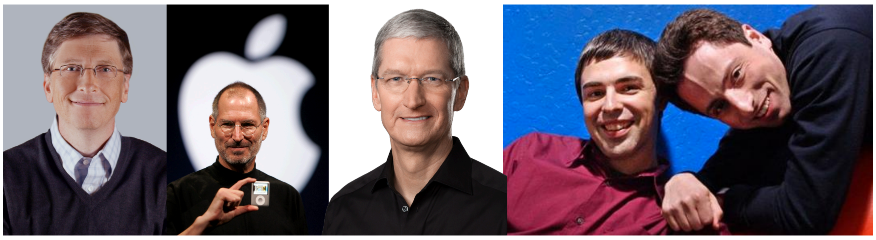
One downside of doing everything yourself is cost. You don't get to participate in the scale economies of those that share technological platforms with each other. Due to the phenomenal pricing power of the Apple brand and the meteoric rise in popularity and sales the last few years this scale disadvantage has become less of an issue. Another downside – or at least risk – is that if you do everything by yourself, you'd better do it right. One key task for Apple is continued innovation. Especially since the iPhone dominates the company's sales and profits today. Innovation has obviously been one of the strong points for the company historically.

The main risk over time may well be brand fatigue. Without the almost religious type of glow around the company created under Steve Jobs the ability to charge the huge price premiums and the ability to negotiate terms with mobile operators that makes them pay most of the price of the iPhone instead of the end consumer might lessen. Any of the two would seriously damage the company's margins. The EBIT-margin in Q2'15 was 38 percent – about four times as high as consumer electronics manufacturers in general.

"Only the Paranoid Survive" – Quote and book title from Andrew Grove, formerly CEO of Intel and technology pioneer "idolized" by Steve Jobs

Knowing that "only the paranoid survive" Apple has worked very skillfully on building in vast amounts of stickiness and high switching costs into their products. Humiliated by their early-period defeat to archrival Microsoft, Apple learnt the importance of building an ecosystem the hard way. And they learnt the lesson well. The strategy that Microsoft used in the enterprise space and which triggered Apples iconic 1984 Big Brother-commercial, Apple is today using themselves in the consumer electronics area targeting the private consumer and fighting it out with its new archrival Google and their Android operating system.

Picture 7.18



Old and new contenders for digital world domination.

Microsoft Windows

Apple iOS

Android

A software ecosystem is a set-up where several independent software developers jointly provide more value to the market than any of them could have done on their own. By developing ancillary products to a platform they create a type of network effect where each additional contribution add to the total value of the platform they are working on to the benefit of all involved. The independent developing companies come to function and co-evolve as one unit on an interactive service and software platform. The underlying software platform that the independent developers are contributing to like Microsoft's operating system Windows or Apples operating system iOS, can be independent of the hardware platform it runs on, like Android, or as in the case of Apple it can apply a proprietary walled garden hardware approach. A "fragmented" versus an "integrated" platform as Apple puts it.

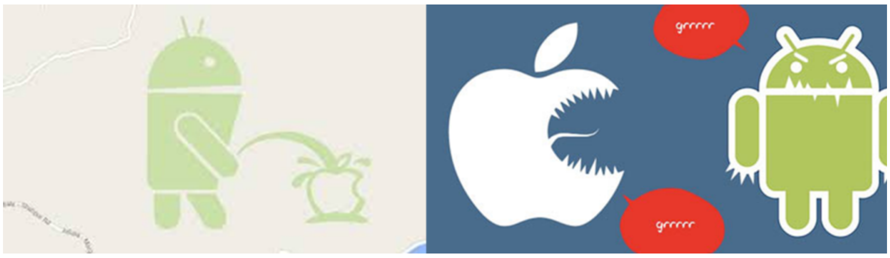
Applications are the primary source of utility for most computing devices and the platform with the best breadth and quality of applications will have a competitive advantage attracting the most customers and by this attracting the most developers further boosting platform functionality in a virtuous cycle. Early on Apple's App Store had a huge advantage over Android in terms of the size of the app-portfolio. Since Android over time became the larger platform in volume (today roughly 85 percent of the number of smartphones and about 60 percent of the number of tablets) and since writing the software for an app is a relatively small feat in terms of software development, this ecosystem advantage has disappeared versus Google - but certainly not towards any other competitor like for example Microsoft with its Windows Phone operating system.

This was not the only tool in Microsoft's old toolbox however. The other main tool was bundling. Microsoft's operating system Windows used to have, and still to a large extent has, a dominating position on the enterprise computer market. For long the equally dominating enterprise software Office did only run on Windows and to be compatible and allow communication with other business users everybody had to use the Office suite and Windows. A network effect mutually reinforced the two products. Another, well-known example is Microsoft's free bundling of the web browser Microsoft Explorer that swiftly killed the incumbent Netscape Navigator and reached more than 95 percent market share in 2002.

Today, Apple is similarly striving to turn most other consumer electronics products into an accessory to its successful, dominant and proprietary iPhone. Through a number of interconnected services like for example a proprietary cloud storage for content like your family photos, like connectivity and seamless services between the Apple products in your home and in your pocket, sharing of iTunes content within a family using the same credit card etc., the consumer finds it increasingly hard not to buy everything from Apple.

Apple combines the walled garden and bundling strategy with adding enough new valuable services to customers so they don't even think about leaving the platform. By this they keep a favorable status in a totally different way than Microsoft did and become much harder to criticize for regulators. This is probably another potential risk for the company – the loss of the good guy-status. The future of Apple after Steve Jobs will be interesting to watch. What we do know is that the contenders, such as Google, will not watch idly.

Picture 7.19



To the left: The unauthorized and inappropriate Android robot placed south of the village Rawapindi, Pakistan on Google Maps. Source: IDG.se

Virtualization and Software Defined Everything

Arguably the IT Hardware environment has always been changing rapidly. The combination of increasing network capabilities making distances less important – and further leading to the emergence of external datacenters - and the rise of the virtual machine are right now dominating change in the enterprise space.

Better Networks
+
Virtual Machines
=
Change

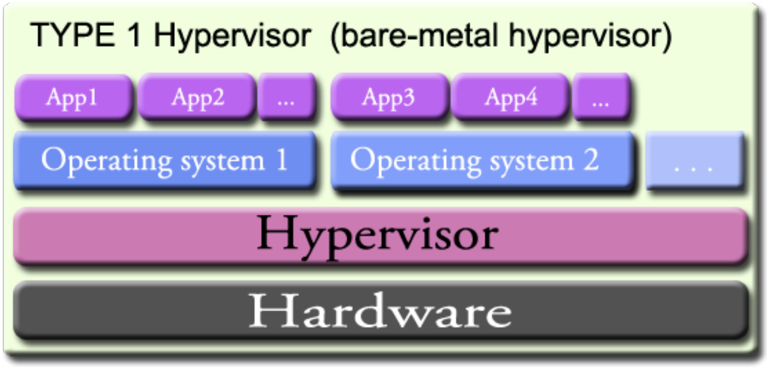
With the rise of the client-server model's dominance in the 1980s and a growing usage of computers, companies for decades bought servers, storage capacity and network equipment en mass to support the needs of their datacenters. Since a server runs application-specific services, one server combining hardware and software was needed for each application.

Often however, since the server was only used for one thing the utilization of it was very low, say 20 percent. The spare computing capacity couldn't be used for anything else. Similarly, networked storage capacity in an EMC storage array or routing capacity in a Cisco router isn't application specific but the solutions bundle proprietary software with often-expensive hardware in a set package. Over time increasingly complex spaghetti networks of high ticket and high maintenance proprietary hardware made up the corporate IT-environment.

The introduction of server virtualization changed all this as it breaks the one-application-per-server paradigm plus de-bundles the hardware and software of servers. It also potentially de-bundles the hardware and software in storage and networking equipment going forward. The agent of change is called a hypervisor and is a piece of middleware that creates and runs virtual machines.

In picture 7.20 the traditional enterprise hardware product combines the gray hardware, the blue operating system and sometimes also the lilac application in one package. With the hypervisor as a middle layer the operating system and application software loses its connection to a specific piece of hardware. Instead the hypervisor creates virtual machines that have the same functionality as previously but that resides on any or several pieces of underlying hardware. VMWare (VM stands for Virtual Machine) has been one of the main driving forces of this virtualization and is now a USD 38 billion company.

Picture 7.20



The hypervisor creates virtual machines. Source: virtualizationsoftwares.com

Today a company can buy general-purpose servers and through software create the required virtual application specific servers spread over the entire computing power of the servers they hold. Server virtualization has as such already disaggregated computing and provided large benefits in terms of equipment utilization rising to about 80 percent, flexible provisioning and dynamic computing capacity allocation, improved disaster recovery, remote programmability etc. Perhaps the main benefit though is lower costs for the datacenter - with the obvious flip side of lower revenues for the equipment vendors. Higher capacity utilization pressures the unit growth and the increased competition from commodity hardware lowers ASP.

SDS – Software Defined Storage

SDN – Software Defined Networking

WAN – Wide Area Network

Even though software defined servers are realities today the parallel products for software defined storage (SDS) and software defined networking (SDN) for MANs and WANs are still in early developing stages. SDN within enterprise LANs are however already used by for example Google in their datacenters.

Some pundits are already discussing a future of "software defined everything" with converged commodity hardware infrastructures and centralized management of virtualized server, storage and networking applications on a joint software control plane. In this future value would migrate to the software stack; a quite prevalent trend the last few decades. Presumably though, not all IT and network processes are equally well suited to function on generic hardware. For example, could a core router in the center of telecoms networks where the hardware is custom-built to handle massive amounts of traffic with highest possible throughput ever run on commodity hardware? Also, MAN and LAN networking equipment doesn't suffer the same low utilization rates as application specific LAN servers – and some networking equipment will always have to be local to allow for, well, networking.

Latency – the time between the source sending a data packet to the destination receiving it (if measured one-way).

We will cover cloud computing in a forthcoming industry primer on software and services but since the implications for the hardware space is potentially profound we'll touch on the topic here as well. When the latency in the networks approach zero, the difference between keeping hardware for a corporate datacenter in a LAN and in a WAN disappears. Distance will not matter much anymore and with the death of distance it suddenly becomes illogical that every corporate office stores their small number of servers, storage arrays and Ethernet switches in their own basement. Microsoft estimates that a 100.000 server datacenter has an 80 percent lower cost of ownership than a 1000 server datacenter due to scale, diversity of demand and the ability to service multiple clients. There are strong economic incitements to consolidate on fewer and larger datacenters. This has paved way for the rapid growth of the massive datacenters of Amazon, Google and Microsoft.

