Cambridge Computer Science Texts · 6

Computing Systems Hardware

M. WELLS

Cambridge Computer Science Texts · 6

Computing Systems Hardware

M. WELLS
Centre for Computer Studies
University of Leeds

Cambridge University Press
CAMBRIDGE
LONDON · NEW YORK · MELBOURNE

Published by the Syndics of the Cambridge University Press
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
Bentley House, 200 Euston Road, London NW1 2DB
32 East 57th Street, New York, NY 10022, USA
296 Beaconsfield Parade, Middle Park, Melbourne 3206, Australia

First published 1976

Printed in Great Britain
at the University Printing House, Cambridge
(Euan Phillips, University Printer)

Library of Congress Cataloging in Publication Data

Wells, Michael.
Computing systems hardware.
(Cambridge computer science texts ; 6)
Includes bibliographies and index.
1. Computers. I. Title. II. Series.
TK7885.W42 001.6'4 75 27263

ISBN 0 521 29034 1 pbk.

Preface

The aim of this text is to introduce the concepts which underlie the structure of computing systems. The first chapter starts with basic principles and explores the important distinctions between encoding with digits and representations of digits; the simple kinds of operation possible with such representations; and the necessity of techniques for suppressing detail when describing a complex system such as a computer. The next four chapters describe some of the equipment currently in use for operations on information, laying emphasis on the economic factors that arise as a consequence of the internal design of individual units. These factors compel the use of devices which are not always those of highest performance in particular situations, and in the final chapter the compromises open to the system designer in his search for the most economic design of a system are discussed.

It is worth reminding the reader that computers are tools, meant to be used by people as part of information processing systems and that when assessing what is the 'cheapest' solution to a problem it is necessary to include all the relevant costs. The fact that this book makes no reference to costs outside the rather narrow confines of the internal details of the computing system should not be taken as an indication of the relative importance of those other costs.

In a subject such as this, where very rapid changes in technology are still taking place, and where the nature of courses dealing with the material also changes, it is difficult to aim a book at an audience; not only is the author aiming at a moving target, but the platform from which he aims is also moving. The book is intended for students in the second year of a course which includes a substantial element of Computer Science. The cautious wording recognises the fact that for most students Computer Science as a subject entire in itself may be a poor choice, and is better taught in conjunction with some other subject.

The reader is assumed to have some prior knowledge of concepts such as 'program', 'compiler', and the equivalent of A-level Physics. In the main, mathematical discussion is kept to a minimum, except for the material at the end of chapter 4 on Data Transmission, which can be omitted. The book is substantially a synoptic view of the technical situation as in 1974, and changes will arise from the advent of new devices, and from the continuing fall in the real price of semi-conductor devices. The reasons for the introduction of new devices are reasonably clear; less clear, but more interesting are the reasons for the fall in prices, and one day a book can be written on the development of the semiconductor industry in the mid-sixties and seventies. To keep pace with such a rate of change is very difficult - one solution, outlined in example 6 of chapter 2 is to have the reader reconstruct some of the factual material. The reader may also wish to discover those places where the forecasts made for future trends in the competitiveness of several devices are wrong!

It is a pleasure to record my indebtedness to all those who have helped in the production of this book; my colleagues Prof. Colin Reeves and Prof. Ewan Page for their helpful criticism; to Carol Gower and Joy Crowther who have patiently typed and retyped drafts; to the publishers; and to my wife without whose encouragement the work would never have been completed.

April 1976 M. Wells

Contents

1 · Basic aspects of digital systems

1.1 *Representations and encodings*

Information is an abstract concept, and from a very early age we are taught to associate symbols with many concepts. We do this to such an extent that there is a strong tendency to equate the symbol with the abstract concept which it represents. Thus we think of 8 as *being* the number eight, rather than as *representing* it; similarly we think of h as being the eighth letter of the alphabet rather than as representing it. In fact of course the abstract number 8 is a rather subtle concept. A very clear account of what a number is can be found in a book *The Psychology of Learning Mathematics* by Richard R. Skemp (Penguin Books, 1971).

We are taught to associate symbols with information so as to facilitate the processing of information. This symbolising of the information may well involve the use of a measure of some kind, where the information we wish to use involves a quantity which is not amenable to counting. For example one would normally count the number of passengers on an aeroplane, although for the designer of the wing a much more relevant factor might be the total weight of the passengers rather than their number. Whereas the number of passengers is simply a number, their weight is expressed in terms of a standard unit such as the kilogram or pound. This rather straightforward notion allows us a convenient classification of information processing systems into those which use quantities of one sort to represent quantities of a different sort, and those which use symbols; there are of course systems which use both methods.

Those systems which use quantities of one sort (usually electrical quantities such as currents or potential differences) to represent quantities of any sort are *analog* systems, while those which use symbols are *digital* systems; systems which use both methods are *hybrid* systems.

