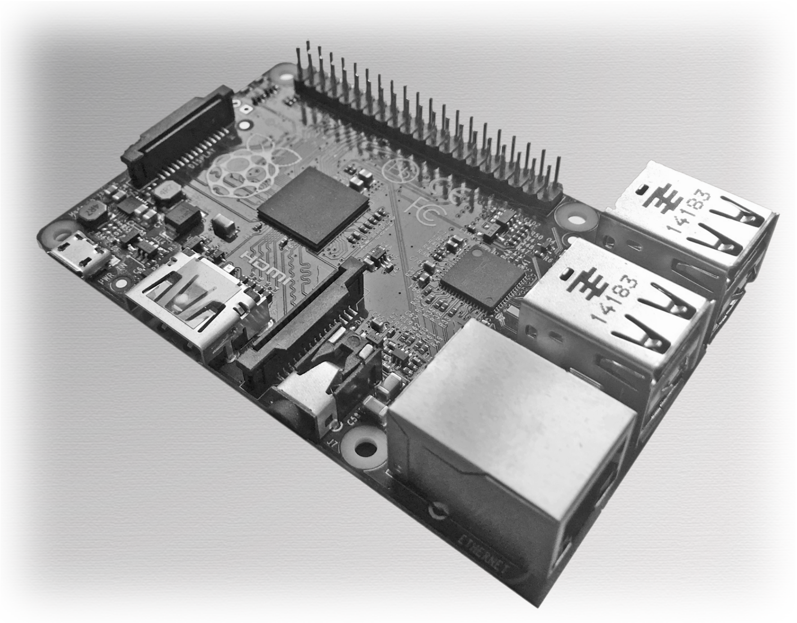
Computers Everywhere 4



A Raspberry Pi Single Board computer, approximately 3 x 2 inches in size. Photo: Christopher Crispin-Bailey 2019.

4.1 A machine with many facets

When we talk about computer systems in general terms, we overlook an important aspect of their design. There is no one single kind of computer system that perfectly suits all possible uses. Indeed, we may not even be aware that computers systems are being used in some situations at all. For instance, if you were asked where to find a computer system in your kitchen, would you immediately think of a microwave oven? The answer is probably no, and yet many everyday items contain computers of some kind within them. But, clearly, these are not the same computers as our laptops or our company data servers.

This underlines an important point: there are many computer systems, with many applications, and using the right kind of system in the right kind of scenario is eminently sensible. Not only that, but it most likely represents the most cost-effective way to solve that particular problem, and cost is always an issue. In this chapter we will try to define some terms of reference and attempt to understand some of the rich variety of computers, and indeed processors, that make up this spectrum of possibilities. To help distinguish between these differences, we often group computers into **general purpose** and **application-specific** domains.

4.2 The general purpose computer

When the first computers were built, back in the days of radio valves, and racks full of spaghetti wiring, computers were built with quite particular purposes in mind. Some of the first digital electronic computers were developed to speed up the calculation of physics calculations, relating to such things as nuclear bomb design, aircraft aerodynamics, cryptography, and only later, as systems became more commercially practical, they were used for processing stock data, financial accounts, and the beginnings of the wider applications we might be familiar with today.

As computers became more sophisticated and powerful, their uses expanded to many more tasks. At this point, many computers began to be designed with the general market in mind: one machine that can do many things for many customers was more cost effective to mass-produce and sell.

But, at the same time, these general-purpose computer systems became less efficient at very specific tasks. There is a saying 'Jack of all trades, and master of none'. This is exactly what general purpose computers are, and therefore another class of specialist computing systems also emerged at this point.

Let us not be too demanding however. Modern computer systems are now so powerful that many specialised tasks can be performed quite adequately on a general purpose computing system, or even a desktop computer. To some degree this is dependant upon the expectations of the user.

It may, for instance, be possible to run a nuclear reaction simulation in a few minutes, and the programmer or program user may not care that it could be done in 10 seconds on a more specialised machine. Nonetheless, there are times when pure computational athleticism in one discipline is the key requirement, and if the choice is between waiting ten months or ten days for a program to complete, then there is an obvious issue with system capability.

At the simplest level of abstraction, computer systems can therefore be divided loosely into two groups. The **general-purpose** machines, as we have just briefly described, and those **application-specific** systems that are engineered to the last detail to be highly efficient at the one, or few, tasks they are designed for.