Datacenters or server farm

Hyperscale cloud datacenters (or server farms) like these are large collections of servers housed by a company in one single facility. External clients connected over networks can then lease server capacity from the center. The big datacenter companies are further the ones that have been leading the development in utilizing both virtualization and novel parallel computing technologies over a large number of white box commodity servers.

Picture 7.21

Server farm. Source: dailymail.co.uk

The offering of an off premise datacenter run by for example Amazon could potentially offer compute, storage and parts of the networking as a subscription service to companies; by this moving applications to the cloud, to be consumed on demand and pooling the hardware that previously was used by all the end clients. We are not yet at this point of "software defined everything" but virtualization is a major change for the IT Hardware subsector at large. It presents opportunity for innovation as well as threats of change for incumbents.

What Should an Investor Think Of?

2 * GDP -> 1 * GDP

During the 1980s and 1990s IT spending grew with about twice the speed of GDP. Nowadays the sector is more mature and roughly grows at the same rate as GDP. About half of the population of the earth now has access to the Internet and mobile phone based Internet access is further rapidly becoming more important than PC-based access. Due to the relentless expansion of the Internet and the cloud, technology usage is shifting to a software application based model, crowding out the old hardware based models. This is a familiar story. Over the last decade software, and to some extent semiconductors, have stolen much of the value-growth of hardware that is becoming increasingly standardized and commoditized.

The pricing power and mass adoption of Apple's products have however lifted the average return on capital for the entire IT Hardware sub-sector. Pre (and excluding) iPhone/iPad the subsector combined an average return on capital roughly on par with the level of the cost of capital with a high cyclicality. IT spending is often discretionary over a year or two. Thus, buyers can postpone purchases during testing economic times. Postponing purchases too long will however hit productivity and open up for competition to eat your lunch. Hence, after a crisis there is often catch up buying.

Another development the last decade has been a functional and geographical split between capital light design and development of intellectual property in the Western hemisphere and capital heavy manufacturing in Asia.

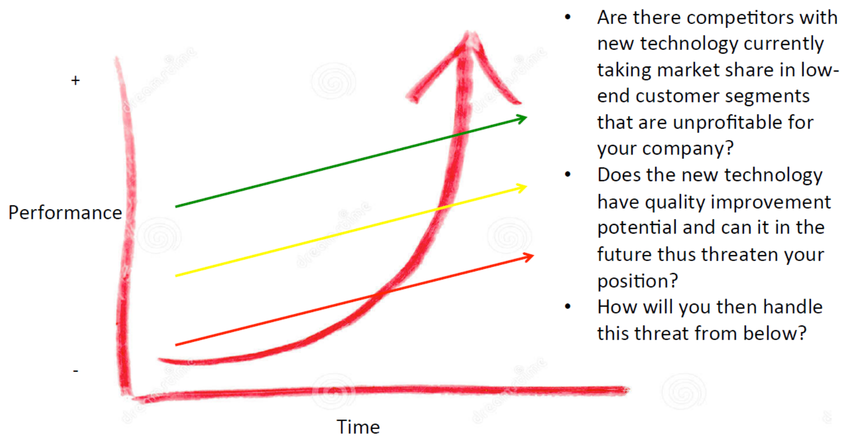
One of the key characteristics of the IT Hardware sub-industry is the short product life cycle of the products. The product that is the shining star of one year can be replaced by a new star the year after. The durability of returns on capital tends to be markedly lower than other industries as disruptive innovation can pull the rug under even the strongest players. These boom-bust-cycles are extremely hard to forecast. Two ways for vendors to try to mitigate this are first to institutionalize an innovative ability and make sure you cannibalize yourself instead of letting others do it plus secondly to add switching costs for the customers.

The IT Hardware sub-industry evolves by incorporating one new disruptive technology after another. It could almost be said that since disruptive technology more often than not originate with smaller non-public companies, investing in public IT Hardware companies is a bet on limited future change.

BYOD – Bring your own device

The pace of disruption has perhaps been marginally slower in enterprise dominated markets compared to retail consumer markets. This still holds, even though there has been a bring-your-own-device trend the last few years where employees have been permitted to bring their own smartphones, laptops, tablets etc. into the workplace IT-environment and thus blurring the line between enterprise and consumer equipment.

Picture 7.22



The threat of disruptive technologies as described by Clayton M. Christensen of Harvard Business School in his book The Innovator's Dilemma

Analyzing barriers to entry, the deeply integrated products supporting mission-critical applications are the stickiest. Within IT Hardware product life cycles vary in length and longer cycles like in enterprise storage, everything else alike, give competitors less chance of stealing a company's customers. Due to the prevalent model of standardized components, contract manufacturing and distribution through standard channels it is rare that companies have proprietary intellectual assets. The brand of Apple, the racks of blade servers and the proprietary file systems of enterprise storage vendors are exceptions. A proprietary file system or rack doesn't necessarily protect you from disruptive innovation though.

The main barrier to entry is instead to try to increase switching costs, making it hard, cumbersome or expensive for customers to change supplier. This can be done by building software content into products, using proprietary interfaces, adding system management tools, analytics etc. If products can be integrated into mission critical functions and become entrenched into the customer's business they can be less easy to replace. Apple is the obvious showcase of a company integrating hardware with an ecosystem of software and services plus adding a strong brand to increase stickiness.

Relative valuation multiples

IT Hardware is dominated by US companies and relative PE-ratios are a key valuation method. However, for faster growing companies without profits, P/S-ratios are used as well. With a slowing IT-spending IT Hardware companies are maturing and valuation metrics are not very different from other sectors. Valuation multiples that often test well in back tests for IT Hardware are P/FCF, EV/FCF and the PEG-ratio.

DCF scenarios

For the new entrants with unknown prospects we would recommend using a DCF with three probability weighted scenarios. Build the DCF to target a point in time, say 10 years into the future, and vary the size of the market, the market share, the margins etc. of the company in three scenarios. Examine which sales growth each scenario implies and try to estimate how probable this kind of growth has been looking at companies in general that have been in a similar situation and then decide on the probability of each scenario.

In terms of financial indicators investors in IT Hardware have historically focused heavily on sales growth, the change in sales growth, book-to-bill, gross margin and the change in gross margin. Due to the short product life cycles the sustainability of the life cycles of companies could be said to be less certain than in many other industries. Strong and escalating sales growth indicates momentum and good current prospects. A high and rising gross margin further indicates good operating prospects, pricing power and technology height.

Often momentum investors

Thus, tech investors are to a larger extent growth and momentum investors than in other industries, looking at factors like trading volume and 3 to 6 month price return as trend following strategies. However, it's seldom positive if sales growth is faster than gross profit growth and in the medium term, than net profit growth. The market typically doesn't like stocks that chase lower-margin revenue.

In any business costs and EPS are important and we shouldn't sketch too much of a caricature of IT companies as something exceptional, but in a sector of short product life cycles the cost focus will naturally be somewhat less. There is less utility from saving costs and reaching temporary high margins if a company in secular decline cannot keep its long term sales at a predictable level.

Due to the creative destruction of new technologies, investing in "turn-arounds" or "fallen angels" in IT Hardware is seldom a winning investment strategy. The history of IT Hardware is filled with disrupted companies like Xerox, Sony, Nokia and Canon. While there are examples of long-term leaders like Apple and IBM who have managed to sustain cross-era leadership this is rare.

Picture 7.23

Nokia's share of Finnish GDP between 1990 and 2012. Source: ETLA

Search for stocks where the company is increasing barriers to entry and the price is right.

The stock market too often prices new niche entrants as if all of them will succeed when in reality only a minority will survive the coming decade. If you as an investor manage to find the one new entrant that succeeds in for example storage virtualization, software defined networking etc. you will profit handsomely. This is however hard to pull off. Some new technologies are fads, some useful technologies can be integrated and disappear into other hardware products like a smartphone and even if everything works out it's hard to know which of all new entrants that will be the winner. We would look to invest in quality companies that are increasing stickiness, innovating for customers with efficient R&D and who try to build collaboration through shared eco-systems and by all this increasing the barriers to entry for competition – provided the price of the stock is right of course!

B) Communications Equipment

Fixed line vs wireless equipment

Products
- Data Networking
- Optical Equipment
- Telecom Equipment

While computers perform calculations on data, i.e. computes; communications equipment allows data to be shared. Products in this sub-industry fall in the segments fixed-line equipment and wireless equipment. Some also suggest a possible product sub-division into 1) data networking dominated by switches and routers with suppliers like Cisco, 2) optical equipment with for example the supplier Ciena and 3) telecom equipment with Ericsson and other vendors. Hence, the customers can be divided into an enterprise segment and telecom operators.

The Products and How They Work

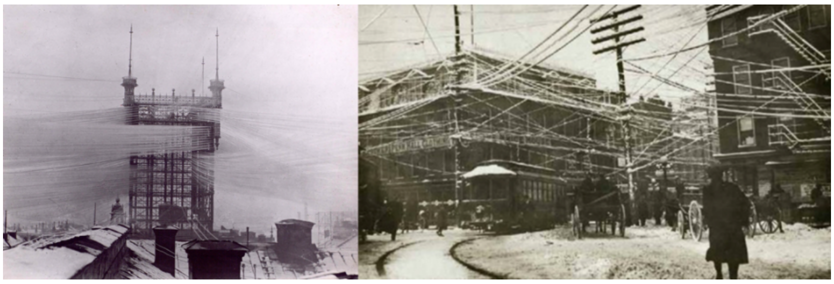
Basic Communications Network Technology

Stripped to its bare essentials a telecoms network consists of nodes connected by links. To this network end user terminals like telephones or computers are connected.

PSTN – Public Switched Telephone Network

Telephony started out as being circuit switched, i.e. opening up a separate dedicated communications line between those communicating that is kept until the communication is ended. This is like having a private traffic lane on the motorway specifically for your car only as soon as you start driving. The capacity of the networks was by today's standards low as the primary service was voice calls over the public switched telephone networks (PSTN), requiring 64 kbit/s.

Picture 7.24



To the left the Telephone Tower in central Stockholm (1887 – 1953) and to the right the city center littered in phone lines. An early example of circuit switched telephony. Source: SvD

Latency – the time between the source sending information like for example a data packet to the destination receiving it (if measured one-way).

