# Asynchronous Design Aspects of High-Performance Logic 

Architectural Modelling of a Bipolar Asynchronous Microprocessor

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER FOR THE DEGREE OF MASTER OF SCIENCE IN THE FACULTY OF SCIENCE

Robert Kelly<br>Department of Computer Science<br>1995

## Table of Contents

Table of Contents ..... 2
List of Figures ..... 5

1. Introduction ..... 11
2. Asynchronous Logic ..... 14
2.1 Delay Modelling ..... 17
2.2 Signalling Protocols ..... 18
2.2.1 Two-Phase Signalling ..... 19
2.2.2 Four-Phase Signalling ..... 20
2.2.3 Data Communications ..... 22
2.2.4 Bundled-Data Interface ..... 23
3. Micropipelines ..... 25
3.1 Control Circuit Elements ..... 28
3.1.1 XOR (Merge) ..... 28
3.1.2 Muller-C (Join) ..... 28
3.1.3 Select ..... 29
3.1.4 Toggle ..... 29
3.1.5 Decision-Wait ..... 30
3.1.6 Arbiter ..... 30
3.1.7 Call ..... 31
3.1.8 Capture-Pass Latch ..... 32
3.2 Control Circuit Examples ..... 32
3.2.1 Event Register ..... 32
3.2.2 Design Example: PARITY FUNCTION ..... 34
4. Verilog HDL ..... 36
4.1 Introduction to HDLs ..... 36
4.2 Introduction to Verilog ..... 37
4.3 Modules ..... 37
4.4 Structural Modelling ..... 38
4.4.1 Design Example: RS Flip-Flop ..... 39
4.5 Behavioural Modelling ..... 40
4.5.1 Compound Statements ..... 41
4.5.2 Process Control ..... 42
4.5.3 Timing Control ..... 42
4.5.4 Design Example: Behavioural Representation ..... 44
4.5.5 Programmable Logic Arrays ..... 45
4.6 Verilog Simulator ..... 46
5. Multi-Level Differential Current Mode Logic ..... 48
5.1 Introduction to Logic Families ..... 48
5.2 Multi-Level Differential Current Mode Logic ..... 50
6. Verilog Modelling of MDCML ..... 55
6.1 Requirement for Accurate Model of System ..... 55
6.2 Determination of Electrical Characteristics of MDCML ..... 55
6.2.1 Output Loading Effects ..... 58
6.2.2 Input Drive Effects ..... 59
6.3 Production of Verilog Model ..... 61
6.3.1 Accuracy Comparison ..... 63
6.3.2 Continuous Assignment ..... 66
6.3.3 Net Delays ..... 67
7. MDCML Asynchronous ARM ..... 69
7.1 ARM Architecture ..... 69
7.1.1 Overview ..... 69
7.1.2 Instruction Set ..... 71
7.2 MDCML Asynchronous ARM ..... 72
7.2.1 Overview ..... 72
7.2.2 Register Bank ..... 76
7.2.3 Memory Interface ..... 83
7.2.4 Address Interface ..... 83
7.2.5 Data Interface ..... 93
7.2.6 Execution Unit ..... 100
7.2.7 Comments on the MDCML Asynchronous ARM Design ..... 106
8. Architectural Modelling ..... 107
8.1 Introduction ..... 107
8.2 Modelling ..... 107
8.3 Features ..... 113
8.3.1 Instantiation Parameters ..... 113
8.3.2 Test Vector Generation ..... 114
8.4 Code Execution ..... 117
8.4.1 Compilation Method ..... 117
8.4.2 Validation Suite ..... 117
8.4.3 Dhrystone Benchmark ..... 118
8.5 Usage ..... 120
8.5.1 Instrumentation ..... 120
8.5.2 Graphical Output ..... 122
8.5.3 Detecting Incorrect Operation ..... 125
8.6 Performance analysis ..... 126
8.6.1 Subsystem Processing Performance ..... 126
8.6.2 Non-symmetrical Propagation Delays ..... 128
8.6.3 Processor-Memory Interaction ..... 129
8.6.4 Internal Pipeline Efficiency ..... 130
8.6.5 Comments on the Performance Analysis. ..... 133
9. Conclusions ..... 134
9.1 Production of the System Model ..... 134
9.2 Current State of the Project ..... 135
9.3 Comments on the Verilog Modelling Environment ..... 135
9.4 Future Research ..... 137
9.4.1 Technology Migration ..... 137
9.4.2 Architectural Design Alternatives ..... 137
References ..... 139
Appendix A: Verilog Model ..... 142

