Proceedings of the Conference on

Computers—Systems and Technology

Tuesday 24th to Friday 27th October 1972

Middlesex Hospital. Medical School, London

Organized by

THE INSTITUTION OF ELECTRONIC AND RADIO ENGINEERS

with the association of

The Institution of Electrical Engineers

The British Computer Society

Published by the I.E.R.E., 8-9 Bedford Square, London WC1B 3RG

Telephone: 01-637 2771  Telegrams: Instrad London WC1
CIRCUIT TECHNOLOGY IN A LARGE COMPUTER SYSTEM.

D.J. Kinniment*, M.Sc., Ph.D., M.I.E.E. and Professor D.B.G. Edwards*

M.Sc., Ph.D., M.I.E.E.

Summary:
In the design of a large high speed computer the size of the system leads to long cable delays for data transmitted between different parts of the machine. This problem and the interconnection of a high speed E.C.L. circuit family in a large computer are discussed. Priority circuits in a large asynchronous system also present problems not usually encountered in smaller machines. Comment is made on future lines of development for technology in high speed computers.

1. Introduction

In the time between the design of the ATLAS computer and the start of the MU5 project (ref.1), considerable advances in circuit and memory technology have been made. The basic ATLAS gate was a discrete component circuit with a propagation delay of approximately 16nSec and a power dissipation of 250mW. These circuits were assembled on plug in printed circuit cards so that the overall packing density was approximately 300 gates per ft³. The main store used a number of stacks of 2µSec ferrite core storage. The corresponding features of the MU5 system are E.C.L. circuits containing 2 – 3 gates per chip with a propagation delay of approximately 2nSec per gate, and four stacks of plated wire store with a cycle time of 260nSec. All interconnections in MU5 are driven as matched transmission lines, this requires relatively high currents in the gate outputs and the average power dissipation is 200mW per gate.

In the period of 10 years between the ATLAS and the MU5 systems, the speed of the circuits has therefore increased by a factor of about 8:1, but the power dissipation per gate has not been significantly reduced. At the same time reductions on physical size due to the introduction of integrated circuit techniques have increased the packing density to over 2000 gates per ft³ in certain areas of MU5. This improvement in volume where the logic circuits are concentrated is not maintained in the overall system when the requirements of cooling and power supply distribution are taken into account, and this effect is more significant in MU5 because of the high power dissipation per unit volume.

An increase in speed of 8:1 in both the logic circuits and the main store implies a corresponding improvement in the system speed, but a comparison of the access times to the store for the two systems given 1.8µSec in ATLAS and 58SnSec in MU5, or a factor of just over 3:1.

These times include the delay time through the current page registers required for address translation in apaged system (ref.2) and for the organisation of four store stacks whose cycle times are overlapped to give a high rate of instructions and operands to the processor.

* Department of Computer Science,
University of Manchester.
The relatively long access time in MU5 is mainly due to the time taken for signals to travel between widely separated parts of the system, and also to a lesser extent the priority decision circuits required in an asynchronous system. While the volume occupied by the circuits has been considerably reduced, the average interconnection length is related to the linear dimensions rather than the volume and an improvement of only about 2:1 in cable delay time has been achieved in practice.

1.1 The MU5 System

A block diagram of the MU5 central processor is shown in Fig.1. This is made up of a number of units each concerned with a part of the activity involved in executing an order. Instructions are obtained from the local store via the store access control (S.A.C.) which contains the current page registers required to translate the virtual addresses used by the processor into the real addresses in the store. Instruction accesses are made in 128 bit groups, stored in an instruction buffer and passed in 16 bit groups to the primary operand unit (PROP) (ref.3). This unit interprets the instruction, and accesses the operand specified directly. This operand may be used by the primary operand unit itself or by the index arithmetic unit (B-ARITH), or indirectly by the secondary operand unit (SEOP) as a data descriptor to access elements in data structures such as arrays, which are processed in the accumulator (ACC).

The paths taken by the address and data in a typical store access for an element in a data structure are identified in the diagram and are as follows:

1. Cable delay from operand buffer to S.A.C. 35nSec
2. Priority logic at input of S.A.C. 65nSec
3. Address translation in C.P.R:s 120nSec
4. Store overlapping priority logic 65nSec
5. Cable delay to store interface logic 25nSec
6. Address buffering in store interface 15nSec
7. Cable delay to store stack 25nSec
8. Store access 125nSec
9. Cable delay to interface logic 25nSec
10. Data buffering in store interface 10nSec
11. Cable to S.A.C. 25nSec
12. S.A.C. 15nSec
13. Cable to operand buffer 35nSec

Cable delays and priority circuits thus represent over half the total access time.