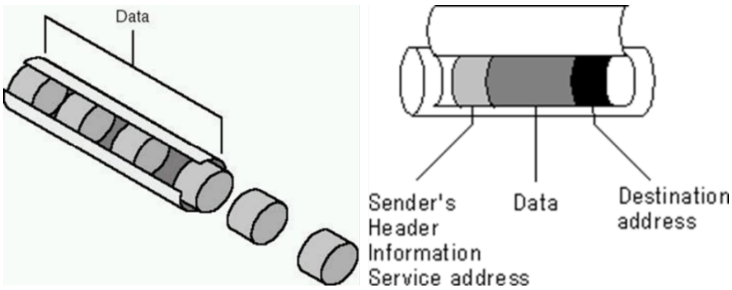
Carrier grade – quality measure demanding "five 9s"

Since an end to end circuit is set up before the call begins, a fixed share of network resources is reserved for the communication and no one else can use those resources until the session is terminated, the quality of circuit switched voice calls was high with minimal delays and low latency. Telecom networks focused on reliability. Hence the expression "carrier grade", referring to a network of very high quality and with an availability of 99,999 percent of the time. Speech is a so called isochronous service, requiring the transmitter and the recipient to be synchronized. Humans subconsciously fill in gaps and correct disturbances in voice communication but we have a hard time handling delays – think of the unintended comedy every time a TV-news reporter tries to report back to the studio over satellite link.

Still, the inefficiency of reserving one circuit for each session and the increased data transmissions paved the way for a transition to packet based switching. Since data traffic often is more "bursty" in its nature compared to the continuous communication of a voice call, packet switching technically makes good sense. Instead of holding a dedicated line open the data in the communication is broken up into small packets that can be sent along a shared line. A circuit switched session takes a pre-determined route through a network. Data packets in packet switching are sent on their way and could all take different ways to reach the recipient. The packets are then re-assembled before presenting an intelligible set of data to the recipient.

In circuit switching there has to be separate signaling to facilitate the set up and termination of a session and also to manage the billing of the session. This was initially done by pulses and tones (think the tone when you lift an old style fixed line telephone). In packet switching the data packets contain several so called headers with the "signaling" information that makes sending the payload from the sender to the receiver possible.

Picture 7.25



An illustrated data packet containing data bits with a description of its content. Source: Microsoft

Although more efficient, with packet switching follows the increased risk that some data packets are lost along the way. Some of the basic parameters of a network are its bandwidth (the usable frequency range for a transmission), the bit error rate and the latency plus the variations in latency. The quality ambition is to distort the data as little as possible when it reaches the recipient. Data is less sensitive to delays but disturbances to the traffic could lead to packet losses creating jitter (timing differences), resent packages leading to latency and worst case to packet losses that render the information worthless. Consequently, the header of the data packet also has to contain information to handle error detection etc.

WAN – Wide Area Network
MAN – Metropolitan Area Network
LAN – Local Area Network

Networks are often divided into so called wide area networks (WAN), metropolitan area networks (MAN or simply metro network) and local area networks (LAN). A WAN is the largest type of network; it could be global and generally consists of several smaller networks. The Internet is a WAN. The MAN is a medium sized network that also could consist of smaller connected networks. The LAN is a smaller computer network that could cover for example an office or a home.

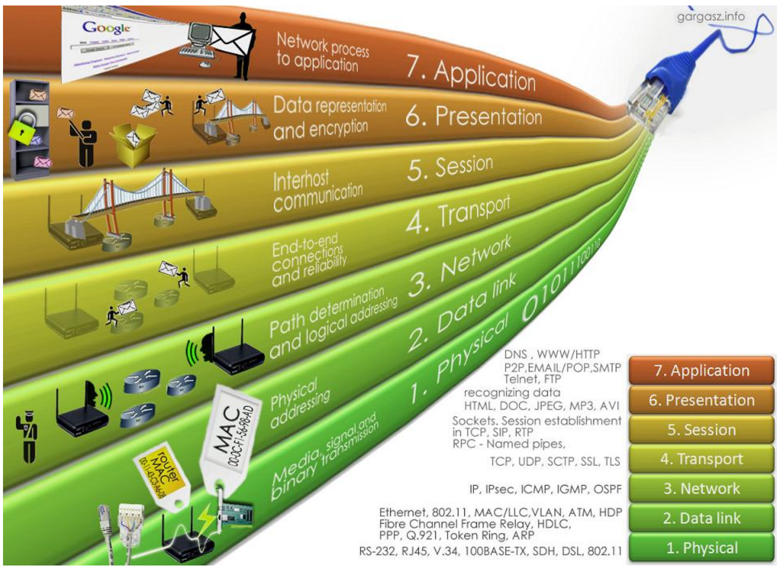
DOCSIS – Data Over Cable Interface Specification

Other categorizations of networks are between enterprise networks, cable networks and telecoms networks and further between fixed telephony networks and mobile telephony networks. We return to the mobile networks later. Cable networks use a technology standard called DOCSIS (Data Over Cable Service Interface Specification) using coaxial copper cables in the distribution network closest to the end consumers.

OSI model – Open System Interconnection model

The data packets cannot find their way over the network to the recipient on their own. A number of technologies are involved. The Open System Interconnection (OSI) model specifies a networking framework into protocols in seven layers (so called protocol stacks or protocol suites). The OSI model is a conceptual framework to better help us understand the complex interactions of packet based communication. Control is passed from one layer to the next. The lower layers of the model work independent of the higher ones.

Picture 7.26



The OSI 7 layer model. Source tutorial.netkromsolution.com

1) The Physical Layer

The physical layer 1 conveys a physical bit stream such as an electrical impulse, a light wave or a radio signal through the network at the electrical and mechanical level. The layer provides the hardware means of sending and receiving data on a carrier media.

2) The Data Link Layer

The data link layer 2, organizes the bit stream so it can be managed. Data packets are encoded from data bits to be sent between nodes and further decoded when they arrive. The layer adds transmission protocol knowledge, management and handles errors in the physical layer. Layer 2 technologies include Ethernet, Frame Relay and ATM.

Ethernet

Ethernet has over time become the dominant layer 2 technology for LANs and MANs. Ethernet divide streams of data into shorter pieces called frames (a type of larger "parcel" to transport several data packets over environments where they couldn't travel themselves). Each frame contains source and destination addresses and error-checking functionality. A frame is different from a data packet. Think of a frame as a car ferry transporting cars that otherwise couldn't travel in that environment. Ethernet governs how and how fast the transmission should be transmitted and when necessary it also handles collision control in network junctions.

3) The Network Layer

The number 3 network layer directs the data networks, provides switching and routing technologies, creating logical paths, known as virtual circuits, to transmit data from node to node. Addressing and packet sequencing are functions in this layer, as well as error handling and congestion control. Layer 3 technologies includes IP (Internet Protocol).

IP – Internet Protocol

MAC – Media Access Control

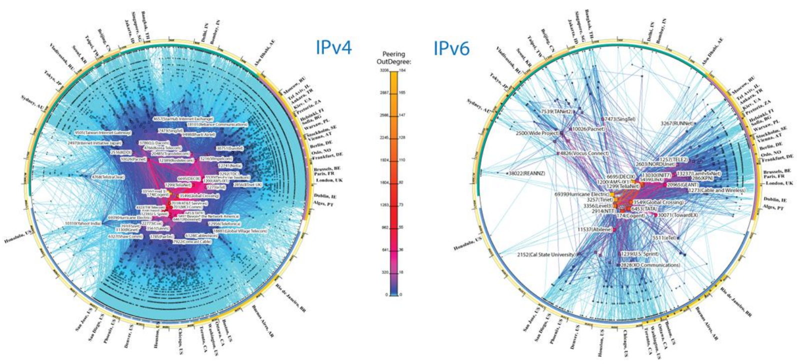
DHCP – Dynamic Host Configuration Protocol

IP is a protocol for assigning addresses to units in networks. IP is then directing the session between two computers with the help of these addresses. When a computer connects to a network it needs to be identifiable to be able to communicate. Instead of using the physical MAC address it uses a logic IP-address which is usually assigned to it by a DHCP-server (Dynamic Host Configuration Protocol). All data packets that are sent in the network contain the IP-address of the recipient and the sender in the header of the packets.

IPv4 vs IPv6

The currently most often used version of the Internet Protocol is IPv4 (i.e. version 4) which has a capacity of generating 2^{32} IP addresses (that is a maximum number of 4.294.967.296 addresses). This version was developed in 1982 and 4 billion addresses must have been seen as more than enough. It was not. Today there is a shortage of IP addresses and the last unused ones were allocated in 2011. Temporary remedies include sharing of addresses and "taking back" addresses that are not being used by a network element (and reassigning a new address if it becomes active again). The new standard is IPv6 that supports 2^{128} IP addresses (we are not going to write that one out – it doesn't fit one line of text). This should be enough to support even the most optimistic Internet of Things-forecast. The problem is that the two versions are not compatible and IPv6 has to be phased in in parallel with IPv4 in a cumbersome process. Early in 2015 about 90 percent of top 500 domains support IPv6 but out of the total Internet web-sites the share is still less than one fifth.

Picture 7.27



Visualization of IPv4 and IPv6 Internet topology. Source: Center for Applied Internet Data Analysis

Domain names

It's not practical to use IP addresses when trying to remember web-sites and the IP-addresses also change. To solve this problem a web-site uses a domain name like www.swedbank.com. A domain name links to one IP address.

4) The Transport Layer

The number 4 transport layer identifies the sender and the recipient, provides reliable transfer of data and is responsible for end-to-end error recovery and flow control. Transport technologies include TCP (Transmission Control Protocol). TCP is the mailman who delivers the data to the IP-address and makes sure the packets arrive in the right order. TCP survey the transmission, makes sure that the sender and recipient are properly verified, labels all the packets and if some of them are missing for the recipient after the transmission, TCP orders the missing packets to be resent.

5) The Session Layer

The session layer, number 5, manages sessions. The layer sets up, coordinates, and terminates sessions between the users at each sending and receiving end. In Internet traffic layer 5 and 6 functions are often made redundant.

6) The Presentation Layer

The number 6 presentation layer adds formatting and encryption. The presentation layer works to transform data into the form that the application layer can accept. This layer formats and encrypts data to be sent across a network, hopefully providing freedom from compatibility problems. Layer 6 technologies include GIF, JPEG and MPEG.