## List of Figures

Two-phase (transition) Signalling. 19
Four-phase Signalling (incorrect operation). 20
Four-phase Signalling (correct operation). 21
Bundled Data Interface. 23
Data Value Constraints. 24
Event Register. 33
Micropipeline Control Block Implementation of Parity Function. 35
Possible Implementation of RS Flip-Flop. 39
Dynamic Power Dissipation. 49
MDCML 3-Input AND Gate. 51
4:1 Multiplexer. 52
Transparent Latch with Reset. 53
2-input AND gate. 56
SPICE model of AND2. 56
2-input AND gate SPICE waveforms. 57
Graphs of Additional Load vs Additional Delay. 59
Graphs of Additional Delay vs Extent of Drive Sharing. 60
Ain after Bin SPICE waveforms. 65
ARM6 Block Diagram. 70
Internal Processor Organisation. 73
Micropipelined Structure of the Execution Pipeline. 75
Register Bank Operation. 76
Asynchronous Register Bank Design. 78
Register Bank Read Cycle Waveform. 80
Register Bank Stalled Read Waveform. 81
Register Bank Write Cycle Waveform. 82
Address Interface Structure. 84
Address Interface Instruction Prefetching Waveform. 89
Address Interface Data Transfer Waveform. 90
Address Interface Block Data Transfer Waveform. 91
Address Interface Branch Waveform. 92
Data Interface Structure. 93
Data Interface Byte Read Waveform. 97
Data Interface Byte Write Waveform. 98
Data Interface Instruction Read Waveform. 99
Execution Unit Structure. 100
Execution Unit Multiply Waveform. ..... 104
Execution Unit Shifted ALU Operand Waveform. ..... 105
MDCML Asynchronous ARM Processor Diagram. ..... 111
MDCML Asynchronous ARM ‘Top-Level’ Verilog Model. ..... 112
Register Bank Display using Verilog gr_regs() system task. ..... 123
Pipeline Occupancy using Verilog gr_bars() system task. ..... 124
Graph of Block Processing Time vs Dhrystone performance. ..... 127
Effect of Non-Symmetrical Propagation Delays. ..... 129
Effect of Memory Speed on Processor Performance. ..... 130
Pipeline Occupancy during Benchmark Execution. ..... 131
Effect of Pipeline Length on Processor Performance. ..... 132

## Abstract

As VLSI process technologies develop and feature sizes shrink, the global clocking schemes currently employed in synchronous systems are beginning to experience difficulties in a number of areas. Asynchronous circuits have a potentially higher performance than synchronous circuits since an asynchronous circuit exhibits average-case performance, in contrast to synchronous systems, which must be specifically designed to accommodate worst-case conditions. However, asynchronous design techniques are not widely understood or developed, particularly in the context of a large, complex system.

Recently, an asynchronous design methodology, namely Micropipelines, has been presented which has proved useful in developing an asynchronous CMOS implementation of an existing commercial RISC architecture. A subsequent project has been initiated to develop architectural modelling and implementation tools for an asynchronous highperformance bipolar implementation of the same target architecture.

This thesis presents the issues involved in asynchronous logic design, the details of the particular asynchronous design methodology employed and an introduction to the architectural modelling environment used in the development of the bipolar asynchronous implementation. The development of the system model is illustrated, with reference to the underlying primitive components and the hierarchical composition of the complete design from asynchronous sub-functions communicating via a well-defined signalling protocol. A demonstration of how the architectural model can be used to generate information regarding the internal operation of the system, which is then used to improve the complete design is given. The suitability of modelling asynchronous systems with the modelling environment employed is discussed.

## Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification at this or any other university or institute of learning.

Copyright in text of this thesis rests with the Author. Copies (by any process) either in full, or of extracts, may be made only in accordance with instructions given by the Author and lodged in the John Rylands University Library of Manchester. Details may be obtained from the Librarian. This page must form part of any such copies made. Further copies (by any process) of copies made in accordance with such instructions may not be made without the permission (in writing) of the Author.

The ownership of any intellectual property rights which may be described in this thesis is vested in the University of Manchester, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the University, which will prescribe the terms and conditions of any such agreement.

Further information on the conditions under which disclosures and exploitation may take place is available from the Head of Department of Computer Science.

## Preface

The author was employed in the motor and light engineering industries before returning to full-time study in 1987. He graduated with a B.Sc. (Hons.) Degree in Computer Engineering from the University of Manchester in 1990. The author was previously employed as a Research Associate on the ESPRIT-funded EDS project, being involved in operating system kernel developments for a message-passing multiprocessor architecture. He is currently employed as a Research Associate in the AMULET group at the University of Manchester on the Transforming Architectural Models (TAM-ARM) project.

## Acknowledgements

The work presented here was carried out as part of the TAM-ARM project funded within the Design Automation section of the DTI/SERC Advanced Technology Programme. The author gratefully acknowledges this support and the technical assistance provided by the industrial project partners: Advanced RISC Machines (ARM) Limited and GEC Plessey Semiconductors.

I would like to express my sincere thanks to my project supervisor Dr. L.E.M. Brackenbury, for her support and encouragement. I would also like to thank members of the original AMULET1 design team: Prof. Steve Furber, Nigel Paver, Paul Day and Jim Garside for revealing the innermost secrets of their creation.

The AMULET research group in the Computer Science department of the University of Manchester provided a stimulating and interesting environment in which to contemplate Life, the Universe and Asynchronous Processor Design - to the remaining group members, both staff and students, I am indebted to you for making it so.

Finally, to my wife, Kathryn, and daughters, Leah and Laura, who have endured so much over the past few months without complaint (well, almost!), I am eternally grateful.

## 1. Introduction

"By combining advances in integrated circuit technology, improvements in compiler design and new architectural ideas, significant performance improvements have been realised in the contemporary design of computer systems. These improvements have only been made possible by bringing together important technological advances with a better empirical understanding of how computers are used. From this fusion has emerged a style of computer design based on empirical data, experimentation and simulation."

These ideas, drawn from probably the most important text on computer design over the past decade - Computer Architecture: A Quantitative Approach [Henn90] - indicate the considerable benefits of producing a model of any proposed prototype system. The model should be capable of being exercised with a realistic workload to provide performance indicators and to enable the effects of design decisions to be explored.

