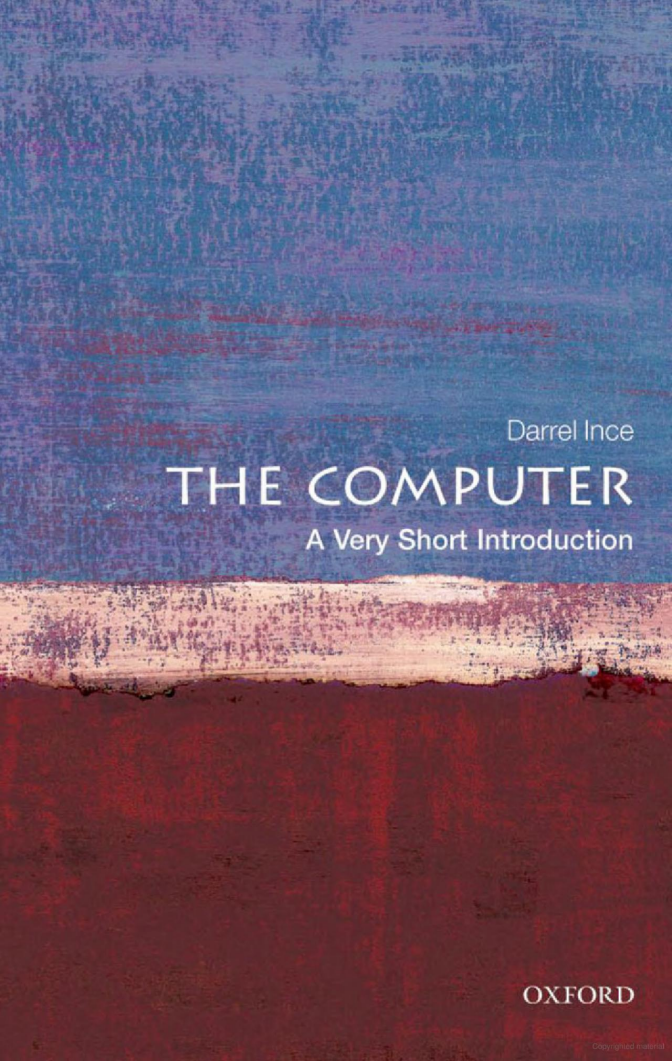

Darrel Ince
THE COMPUTER
A Very Short Introduction
OXFORD

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 2011

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain on acid-free paper by
Ashford Colour Press Ltd, Gosport, Hampshire

ISBN: 978–0–19–958659–2

1 3 5 7 9 10 8 6 4 2

Contents

This page intentionally left blank

List of illustrations

This page intentionally left blank

Chapter 1
The naked computer

Introduction

One of the major characteristics of the computer is its ability to store data. It does this by representing a character or a number by a pattern of zeros and ones. Each collection of eight zeros and ones is known as a 'byte'; each individual one or zero is known as a 'bit' (binary digit). Computer scientists use various terms to describe the memory in a computer. The most common are the kilobyte, the megabyte, and the gigabyte. A kilobyte is 10^3 bytes, a megabyte is 10^6 bytes, and a gigabyte is 10^9 bytes.

The first computer I used was an Elliot 803. In 1969, I took a computer-programming course at my university which used this computer. It was situated in a room which was about 40 foot by 40 foot, with the hardware of the computer contained in a number of metal cabinets, each of which would fill almost all of the en-suite bathroom I have at home. You submitted your programs written neatly on special paper to two punch-tape operators, who then prepared a paper-tape version of the program. Each row of the paper tape contained a series of punched dots that represented the individual characters of the program.

The program was then taken to the computer room, the tape read by a special-purpose piece of hardware, and the results displayed on a device known as a Post Office Teletype; this was effectively a typewriter that could be controlled by the computer, and it produced results on paper that were barely of better quality than toilet paper.