In order to overcome this problem, and also to help achieve the design aim of a performance improvement at least a factor of 20 over ATLAS, three separate buffer stores have been provided in the system to
reduce the access time for the different types of data used. One
buffer is provided for instructions one for PROP named variables (ref.4),
and one for array elements, named variables and literals used in the
accumulator. Studies of the patterns of accesses made by selected programs
show that the majority of these accesses are for named
variables within a routine, and that only a small number of these variables
are in use at a given time (ref.5). The buffers can
therefore be relatively small and it is estimated that a name store
with a capacity of only 32 variables will be sufficient to hold the values
required by over 99@o of named primary operand accesses (ref.6). The rate of
execution of instructions is also improved by means of a pipeline type of
construction (ref.3) in which several partially completed orders are in
progress concurrently.

The use of pipeline techniques allows an average rate of one instruction
executed every 120nSec and a peak rate of one every 40nSec. Most operands
and instructions can be obtained from the small buffer stores situated
close to the point of use so that the access time to each
type of data is relatively short and the number of requests to the
local store is reduced to an average of one 64 bit access every 210nSec for
instructions and one 64 bit access every 800nSec for operands.

2. Circuit Family and Interconnections

The range of integrated circuit devices used in MU5 is shown in Fig.2 which
also indicates the number of each type used. The pipeline
type of construction is reflected here by the relatively large number
of dual flip flop devices, mainly used as storage registers in the various
stages of the pipeline. Between pipeline stages gating of data, and
decoding etc. is mainly performed by multi input and/or gates of
type shown in column 5. The most complex devices used are the
16 bit random access memory, and 8 bit associative memory chips used
in the associatively addressed buffer stores (ref.4), and translation
circuits from ECL levels to levels compatible with the associative devices
account for most of the flat packs. discrete transistors and diodes.

The majority of the integrated circuits in the system are mounted
on plug in printed circuit modules 1.6" x 2.1" with 20 pins (ref.7).
Up to 200 of these modules can be interconnected by means of a single
12 layer printed circuit platter as shown in Fig.3. The packing
density of circuits on these modules however is relatively poor and
commonly used complex function macros such as adders (ref.8) and a
32 bit associative store have been designed on 1.6" x 4.4" and
3.0" x 4.4" boards. These are also shown in Fig.3. The platters in
the MU5 system measure approximately 13" x 16" and up to 33 platters
can be contained within a logic bay. Within a bay 24 platters are mounted
on moveable doors to allow easy access for maintenance and
a further 9 are mounted on a fixed central plane. This arrangement gives
relatively short runs for interconnections whose source and destination are
situated on the same platter but interconnections crossing platter
boundaries must travel an average of 12", and those passing from a fixed
plane to a door, or from one side of a door to
the other travel an average of 8 feet along coaxial cables. A histo-gram
of the distribution of interconnection lengths in the MU5 system, measured
from the output pin of the source circuit to the input pin of the
destination circuit is shown in Fig.4.
FIG. 5. ECL FLIP FLOP SETTLING TIME

FIG. 6. PERFORMANCE OF ECL FLIP FLOPS
Since a maximum of 3 integrated circuits can be mounted on a 20 pin module or 7 circuits on a 40 pin module relatively few interconnections are between circuits on the same module, and a typical connection travels a distance of about 1" to reach the platter, 2" on the platter and a further 1" from the platter to its destination on a second module.

Within the MU5 system approximately 80% of all interconnections are between integrated circuits on the same platter. 13% travel between adjacent platters, 5% go through cables to other platters in the same bay and 2% travel from bay to bay, distances of up to 50ft.

ECL circuits with typical edge speeds of 2nSec are used, driving interconnections which are typically 4" in length but vary up to 50ft, and average 6" between gates on the same platter. The relatively long delay time of these interconnections dictates the use of a matched transmission line approach to minimise the possibility of reflections. A series matching technique has been used in MU5 in which the ECL gates can drive two matched 75Ω lines from each output, and each line is capable of driving up to two inputs at the receiving end. Associated with every output is a group of three resistors, the output load resistor and two series matching resistors. These are fabricated on a ceramic chip in thick film technology and over 45,000 groups are required in the system. Because of the capacitance represented by each input and the series resistance of the line matching resistor the rise time at the input gate is degraded. and an effective extra delay of 0.6nSec per load introduced. An average gate thus introduces a delay of 2nSec due to propagation through the ECL circuit itself, 1nSec transmission time along the 6" interconnection path and a further 1nSec delay from the input loading. Typical delay per gate in a system is therefore approximately 4nSec. and coaxial cables up to 50ft long are driven without difficulty by the circuits.

3. Asynchronous timing

Each of the units shown in Fig.1 has its own internal timing, functions and data being passed from one unit to another when the sending unit has the data available and when the receiving unit is not busy. This type of operation in which data transfers take place asynchronously generally allows the system to operate at a greater speed than a completely synchronous system where transfers can only take place at fixed times. but can lead to difficulties in the circuits controlling the transfers.

This problem occurs in a number of places in the system but can be illustrated by the paths from the three buffer stores through the store access control to the local store. Here three different units may request a store cycle at any time. and a decision must be taken by the store access control as to when the request can be accepted, so that the unit issuing the request can free its output address and data buffers for other activities.