The work presented in this thesis is concerned with the architectural modelling of an Asynchronous Bipolar Microprocessor. The prototype processor design is derived from AMULET1 [Furb94a], an asynchronous CMOS version of the ARM RISC microprocessor. Although the AMULET1 architecture was not the first asynchronous microprocessor [Mart89], it is the first to overcome the difficult implementation areas of handling interrupts and exact exceptions, and providing multi-cycle instruction support. The implementation technology is based on a high-performance, differential, current-mode logic family developed by GEC-Plessey Semiconductors. As outlined above, a simulation model is desirable before implementing a prototype system. When, as is the case
with this work, a novel system architecture produced using an unfamiliar design methodology is to be implemented on a new advanced bipolar process, then extensive simulation becomes essential.

There are several objectives of this thesis. The first is to introduce the reader to the issues involved in asynchronous logic design in Chapter 2 and to the specific asynchronous design style used for the project in Chapter 3. The second is to familiarise the reader with the chosen system modelling language in Chapter 4 and the high-performance bipolar technology used to implement the prototype system in Chapter 5. The next objective is to show how the system model components are constructed based on the circuit characteristics of the underlying implementation technology. This is presented in Chapter 6. A further objective, achieved in Chapter 7, is to introduce the ARM architecture and explain the operation of the asynchronous implementation. The final objective is to show how the modelling environment is used to incorporate the design methodology and to demonstrate how the information produced by the model may be used to improve the design of the system in Chapter 8.

The structure of the remainder of this thesis is as follows:

Chapter 2 explains the domination of synchronous design techniques in current electronic circuit synthesis. The problems with synchronous design, which are generating renewed interest in asynchronous design styles, are noted. An introduction to asynchronous logic is given, along with the signalling protocols used, and some of the issues involved in delay modelling are considered.

Chapter 3 gives an introduction to the particular asynchronous design methodology used in the development of the Asynchronous Bipolar microprocessor. Examples of the control circuit elements used are included.

Chapter 4 presents the modelling environment and demonstrates some of the language constructs and the hierarchical structure capabilities. An indication of how time is managed while exercising the model is given.

Chapter 5 introduces the differential bipolar technology employed to implement the prototype system. The circuit operation is explained by considering some gate function examples.

Chapter 6 shows how the architectural models of the bipolar logic gates and functions are developed based on the circuit simulations of the equivalent transistor models of the basic gates. The effects of gate output loading and input drive characteristics are explored.

Chapter 7 outlines the ARM target architecture and the instruction set. The structure of the asynchronous bipolar architecture is then presented with detailed examination of the major functional units, namely the Register Bank, Address Interface, Data Interface and Execution Unit. Simulator output waveforms are included to demonstrate the operation of the units.

Chapter 8 illustrates how the architectural model of the asynchronous ARM was developed in the modelling environment using a hierarchical, modular structure. Some of the features of the modelling language are then elaborated and some examples of the modelling tools that have been constructed are demonstrated. Various executable programs used to validate the architecture and measure performance are presented. An illustration of how the system model is used to gain information regarding the operation of the design and subsequently, how this information is used to suggest system design enhancements is given.

Chapter 9 summarises the current state of the project and draws together the conclusions resulting from this work, discussing the applicability of the Verilog to the architectural modelling of asynchronous systems. Future research areas, continuing on from this work are suggested.

The Appendix contains the complete hierarchical Verilog model of the MDCML Asynchronous ARM including the functional subsystems, asynchronous control elements and standard logic gate primitives.

## 2. Asynchronous Logic

Computer technology has evolved rapidly over the past few decades and the demand for even higher performance machines seems set to continue as computing solutions to new, more complex and computationally intensive problems emerge.

Synchronous design techniques have dominated the field of digital logic synthesis during this development period. This supremacy has been brought about for several reasons:

The concepts required to create a synchronous solution to the production of a logic circuit are easily understood - the designer simply defines the combinatorial logic necessary to perform the required function and then surrounds it with latches which are enabled with a common clock. In a large design, the entire system is then a composition of subsystems communicating by passing data values between the clock-controlled registers.

The global clock fulfils two system functions - the clock transitions define the successive instants at which the system state changes can occur and the clock period is sufficient to account for the logic and wire delays. Since the clock period is specified to be greater than the slowest combinatorial path that could occur during the computation, circuit hazards and feedback problems can generally be ignored [Seit80].

- By neglecting the effects of clock skew - the time difference between the arrival of the global clock signal at different points in the system - the total system state, when considered at the end of the clock period, is assumed to be deterministic and discrete, changing only at the edges of the system clock.

The synchronous design style is well-understood and formalised and is therefore readily accessible to potential digital logic engineers, the preponderance of synchronous circuits is then reinforced when these new engineers become productive.

Also, widely available standard components, which are well-specified and documented, have been positively developed for use in the synchronous style.

Verification of the correct operation of a synchronous design simply involves checking the setup and hold times of the outputs of the combinatorial logic sections of the design to ensure that they meet the requirements of the clocked registers.

CAD tool support has also been developed, in parallel with the synchronous design concepts, which manage much of the timing verification involved.

- Testing is also a much easier proposition in a synchronous circuit since many techniques including, for example, Scan Paths and BILBO (Built-In Logic Block Observation) are well-developed.

Recently, however, significant interest has arisen in the field of asynchronous logic design. This interest may be as a consequence of the problems associated with the global clocking strategy becoming more acute, a recognition that the formal techniques for handling asynchronous behaviour and the automatic synthesis potential of asynchronous circuits are now worth exploiting, or that inspiration has been generated by recent publications in the field, most notably the 1988 Turing Award Lecture on MICROPIPELINES given by Ivan Sutherland [Suth89].