7) The Application Layer

HTTP – Hypertext transfer protocol

The final seventh application layer supports application and end-user processes. Communication partners are identified, quality of service is identified, user authentication and privacy are considered. Everything at this layer is application-specific. Layer 7 initiates a request or accepts a request for a session. Applications include www-browsers and HTTP (hypertext transfer protocol).

ITU – International Telecoms Union

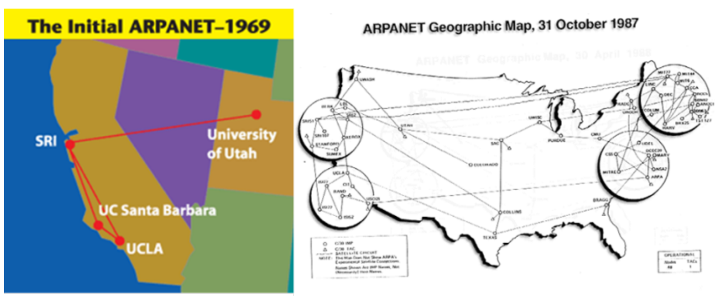
ETSI – European Telecommunications Standards Institute

Many of the above technologies are set by or at least heavily influenced by the standardization made by various standardization bodies. Since networking aims to set up communication between parties there is an obvious benefit to using compatible equipment and hence to standardization of how the equipment is constructed. Transparent standards also open up for competition between different vendors. The downside of a standardization process is long lead times and a certain amount of bureaucracy. The main telecoms body is the International Telecoms Union (ITU), which is also the oldest organization within the UN. In Europe the European Telecommunications Standards Institute (ETSI) is one of the heavyweights.

ARPA – Advanced Research Projects Agency

When discussing the OSI-stack the Internet takes a prominent place despite being a relatively young network. The Internet goes back to the US Department of Defense backed agency Advanced Research Projects Agency, ARPA (nowadays called DARPA, where the D stands for Defense). In the 1960s ARPA created the ARPANET, a network of computers connecting universities and research institutions. The network lacked a standard method for interfacing until the 1970s when Robert Kahn and Vint Cerf developed TCP/IP which allowed computers from different networks to communicate with each other.

Picture 7.28



The original ARPANET with four connected users in 1969. Source: SRI International

The information on the network was however in different forms and this made it difficult to organize and find. In 1989 Tim Berners-Lee at CERN in Switzerland proposed a hypertext markup language as the standard, came up with the name World Wide Web (WWW) and got help from a group of students at the University of Illinois to program the first web browser called Mosaic. Its user friendliness lead to a quick adoption of the World Wide Web and by the mid-1990s the Internet revolution kicked into high gear.

www – World Wide Web

Routers and Switches

We now turn to the many important nodes in the fixed line network, including optics and starting with data networking.

Routers and switches are the nodes that make sure that the packets end up at their intended destination. A switch is a device that filters and forwards packets between network segments. Switches operate mainly at the data link layer (layer 2) of the OSI model and therefore support any layer 3 packet protocol (like IP). A router is a device that forwards data packets along networks working at the number 3 network layer. Routers use headers and so called forwarding tables to determine the best path for forwarding the packets. Routers and switches can be small and applied in a datacenter/LAN environment or massive core equipment for WANs.

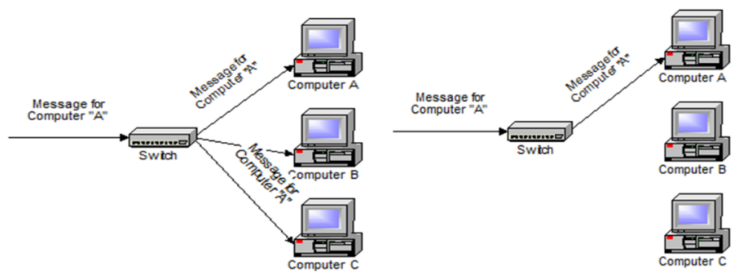
Switch layer 2

Router layer 3

A switch directs incoming frames to the appropriate port based on their destination MAC address. Initially, a switch knows nothing and simply passes on incoming messages to all ports. But by paying attention to the traffic that comes across it, a switch can learn where nodes with particular addresses are located. That means that subsequent messages for a recipient need only be sent to that one port. Switches learn the location of the devices that they are connected to almost instantaneously. The net result is that most network traffic only travels to where it was intended rather than to every port. On busy networks, this can make the network significantly faster.

MAC – Media Access Control

Picture 7.29



The functionality of a switch. Source: www.askleo.com

Routers come in all shapes and sizes, from small, four-port broadband routers to large, powerful devices that drive the Internet itself. It contains an operating system called firmware. A simple way to think of a router is as a computer that can be programmed to understand, possibly manipulate, and route the data that it's being asked to handle. As far as simple traffic routing is concerned, a router operates exactly as a switch, learning the location of the computers on its connections and routing traffic only to those computers. A router is however smarter than a switch and it not only sends traffic in a network, it can also send traffic between networks.

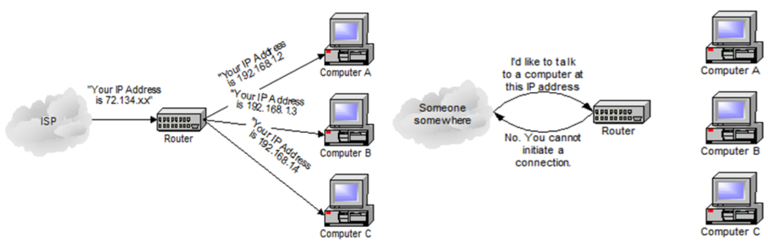
To send packets through a network routers connect to other routers. Data packets are then sent from router to router until it reaches its final destination. To be able to do this the router builds a list of the preferred routes between various parts of the network called a lookup table by communicating with other servers. When a packet arrives to the router, the router will send the packet to its processor to locate the destination IP address in the header, and then compare it to the lookup table, decide on the best path to the destination and forward the packet to the next router that is on that path.

DHCP – Dynamic Host Configuration Protocol

NAT – Network Address Translation

Routers perform at minimum two additional and important tasks: Dynamic Host Configuration Protocol (DHCP) and Network Address Translation (NAT). DHCP is the means of which dynamic IP addresses are assigned. NAT is the way that the router translates the IP addresses of packets that cross the Internet/local network boundary.

Picture 7.30



The functionality of a router. Source: www.askleo.com

The router can additionally act as a firewall and protect from malware that might spread by trying to connect to your computer over the network. With all the added functionality and handling there will inevitably be situations where the throughput of a router is too low compared to the traffic causing congestion and latency, i.e. slower networks.

Picture 7.31

Come on, you don't need a source for this one…

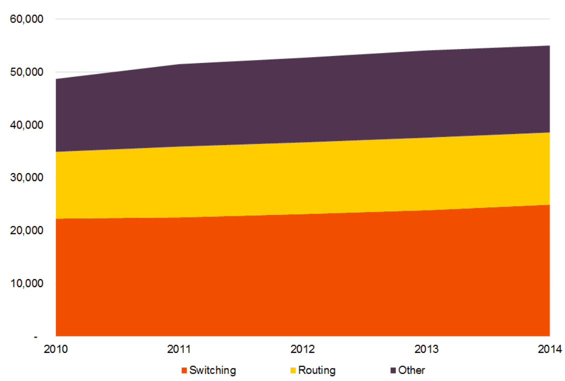
QoS – Quality of Service

With increasing amounts of traffic the need arises for prioritizing the more important data packets over others. Quality of Service (QoS) is thus a router functionality that prioritizes some traffic over other. Transmissions like for example voice over IP that require an uninterrupted data flow gets the inner track and the other traffic will have to wait in line.

DPI – Deep Packet Inspection

This trick is performed by so called deep packet inspection (DPI). Instead of simply examining a data packets destination address the router also inspects the content of the other headers to figure out the packet's type of content. DPI enables improved network management, better user service and added security functions. It also enables data mining and surveillance. Since the technology can be used to prioritize the traffic of those that pay more than others it has spurred a debate around so called net neutrality, where those who are afraid that the telecom operators will make the Internet a less free place team up with multinationals like Google who want to make sure that the telecom operators aren't allowed to charge more from "over the top companies" (OTT) who deliver Internet content and services on top of the operators' networks.

Picture 7.32



Sales of switches and routers in USD mn. Source: IDC, Bloomberg

Modems

Modem – modulator–demodulator

The task of a modem is to modulate a signal from an analog sine curve like a sound wave or a radio wave to an analog binary signal and vice versa. Hence, the name modem (modulator-demodulator) and the squeaking sound of the modems of the 1990s. The need for modulation arises as air and copper lines generally cannot transmit binary signals.

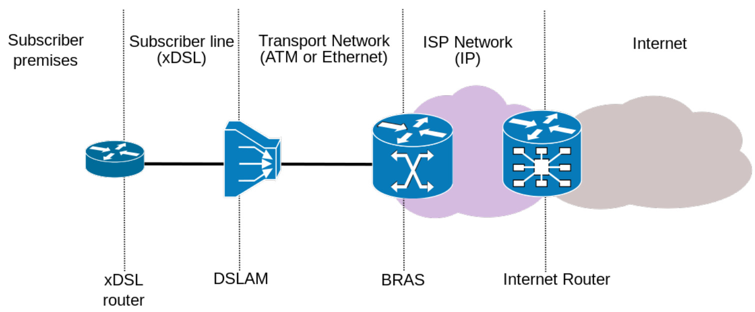
ADSL – Asymmetric Digital Subscriber Line

DSLAM – Digital Subscriber Line Access Multiplexor

VDSL – Very high-rate Subscriber Line

An ADSL router (Asymmetric Digital Subscriber Line) is a modem technology using the copper connection of the telephony network. In contrast to the 1990s modem technology, ADSL is always connected and it transmits data on a higher frequency than the voice phone call. Thanks to the higher frequency the capacity increases and the data transmission can share the copper line with a phone call without them disturbing each other. The capacity of ADSL not only depends on the transmission frequency but also on the quality of the copper line and the physical length of transmission cable. The longer the distance between the modem and the closest telephone exchange's DSLAM (Digital Subscriber Line Access Multiplexor), the larger the noise level and the lower the capacity. VDSL (Very high-rate Digital Subscriber Line) is a high capacity but short distance version of ADSL.

