Computers in microbiology

a practical approach

Edited by

T N Bryant

J W T Wimpenny

Computers in microbiology

a practical approach

Edited by

T N Bryant

Medical Statistics & Computing,
University of Southampton,
Southampton General Hospital,
Southampton SO9 4XY, UK

J W T Wimpenny

Department of Pure and Applied Biology,
University of Wales,
College of Cardiff,
Cathays Park,
Cardiff CF2 1XH, UK

IRL PRESS
at
OXFORD UNIVERSITY PRESS
Oxford New York Tokyo

IRL Press
Eynsham
Oxford
England

First published 1989

British Library Cataloguing in Publication Data

Computers in microbiology.
1. Microbiology. Applications of computer systems
I. Bryant, T.N. II. Wimpenny, J.W.T. III. Series 576'.028'5

Library of Congress Cataloging in Publication Data

Computers in microbiology: a practical approach/edited by T.N. Bryant and J.W.T. Wimpenny.
p. cm.
Includes bibliographies and index.
1. Microbiology – Data processing. 2. Microbiology – computer simulation. I. Bryant, T. N. II. Wimpenny, J. W. T. (Julian, W. T.)
[DNLM: 1. Computers. 2. Microbiology – methods. QW 26.5 C738]
QR69.D35C66 1989
576'.028'5—dc19

ISBN 0 19 963014 3 (hardbound)
ISBN 0 19 963015 1 (softbound)
Previously announced as:
ISBN 1 85221 084 2 (hardbound)
ISBN 1 85221 086 9 (softbound)

Typeset and printed by Information Press Ltd, Oxford, England.

Preface

Computers are now a part of our lives, almost as familiar as the telephone and the microwave oven. Microbiologists are not divorced from this general trend and their attitudes to this technological revolution is quite varied. Some have never programmed a computer in their lives and view the keyboard with a certain suspicion even now. Others are converts, happily exploring the power and the puzzle that computers provide. Many have used computers for most of their professional lives.

This book was written by microbiologists who are experts in their fields for other microbiologists. Hopefully each of the groups described above will gain something from these articles.

The areas covered in the book discuss the application of computers to a number of fundamentally important areas in microbiology. These include clinical microbiology, data analysis, fermentation measurement and control, image analysis, modelling and simulation, taxonomy and systematics, and finally teaching. Each chapter illustrates how computers can be used to solve microbiological problems and is not concerned with computer programming *per se*.

T.N.Bryant
J.W.T.Wimpenny

Acknowledgement

Trademarks/Owners

IBM/IBM Inc.
Apple. Macintosh/Apple Computers Inc.
Turbo Pascal/Borland International Inc.
Microsoft/Microsoft Corp.

All other brand and product names are registered trademarks of their respective holders.

Contributors

R.M.Atlas
Department of Biology, University of Louisville, Louisville, KY 40292, USA

S.Bascomb
Department of Medical Microbiology, Wright Fleming Institute, St Mary's Hospital Medical School, London W2 1PG, UK

N.Bratchell
Institute of Food Research, Bristol Laboratory, Langford, Bristol BS18 7DY, UK

T.N.Bryant
Medical Statistics & Computing, University of Southampton, Southampton General Hospital, Southampton SO9 4XY, UK

K.A.V.Cartwright
Public Health Laboratory, Gloucester Royal Hospital, Great Western Road, Gloucester, GL1 3NN, UK

P.J.H.Jackman
National Collection of Yeast Cultures, Division of Genetics and Microbiology, AFRC Institute of Food Research, Norwich Laboratory, Colney Lane, Norwich, Norfolk NR4 7UA, UK

H.J.H.MacFie
Institute of Food Research, Bristol Laboratory, Langford, Bristol BS18 7DY, UK

P.G.G.Miller
Department of Microbiology, University of Liverpool, Liverpool L69 3BX, UK

P.Morgan
Shell Research Ltd, Sittingbourne Research Centre, Sittingbourne, Kent ME9 8AG, UK

A.G.O'Donnell
Department of Agricultural and Environmental Science, The University, Newcastle upon Tyne NE1 7RU, UK