We shall examine some of the parameters and attributes of these systems in a little more detail in the next sections. There are further divisions of course, and we will highlight these along the way.

4.3 General purpose computers

It is safe to say that almost everyone has encountered a general purpose computer at some point. The classical examples of these systems include the desktop computer and the laptop. These machines typically have the characteristics shown in the panel following.

We may wonder what 'a large amount of memory' actually is. Well, this is not so easy to answer, as memory capacities have been steadily increasing for many decades, so this is not a 'stand-still' number. Currently (2021) a computer system with 8 Gigabytes of memory is considered fairly ordinary. Some laptop and desktop users may have as much as 32 Gigabytes of main memory, where they have particular needs for certain kinds of software (in fact they are customising their general purpose computers to be more application specific).

Features of General Purpose Computers

- A large amount of memory (large in terms of the everyday user), that allows a wide range of tasks to be performed,
- A large amount of local disk storage (again, large in everyday terms),
- Built from readily available, off-the-shelf, mass-market components,
- Can run a variety of operating systems,
- Can run a wide variety of software,
- Has lots of different kinds of connections for peripheral devices.
- Is relatively inexpensive.

As far as disk storage is concerned, there is a similar story. Laptops and desktop computer systems with 1 Terabyte of disk storage are becoming increasingly the norm, and certainly 500 Gigabytes would be considered very standard at this moment in time.

A key point here is 'relatively inexpensive'. These computer systems could of course have much bigger memories and much larger disk capacities, but at a significant cost, which is not justified for general purpose computing.

4.3.1 The computer in your hand.

These days, it is not just desktops and laptops that are general purpose computers. In recent years, **smartphones** and **tablet** computers have become widely popular and powerful enough to be used in similar ways: these are also general purpose computers. The interface is somewhat specialised (a **touchscreen** rather than keypad and mouse), and the size, weight and power source place some constraints on the design, but in overall terms these are very much general purpose devices, if perhaps perhaps optimised toward media-rich applications.

Ask yourself how often you make an actual voice-call on your smartphone? Does this amount to as much as 5% of the time you use it? One might suspect possibly even less. Meanwhile, you may well use the same device for games, diaries, home banking, news, social media, web-surfing, navigation, shopping, and much more. These are certainly general purpose devices in the modern sense.

The only difference here is that the kind of circuitry required for these devices must be very compact and very power efficient. As a result of this, the emergence of mobile processors has been observed in the chip industry in recent years. These processors are typified by integrating many functions onto a single chip[37] alongside the usual CPU hardware (functions that in a desktop may be provided by physically separate chips or circuit boards). Meanwhile, the power consumption of these chips is optimised for the kind of uses that a hand-held device most needs.[38]

4.3.2 The hidden computer

Another type of computer you may well encounter frequently in everyday life is a computer which we refer to as an **embedded system**. These may be the most numerous computer systems in the world, possibly even exceeding mobile processors in smartphones and devices, though often using similar processor families.

Embedded systems are effectively computer systems embedded in devices in such a way that they perform a set of functions, without a person having full access to the computer system as a computer user. The example given earlier was the microwave oven. This common kitchen utility frequently has a digital display, a keypad, and a relatively simple processor chip inside the control panel, to operate all of the internal parts of the system that make the oven work. It follows algorithms in order to perform cooking tasks according to the settings provided via the keypad, and it displays information on the display. The processor will also have some memory (generally a small memory built into the same chip). So, referring back to Chapter 2, your microwave oven may well include a von Neumann architecture (CPU, Memory, Input, and Output).

However, it is not just microwaves. Consider the following list, which is certainly only the tip of the iceberg:

Typical places to find embedded systems:

In a microwave oven, washing machine, dishwasher,
On the digital lock on your office door,
Inside an office printer,
At the self-service supermarket checkout,
Inside your car engine control unit, and key-fob
In a central heating control dial,
In an electric toothbrush,
Inside your flat-panel tv.

Figure 4.1: A few of the many examples of devices typically containing some form of CPU, some less obvious than others.

[37] Often referred to as a system on a chip or SOC.

[38] Some examples of these include **A11 Bionic** (APPLE), **Snapdragon 845** (QUALCOMM), **Exynos 9810** (SAMSUNG) and **Hisilicon kirin 970** (HUAWEI).