The storage capacity of computers is measured in bytes; the Elliot computer had 128 thousand bytes of storage. It used two cabinets for its memory, with data being held on small metallic rings. Data were fed to the computer using paper tape, and results were obtained either via paper or via a punch which produced paper tape. It required an operator to look after it, and featured a loudspeaker which the operator could adjust in volume to check whether the computer was working properly. It had no connection to the outside world (the Internet had yet to be invented), and there was no hard disk for large-scale storage. The original price of the first wave of Elliot 803s was £29,000, equivalent to over a hundred thousand pounds today.

While I am writing this chapter, I am listening to some Mozart on a portable music device known as an MP3 player. It cost me around £180. It comfortably fits in my shirt pocket and has 16 gigabytes of memory – a huge increase over the capacity of the only computer at my old university.

I am typing the book on a computer known as a netbook. This is a cut-down version of a laptop computer that is configured for word processing, spreadsheet work, developing slide-based presentations, and surfing the Internet. It is about 10 inches by 6 inches. It also has 16 gigabytes of file-based memory used for storing items such as word-processed documents, a connection to the Internet which downloads web pages almost instantaneously, and a gigabyte of memory that is used to store temporary data.

There is clearly a massive difference between the Elliot 803 and the computers I use today: the amount of temporary memory, the

amount of file-based memory, the processing speed, the physical size, the communications facilities, and the price. This increase is a testament to the skills and ingenuity of the hardware engineers who have developed silicon-based circuits that have become smaller and more powerful each year.

This growth in power of modern computers is embodied in a law known as 'Moore's law'. This was expounded by Gordon Moore, the founder of the hardware company Intel, in 1965. It states that the density of silicon circuits used to implement a computer's hardware (and hence the power of a computer) will double every two years. Up until the time of writing, this 'law' has held.

The computer has evolved from the physical behemoths of the 1950s and 1960s to a technological entity that can be stored in your jacket pocket; it has evolved from an electronic device that was originally envisaged as something only large companies would use in order to help them with their payroll and stock control, to the point where it has become an item of consumer electronics as well as a vital technological tool in commercial and industrial computing. The average house will contain as many as 30 computers, not only carrying out activities such as constructing word-processed documents and spreadsheet tables, but also operating ovens, controlling media devices such as televisions, and regulating the temperature of the rooms.

Even after 70 years, the computer still surprises us. It surprised Thomas Watson, the founder of IBM, who predicted that the world only needed about five computers. It has surprised me: about 20 years ago, I saw the computer as a convenient way of reading research documents and sending email, not as something that, combined with the Internet, has created a global community that communicates using video technology, shares photographs, shares video clips, comments on news, and reviews books and films.

Computer hardware

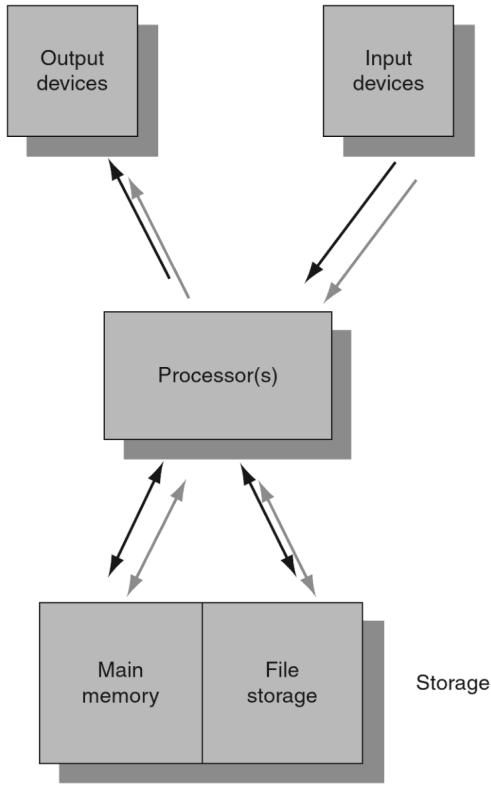
One aim of this book is to describe how the computer has affected the world we live in. In order to do this, I will describe the technologies involved and the applications that have emerged over the last ten years – concentrating on the applications.