Picture 7.33



xDSL network diagram. Source: Wikipedia

ATM – Asynchronous Transfer Mode

The DSLAM in the telecom exchange station is firstly a collection of modems connecting to the individual ADSL routers. To save network capacity the DSLAM collects the data from several sending modem ports and aggregates the traffic into one composite signal through so called multiplexing (will be further described below). In this second capacity it works as an OSI layer 2 network switch using Frame Relay or ATM frames and IP packets to set up a datalink with the core network.

Optical Equipment

Optical networking is a means of communication that uses light as the carrier of data among nodes in a network. With reference to the above-presented OSI-stack we are now in the physical layer one. We are basically talking about lasers flashing light through glass fiber. A transmitter directs the optical equipment to turn the light on and off generating a light signal. The light source is often a laser diode. A receiver in the other end takes the signals, decodes them and sends an electrical signal to a computer, a TV or whatever end equipment the signal was intended for.

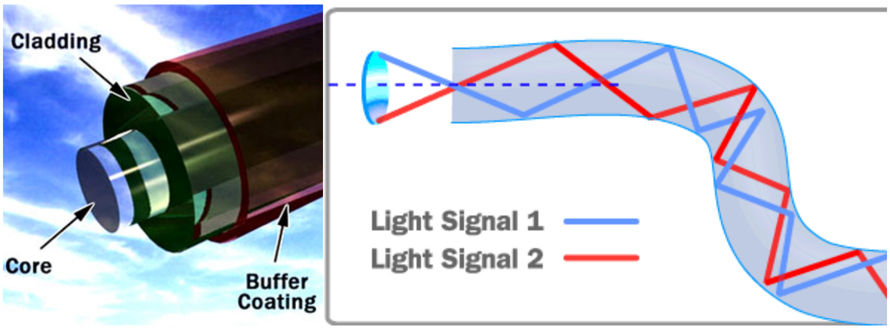
Speed of light ~ 300.000 km/s

The traveling speed of the carrier signal is obviously of vital importance for the speed that a network can transmit data. In a vacuum all electromagnetic radiation like a radio wave or a beam of light travels at the same speed. The speed is the fairly impressive 299.792.458 m/s, which equals 7.5 laps around the earth per second. Unfortunately, in other denser media than vacuum the speed will slow down and it will vary between types of electromagnetic waves and the types of environment it travels through. In general air will present a much noisier environment for wireless communication than the more controlled and disturbance free environments will for fixed telephony and hence the latter should be able to provide the speedier solution. Light traveling through purified glass is currently the solution closest to the ideal vacuum environment and fiber is therefore the "copper lines" for the next century. And the glass fiber is also much cheaper than cooper cables.

"Sending Morse code"

Fiber optic cables are strands of optically pure glass as thin as a human hair. The optical modulation of the carrier signal is in its simplest form to turn it on or off. This means that the laser light used for transmission is either on or off depending on the logical state 1 or 0 of the data signal and thus the optical network carries digital information. It's like sending a Morse code of dots and dashes with a flashlight in the night. The fiber consists of a core of glass where the light travels, an outer optical material called cladding that works like a mirror reflecting the light back to the core and a buffer coating made of Kevlar and plastic that protects the fiber from damage and moisture. The glass is so pure that if you made a window out of it you could make it several miles thick and you could still see through it.

Picture 7.34



The buildup of a fiber cable and its internal reflection. Source: howstuffworks.com

The light travels in the core by constantly bouncing from the claddings internal reflection and finally terminates in a port. Since the cladding doesn't absorb much light the wave can travel great distances. How long depends on the purity of the glass and the wavelength of the transmitted light. The transmission ability of a copper cable will benefit from being thicker. In optics the reverse is true. A thinner core will give the light wave less opportunity to bounce around between the walls and as the light will travel in a more straight line it will travel longer.

Dark fiber

A side note, you sometimes run into the cool sounding concept "dark fiber". This simply means fiber optic cables that are laid into the ground but are being idle as no light source is connected to the ends of the fiber.

SDH – Synchronous Digital Hierarchy

SONET – Synchronous Optical Networking

WDM – Wave Division Multiplexing

DWDM – Dense WDM

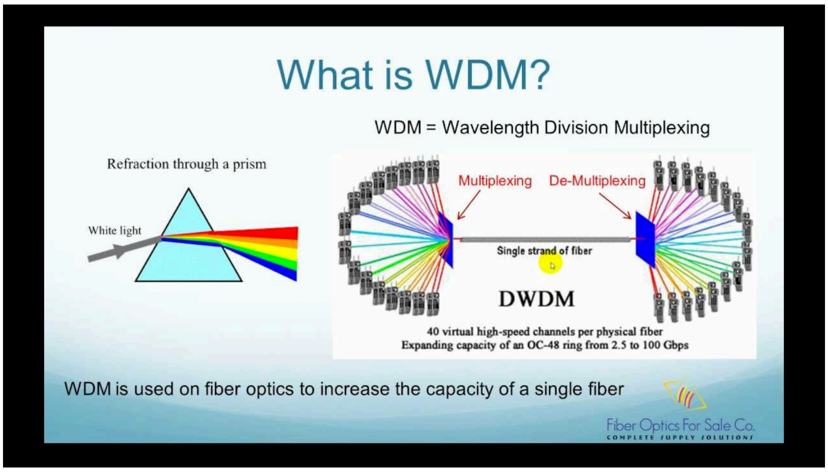
The optical equipment market can be segmented into two main buckets; 1) the legacy equipment using SDH/SONET (synchronous digital hierarchy/synchronous optical networking) transport protocols designed to transmit circuit-switched voice and 2) wave division multiplexing equipment (WDM or DWDM for Dense WDM). We will concentrate on the latter.

Lambda - wavelength of light transmitting data

Transmode

The invention of WDM meant that multiple individual light waves of different wavelengths, i.e. colors, originating from different users were combined into one stream that could be sent through one single fiber. One individual wavelength of light transmitting data is called a lambda. This splitting of the light into several simultaneously transmitting carrier waves allowed for a massive boost in capacity compared to the legacy SDH/SONET technology. Conventional WDM provides up to 16 carrier channels on the same fiber. DWDM uses denser channel spacing and can offer for example 80 carrier channels. Theoretically the maximum numbers of carrier waves are several thousands. Transmode is a supplier of both WDM and DWDM equipment.

Picture 7.35



WDM splits the white light into several carrier signals. Source: FO4SALE.com

ROADM – Reconfigurable Optical Add-Drop Multiplexer

The problem with bundling several signals and sending them to the same place is that some of them might be supposed to be heading in a different direction after traveling halfway. With a ROADM (reconfigurable optical add-drop multiplexer) it is possible to selectively remove or add traffic along the way. “Add” and “drop” refer to the capability of the device to add one or more new wavelength channels to an existing multi-wavelength WDM signal or to remove one or more channels. This can be done without converting the signals on all the WDM channels to electrical signals and back again to optical signals.

In general, the optical-electrical conversions in a network are a major source of latency and bottlenecks due to the lower bitrates of electrical signals. For example, when traffic is routed by a router, optical signals have to be converted to electrical signals to be inspected and when the router has decided where to send the packets they are converted back to optical signals.

The WDM market can further be understood by segmenting the market into metro versus long haul. With advanced modulation schemes and amplifiers metro systems can transmit over distances up to about 1,200 kilometers and long haul systems typically extend this distance to 4,000 kilometers or more. Amplification is done by an optical regenerator that consists of a special doped coating on the glass fiber that is pumped with energy from a laser. When a degenerated weak signal comes into the doped coating the energy from the laser is used to generate a new stronger light signal.

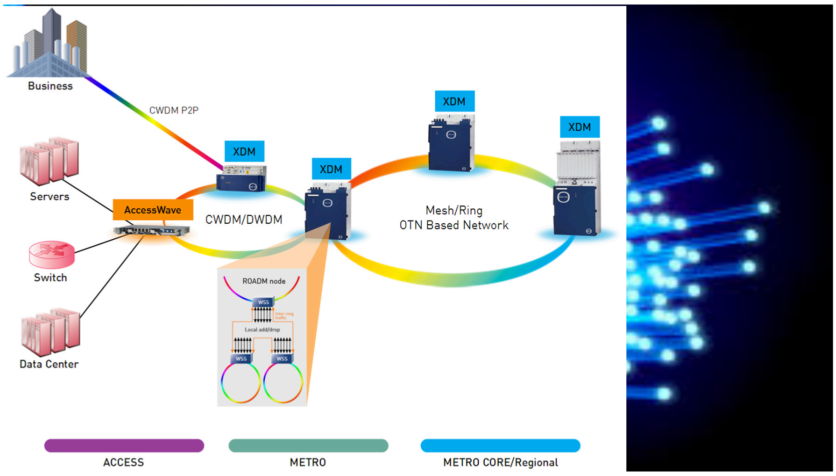
In the picture below the last mile access network is shown to the left. The multichannel signal is de-multiplexed to a single optical wavelength per designated user. One solution to accomplish this is through a so

PON – Passive Optical Network

AON – Active Optical Network

called passive optical network (PON) in which unpowered optical splitters are used to separate the single optical wavelengths to serve a user (think of a prism working in reverse). The alternative active optical networks (AON) instead use powered switching equipment. PONs are efficient, cheap, have low maintenance while AONs offers good interoperability with Ethernet and good flexibility.

Picture 7.36



Optic equipment applied in access networks and metro networks. Source: sagatelecom.com

In the 1990s technology and telecoms boom one of the straws that broke the back of many alternative telecom carriers was actually DWDM as suddenly the capacity of the networks expanded heavily without the same quick expansion of demand.

Mobile Telephony

Routers, switches, DSLAMs and optical equipment are important parts of what could be said to be the fixed line network. We now turn to the wireless network.

AT&T – American Telegraph & Telephone

AMPS – Advanced Mobile Phone Service

American history writing tells us that the first known mobile phone call was made by Martin Cooper at Motorola, who in 1973 set up a base station in New York and called his rivals at AT&Ts Bell Labs. In 1983 mobile telephony was launched commercially in the US using the analog standard AMPS with Motorola's DynaTAC-phone which had the size of a brick.

In reality the Swedes were way ahead of the Americans and to give some background to the prominent position of the Nordic countries in mobile communications we thought we would take a short detour into mobile phone history. In 1949 the Swedish government agency Telestyrelsen (now a part of Post- och Telestyrelsen, PTS) gave Ragnar Berglund and Sture Lauhrén the task of developing a circuit switched mobile network and in 1950 Laurén could call "Fröken ur" from his car making the world's first mobile phone call.