As VLSI process technologies develop and feature sizes shrink, the global clocking schemes currently employed in synchronous systems are now beginning to experience severe difficulties in the following areas:

Since the clock signal controls the entire system, it must be distributed across the entire chip. This requirement for large scale clock driver circuitry is expensive - in current high performance microprocessors a considerable proportion of the silicon area used and power dissipation required is given over to the global clock logic [Dobb92].

The design effort needed for the clock driver circuitry, and consideration of the effects of clock skew, is non-trivial. It is becoming increasingly difficult to maintain the clock skew within reasonable bounds across all process, temperature and circuit operational speed parameters and may result in the clock period being extended. In current leading-edge synchronous microprocessor designs, a significant proportion of the clock period is used to account for the effects of clock skew.

The circuit modifications required when a relatively small subsection of the system is changed may have ramifications across the entire chip design.

The global clock period must allow for the worst-case logic delay, even though, if the system is not operated in an extreme environment, the worst-case delay may never actually occur. The resulting performance is then reduced as the system is effectively idle during the time between the outputs of the combinatorial logic settling and the arrival time of the (worst-case period) clock.

It has long been recognised by logic circuit designers that asynchronous circuits have a potentially higher performance that synchronous circuits, since an asynchronous circuit exhibits average-case performance (the processing commences as soon as the new input data arrives - the time required to complete the computation execution being dependent on the actual input data values). A synchronous ALU, for example, must be particularly designed to allow for worst-case execution time irrespective of the actual input data values presented to the circuit.

In general, arbitration is required when several sources compete for the same service (or resource), since the service request signals may arrive at the shared resource at any time. In a synchronous system, asynchronous inputs are synchronized to a local clock, allowing metastable effects to be (hopefully) resolved in a limited period. An asynchronous system can wait an arbitrary time for arbitration to occur before making a clear decision. As a result arbitration is inherently more robust and reliable.

However, the asynchronous design framework is unfamiliar to established engineers. The basic 'building blocks' of asynchronous logic synthesis need to be developed, since
currently the components are unfamiliar and unoptimised. Also, the circuit size of an asynchronous solution, relative to the equivalent synchronous design, is possibly increased (in part, due to the unoptimised basic components); although this may be offset by the non-trivial requirement for clock-driver circuitry for larger systems.

The existence of circuit races or hazards causes a further complication in an asynchronous design. Fundamental Mode operation ${ }^{1}$ must be employed, or various assumptions must be made regarding the relative delays or speeds of the circuit component elements. The testing of asynchronous circuits also causes problems. Sequential circuits are very difficult to test and techniques have not yet been fully developed to test asynchronous combinatorial logic. No high-level method has yet been produced to assist in checking the liveness (absence of deadlock) of a design.

Several methodologies have been developed to synthesize asynchronous circuits, some are based on enhancements to Petri nets [Pete81, Moln92], others are compilationbased on high-level languages [Mart90, Brun91] developed from CSP [Hoar85]. Surveys of asynchronous design methodologies and techniques can be found in [Gopa90, Hauc93]. Some of the asynchronous design terminology that may be encountered in the text will now be explained. This relates to the modelling of signal propagation delays and the mechanisms used for communication between asynchronous subsystems.

### 2.1 Delay Modelling

In a BOUNDED DELAY model, it is assumed that the delays in the circuit elements and wires are known (or at least have some upper bound). When input signals are applied to a circuit, then after a particular time interval has elapsed the output signals are known to be valid. Note that this is also the delay model used for synchronous designs.

DELAY-INSENSITIVE circuits use a contrasting model to that used in bounded delay circuits; it is assumed that all signal delays in both elements and wires are unbounded.

[^0] change.

No matter how long the circuit waits, there is no guarantee that an input signal will be received. Circuits designed in this style must include functions to detect when a new input value actually arrives.

The SPEED-INDEPENDENT model is a weaker form of the delay-insensitive paradigm, in that it is assumed that the element delays are unbounded but the interconnection wires have zero delay.

### 2.2 Signalling Protocols

Communication between modules or subsystems in an asynchronous environment is achieved by employing a commonly agreed set of control signals (and some associated operational rules) which are passed between adjacent modules. The method usually involves detecting an 'event' on the control signals, eg. a change in the voltage level of the interconnecting wire.

In order to construct asynchronous systems by the composition of individual subsystems, where each performs a specific (and different) function, a general signalling protocol is required. This protocol will operate between the various modules without any regard to the internal processing rates of individual modules, or of the actual signal propagation delays of the communication links. This can be achieved by placing no restrictions on the timing of the signals involved in the communication protocol. Only the sequence of the control signal transitions is significant.

The basis for some of the simplest protocols involves the use of two wires connected between adjacent modules: a REQUEST wire and an ACKNOWLEDGE wire.

Asynchronous systems usually employ one of two communication protocols: twophase (or transition) signalling or four-phase signalling.

### 2.2.1 Two-Phase Signalling

In this protocol, any transition between the two logic levels, a HIGH to LOW transition or a LOW to HIGH transition, is accorded the same meaning.

A transition may also be referred to as an EVENT, hence the alternative name for this protocol is 'event signalling'.

Two-phase signalling operates between two modules in the following manner:


Figure 1 : Two-phase (transition) Signalling.

