# 9 Digital Technologies and Data Systems

While computing technologies are young by comparison with other efforts of human ingenuity, their origins go back to many early chapters in the human quest for the understanding and realization of mechanical aids to knowledge.

Giuseppi Primiero (2020, 7)

Computing is symbol processing. Any automaton capable of processing symbol structures is a computer . . . We may choose to call such symbol structures information, data, or knowledge, depending in the particular 'culture' within computer science to which we belong.

Subrata Dasgupta (2016, 121)

Data science is the process by which the power of data is realised, it is how we find actionable insights from among the swathes of data that are available.

David Stuart (2020, xvi)

## Introduction

Technology, from the Greek *techné*, meaning art, skill or craft, is usually taken to mean the understanding of how to use tools, in the broadest sense of the word. The term *information technology* was first used in the 1950s to describe the application of mechanised documentation and then-new digital computers, and became widely used in the 1980s to describe the wider-spread use of digital technology (Zorkoczy, 1982).

Information technology is usually associated with computers and networks. But, in a wider sense stemming from the original meaning of the word, the technologies of information include all the tools and machines which have been used to assist the creation and dissemination of information throughout history, as discussed in Chapter 2; from ink and paper, through printing to microforms, document reproduction and photocopying, and mechanised documentation technologies such as card indexes, punched cards, edge-notched cards, and optical coincidence cards. Krajewski (2011) examines the idea of card index files as a 'universal paper machine', in a sense the forerunner of the computer, between the 16th and 20th centuries.

Our focus here is on digital technologies, echoing the view expressed by Gilster (1997) that all information today is digital, has been digital or may be digital. This chapter covers the basics of digital technologies and the handling of data; the following chapter deals with information systems.

## Digital technologies

We will describe these aspects only in outline; see Ince (2011) and Dasgupta (2016) for more detailed but accessible accounts, and Primiero (2020) for a more advanced summary; for accounts of the historical development of the computer, see Ceruzzi (2012) and Haigh and Ceruzzi (2021).

Any digital device represents data in the form of *binary digits* or *bits*. Patterns of bits, *bit strings* with bits conventionally represented by 0 and 1, may represent data or instructions. A collection of eight bits is known as a *byte*. Quantities of data are represented as multiples of bytes, for example: a kilobyte, defined as $2^{10}$ bytes, i.e. 1024 bytes.

Any character or symbol may be represented by a bit string, but this requires an agreed coding. The most widely used code since the beginning of the computer age was the American National Standards Institute's ASCII (American Standard Code for Information Interchange) 7-bit code, but it is limited to the Latin alphabet, Arabic numerals and a few other symbols. It has been supplanted by Unicode, which can handle a wider variety of symbols and scripts. It can do this by using a 32-bit code; the more bits in the code, the more different symbols can be coded. Codes provide arbitrary representations of characters; for example, the ASCII code for the letter L is the bit string 1001100.

The idea that a general-purpose computing machine might be possible stems from the theoretical studies of the British mathematician Alan Turing (1912–54), whose concept of the *Turing machine*, developed in the 1930s, showed that a single machine limited to a few simple operations (read, write, delete, compare) could undertake any task if appropriately programmed.

The basic architecture for a practical digital computer was set out by John von Neumann (1903–57), the Hungarian-American mathematician, in the 1940s (Figure 9.1). His design, which gave the first formal description of a single-memory stored-program computer, is shown in Figure 9.2.