In 1956 a commercial mobile network called MTL was launched in Stockholm and Gothenburg by state owned Televerket. MTL and its successor MTB that also extended to Malmö in 1983 had a total of 660 subscribers. The main problem apart from offering an exuberantly expensive service was that the local mobile networks only could communicate internally. To lower the prices, get interoperability and better coverage a Nordic standardization effort was initiated for a network called MTC. This network would soon change name to NMT(Nordiskt mobiltelefonisystem).

NMT – Nordiskt mobiltelefonisystem

In 1979 NTT (Nippon Telegraph and Telephone Corporation) launched a commercial network using a proprietary standard and in 1981 Comvik was the first to launch a service in Sweden using NMT – a week before Televerket launched their service. The standard was developed in a group led by Thomas Haug from Televerket and NMT built on the older technology developed by Televerket with Östen Mäkitalo from Televerket Radio conducting the system tests.

NTT – Nippon Telegraph and Telephone Corporation

Partly Ericsson owned, Svenska Radioaktiebolaget started to sell NMT base stations and other network equipment. The company became fully owned by Ericsson in 1981, subsequently changed its name to Ericson Radio Systems and with early pioneers like Åke Lundqvist over time became the dominating business of the telecom equipment company. Since NMT was developed to be a multinational standard from start and not a proprietary service, the technology became the dominating first generation standard globally, giving the Nordics a head start in the area and the second generation Nordic-developed technology GSM (Groupe Spécial Mobile, after the name of the group that specified the standard) at its peak had a 80 percent global market share.

Ericsson Radio Systems

The technology generations of the mobile networks have been named 1G, 2G and so on meaning first generation, second generation etc. Below we show a map over some of all the mobile technologies globally starting with 2G. US D-AMPS and Japanese PDC migrated to the European GSM-track. In 3G China launched a proprietary technology called TD-SCDMA but this never really took off. GPRS (General Packet Radio Service) was the first packet based mobile communication and its heir EDGE (Enhanced Data Rates for GSM Evolution) today functions as a fall back for data traffic when the 3G or 4G network coverage is missing. GSM migrated to UMTS (Universal Mobile Telecommunications System) with its brand new air interface

D-AMPS – Digital AMPS

PDC – Personal Digital Cellular

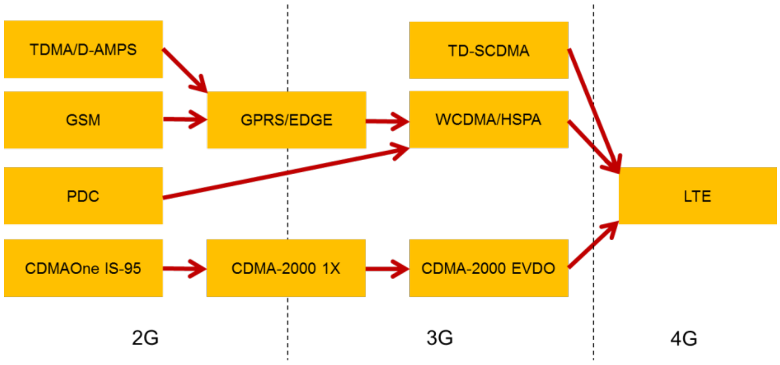
GPRS - General Packet Radio Service

EDGE - Enhanced Data Rates for GSM Evolution

UMTS - Universal Mobile Telecommunications System

WCDMA - Wideband Code Division Multiple Access technology WCDMA (Wideband Code Division Multiple Access).

Picture 7.37



Mobile technology generations. Source: Swedbank

Each generational transition has meant higher spectral efficiency measured as more transmitted bps/MHz. However, the progress of any radio access technology is bound by the laws of physics and the spectral efficiency cannot move beyond the limit of the so called Shannon's Theorem as defined by the brilliant Claude Shannon, commonly seen as one of the founders of information and communications theory. To reach a recipient a signal has to be a certain times stronger than the noise level of the environment (so called signal-to-noise ratio measured in decibels, SNR). The upper transmission bound of a radio link is a function of the available bandwidth and the SNR.

SNR – signal-to-noise ratio

Picture 7.38

Shannon's Theorem

$$C = B * \log2(1 + S/N)$$

where C is the achievable channel capacity, B is the bandwidth of the line, S is the average signal power and N is the average noise power.

Shannon's Theorem. Source: Freie Universität Berlin, Fachbereich Mathematik und Informatik

CDMA – Code-Division Multiple Access

LTE – Long Term Evolution

OFDMA - Orthogonal Frequency-Division Multiple Access

MIMO – Multiple Input Multiple Output

64QAM – 64 Quadrature Amplitude Modulation

The two still standing mobile technology contenders, the US CDMA-track and the European GSM-track partially started to merge already in 3G but in 4G the world will at last have one unified mobile technology with LTE (Long Term Evolution). LTE does as all new generations offer higher speeds. The spectral efficiency increase through the use of 1) OFDMA (orthogonal frequency-division multiple access) - a wireless access method that is relatively more resistant to radio wave interference, 2) MIMO (multiple input multiple output) which is a technique to use multiple antennae at both the transmitter and the receiver and 3) 64QAM (quadrature amplitude modulation) which is a modulation scheme with higher modulation rate (see below).

LTE also offers significant improvements in latency plus a migration to packet switched transmission of voice calls over IP (sometimes called VoLTE, Voice over LTE as a paraphrase to the expression VoIP, voice over IP). VoLTE will open up for significantly better sound quality during mobile conversations through increased sampling rates. Within all the above standards there have been numerous intermediate versions. LTE is being upgraded to LTE Advanced which theoretically can support speeds of up to 1 Gb/s.

VoLTE – Voice over LTE

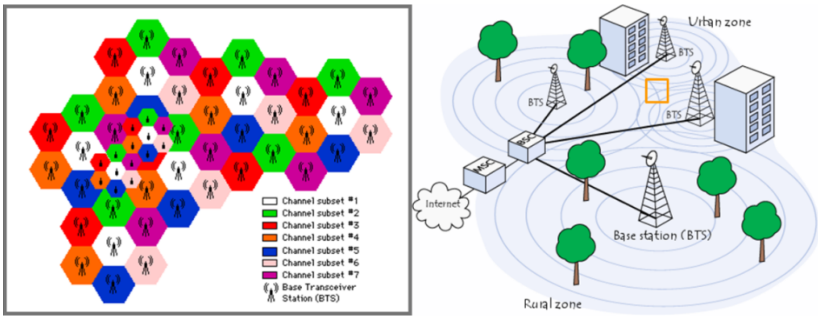
A mobile network consists of radio cells that cover a geographical area. Each cell belongs to one site that could contain one or more cells. Each cell has a base station and each base station has one or several pieces of radio equipment (in GSM called transceivers, TRXs) receiving and transmitting signals to the mobile phones. To minimize interference a distance has to be maintained between cells using the same frequencies. In the below picture 7.39 to the left a specific frequency is symbolized by a specific color. In areas with congested traffic it is possible to increase capacity by splitting one cell into several smaller cells with lower transmission power reusing the same frequencies as in the larger cells.

TRX – Transceiver in GSM

When a mobile user moves from one cell to another he must be able to retain the communication or the call will be dropped. The process by which this occurs is called handover. Network elements monitor the position of the user and the quality of the link from the mobile phone to the base station. At the same time the availability and quality of nearby channels in neighboring cells are monitored. The user is then switched to the new cell when the quality of its signals is higher than those in the old cell.

In real life a cell is not hexagon shaped but more circular. This means that there could be gaps between cells where the coverage is poor as illustrated by our orange square in the picture. In WCDMA and LTE there is a phenomena called cell breathing where the size of the cell shrinks when the traffic load is heavy, potentially creating gaps in the coverage. The picture 7.39 to the right also illustrates the transmission between the site towers and the core network. The boxes in the picture are those of a GSM-network but the logic is similar in later generation technologies. The transmission links are nowadays either fiber or radio links.

Picture 7.39



Architecture of a cellular network. Source: MSPVL Polytechnic College, Pavoorchatram for the left picture and Kioskea.net for the right picture

In licensed wireless communications such as mobile phone networks each telecom operator, often through placing the winning bid in a public auction, has the right to transmit over a certain range of the electromagnetic spectrum. Different parts of the electromagnetic spectrum with different frequencies have varying ability to penetrate the earth's atmosphere and also different ability to penetrate dense objects. The most suitable part of the spectrum for transmission of communication through air is the radio spectrum ranging from 3 kHz to 300 GHz but there are variations within this spectrum as well.

Radio spectrum – 3 kHz – 300 GHz

The lower the frequency the better as the radio waves higher up in the spectrum are so miniscule that they are very easily disturbed by the environment. Radio waves with higher frequencies than 2 GHz have difficulties penetrating objects like the concrete wall of a building or passing through a tree and they are also attenuated by the oxygen molecules and water vapor of the air. Above 3 GHz the wave requires free line of sight between sender and receiver for satisfactory transmission. This is contrary to the low frequency radio spectrum called VLF (Very Low Frequency) that transmits over vast distances and is used for example by maritime radio and navigation systems.

VLF – Very Low Frequency

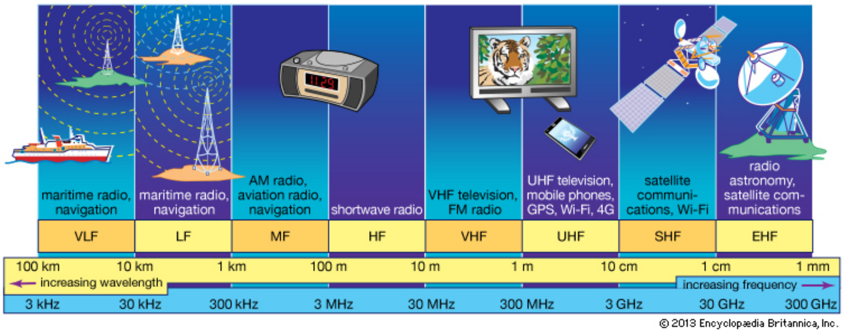
However, higher frequencies can also deliver higher capacity than lower so there is an advantage for an operator to be able to cover rural areas with for example 900 MHz and complement this with urban coverage using for example 18000 MHz. Further, unconstrained by the atmosphere, radio waves with the highest frequencies are used for satellite communications and astronomic observations.