The sender generates an event (transition) on the REQUEST wire. At some point in time later, the receiver detects the request transition and indicates that it has received the request signal by generating an event on the ACKNOWLEDGE wire. The sender eventually receives the acknowledge event, signifying that the receiver is ready to receive another request.

The arrows on the diagram indicate the constraints on the sequence of events allowed on the control signals used in the protocol. The THICK arrows show the constraint imposed by the receiver: an acknowledge event cannot be generated until a request has arrived. The THIN arrows show the constraint imposed by the sender: the sender cannot generate another request until the previous request has been acknowledged by the receiver. The fact that each of the modules regulates the sequencing of one of the control signals indicates that the correct operation of the inter-module communication path will only occur when both sender and receiver obey the protocol rules.

Because BOTH edges are used in the two-phase scheme - the actual logic LEVEL of a particular control signal does not assume any significance - it provides the capability of
increasing the performance of communication protocols above that of conventional signalling methods, since every change in the signal carries some information content.

Initially, the concepts of transition signalling may be difficult to assimilate into the mindset of the conventional logic designer, since the two-phase circuits must be symmetrical with respect to the high and low logic levels of the control signals.

### 2.2.2 Four-Phase Signalling

Four-phase (or 'Return to Zero') signalling is characterised by the control signals being active when in the HIGH (logic ' 1 ') state and then being required to return to the LOW (logic ' 0 ') state before subsequently becoming active again.

The protocol could take the following form:


Figure 2 : Four-phase Signalling (incorrect operation).

The sending module raises the REQUEST line to its HIGH (active) state and after a short time interval deactivates the signal by taking it LOW again. The receiver, having detected the request line entering its active state, produces a response by briefly raising the ACKNOWLEDGE line to its HIGH state.

However, the protocol in this form may result in communication failure since, if the sender has a comparatively faster circuit operation than the receiver, the sender may raise then quickly lower the request line to produce a very narrow 'pulse' which the receiver may be unable to detect.

Correct protocol operation is enforced by requiring the sender to continue holding the request line in its active (HIGH) state until 'request reception' is indicated to the sender by the receiver raising the acknowledge line into its HIGH state. The request line is then deactivated, allowing the receiver to subsequently deactivate the acknowledge line.

The properly functioning four-phase protocol is then:


Figure 3 : Four-phase Signalling (correct operation).

Again, each of the modules taking part in the communication imposes constraints on the sequencing of the control signal transitions. The THICK arrows indicate the constraints enforced by the receiver: the acknowledge line can only enter its active state after the request line is activated and can only be deactivated after the request line is deactivated. Similarly, the THIN arrows show the constraints imposed by the sender: the sender must not 'remove' the request signal until the receiver acknowledges that it has 'seen' the request and a subsequent request must not be generated until the acknowledge has entered its inactive state.

The four phases of the protocol can be observed by noting the four possible combinations of the control signals:

Request LOW, Acknowledge LOW - Inactive

Request HIGH, Acknowledge LOW - Requesting

Request HIGH, Acknowledge HIGH - Acknowledged

Request LOW, Acknowledge HIGH - Request cleared, Acknowledge to clear

Four-phase signalling may be more familiar to current logic designers since each phase of the protocol may easily be determined by examining the logic LEVELS of the control signals.

Also, four-phase signalling is easier to implement because of the widely-available standard components which have been developed to manage logic levels.

### 2.2.3 Data Communications

In addition to the signalling protocols used to indicate control actions, outlined above, a mechanism for passing data values between modules is required.

The simple REQ/ACK scheme can only signal events. In order to transmit data values a method of differentiating between two alternative events (sending a ' 1 ' and sending a ' 0 ') must be employed. This method could be extended, by using two sets of REQ/ ACK pairs, into a four-wire per bit signalling system where each pair is used to communicate a particular bit value: Req0/Ack0 is used to send and acknowledge zeros, Req1/Ack1 is similarly used for ones. In the simplest system, consisting of only four wires, multiple bit values, bytes or words, are sent in bit-serial fashion.

The number of wires required, per bit, may be reduced to three by noting that the two acknowledge wires Ack0 and Ack1 may be combined into one common acknowledge wire, Ack.

This idea of a common acknowledge wire can be used for the communication of multiple bit 'words'. Two request wires, R0 and R1, are provided for the transmission of each bit (a technique also known as DUAL-RAIL ENCODING) and the common word acknowledge signal is returned only when a transition has occurred on one of the request wires for each of the bits of the word.

### 2.2.4 Bundled-Data Interface

Although the previous schemes provide a robust communication technique in an environment where signal propagation delays are unpredictable, the cost in terms of number of signal wires needed is high. This is especially the case when the communication is over a relatively long distance. There is also a cost in terms of the signal detection/completion circuitry required when dealing with multiple bit data 'words'.

The bundled data interface seeks to significantly reduce the number of signal wires, particularly for large bit-width data values, to just one data wire per bit (as in conventional synchronous 'bus' structures). This set of signal wires is collectively known as a BUNDLE. For each wire, the logic level indicates the value. In addition, a request/acknowledge pair of control wires is needed per data word.

Assuming that two-phase (transition) signalling is used on the req/ack control wires, the data values are transmitted in the following manner:


Figure 4 : Bundled Data Interface.

The sender places the n -bit value onto the data wires (bus) and then generates an event on the REQUEST line. At some later time, the receiver will detect the arrival of the request event which will indicate that the data bus is holding the correct transmitted value. The receiver will then latch the data value before generating an acknowledge event back to the sender. The sender is then free to remove the current data value and set up the next value for transmission.