That is quite a list! Yet our list barely scratches the surface. It was once claimed that the average home would probably never require a computer, and indeed it would be hard to envisage any family having more than one. Yet we each seem to have tens of computers in our homes, and we don't even know it. Naturally there is a downside to all of this: all of these devices consume power, create radio emissions, and many are potentially hack-able under the right conditions.

Apart from these 'every-day' cases, there are also embedded systems that are highly engineered for very demanding applications, some extremely small in size as in Figure 4.2. These systems have to be designed and programmed in particular ways to ensure that they operate reliably under all relevant conditions. This includes being tolerant of fault conditions (**fault tolerance**), having the ability to revert quickly to a correct state after detecting a fault, being able to respond within very tight timescales to specific events (known as **real-time computing**), and being resilient to security threats such as hacking and viruses.

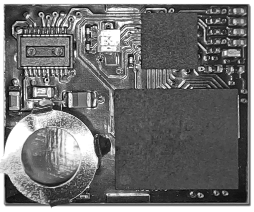


Figure 4.2: The NAT (Neural Activity Tracker). This is a miniature EEG sensor measuring only 22x18x8mm. (Cybula Ltd. and University of York).

Examples of this include aircraft control systems, missile guidance, self-driving cars and driver assist capabilities, medical devices such as insulin pumps, medical measurement devices, Dental X-ray machine, and many other 'safety-critical' cases (where any incorrect behaviour simply cannot be tolerated due to the consequences).

4.4 Being specific

When a computer system has enough specialisation to cease to be considered a general purpose system it becomes application specific. As one might expect, there are many examples of this. Many embedded systems are to some extent application specific. They may be designed to fulfil a particular task well at low cost (microwave oven controller), they may be designed to fulfil some very demanding requirements: self-driving cars and safety-critical scenarios for example. Other systems may simply be designed to perform a particular kind of computational task very efficiently. Examples of computationally specialised systems include systems designed to solve particular scientific problems.

Let us consider some examples:

- **Weather forecasting:** requires a huge number of complex mathematical computations to be performed in a very tight timescale. There is no point in forecasting tomorrow's weather if it takes three days to get an answer. This kind of system would likely have a large number of processors to handle the volume of computations, and a large memory distributed amongst those processors.

- **Web-search Engines:** Designed to handle high volumes of small simple requests to deliver blocks of data (web page content), these systems have lots of hard disk space, high performance connectivity to networks, and so on.
- **Artificial Intelligence (AI):** A relatively new computing demand, artificial neural networks (ANNs) have been around for a long time as a computing concept. However, processors designed specifically to process ANN tasks are a recently maturing field. A few companies have developed dedicated Neural Network processor architectures, or 'AI engines', which require very specialised systems to be built around them. An example is the GOOGLE TPU designed for large-scale artificial neural network systems[39].
- **Digital Signal Processors:** This is a class of processor that is very efficient at performing specific mathematical operations on signals. They can be used in sensor systems, radar, biomedical imaging, audio processing (noise cancelling headphones for instance), and other applications. Often these processors have specialised instruction sets[40]. DSP's are also optimised to permit pipelining of complex instructions to maximise algorithmic throughput.
- **Space Flight:** Computer systems designed for space hardware such as satellites, space probes, and manned spacecraft are a very specialised form of computer system. They have to be very robust, in particular having resilience to radiation and cosmic rays that are encountered in space[41], and can easily crash a computer system or corrupt its data memory. They also often need to adhere to very precise safety-critical design requirements There is no room for error in space.
- **Nuclear Applications:** Many issues experienced in space-flight, satellites, etc, are also relevant to the nuclear industry. Building robots that can operate in the harshest nuclear environments, such as the Chernobyl nuclear site, is extremely demanding on processor and electronics reliability.
- **Graphics, media, and gaming:** Many computer systems can be adapted to be high performance gaming machines. However, in order to really perform at the top of their game (no pun intended), they need some rather specialised system components. They often use special kinds of hard disks (solid state drives[42]) that are currently quite expensive, but very fast. They also use specialised graphics modules, known as graphics cards, which incorporate specialised processors known as GPUs (Graphical Processing Units). In

[39] GOOGLE TPU: Tensor Processing Unit, a specialised AI processing engine.