The radio spectrum is a finite resource and large parts of it are reserved for military and governmental use. Mobile networks transmit over the UHF-part of the spectrum (Ultra High Frequency, 300 MHz to 3 GHz). Being first to the party 1G-technologies like NMT transmitted over the excellent 450 MHz frequency. Several of the later generation

UHF – Ultra High Frequency

technologies have been forced to use frequencies above 2 GHz which means shorter transmission ranges. Shorter transmission means smaller cells, with the subsequent need for more base stations to cover the same area and also worse indoor coverage – causing a need for specific indoor small cell systems. The relationship between frequency and cell size is almost linear where you need twice the amount of cells when moving from 900 MHz to 2.1 GHz.

Picture 7.40



The radio spectrum part of the electromagnetic spectrum. Source: Britannica

The size of the cell is further affected by the output power, i.e. the power with which the signal is sent. A small cell solution for indoor coverage is simply a low powered base station. The output powers of base stations and of mobile phones are regulated for safety reasons.

All the technologies in the previous picture 7.37 are air interface technologies that combine a number of choices between frequencies, duplexing methods, modulation schemes, multiplexing types etc. Below we shortly touch on these topics. Further, most of the technologies also specify the network architecture for the transmission network from the base stations to the core network.

We start with duplexing; in most types of communication there is a need to communicate two ways (ask your spouse…). There are two main types of duplexing technologies in mobile telecoms, allowing the simultaneous sending and receiving of data. The most common one is FDD (frequency-division duplexing) where the network uses different frequencies for the uplink sending data from the end user terminal and for the downlink sending data to the end user terminal. The other solution is TDD (time-division duplexing) that only uses one frequency but instead divides this into time slots. Some of the time slots are then used for the uplink and others for the downlink. A satellite broadcasting TV-signals is instead an example of simplex, i.e. communication only going in one direction.

FDD – Frequency-Division Duplexing

TDD – Time-Division Duplexing

A radio wave is a continuous sine curve. How is it possible to use this wave as a carrier of communication? The simple answer is, by manipulating its form and shape in a way that can be understood by the receiver. This is called modulation. Looking at picture 7.4 a radio wave can be compressed to become thinner and this is called amplitude modulation, it can be pulled apart to have varying wavelengths which is called frequency modulation or it can be made to skip a half beat called phase modulation.

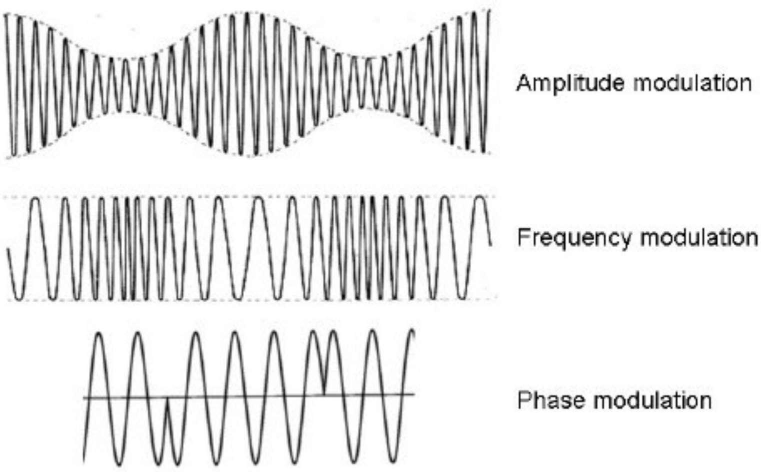
Amplitude – measure of the height of an electromagnetic wave, which indicates the strength of the signal.

Frequency – the number of oscillations per second for an electromagnetic wave.

Wavelength – the distance between two consecutive maxima or minima of the wave. The higher the frequency the shorter the wavelength.

Phase – the angle of the wave.

Picture 7.41

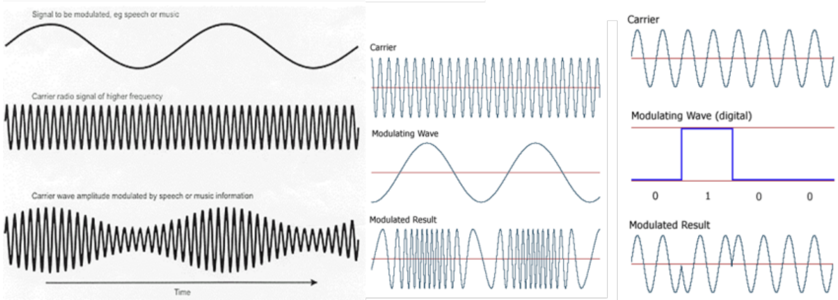


Different modulation schemes. Source: Ofcom

Modulation rate – symbols per second

The below pictures show how a carrier wave combines with a modulating signal to produce different types of modulations. The modulating signal could be a sound wave or a binary code. The modulation is performed by a modulator and the demodulation separating the signal from the carrier wave on the receiving side is done by a demodulator. The so called modulation rate is the rate at which a carrier wave is purposely varied per second, i.e. the number of symbols per second. One variation or symbol, is not the same as one binary data bit even though they are related. With higher frequency the number of oscillations per second increases and by this allows for more symbols per second. However, higher frequencies are also more sensitive to disturbances blurring the modulation and lowering the throughput.

Picture 7.42



Effects of signal on carrier radio signal in different modulation schemes. Source: Ofcom & ITU

Multiplexing

- TDMA - Time Division Multiple Access
- CDMA - Code Division Multiple Access

Of the technologies in picture 7.37 D-AMPS, GSM and PDC are so called TDMA-technologies (Time Division Multiple Access). The CDMA-track, WCDMA and LTE are CDMA-technologies (Code Division Multiple Access). The two are multiplexing technologies that combine multiple signals or data streams into one signal over a shared medium to save bandwidth. Multiplexing is performed by a multiplexor and at the receiving end there is a demultiplexor separating the data streams (or simply a “mux” and a “demux”). As is evident by the name, TDMA just like TDD use time slots on a single frequency but instead of ensuring duplex connection the goal with TDMA is simply to combine several different signals into one by slotting them into the time slots.

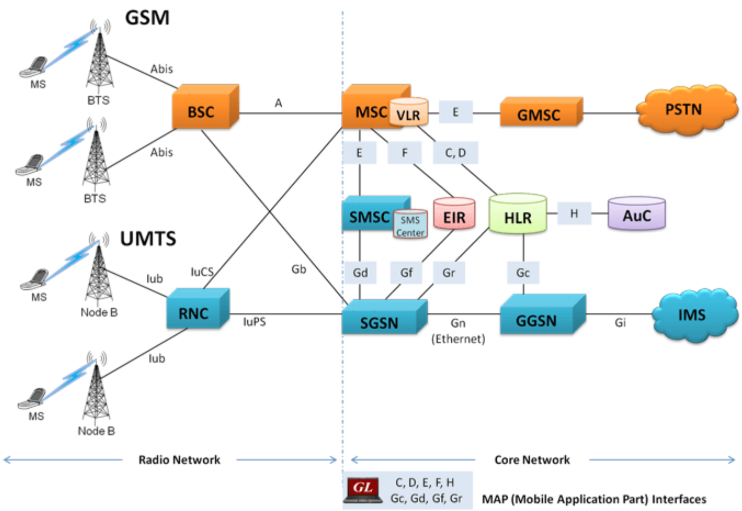
CDMA assigns a unique code to each end user to modulate their signal and then transmit all the users’ signals spread over a shared channel. The separation of the signal in the receiving end will be done by correlating the signal with the coding of the desired sender and extracting the signal. An often used analogy is that of a busy room where people want to talk to each other simultaneously. In TDMA people take turns speaking (time division). In CDMA people speak in different languages (code division) understanding those who speak their own language and tuning out those who speak the other ones.

A mobile network is more than the radio transmission. It consists of a number of “boxes” performing various duties and the connections between the boxes. LTE has a so called flat architecture where the functionality of the transmission network equipment has been moved to the base stations or the core network.

Picture 7.43

Boxes…

RNC - Radio Network Controller
SGSN - Serving GPRS Support Node
BSS - Base Station Subsystem
RAN - Radio Access Network
IMS - IP Multimedia Subsystem
SMSC - Short Message Service Center
HLR - Home Location Register
AuC - Authentication Center
SIM - Subscriber Identity Module
EiR - Equipment Identity Register
OSS - Operations Support Systems
BSS - Business Support Systems



Network elements in GSM and UMTS. Source: indiaprwire.com

BTS – Base Tranceiver Station

Due to its importance to Ericsson the last two decades the base station is perhaps the most well-known network element (not counting the mobile phone). In the GSM-network the base station is called BTS (Base Tranceiver Station, where "transceiver" combines the words transmitter and receiver) and in later technologies it is called node B (eNode B in LTE). In principle a base station works like a giant mobile phone. It's got 1) one or several antennas placed high up in a mast or a roof, 2) a radio unit handling filtering, amplification, duplexing etc. enabling the transmission of the physical signal and 3) a baseband that handles modulation and demodulation, encoding and decoding etc., managing the transmitted data as well as power management.

Base station virtualization

A side note: at times Ericsson talks about the current virtualization of the base station scaring the bejesus out of US tech investors. This is however not equivalent to the virtualization that has taken place in the server area, where application specific hardware to some extent has been replaced by commodity hardware and overlay management-software creating virtual dedicated servers. Instead, what Ericsson is referring to is a physical split between the radio unit and the baseband unit where the former is moved from the ground to the antenna. First, this shortens the distance the weak and error prone analog signals have to travel between the antenna and the radio unit. Secondly, it allows a number of radio units to share the same centrally placed baseband unit.

5G launch 2020

Before we leave the topic of licensed wireless communications we thought we would take a sneak peek at what 5G might bring when it is launched around 2020. First, it must be said that the standardization of 5G hasn't started yet. Still, there seems to be a rough consensus over what direction to take. 5G will continue to use later enhanced versions of LTE for macro coverage. This ensures backward capability with earlier equipment.