J.I.Prosser
Department of Genetics and Microbiology, University of Aberdeen, Marischal College, Aberdeen AB9 1AS, UK

M.C.Whiteside
Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

J.W.T.Wimpenny
Department of Pure and Applied Biology, University of Wales, College of Cardiff, Cathays Park, Cardiff CF2 1XH, UK

Abbreviations

ADC	analogue-to-digital converter
CAI	computer-assisted instruction
CCD	charge-coupled device
CDSC	Communicable Disease Surveillance Centre
c.f.u.	colony-forming units
CPU	central processing unit
DAC	digital-to-analogue converter
DAPI	4',6-diamidino-2-phenylindole
DDC	direct digital control
DEFT	direct epifluorescent filter technique
DFA	discriminant function analysis
DHA	District Health Authority
DOT	dissolved oxygen tension
DTP	desk-top publishing
ELISA	enzyme-linked immunosorbent assay
EM	electron microscopy
GLC	Gas–liquid chromatography
GLIM	Generalized Linear Interactive Modelling
HIV	human immunodeficiency virus
HPLC	high-performance liquid chromatography
IA	image analysis
ID	identification
K	Kilobyte
LAN	local area network
LAT	lock-up table
MANOVA	multivariate analysis of variance
Mb	megabyte
MFLOPS	million floating point operations per second
MIC	minimum inhibitory concentration
MIPS	million instructions per second
MLP	Maximal Likelihood Program
MPU	microprocessor unit
NLQ	near letter quality
OTU	operational taxonomic unit
PAD	packet assembler disassembler
PCR	principal components regression
PDL	page description language
PHLS	Public Health Laboratory Service
PID	proportional integral derivative
PIO	parallel input/output
PLS	partial least squares
RAM	random access memory
R–K	Runge–Kutta
ROM	read only memory
SSC	supervisory setpoint control
WORM	write one read many

Contents

CHAPTER 1

Introduction

T.N.BRYANT and J.W.T.WIMPENNY

1. INTRODUCTION

Computers are at the centre of a modern technological and information revolution. Their speed, power, flexibility and sheer value for money are changing almost every aspect of our lives, hopefully for the better.

Microbiology, like almost any other field of endeavour, stands to gain from this revolution. Microbiologists, as do scientists in other fields, vary in their response to the phenomenon of computing. A few enthusiasts are familiar with the theory of computer operation. They understand the operating system and are able to design and implement programs of varying degrees of sophistication. Others are excited by the possibilities of applying these techniques to their work but lack both the experience and the confidence perhaps to get started. Another group eschew the field for one of a number of reasons. This book is aimed at all these classes. We hope that experienced computer users will gain at least some information from these pages whilst the reluctant and not so reluctant debutante will find that taking the plunge is really not so hard.

The application of computers in any field roughly divides into two parts. First, general computing techniques not confined to any one discipline, and second, specific applications used mainly or wholly by practitioners in the field. Since the first area (which includes, e.g. word processing), can actually take up most of the scientist's computing time we plan to discuss these topics briefly in this introductory chapter. Specifically microbiological subjects will be treated separately in the relevant chapters.

2. WHAT IS A COMPUTER?

Many readers already have a working knowledge of computers and so may skip this brief introduction. If fuller details are needed the bibliography at the end of this chapter should be consulted. Most people have some idea of what computers are and what they can do, thus a computer is, 'a machine which, under the control of a stored program, automatically accepts and processes data, and supplies the results of that processing' (British Computer Society, 1).