In this book we shall concentrate almost exclusively on digital systems, which manipulate symbols. This manipulation is (or ought to be) done in such a way that if we were to take some data about the real world and apply to this data some operation to produce a result, then the symbols which result if we perform on our digital system a representation of this operation on a representation of our original data will be a representation of this result. We can see this taking place in figure 1.1.

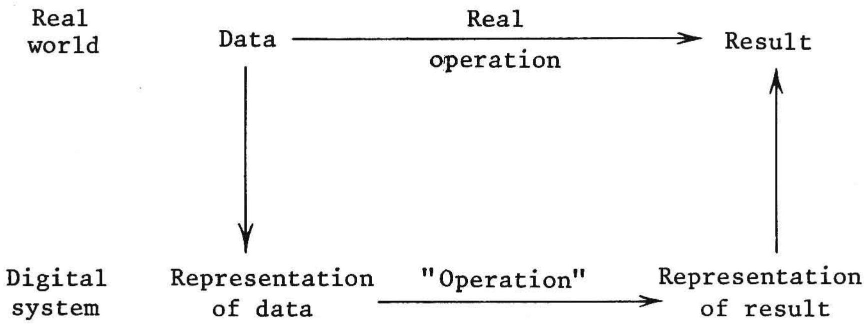

Figure 1.1 Operating on representations of data.

It is not in general possible to proceed directly from the data to an internal representation as figure 1.1 suggests. One reason is that the engineering of a digital system is a compromise between the requirements of extreme reliability and extremely high speeds of operation. The resolution of these requirements leads to the number of different symbols that can be represented within the system being as small as possible. This number of different symbols cannot be one and must obviously be an integer (notice that the absence of a symbol is itself a symbol). If you do not believe this, try to make sense of this, inwhichallthespaces (i.e. absences of symbols) havebeensuppressed!

It is instructive to consider how this situation arises, in rather general terms. A digital device has various stable states which can be recognised one from another. If we plot a graph of the potential energy of the device as a function of a 'displacement' the graph will have a number of local minima, each corresponding to a stable state of the device. The term 'displacement' here is used in a rather general way; typically for an electrically operated device the displacement would be the potential difference between two parts of the device, while for a mechanical device it might be the length of a spring or the angle between two ends of a component in torsion. Whatever the displacement, there will be a minimum for each stable state, and these minima will be separated by maxima in the energy/displacement curve. (Figure 1.2. See page 4.)

Switching the device from one state to another involves forcing the displacement from one value to another and this in turn involves performing work in climbing the potential energy hump between the two states. If W is the height of this potential hump and P is the maximum input power available then the switching time is in the order of W/P. We can make the device operate more quickly either by reducing W, or by increasing P. However, if we make W too small the device will become unreliable. In any system noise is always present, and this has the effect of introducing an uncertainty as to the potential energy of the device. If this uncertainty in energy is ΔW, then it is a standard result in thermodynamics that the system may switch from one state to another with a probability which is dominated by a term of the form $e^{-W/\Delta W}$. Since we have rather little control over ΔW, we can minimise this probability by increasing W. The choice of W is thus a compromise between our wish for maximum immunity to noise, and minimum operating time.

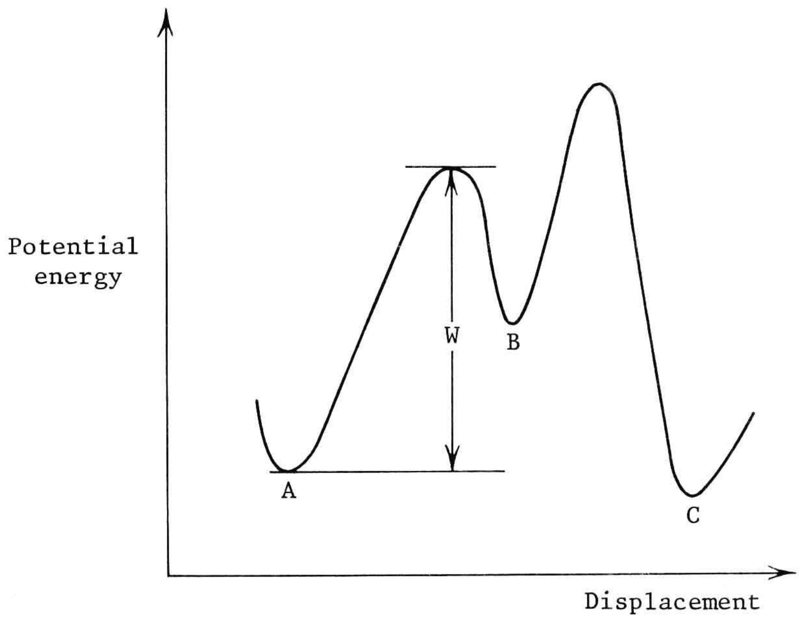

Figure 1.2 Energy/Displacement curve for a device with three stable states A, B, C.