Note that for correct operation, there is an implied assumption that the data value arrives at the receiver before the request event i.e. in the same order as they were generated by the sender. More formally - "The sequence relationships of events in a bundle are the same at the sender and the receiver." [Suth86]. This is a timing constraint on the use of the bundled-data interface and the logic circuit designer must ensure that this timing relationship is satisfied.

Also note that the sender may not change the data value, once it has generated a request, until it receives an acknowledgment from the receiver. From the point of view of the receiver, the data is only valid from the time of reception of the request event until the acknowledge is generated.


Figure 5 : Data Value Constraints.

## 3. Micropipelines

Pipelining is used in computer architectures to provide increased processing rates through the use of concurrency [Kogg82]. A large computation is divided into a series of operations, which execute in parallel.

Pipelines may be clocked (synchronous) or event-driven (asynchronous). In both synchronous and asynchronous pipelines, the throughput - the number of data items processed in a given time interval - is limited by the computational rate of the slowest subsystem (module) in the pipeline. However, the latency - the time taken for an individual data item to pass through the complete (empty) pipeline - of the synchronous and asynchronous pipelines differs.

For the synchronous case, the latency is calculated to be the number of pipeline stages multiplied by the processing time of the slowest element; the clock period must be specified to accommodate the slowest element, even though all other elements may be capable of sustaining much higher clocking rates.

The latency of an asynchronous pipeline is calculated to be the sum of the processing times of each element. This latency can be significantly less than that of the synchronous case if there is a wide range of processing rates for the component elements.

The lower latency of asynchronous pipelines may be exploited in, for example, processor instruction execution pipelines where the pipeline is frequently flushed when a branch instruction is executed.

Pipelines may also be categorised as ELASTIC or INELASTIC. For an inelastic pipeline, the input and output data rates must match, implying that the total amount of data contained in the pipeline is fixed. When an inelastic pipeline contains no processing elements, i.e. each stage consists of a storage element only, it acts like a simple SHIFT REGISTER. In contrast, the input and output rates of an elastic pipeline, however, may differ momentarily and therefore the amount of contained data is variable. An elastic pipeline without processing elements is a FIFO (First-In, First-Out).

FIFOs provide an important buffering function between systems acting at variable processing rates. The implementation of elastic FIFOs is difficult in a synchronous model: each stage must have a full/empty flip-flop, and each stage must be provided with full/empty information about the previous and successor stages. A particular stage receives a new data value if, at the appropriate clock transition, the stage is EMPTY and the previous stage was FULL. The stage can then pass on the data value if, at a subsequent clock transition, the next stage is EMPTY. The current stage can then make itself available to receive a new data value by changing its state flip-flop from FULL to EMPTY.

Also, since the clocking rates at the input and output of an elastic pipeline may be different, some form of arbitration and synchronisation will need to be provided between the FIFO and the systems connected to it. An asynchronous implementation removes the requirement for arbitration by allowing the input and output processes of each stage of the FIFO to operate at their own pace.

In 1988, at the Turing Award Lecture, Ivan Sutherland put forward a modular approach to the design of computer systems using asynchronous logic. His idea was based on the use of simple data processing pipelines whose stages operate asynchronously. He termed these 'MICROPIPELINES' [Suth89]. A micropipeline is an elastic, boundeddelay, event-driven system using transition signalling and the bundled-data interface. A micropipeline without processing, a simple elastic FIFO, can be constructed from a basic component known as an Event Register (see Section 3.2.1).

One of the benefits of micropipelines is that the registers, used to hold the data values as they flow through the pipeline, can be used to filter out hazards. This is achieved, in a similar manner to that used in synchronous designs (where the clock period is sufficiently long to account for hazards), by locally delaying the request output signal until all data values are stable. Also, any combinational logic structures can be used 'between' the pipeline registers including existing circuits used in synchronous designs.

As discussed in Chapter 2, the bundled-data interface is useful for data communications since it reduces the number of data wires required to transmit a value, particularly for large numbers of bits. In implementing an asynchronous 32 -bit processor, for example, the inter-module completion-detection circuitry required if the data was transmitted using the ' 2 wires per bit' protocols would be prohibitively large.

Micropipelines offer the opportunity to construct complex systems by the hierarchical composition of simpler modules. The two-phase signalling protocol allows modules of widely-differing performance to be easily integrated into a complete, correctly functioning, system. The data-driven execution rates of the individual asynchronous modules allow the benefits of average-case performance. The micropipeline approach also provides the facility to replace a particular module with one of a higher performance without impacting on the correct operation of the total system (as would be the likely case with a synchronous global clocking scheme).

In the context of VLSI technology, the design cost of large systems both in terms of time and effort is beginning to outweigh the combined fabrication and production costs of the final integrated circuits. Since an 'ad hoc' design style is impractical for large scale circuits, micropipelines provide a basis for an asynchronous design methodology for the construction of such systems. Pre-synthesised modular solutions to standard problems, packaged in an asynchronous design library, can then be interconnected using the transition signalling protocol.

The circuit designer simply ensures that each module conforms to the interface protocol and need not be fully conversant with the internal intricacies of the asynchronous 'cells'. The inefficiencies of this 'standard module' approach may be negligible when compared to the extra design cost of a full custom approach.

### 3.1 Control Circuit Elements

The transition signalling control circuits used to coordinate the activities of micropipelines may be constructed from a standard set of 'event logic' modules [Suth86]. Some of the more widely-used modules are presented below.