In this particular example the requests are first staticised in flip flops. On the basis of what type of requests are outstanding, and the state of the pipeline, a combinational circuit indicates which of these requests need to be serviced. This process takes a time of
15nSec from the initial receipt of a request, and 5nSec later the strobe for three decision flip flops is set. When the unit next becomes free the strobe is removed and the state of the decision flip flops indicates which reaquests require action at that time.

Because a second request may occur a short time after the first request, it is possible for the inputs to the decision flip flops to change state just before the strobe is removed, leaving the outputs midway between a '0' and a '1' level. Under these conditions the time taken for the flip flop output to reach a constant level may be long compared with its normal propagation delay as shown in Fig.5, and it is possible for subsequent circuits designed to select the highest priority request for servicing, to give an output inconsistent with either '0' or '1' inputs during this period.

Clearly sufficient time must be allowed here for the decision flip flops to settle to a constant level or failure of the control circuits may occur. The settling time of the circuit used here is a function of its gain-band width product, and the displacement of the output from the mid-level at the time the strobe is removed. In the case of an output starting from the exact mid level, it is possible for an infinitely long settling time to be required, but the probability of such an occurrence is extremely low.

A graph of the mean time between failures of a typical control system against time allowed for settling is shown in Fig.6. This indicates that an ECL flip flop with a propagation delay of 2.2nSec requires over 30nSec settling time for the failure rate to be reduced to an acceptable level and it can be shown that the mean time between failures is given by

\[
\text{M.T.B.F.} = \frac{T_1 T_2 e^{t/\tau}}{\tau}
\]

where

- \( T_1 \) = mean time between strobes
- \( T_2 \) = mean time between requests
- \( \frac{2\pi}{\tau} \) = gain-band width product of flip flop
- \( t \) = settling time allowed

Since the number of decision flip flops required here is relatively small considerable advantage can be gained in the system by using a special circuit with a higher gain bandwidth product than the standard device, and a flip flop with 1.8nSec propagation delay requires only 20nSec settling time here.

The total priority circuit time from the receipt of a request to its acceptance into the input buffer of the S.A.C. using the faster flip flop is 65nSec and together with the cable delays still adds significantly to the communication time between the processor and the store.

4. Future Developments

Families of logic circuits are now available with propagation delays as low as 1nSec/gate, but a corresponding improvement in system performance will not necessarily be obtained by the use of these circuits alone. Other techniques must also be introduced to improve the packing density of circuits and reduce the interconnection delay between gates. In
addition to the interconnection delay problem, system speed is often critically dependent on a relatively few areas such as fast priority circuits, and accurate clock pulse drivers. In order to obtain the maximum performance in the system these control circuits should be faster than the standard circuits used in data paths.

The packing density of gates in a system can be improved by increasing the number of gates per integrated circuit package. Whilst the number of gates per package can be increased substantially by the use of medium and large scale integration techniques, the number of connections required to other packages is also a function of the number of gates in the package, and has been quoted (ref.9) as

\[ p = 3.5 \, n \]

for a high performance system, where \( p \) is the number of pins and \( n \) the number of gates. This expression correlates well with experience in the MU5 system, and indicates that a system partitioned into 30 gate blocks could be accommodated in a number of 40 pin dual in line packages. A package of this size, mounted on a multi layer printed circuit board would require at least 2 square inches of board space. This area is comparable to the area of an MU5 module and represents an improvement of about 10:1 in gates/unit volume over the MU5 technology. An improvement of this magnitude cannot be maintained throughout the system since complex circuits will not be available for each individual system requirement and a large proportion of simple gates are needed in order to optimise the design.

The reduction in interconnection path length obtained by the use of this scale of integration will therefore be approximately 2:1, and a significant part of the system delays will still be provided by the interconnections.

In order to reduce the interconnection delays to a level compatible with 1nSec gates, the packing density must be further increased, but the use of dual in line packages to accommodate the semi conductor chips presents a severe limitation on the number of interconnection pins available. Packages containing 300 gates would typically require 250 pins, and the area occupied would not be significantly less than 10 individually packaged 30 gate arrays. Improvements in the packaging leading to a greater number of interconnection pins per unit area are required before full use of the potential of large scale integration can be made.

The use of more complex semi conductor devices in a high performance system requires a relatively large number of different circuit types, but the small number of each type used and the long production cycle involved for minor design changes make this approach economically unattractive.

Thick film hybrid circuit modules interconnecting a standard range of unpackaged semi conductor chips are another approach which allows high packing density and a relatively short production cycle for modification. This technique also allows the use of semi conductor devices in volume production, and the possibility of tailoring each circuit block to the system requirements with acceptable production costs. Developments in this area may provide the best technology for future large fast computer systems.
5. References


6. Acknowledgments

The MU5 project is supported by the S.R.C. and by I.C.L. The authors would like to thank all members of the University and I.C.L. involved in the project for many helpful discussions and particularly Professors T. Kilburn and F. H. Sumner.