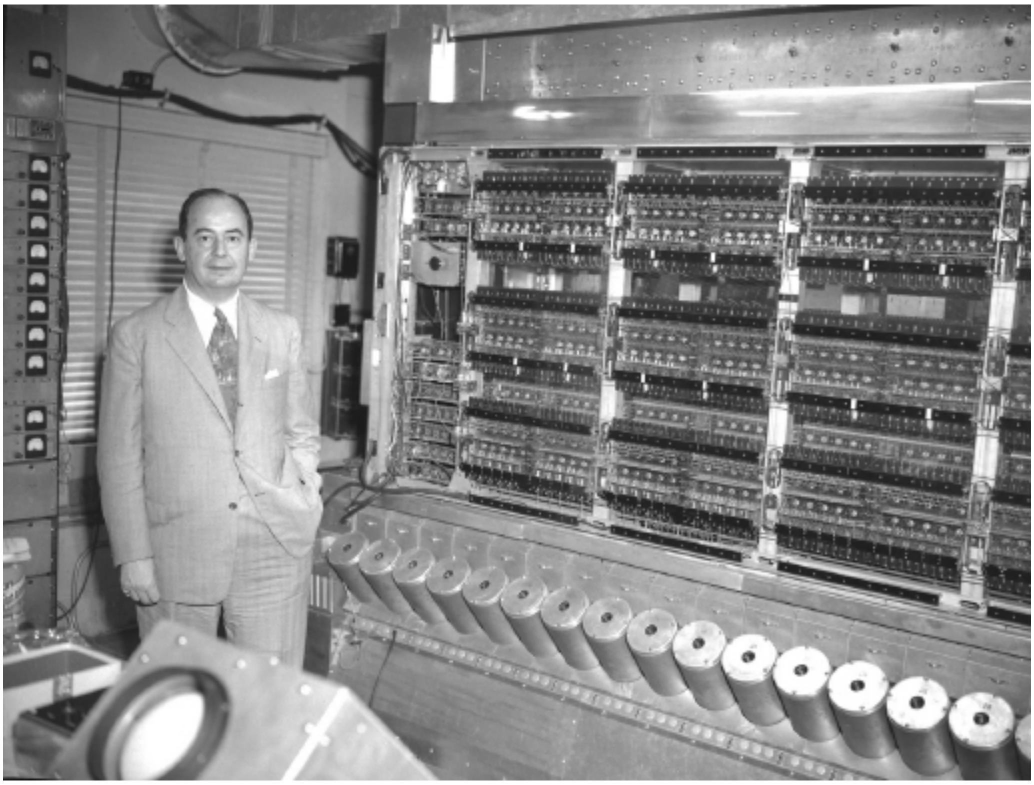

**Figure 9.1** *John von Neumann* (Alan Richards photographer. From the Shelby White and Leon Levy Archives Center, Institute for Advanced Study, Princeton, NJ, USA)

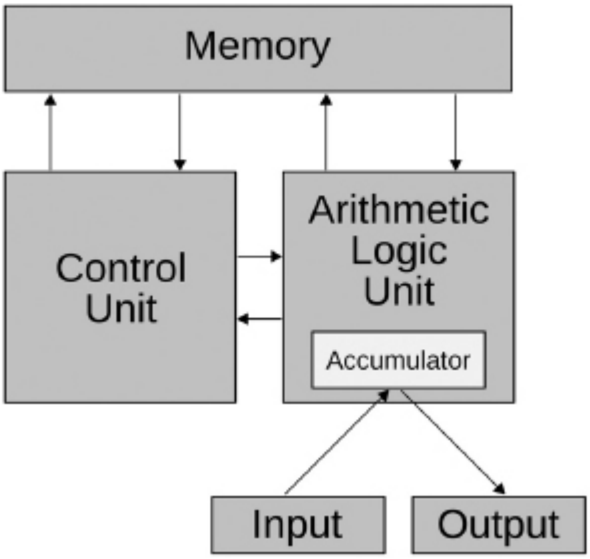


**Figure 9.2** *Von Neumann computer architecture* (Wikimedia Commons, CC BY-SA)

This architecture is general purpose, in the sense that it can run a variety of programs. This distinguishes it from special-purpose digital computers which carry out only one task, as in digital cameras, kitchen appliances, cars, etc. A von Neumann machine loads and runs programs as necessary, to accomplish very different tasks.