First, the basic architecture of the computer; I will describe this architecture in a little more detail in Chapter 2. This is shown in Figure 1. The schematic shown in this figure describes both the earliest computers and the newest: the basic architecture of the computer has not changed at all over 60 years.

At the heart of every computer is one or more hardware units known as processors. A processor controls what the computer does. For example, it will process what you type in on your computer's keyboard, display results on its screen, fetch web pages from the Internet, and carry out calculations such as adding two numbers together. It does this by 'executing' a computer program that details what the computer should do, for example reading a word-processed document, changing some text, and storing it into a file.

Also shown in Figure 1 is storage. Data and programs are stored in two storage areas. The first is known as main memory and has the property that whatever is stored there can be retrieved very quickly. Main memory is used for transient data – for example, the result of a calculation which is an intermediate result in a much bigger calculation – and is also used to store computer programs while they are being executed. Data in main memory is transient – it will disappear when the computer is switched off.

Hard disk memory, also known as file storage or backing storage, contains data that are required over a period of time. Typical entities that are stored in this memory include files of numerical data, word-processed documents, and spreadsheet tables. Computer programs are also stored here while they are not being executed.



1. The architecture of a computer

There are a number of differences between main memory and hard disk memory. The first is the retrieval time. With main memory, an item of data can be retrieved by the processor in fractions of microseconds. With file-based memory, the retrieval time is much greater: of the order of milliseconds. The reason for this is that main memory is silicon-based and all that is required to read data there is for it to be sent along an electronic circuit. As you will see later, hard disk memory is usually mechanical and is

stored on the metallic surface of a disk, with a mechanical arm retrieving the data.

Another difference between the two types of memory is that main memory is more expensive than file-based memory; consequently, there is usually far less main memory in a computer than file-based memory (I have a laptop that has 3 gigabytes of main memory and the file-based memory contains 500 gigabytes of storage).

Another set of components of a computer are input devices. These convey to the computer what the user requires of the programs executed by the computer. The two devices that you will have met most frequently are the keyboard and the mouse. There are, however, a number of other devices: touch screens that you find on iPods and satellite navigation systems and pressure monitors found as part of the instrumentation of a nuclear power station are two further examples.

The final component of a computer is one or more hardware devices that are used to display results. There are a variety of such devices. The most familiar to you will be the computer monitor and the laser printer; however, it can also include advertising displays found at events such as football matches, the console that displays flight data on the instrumentation found in the cockpit of a plane, the mini-printer that is used to produce a supermarket receipt, and the screen of a satellite navigation device.

The working definition of a computer that I shall use within this book is:

> A computer contains one or more processors which operate on data. The processor(s) are connected to data storage. The intentions of a human operator are conveyed to the computer via a number of

input devices. The result of any computation carried out by the processor(s) will be shown on a number of display devices.

You may think this statement is both pedantic and self-evident; however, I hope that you may see as this book unfolds that it has a number of radical interpretations.

Before leaving this section, it is worth looking at another indicator of the growth in power of computers. In their excellent book *The Spy in the Coffee Machine*, O'Hara and Shadbolt describe the progress made in computer-based chess. To be good at chess requires you to look ahead a number of moves and evaluate what your opponent would do for each of these moves, and then determine what move you would make for each of these moves, and so on. Good chess players hold lots of data in their heads and are able to carry out fast evaluations. Because of this, the computer has always been seen as potentially a good chess player.

The chess programs that have been written effectively store lots of moves and countermoves and evaluate them very quickly. O'Hara and Shadbolt describe how in 1951 a computer could only think ahead two moves, in 1956 a computer could play a very restricted game of chess on a smaller board but would take upward of 12 minutes to make a move. However, in 1997 a computer beat the world champion Gary Kasparov. This progress is partly due to improvements in software techniques for game playing; the major reason though is that computers have become faster and faster.