As well as switching devices from one stable state to another, we must also be able to determine the state of a particular device. To do this we must determine the value of the displacement, and decide to which of the possible displacements this value corresponds. We must take into account two effects which serve to widen the displacement associated with a state from the discrete values indicated by the positions of the minima on figure 1.2 to a range of values. The first of these is again noise, which introduces an uncertainty in the energy level, and so into the value of the displacement for any particular device. The second is the effect of component tolerances. The devices when manufactured will not all be identical and the energy/displacement curve for a given device will show small deviations from the curve for any other. When manufacturing devices there will be some form of quality control to ensure that the curves for all the devices agree to within some predetermined limits, which are the 'tolerances' mentioned above. The positions of the minima for a particular device may lie anywhere within these tolerance limits.

Both noise and tolerance limits widen the range of displacements which correspond to the energy minima from a single point to a probability distribution. It is essential that under no circumstances should the tails of any of these distributions overlap, for if they do wrong decisions as to the states of these devices will be made. The risk of this occurring is reduced by making the displacement from one state to the next as large as possible when compared with the spread of displacements possible for a particular state. Since the maximum possible displacement is limited, we can maximise the separation by using the smallest possible number of states and arranging for these to occur as near as we can to the ends of the range of possible displacements.

Thus virtually all modern digital systems use only two internal symbols, disguising this fact with varying degrees of success. It is therefore necessary to proceed in two steps:

(1) encoding, i.e. mapping the set of symbols that we wish to handle on to groupings of the (two) symbols the system can handle;

(2) representing, i.e. transforming the representation of the symbols the system can handle into a form convenient for the system.

Our simple picture of figure 1.1 now takes on a slightly more complex appearance, figure 1.3. (See page 6.) The data of the original problem are first encoded (using for example a card punch) and this encoding is then transformed into an internal

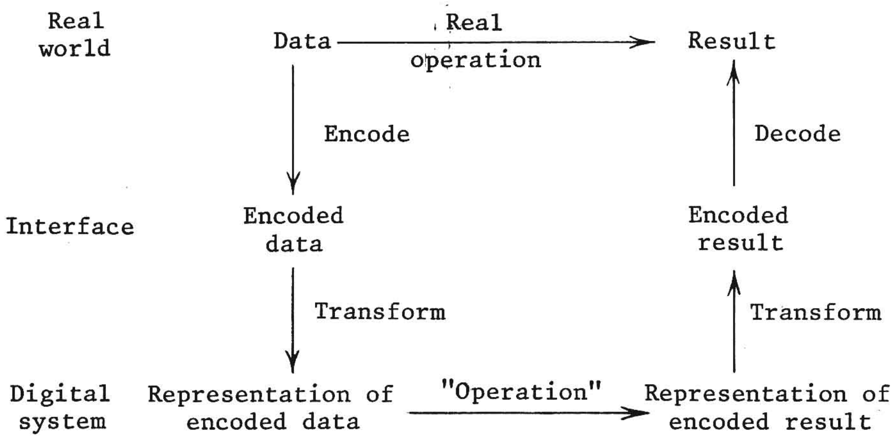

Figure 1.3 Operating on representations of encoded data.

representation (using now a card reader) in which form the data are processed. This internal representation can now be transformed again (using a card punch driven by the digital system) and then converted to a result (by interpreting the cards). It is of course arguable that by interpreting the cards we do *not* produce the result but merely yet another representation. It is also important to realise that what we have done is solely concerned with information, NOT with reality If this point is obscure, consider applying the abstract operation of multiplication to the abstract data the numbers three and five; unless a mistake is made at some stage the system will produce a representation of fifteen. Now by contrast consider applying the operation of 'boiling' to 'a panful of potatoes'; the results of the manoeuvrings of figure 1.3 will NOT be a panful of boiled potatoes. Only the real operation in the real world will bring about this result. In such a simple case this point is obvious. However, as systems become more complex it is only too easy to confuse the representation of the situation which is maintained by the digital system with the real situation as it exists in the real world, and to suppose that actions within the real world will automatically bring about changes in the representation within the digital system or even more dangerous that changes within the representation will cause changes in the real world.

1.2 *Operations within digital systems*

As we have just seen a digital system holds physical representations of encodings of symbolic information, each symbol being represented by a grouping of several of the rather small number of distinct symbols possible within the digital system. If we take information in this way we find that there are five possible things that we can attempt to do with the information:

(0) leave it alone;

(1) retain it - i.e. keeping its position in space fixed, move it through time;

(2) change its representation - i.e. without changing the encoding, change the way in which the individual elements of the encoding are represented;

(3) transmit it - i.e. keeping its position in time fixed, move it through space;

(4) manipulate it - i.e. without changing the representation change the actual symbols that are encoded.

It is plain that option (0) is not really an option at all. If we leave the information alone the information will simply disappear, and the only meaningful possibilities are options (1) to (4).

It is also plain that these possibilities are idealisations of what can be achieved in that our abilities in certain cases are limited by laws of nature. We can (as yet) move