The heart of the computer, the *processor*, often referred to as the *central processing unit* (CPU), carries out a set of very basic arithmetical and logical operations with instructions and data pulled in from the *memory*, also referred to as *main* or *working* memory. This sequence is referred to as the *fetch–execute cycle*. Two components of the processor are sometimes distinguished: an *arithmetic and logic unit*, which carries out the operations, and a *control unit*, which governs the operations of the cycle.

While programs and data are being used they are kept in the memory. While not being used they are stored long-term in *file storage*. Items in memory are accessible much more rapidly than items in file storage, but memory is more expensive, so computers have much more file storage than memory.

Data comes into the computer through its *input devices* and is sent into the outside world through the *output devices*. The components are linked together through circuits usually denoted as a *bus*, sometimes referred to more specifically as a *data bus* or *address bus*.

All of these components have undergone considerable change since the first computers were designed. Processor design has gone through three main technological stages. The *first generation* of computers used valves as processor components, and the *second generation* used transistors. Computers of the *third generation*, including all present-day computers, use circuits on 'silicon chips', by very-large-scale integration (VLSI), which allows millions of components to be placed on a single computer chip, thus increasing processing speed.

The storage elements of the computer's memory comprise regular arrays of silicon-based units, each holding a bit of data, and created using the same VLSI methods as processors. Earlier generations of computers used *core storage*, with data held on tiny, magnetised elements arranged in a three-dimensional lattice.

File storage, holding data and programs that are not needed immediately, has always used magnetic media, which can hold large volumes of data cheaply, provided that quick access is not required. Earlier generations of computers used a variety of tapes, drums and disks; current computers use *hard disks*, rapidly spinning magnetisable disks with moveable arms with *read /write heads*, to read data from, or write data to, any area of the disk. Regardless of the technology, each area holds one bit of information, according to its magnetic state.

Input devices fall into three categories: those which take input interactively from the user; those which accept data from other digital sources; and those which convert paper data into digital form. The first category comprises the venerable 'QWERTY' keyboard, originally developed for typewriters in the 1870s, together with more recently developed devices: the mouse and other pointing devices and the touch-screen. The second category comprises the silicon-memory data stick, replacing the various forms of portable magnetic *floppy disks* used previously, and the ports and circuits by which the computer communicates with networked resources. The third category is that of the scanner, digitising print and images from paper sources.

Output devices are similarly categorised; display screens, which allow user interaction through visual and sound output; devices which output digital data, such as data sticks, network circuits and ports; and devices which print paper output, typically laser or inkjet printers.

The most recent development is the trend towards mixed reality and immersive environments, which may require specific input/output devices (Greengard, 2019; Pangilinan, Lukos and Mohan, 2019). *Virtual reality* (VR) provides immersive experiences in a computer-generated world, often using head-mounted display devices, such as Oculus Rift. *Augmented reality* (AR) overlays computer-generated information, often visual but sometimes multisensory, on the real world. *Mixed reality* is the combination of real and virtual environments, a spectrum with AR at one end and VR at the other. These are of increasing relevance to the provision of information services (Robinson, 2015; Varnum, 2019; Dahya et al., 2021; Kroski, 2021).

## Networks

Virtually all computers are connected via some form of network to others, to enable communication and information access and sharing. Since the 1990s the internet and the World Wide Web have become ubiquitous, such that it has become difficult to think about digital technologies without these at centre stage.

The growth of networked computing has been driven by three factors: communications technology, software and standards. In terms of technology, older forms of wired network, originating in the twisted copper-pair cables used for telegraph systems from the mid-19th century, have been succeeded by fibre-optic cables and various forms of wireless transmission. These allow much faster transmission speeds and greater information-carrying capacity; such systems are described loosely as *broadband*. Software systems have improved the efficiency and reliability of transmission greatly: an important example is *packet switching*, by which messages may be split up and their constituents sent by the fastest available route, being recombined before delivery.

Standards are fundamentally important, so that different forms of network, and different kinds of computers connected to them, can communicate effectively. The internet,