The Internet

Computers do not operate in isolation: most are connected to a computer network. For most computers, this will be the huge collection of computers and communication facilities known as the Internet; however, it could be a network that controls or monitors some process, for example a network of computers that

keep a plane flying, or a network of computers used to monitor the traffic flow into and out of a city.

The Internet has had a major effect on the way computers are currently being used; so it will be worthwhile looking briefly at how it interacts with a typical computer – say the PC that you use at home.

The Internet is a network of computers – strictly, it is a network that joins up a number of networks. It carries out a number of functions. First, it transfers data from one computer to another computer; to do this, it decides on the route that the data takes: there is a myth that when you carry out some activity using the Internet, for example downloading a web page, the connection between the computer holding the page and your computer is direct. What actually happens is that the Internet figures out a route that the data takes via a number of intermediate computers and then routes it through them. So when you see a web page displayed on your computer, that page may have been split into blocks of data, with each block having travelled through a number of continents and traversed a number of intermediate computers belonging to companies, universities, charitable organizations, and government organizations.

The second function of the Internet is to enforce reliability. That is, to ensure that when errors occur then some form of recovery process happens; for example, if an intermediate computer fails then the software of the Internet will discover this and resend any malfunctioning data via other computers.

A major component of the Internet is the World Wide Web; indeed, the term 'Internet' is often used synonymously with the term 'World Wide Web'. The web – as I shall refer to it from now on – uses the data-transmission facilities of the Internet in a specific way: to store and distribute web pages. The web consists of a number of computers known as *web servers* and a very large

number of computers known as *clients* (your home PC is a client). Web servers are usually computers that are more powerful than the PCs that are normally found in homes or those used as office computers. They will be maintained by some enterprise and will contain individual web pages relevant to that enterprise; for example, an online book store such as Amazon will maintain web pages for each item it sells.

The program that allows users to access the web is known as a *browser*. When you double-click the browser icon on your desktop, it will send a message to the web asking for your home page: this is the first page that you will see. A part of the Internet known as the Domain Name System (usually referred to as DNS) will figure out where the page is held and route the request to the web server holding the page. The web server will then send the page back to your browser which will then display it on your computer.

Whenever you want another page you would normally click on a link displayed on that page and the process is repeated. Conceptually, what happens is simple. However, it hides a huge amount of detail involving the web discovering where pages are stored, the pages being located, their being sent, the browser reading the pages and interpreting how they should be displayed, and eventually the browser displaying the pages.

I have hidden some detail in my description. For example, I have not described how other web resources such as video clips and sound files are processed. In a later chapter, I will provide a little more detail. At this point, it is just worth saying that the way that these resources are transferred over the web is not that different to the way that web pages are transferred.

The Internet is one of the major reasons why computers have been transformed from data-processing machines to a universal machine that can, for example, edit music files, predict the

weather, monitor the vital signs of a patient, and display stunning works of art. However, without one particular hardware advance the Internet would be a shadow of itself: this is broadband. This technology has provided communication speeds that we could not have dreamed of 15 years ago. Most users of the Internet had to rely on what was known as a dial-up facility which transferred data at around 56 kilobits of data a second. When you consider that the average web page size is around 400 kilobits, this means it would take around 7 seconds for a web page to be displayed in your browser. In the 1990s, companies used dedicated communications hardware to overcome this lack of speed.

Unfortunately, the average user was unable to do this until broadband became generally available.

Typical broadband speeds range from one megabit per second to 24 megabits per second, the lower rate being about 20 times faster than dial-up rates. As you will see later in the book, this has transformed the role of the home-based computer.

Software and programs

The glue that binds all the hardware elements shown in Figure 1 together is the computer program. When you use a word processor, for example, you are executing a computer program that senses what you type in, displays it on some screen, and stores it in file-based memory when you quit the word processor. So, what is a computer program?

A computer program is rather like a recipe. If you look at a recipe in a cookbook, you will see a list of ingredients and a series of instructions that will ask you to add an ingredient, mix a set of ingredients together, and place a collection of ingredients in an oven. A computer program is very much like this: it instructs the computer to move data, carry out arithmetic operations such as

adding a collection of numbers together and transfer data from one computer to another (usually using the Internet). There are, however, two very important differences between recipes and computer programs.