[40] An example is the Multiply Accumulate operation, which performs the equivalent of A = A + (B x C). This is a frequent operation in signal processing. A well known example is (n x k)+(m x j), which is used to mix two signals together. This can be achieved by two MAC operations.

[41] These are sometimes referred to as 'rad-hard' systems.

[42] Solid state drives, or SSD's are becoming more common as a general system component. However, typical capacity versus cost is still nowhere near that of traditional disk units. This will change in time.

effect, the graphics task is handled by a very specialised computer system all of its own (the graphics card).

4.5 The impact on processor technology.

What we have observed, in the preceding sections, is that there are many different kinds of computer system. And they have quite varied requirements. As a result we find that processors, being the heart of the machine, are also quite varied in their capabilities, attempting to match processor architectures to computer system requirements. There are far too many processors in the marketplace at present to list them all, but we can make some attempt to list a few examples of each kind, examples of which are given in Table 4.1.

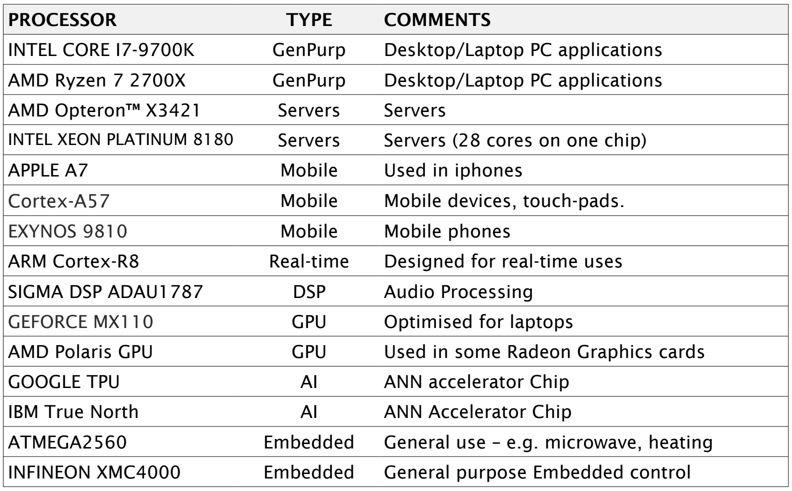
4.6 Mainframes and supercomputers

There was a period, probably between the late 1960s and throughout the 1970's, when any computer system was still a considerable investment for a large company. Consider an insurance company, where there might be 100 employees, all wishing to use computer resources from time to time on a daily basis. The cost of putting a computer on every desk would have been prohibitive in those days, and the desks would need to be very large! Instead, the concept of a mainframe, and mini-mainframe computer had already become established to solve this requirement economically.

The concept was straightforward: multiple users could access the same computer (probably set up in a convenient basement in the company tower block), as if at the same time. Typically access would be via a terminal (a screen and keyboard with very minimal hardware[43]). These were often known as **dumb terminals** since they could do nothing on their own. Instead, they were simply remote interfaces to the 'real' computer system. This was also an early form of networked computing infrastructure.

Consequently, any user could run their application and perform a task right at their desk. In order to give the impression that everyone had equal access, the concept of time-slicing was utilised. Effectively, if every user had one millisecond of CPU time every second, and the terminals handled the local keyboard and display tasks, the impression was given that all 100 users were using the computer's resources simultaneously.

[43] The VT100 terminal is considered a classic example, and is still emulated in some software terminals.



PROCESSOR	TYPE	COMMENTS
INTEL CORE I7-9700K	GenPurp	Desktop/Laptop PC applications
AMD Ryzen 7 2700X	GenPurp	Desktop/Laptop PC applications
AMD Opteron™ X3421	Servers	Servers
INTEL XEON PLATINUM 8180	Servers	Servers (28 cores on one chip)
APPLE A7	Mobile	Used in iphones
Cortex-A57	Mobile	Mobile devices, touch-pads.
EXYNOS 9810	Mobile	Mobile phones
ARM Cortex-R8	Real-time	Designed for real-time uses
SIGMA DSP ADAU1787	DSP	Audio Processing
GEFORCE MX110	GPU	Optimised for laptops
AMD Polaris GPU	GPU	Used in some Radeon Graphics cards
GOOGLE TPU	AI	ANN accelerator Chip
IBM True North	AI	ANN Accelerator Chip
ATMEGA2560	Embedded	General use – e.g. microwave, heating
INFINEON XMC4000	Embedded	General purpose Embedded control

Table 4.1: A variety of processors with specific/-generic application categorisations.