### 3.1.1 XOR (Merge)



The Exclusive-OR (or non-equivalence) gate provides an 'OR' function for transition signals. An output transition (event) occurs in response to a transition arriving at any of the inputs. This module is also known as a MERGE element.

### 3.1.2 Muller-C (Join)



The Muller-C element serves as an 'AND' function for events. A transition occurs on its output only after a transition has occurred on each of its inputs. In logic level terms, when the input levels match, the output assumes the same logic level as the inputs, otherwise the output holds its previous level. A reset input may be added to force the output to a defined initial state. The standard AND logic symbol with a large ' C ' inside is used to represent the Muller-C element, which is also known as a JOIN or RENDEZVOUS element.

### 3.1.3 Select



The Select element 'steers' an input transition to one of the two outputs depending on the Boolean value of a second, 'select', input. The select Boolean value must be valid before the input transition occurs. This is, effectively, a bundling constraint on the IN (event) input. The module is NOT delay-insensitive because of this requirement. Furthermore, while the Select element is essentially an event-triggered device, the logic level of the select input is significant.

### 3.1.4 Toggle



RESET

In a similar manner to the Select element above, an input transition of the Toggle element is steered to one of the two outputs. However, the output event is produced alternately on the two outputs in response to an input transition. Following a Reset signal, the first output to receive an event in response to an input event is marked with a heavy dot (see diagram), the outputs are then known as Dot and Blank (i.e. no dot).

### 3.1.5 Decision-Wait



A Decision-Wait element has two sets of input signals and produces an output event when one event in each input 'set' has been received. For example, an event on input Y and an event on either X1 or X 2 will produce an output event on either Q1 or Q2 respectively. Note that for correct operation, only one input event can be received on an X input ( X 1 or X 2 ) for each event received on the Y input before the appropriate output transition occurs.

### 3.1.6 Arbiter



An Arbiter is used to guarantee mutually exclusive access to a shared resource for two competing independent requests. The arbiter chooses only one of the active input requests and allows only the corresponding output grant signal event to occur. When the arbiter is already in use by a requester, a second requester is inhibited until the "request done" acknowledge event is received (DONE1 or DONE2, depending on the currently active requester) indicating that the active requester is releasing control of the arbiter. The arbiter will then issue a grant signal (event) to the second requester.

Although the input requests can occur at any time, even simultaneously, the output grant signals are guaranteed to be mutually exclusive or serialized.

### 3.1.7 Call



The Call element provides the functional equivalent of a "subroutine call". A request event for access to a common hardware function is received on one of the two request inputs, R1 or R2, which will subsequently generate a 'subroutine' request event on the Rsub output. When the subroutine function has completed, indicated by the arrival of an event on the Dsub (subroutine done) input, the Call element generates an output event on the appropriate 'request done' output (D1 or D2, depending on the active requester).

For correct operation, the full Request / Subroutine Request / Subroutine Done / Done cycle must complete before the next Request occurs and therefore the two input request signals, R1 and R2, must be mutually exclusive. For a circuit topology where R1 and R2 cannot be guaranteed to be mutually exclusive, the input requests may be routed to the Call element via an Arbiter.


### 3.1.8 Capture-Pass Latch



The Capture-Pass latch is a storage element for use in an event-controlled system; unlike a traditional level-sensitive latch in which the high and low states of the clock/ enable signal indicate a different function, the event-controlled latch must provide equivalent responses to rising and falling transitions.

When the Capture and Pass inputs are in the same state (either both high or both low), the latch is in the PASS state: the output of the latch follows any change in the input value. When the Capture 'event' occurs, the latch will become insensitive to changes in the input data and will CAPTURE (store) the current input value, resulting in the output value being held stable. After a subsequent Pass 'event', the element will become transparent and the output will again follow the input.

For the Capture-Pass latch to operate correctly, the Capture and Pass events must alternate.

### 3.2 Control Circuit Examples

### 3.2.1 Event Register

As mentioned previously, an asynchronous FIFO can be constructed from a basic component known as an Event Register.

An Event Register uses a two-phase signalling protocol on its input and output control circuits and incorporates an event-controlled storage element for the associated bun-dled-data value.


Figure 6 : Event Register.

Event Registers with 32-bit data values are used extensively throughout the Asynchronous ARM design.

The operation of the Event Register is as follows:
i) Initially, assume all signals are $L O W$ and the Capture-Pass latch is in the PASS (transparent) state.
ii) An input data value is supplied followed by the arrival of a ReqIN event at the Muller-C and, because of the input inversion of the $L O W$ state of the other input, an output event is generated from the Muller-C.
iii) The Muller-C output event causes the Capture-Pass latch to enter the CAPTURE state: it latches the data value presented on its input.
iv) Once the Capture-Pass latch has captured the data, an AckIN event is sent to the 'previous' stage (the previous stage can now prepare a new
input data value) and a ReqOUT event is sent to the 'next' stage of the pipeline.

Note that the Event Register will only capture the data in response to a ReqIN event if the stage is currently EMPTY. That is, if an acknowledge from the output stage is not pending from a previously generated ReqOUT event to the next stage.

Also, the Capture-Pass latch must be in the PASS state (Dout valid) before a Capture event occurs and, since the ReqOUT event is equivalent to the Capture event, the Dout data value must be valid before the ReqOUT event. The output interface of the Event Register therefore obeys the data bundling constraint.