2.1 A simple computer

The diagram (*Figure 1*) is a simple representation of a computer showing the hardware or the physical components that are needed. The essential element of any computer is the central processing unit (CPU). The CPU carries out two basic functions, it stores information and it acts as a calculator. The CPU is a very simple brain analogue and in the same way that the brain can only function in the context of the rest of the body,

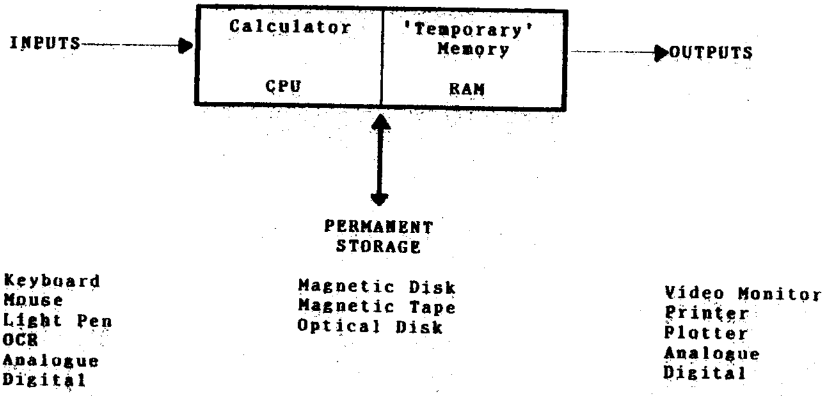

Figure 1. A diagrammatic representation of a computer.

the CPU requires 'peripherals' to perform useful work. The latter include the provision of power and the ability to communicate with a wide range of input and output devices. Although the CPU contains memory this is typically 'dynamic' and once the electrical power is switched off its contents are lost.

Permanent memory or 'backing store' is usually provided by magnetic disc or magnetic tape although optical discs based on laser technology are now available. Discs are either 'floppy' (also called discettes) or 'hard' (sometimes referred to as Winchester discs). Floppy discs are removable, lightweight, flexible structures that can be sent through the post. Hard discs are rigid and they operate within a sealed unit. With few exceptions they are not removable. Because they are rigid, hard discs have closer geometrical tolerances leading to higher packing densities and rates of data transfer that may be 10 times that possible with floppy discs. Optical media, though capable of storing gigantic amounts of information, are primarily for data that, once created, does not need altering. Such devices include WORM drives. The acronym stands for write once read many (times).

2.2 Types of computer

In principle the organization of the CPU is the same for all computers, be they micro-, mini- or mainframe. The classification of computer types is very blurred, the British Computer Society defines a mainframe computer as, 'a computer with a variety of peripheral devices, large amount of backing store and a fast processing unit. The term is generally used in comparison with a smaller or subordinate computer' (1). Clearly the terms used in this definition are relative since the modern microcomputer is as powerful as the mainframe system of 10 years ago. Perhaps a better definition of a mainframe computer is that it costs a great deal of money and it needs an air conditioned environment and operators to run it! The term minicomputers is vague and perhaps only persists now because historically some computer firms established a niche selling small laboratory-based computers which, to distinguish them from their costlier brethren, were christened minicomputers. Today both mainframe and minicomputers share another attribute. They are designed and optimized to share processor power between a number of different end-users or tasks. The power and the speed of these systems, allow each user his own 'virtual' computer even though many other users could be connected simultaneously. For some applications this concept of the virtual computer could be unsatisfactory. Thus a few processor-intensive numerical simulations or a system operating near its design capacity could become intolerably slow. Highly interactive applications such as word processing, especially using screen editing systems, become impossible to use properly under these conditions and perform much better on a microcomputer dedicated to a single user. As a result word processing programs on mainframe computers remain at the dinosaur stage in evolution! The mainframe computer comes into its own in a number of important areas. It can run lengthy numerical programs in batch mode as a background job or overnight; it offers message and file transfer and electronic mail facilities between its own users and between users at other sites throughout the world and last, but not least, it offers comprehensive archiving and backup facilities as a routine service.

One feature that distinguishes microcomputers from the rest is that they have become consumer durables traded like TV sets from high street stores. They have become desk-top artefacts not out of place in an office or a living room. They need neither sophisticated power supplies nor air conditioning and their computational power, however it is measured, is still increasing at an astonishing rate.

This rapid increase in power and availability means that a great deal of time and application is needed to assimilate the necessary information to use it to its full potential. On the other hand power and sophistication also means that 'user-friendliness' can be designed into operating systems and commercially available software so that comparative novices can achieve results quickly. This is quite in contrast to the situation on mainframe and on some minicomputers, where it is expected that computer staff will customize applications plus documentation for a generally computer literate population.