Actually, a similar principle applies today, although the terminology and technology have changed. We talk about cloud-computing, thin-clients (the modern equivalent of a dumb terminal), and servers (akin to a mainframe). We will find out more about these later.

Going in the opposite direction, there were some computing domains where sharing out small slices of compute time to lots of general purpose users was not the primary goal. Instead, the demand was to run hugely complex computational tasks, with massive amounts of data, and to do so as fast as possible. A system designed to fulfil that kind of function is known as a supercomputer. These are the giants of computer systems, with disk storage measured in Petabytes [44] , hundreds if not thousands of processors, and terabytes of memory. Of course, such a precious resource can be in demand, and there may still be a small group of users competing for access. This requires a job-scheduling strategy.

It may be worthwhile defining what exactly we mean by fast, in the context of supercomputers. From the user perspective, the problem is to complete a hugely complex simulation or calculation within a reasonable time frame. For a PhD researcher, who has only three years to complete a PhD, a month may be a long time for one experiment. However, for an astrophysicist working on mathematical problems related to the physics of black holes, three or 6 months might be acceptable. For a weather centre, running a simulation in a few hours is probably the maximum that can be tolerated, otherwise the weather forecast would be issued the day after it has happened!. Some problems are so demanding that they can take several years to run a complete processing task, even on

[44] 1000s of Terabytes, or more accurately 1 Binary Petabyte equates to 1024 Terabytes or 2^{50} bytes.

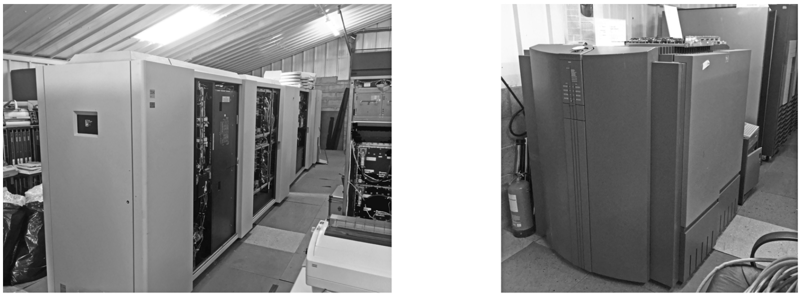


Figure 4.3: Supercomputers. Photos showing (left) an older IBM mainframe and (right) late 1980s/90s CRAY EL98. With thanks to The Jim Austin Computer Museum (2019). The Cray EL98 had an entry level price of around $340,000, and achieved up to 1 GigaFlops.

some of the world's most powerful supercomputers. And of course this consumes huge amounts of power, and requires huge amounts of cooling, deep pockets, and patience.

Because supercomputers are so expensive, some groups, such as the SETI Institute[45] have started initiatives where compute resources are crowd-sourced: By taking a little bit of spare CPU capacity from every participant's desktop, often when those computers are sitting idle, they can aggregate huge amounts of computing power and achieve computing tasks that would otherwise require a dedicated supercomputer. Of course this relies upon goodwill and participant enthusiasm. For interest, Figure 4.3 shows a 1970's mainframe system, and a 1990's era supercomputer.

In some ways, supercomputers are seen as the pinnacle of current computational achievement, certainly for scientific computing tasks, and there is strong commercial competition to be top. In order to keep track of this progress, the Top-500 supercomputer index is updated regularly and lists all of the world's fastest supercomputers[46].

Not surprisingly, the kind of processors used in these systems has also become very specialised. Specialised data processors known as vector processors, processors that contain thousands of simple number crunching elements (cellular arrays), and the adoption of graphic processing units[47] for scientific computation are all examples. Often these chips can be very expensive.

4.7 The internet of things

A recurring theme over the past 30 years has been the idea of a world in which computers are everywhere, performing all kinds of functions, and

[45] SETI: Search for Extra Terrestrial Intelligence, www.seti.org

[46] web URL: https://www.top500.org This site also has regular news and technical features on the latest developments.

[47] GPU's have a class of computational capabilities that are highly mathematical, and often have high degrees of parallelism in terms of their computational units. This makes it possible to map complex scientific problems onto their hardware and achieve high throughput.