The first difference is size. While a typical recipe might contain about 20 lines of text, computer programs will contains hundreds, thousands, or even millions of lines of instructions. The other difference is that even a small error can lead to the catastrophic failure of a program. In a recipe, adding four eggs instead of three may result in a meal with a slightly odd taste or texture; however, mistyping the digit 1 instead of 2 in a million-line program may well result in a major error – even preventing the program running.

There are a variety of programming languages. They are categorized into high-level and low-level languages. A high-level language, such as Java or C#, has instructions that are translated to many hundreds of the individual instructions of the computer. A low-level language often has a one-to-one relationship with the basic computer instructions and is normally used to implement highly efficient programs that need to respond to events such as the temperature of a chemical reactor becoming critical.

Every few days the media features a story about a software project that has overrun or is over budget or a computer system that has dramatically failed. Very rarely can these failures be attributed to a failure of hardware. The failures occur for two reasons. The first reason for a malfunctioning of an existing computer system is technical error, for example a programming error that was not detected by testing. The second reason is due to managerial failings: projects that overrun or dramatically exceed their budgets tend to occur because of human factors, for example a poor estimate of project resources being produced or a customer changing their mind about the functions a system should implement.

My own view is that given the complexity of modern computer systems, it is hardly surprising that projects will be late and that errors will be committed by developers.

Book themes

The first theme of the book is how hardware advances have enabled the computer to be deployed in areas which would have been unheard of a decade ago. The circuit that a computer processor is deposited on can be easily held in the palm of one hand rather than in a large metal cupboard. A memory stick containing 16 gigabytes of data can easily be attached to a key-ring. Moore's law implies that the computational power of a computer processor doubles every two years. You can now buy hard disk storage of 500 gigabytes for under £60. There are a number of implications. The first is that in the past decade computers have been able to do things few people dreamt of in the 1990s, for example British Telecom's Vision programme that brings television over the Internet. The second is that the reduction in size of computer hardware has enabled them to be physically deployed in environments which would have been impossible a few years ago.

The second theme is how *software* developers have taken advantage of advances in hardware to produce novel applications. An example of this is that of MP3 players such as the Apple iPod. The iPod, and other devices such as the Sony Walkman, obviously rely on advances in hardware. However, they also rely on a software-based technique which, when applied to a sound file, compresses it so that it occupies 10% of its original size without an appreciable decline in sound quality.

A third theme is how the Internet has enabled computers to be connected together in such a way that they behave as if they were just one big computer. This is embodied in an idea known as 'cloud computing' whereby data, rather than being stored in a

local database, are held in a number of computers connected to the Internet and may be accessed by programs that can be developed by Internet users who have relatively low-level programming skills.

Allied to this idea is that of the Internet as a huge resource of data which the computer user can tap into. This includes data such as that published by the British government's data.gov.uk and US government's Data.gov programs, but also data that have been contributed directly or indirectly by users of the Internet. For example, there are sites that enable you to home in to your town or village and examine the broadband speeds that are being experienced by your neighbours, the data that these sites contain having been donated by the users of the site.

A fourth theme is how the Internet has provided creative facilities that were only available to professionals. For example, computer hardware advances, software advances, and advances in the technologies used to create video cameras mean that anyone can become a film director and display their results on the Internet. A computer user can now buy hardware and software for less than a thousand dollars that enables them to reproduce the features of a recording studio of the 1990s.

A fifth theme is how advances in computer processor hardware have enabled number-crunching applications which, until a few years ago, were regarded as outside the realm of computation. Moore's law implies that computer processors become twice as powerful as they were in the previous year. The consequence of this is that over the past decade, processors have become over a thousand times more powerful and, combined with other hardware improvements such as the increased speed of data-storage devices means that, for example, simulations involving the natural world – for example, simulations of hurricanes – can now be easily carried out without deploying powerful supercomputers.