2.3 Representing information in a computer

The basic element used to hold information in a computer is the binary digit or bit. Like a switch this can have either of two states, off or on, represented in binary form as 0 or 1. Computers manipulate collections of bits. A group of 8 bits is known as a byte which represents 256 combinations ranging from 00000000 to 11111111 (0–255 in decimal). Since it is tedious to work in binary notation the hexadecimal notation using numbers to the base 16 is used instead. Here numbers up to nine are represented by 0–9 as usual whilst 10–15 use the letters A–F. This reduces the number of digits used to express a number and is more appropriate than decimal notation especially when dealing with bytes and binary data. Thus the decimal number 197, is 11000101 in binary and C5 is hexadecimal. Information may be manipulated as a single byte or as a combination of bytes called a word. The size of computer word can vary according to the type and make of mainframe systems where the word may be 32, 36 or even 60 bits long. Computers can obviously manipulate textual information, the 256 possible permutations of 8 bits are more than enough to allow a single byte to represent upper and lower case letters, digits, punctuation marks and so on. The most common code used to represent characters is called ASCII, or American Standard Code for Information Interchange. ASCII provides an easy way, although not the only

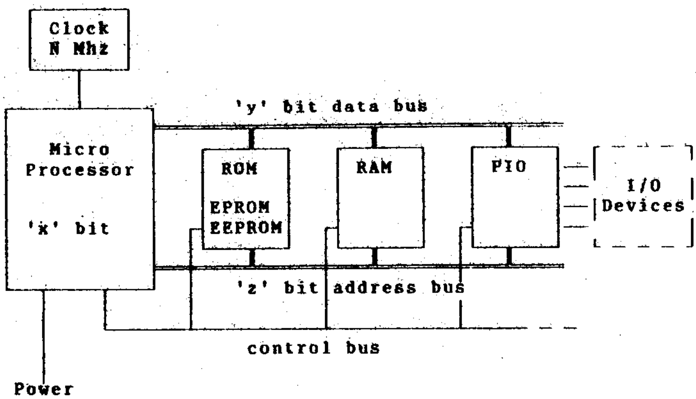

Figure 2. The main components of a microcomputer.

one, of allowing computers and peripheral devices to exchange information. Storage capacity in computing is frequently expressed as thousands of bytes or kilobytes, this is not exactly 1000 but 1024 bytes in decimal since it actually corresponds to the binary expression 2^{10}. Computer capacities are therefore larger than one might at first expect: for example 640 kilobytes (K) of RAM actually means 655 360 bytes of memory whilst 20 megabytes quoted for the capacity of a disc means that 21 309 440 bytes can be stored on it.

2.4 A typical microcomputer

The main components of a 'typical' microcomputer are shown in *Figure 2*. In principle the organization of larger computers is quite similar, though commonly several components are used to create the functional equivalent of a single component in a microcomputer. A typical microcomputer is based around a microprocessor unit, or MPU, which is a CPU without the memory, manufactured as a single integrated circuit. The activities of the MPU are synchronized by a clock whose frequencies range from 1 MHz upwards, depending on the type of microprocessor. All systems have a limited amount of permanent ROM or read only memory: this contains instructions to bring the machine to life when it is first switched on (booting it up in computer jargon). It usually provides some basic operating facilities such as the ability to delete or print information and a language such as BASIC in some systems. The information in ROM cannot be altered. A variation of this is the EPROM (erasable programmable read only memory) and sundry other similar devices which may be present in some systems. RAM, standing for random access memory, is the main memory in a computer. Its contents are volatile and lost when the power is switched off. Ram memory is used to store programs and all the necessary data required for the program to operate. In many computers it also contains the operating system (a specific set of programs) which is read into memory from a disc on booting up the machine. Input and output devices communicate with the MPU using a parallel input/output device or PIO and it is the latter which provides the link to the outside world. The transfer of information between