However, the aim is to step away from the one-size-fits-all type of network of previous generations. To be able to provide more tailored services to a large variety of user cases 5G will look to utilize a mesh of several radio networks using complementary frequencies. These could be both licensed and unlicensed. For applications like driverless cars the latency will potentially be paramount while the transmission range could be of less importance. In this case solutions built on using licensed frequencies ranging from 10 to 30 GHz would be suitable. World Radio Conference (WRC) is expected to release these types of frequencies on the market in 2019. Other applications with other types of demands will be served with another mix of coverage area, duration, capacity, speed, latency, robustness, security and availability.

To be able to provide this type of flexible services, the network construction will also have to be more flexible. The tools to manage this will center on software-defined networking and virtualization of network functions allowing networks to be broken up into modules plus flexible radio-access solutions that can adapt to different user cases and frequencies. LTE and the other radio solutions will have to work in parallel utilizing for example dual-connectivity where a device maintains simultaneous connectivity with two radio access solutions servicing different needs. Other key focuses are energy saving measures by minimizing the signaling in the networks unrelated to sending a payload.

Unlicensed Wireless Technologies

Typically a mobile operator uses something like 40 MHz of licensed spectrum for a service. This amount of bandwidth pales compared to the amount of available unlicensed spectrum that could be 10 times larger. In the late 1990s the standardization body the Institute of Electrical and Electronic Engineers developed a standard for short range unlicensed wireless communication that got the unsexy name IEEE 802.11. Several technology generations later this is what we call Wi-Fi (wireless fidelity) or sometimes WLAN (wireless local area network) – although the latter is in reality a broader concept. Wi-Fi transmits on the 2.4 GHz or 5GHz frequencies and replaces the network cable as a convenient way of hooking up to a fixed network.

IEEE - Institute of Electrical and Electronic Engineers

Wi-Fi – Wireless Fidelity

WLAN – Wireless Local Area Network

Wireless N (IEEE 802.11n) is currently the most frequently used version and it can under ideal conditions deliver speeds of 300 Mb/s. Unfortunately, with Wi-Fi the conditions are seldom ideal. The 2.4 GHz spectrum is a busy place. Several Wi-Fi networks often compete for the capacity together with microwave ovens, cordless mice, video or music streaming devices, baby monitors, Bluetooth units and God knows what else, all sharing the same spectrum. The theoretical speeds are even higher on the 5 GHz frequency but due to the high frequency of the radio wave the transmission range is much shorter.

Bluetooth

NFC – Near Field Communication

Other short range unlicensed wireless technologies include Bluetooth and near field communication (NFC). Bluetooth which was invented by Ericsson and is named after the Danish king Harald Blåtand is used to connect units like computers, mobile phones etc. The transmission range of different Bluetooth technologies varies from 1 meter to 100 meters.

NFC is used when entering the subway, in the key ring used at the workplace entrance or in the card opening the hotel door. Recently NFC-technology has started to be included in smartphones to allow them to also function as payment cards, entrance cards etc. The range of the technology is only about a centimeter or two. Unlike Bluetooth and Wi-Fi, NFC doesn't need a transmitting technology for its communication – it just needs to be really close to the other unit. In for example a card used to enter the subway there is a small NFC-tag in the card. This tag has no battery making it very low maintenance. The energy needed to read the tag is supplied from the reader unit.

The Market and the Business Model

There are in effect two communications equipment markets where the first is serving enterprise customers and mostly depends on the enterprise IT budget and the second is serving telecom operators living of the equipment part of their capex budget. Telecom operators spend about the equivalent of 15 percent of their sales on capex and out of this perhaps a third is communications equipment.

The economic cyclicality spreads to the communications equipment manufacturers through the capex budgets of large enterprises and telecom operators. The growth of the communications equipment market is as such tied to GDP growth but technology upgrade cycles play a larger part than with IT hardware. This has been the most obvious in wireless technology where every generation of network technology has required new investments in hardware and software.

The market for fixed telephony equipment contains several product niches with separate vendors, even though players like Huawei and Alcatel have very broad product lines and often offer one stop-shopping. The market for wireless equipment is considerably more consolidated with a few large suppliers. The contenders in the space have however changed massively over the last one and a half decade. Late 1990s there was a dozen credible vendors. The TMT-crash plus the emergence of the Chinese competitors Huawei and ZTE have forced all the North Americans to hand in the gloves and several of the remaining Europeans have merged.

Ericsson

In this respect Ericsson has managed to navigate very choppy waters in a skillful way, actually gaining market share in the transformation. Everything else alike, in a business where the fixed R&D expenses are high and prices has seen pressure it is rational for markets to consolidate.

Equipment for mobile networks has several proprietary interfaces. A vendor's equipment is as such not compatible with another vendor's and thus a manufacturer gets a contract to supply a radio network in one coherent region. The trick for the vendor is to get the foot into the door. The bidding process for new wireless network installations is highly competitive and the winner often suffers a drag on margins. Suppliers take an upfront hit on margins rolling out the initial network to be able to gain more profitable software and hardware upgrade business later.

SDN – Software defined networking

The data networking market is entering into a transformation to software defined networking (SDN) that will change LAN switching and potentially down the road also effect WAN equipment suppliers. The speed and magnitude of changes are obviously hard to predict but since the improvements in efficiency and in total cost of ownership are real the case is compelling. As mentioned previously (see the section Virtualization and Software Defined Everything) the traditional datacenter entertained a number of proprietary equipment in the building blocks compute, storage and networking, all operating in silos with low utilization.

By a virtualization process the underlying hardware and its operating software are separated allowing automated and policy based provisioning and managing of the underlying hardware. This has already happened in compute where software defined compute (SDC) is a fact. The processes of enabling software defined storage (SDS) and SDN are in their early stages. The vision for many is a fully virtualized datacenter – a software defined datacenter (SDDC) only using software defined infrastructure (SDI). Sorry for all the acronyms but when an area is hot they tend to pop up.

SDC – software defined compute
SDS – software defined storage
SDDC – software defined datacenter
SDI – software defined infrastructure

The hyperscale cloud datacenters of Amazon, Google, Facebook and the likes are the forerunners in this process and have used SDN in their internal LAN environments for some years. The separation of the software control plane and the hardware data plane improves utilization of the hardware as several virtual machines (VM) can populate the same hardware and thus decrease volume need for "boxes"; and the separation of hardware and software facilitates a decoupling of the faster software innovation cycle from the slower hardware cycle. Further, the hardware can now be cheaper white boxes putting pressure on pricing. Given their scale the cloud datacenters have built proprietary operating systems, for example managing Gmail in the case of Google, running on large numbers of white boxes.

VM – virtual machine

Obviously this is a threat for the incumbent providers of LAN switching equipment, like for example Cisco, Juniper and Arista, who sell products bundling software and often quite expensive hardware. Further, through new software (OpenStack etc.) in a so-called orchestration layer, improved functionality is added for simultaneously managing virtualized compute, storage and networking, i.e. synchronizing the management of the entire datacenter. This could mean dramatic cuts to the administrative staff costs of the datacenter but also less revenues for IT consultants.

Apart from the hyperscale cloud datacenters a number of telecom operators, having massive datacenters as well, have started SDN-trials for their LANs. The snowball has started to roll even if it's contained to the LAN environment so far. In a recent survey performed by SAP and Oxford Economics the respondents among a broad set of sectors, 99 percent said that cloud computing was a part of their IT-strategy and 69 percent claimed to plan to invest heavily.

SDN is not only a negative for communications equipment providers since interconnection between virtualized and cloud based datacenters will increase and by this also increase the amount of network communication overall. Neither is SDN without problems. There have been scaling problems as the handling of the LAN data switching is centralized to a single overlay software function. This problem is obviously massively larger when moving into a WAN environment.

When the server industry saw this parallel development starting about a decade ago the incumbent providers were reluctant and slow to respond as they didn't want to cannibalize their business and VMWare was able to corner the market. This doesn't seem to be the case in networking as Cisco is head to head to VMWare in product development and Juniper isn't too far behind either. This doesn't prevent the cannibalization but it's at least better to cannibalize yourself than letting someone else do it. It will be an interesting market battle to follow the next few years.

What should an investor think of?

Focusing on the sales of wireless equipment, the revenue growth has been depressingly low the last decade and a half; and this is despite an exponential growth in network traffic. The reason is that looking at the equipment handling the traffic the equipment price per transmitted bit has gone down to negate most gains from volume growth. Why is this? A side answer is that Moore's law makes any partly semiconductor based equipment cheaper to produce over time, but the main reason for this is that the funds that flow into the value chain, i.e. the revenues of the telecom operators globally are growing at a pedestrian pace. Telecom service is a very public and politicized business and as such has special sector regulation; much like the power utility sector. The regulators simply haven't allowed the operators to benefit from the traffic expansion.

The funds going into the value chain aren't growing

If the funds available to the equipment sector's customers aren't growing, they cannot spend an increasing amount of money on equipment. That is, if the equipment companies can't force them to spend a higher proportion of their sales on capex. In other words the relative bargaining power is important. Sadly, for the sake of the wireless equipment vendors the telecom operators are huge organizations with plenty of clout. To top it off the arrival of the new Chinese vendors that were more interested in gaining market share than in the medium term level of margins and returns on invested capital didn't help the situation.

What could change to the better for the wireless equipment vendors? Either the operators are allowed to have some increased pricing power towards their customers or the operators loose some of their bargaining power towards their suppliers. Since the modern society is very dependent on telephony, the Internet and on networking it is in the regulator's interest to promote investments in the networks. This can only be done if the telecom operators are allowed to make a decent return on investment. Thus, the regulators cannot only look to the short term crowd pleasers of regulated price decreases.

It is also reasonable to envision that now when the Chinese companies have become relatively large, they should start protecting what they've got instead of inserting massive price pressure to take market share. Declining prices now hurt them as well. If this were to happen the wireless equipment space could become a rather nice place to be in as the number of players now is down to less than half a dozen. This later scenario is however a one trick pony – when the increased bargaining power has caused a step change in the telecom operator's capex-to-sales the effect is done. The best scenario is no doubt that the growth of the telecom operators would shift up as it would benefit the entire value chain longer term.

The sale of wireless equipment is rather choppy with large contracts that could come into or fall out of a quarter and where different contracts have hugely varying margins. Combine this with the industry's high fixed R&D costs and the result is a quite unpredictable bottom line in the short run. The high fixed R&D costs also make economies of scale critical to survive in the long run. Further, services have become a much larger part of the industry's revenues as the equipment manufacturers take over parts of telecom operator's network management. The higher mix of recurring service revenues has helped to stabilize sales somewhat.

Happy investing.