### 3.2.2 Design Example: PARITY FUNCTION

A dual-rail encoded parity function using a transition signalling protocol was presented to the IFIP Working Conference on Asynchronous Design Methodologies (April 1993, Manchester) by Charles Molnar and this will be used as a design example:


The circuit will receive an input signal as an event on one of the two input wires, $\mathrm{I}_{0}$ or $\mathrm{I}_{1}$, depending on whether a ' 1 ' or a ' 0 ' is indicated. The circuit must then provide an output event on one of the two output wires, P 0 or P 1 , to indicate the cumulative parity of all of the inputs received up to that point.

A general, high-level, formal method of synthesizing Micropipeline control circuits does not, as yet, exist. A pragmatic approach must therefore be taken to derive a design for the parity function circuit using the Micropipeline control blocks outlined previously in Section 3.1.

Assume that after a global reset, all the interface signals (inputs and outputs) of the parity function are $L O W$.

It can be noted that the accumulated parity of the received inputs is actually given by the logic level of the $\mathrm{I}_{1}$ input. If $\mathrm{I}_{1}$ is $H I G H$, the accumulated parity is ' 1 ', if $\mathrm{I}_{1}$ is $L O W$, the parity is ' 0 '.

Events occurring on the $I_{1}$ input cause alternating output parity signals to occur, since an $I_{1}$ event indicates the arrival of a ' 1 ' and this will cause the accumulated parity to change. A TOGGLE element can be used to alternate the parity output events when the $\mathrm{I}_{1}$ input event indicates another ' 1 ' has been received.

Events occurring on $\mathrm{I}_{0}$ (indicating a ' 0 ' has arrived) cause an output event on $\mathrm{P}_{1}$ or $\mathrm{P}_{0}$ depending on the current accumulated parity value. That is, the output to be activated is indicated by the logic level of the $\mathrm{I}_{1}$ input signal. A SELECT block can be employed to 'steer' the $\mathrm{I}_{0}$ input event to the appropriate parity function output based on the logic level of $I_{1}$.

Two XOR elements are used to merge each of the separate sources of the $\mathrm{P}_{1}$ and $\mathrm{P}_{0}$ output events (from the Toggle and Select blocks) onto the actual parity function outputs. The Micropipeline control block implementation ${ }^{1}$ is shown in Figure 7.


Figure 7 : Micropipeline Control Block Implementation of Parity Function.

[^1]
## 4. Verilog HDL

### 4.1 Introduction to HDLs

When viewed at its lowest level, a digital system, particularly in the context of a VLSI implementation, may consist of several hundreds of thousands of primitive components. These components may be transistors or simple logic gates. At a higher level, these elements may be logically grouped into functional units such as Arithmetic Logic Units (ALUs), cache memories and Floating Point Units (FPUs) [Thom92].

Hardware Description Languages (HDLs) have been developed to assist the design process of such systems in managing the complexity involved in the synthesis of complex digital systems [Hart87]. The system may contain a large number of elements and a wide range of logical and physical implementation abstractions, in order to give a total overview of the system.

Initially, a conceptual idea of the required logical system is coupled with a set of constraints (relating to performance, power requirements, circuit size etc.) that the implemented system must meet and a set of primitive components from which to construct the system. The creative design process is an iterative operation of either manual composition or automatic synthesis of alternative solutions, which are then compared against the given system constraints. Normally, the design is partitioned into smaller sub-units, in the classical engineering technique of "divide and conquer" (or top-down design), and each sub-unit is then further divided until the complete system is specified in terms of the known primitive components [Brow91].

### 4.2 Introduction to Verilog

The Verilog ${ }^{1}$ Hardware Description Language [Veri92] describes a digital logic system as a collection of textual-based models that define the functionality of the component sub-units and connections to those sub-units. The language accommodates a wide range of levels of abstraction:

ALGORITHMIC - the component's operation is expressed in high-level (program-like) language constructs.

REGISTER TRANSFER LEVEL (RTL) - the flow of data between registers is described.

GATE LEVEL - the system is defined in terms of logic gate primitives and their interconnections.

SWITCH LEVEL - for low-level design, particularly for MOS implementation, the system may be described in terms of transistors and storage nodes.

The language supports the early conceptual stages of design with its behavioural levels of abstraction (algorithmic and RTL), and the later implementation stages with its structural levels of abstraction (gate and switch levels).

During the design process, behavioural and structural constructs can be mixed as each of the sub-systems is designed. Hierarchical constructs are also provided to allow the system designer to control the complexity of the description.

### 4.3 Modules

Verilog describes a digital system in the form of a set of MODULES. The logical structure of each module is expressed either in logic gate (or MOS primitives) terms or as a behavioural representation.

[^2]A module definition includes declarations of the external interface presented to other modules and any internal state used by the module. The external interface is defined in terms of PORTS, which are specified in parentheses after the module name. Ports may be declared to be INPUTS, OUTPUTS or bidirectional INOUTS. A module body contains either behavioural statements which specify the functionality of the module, or statements which create instances of other user-defined modules or logic gate primitives. By allowing module definitions to instantiate other modules, a hierarchical description of the system can be specified. The use of a hierarchical modular approach accommodates the "bottom-up" and "top-down" design styles.

### 4.4 Structural Modelling

A structural representation of a functional unit is achieved using gate and/or switch level modelling. A set of 26 standard gate-level primitives are incorporated and these can be extended by employing user-defined primitives. This provides a compact and efficient way of describing an arbitrary block of logic.

The Verilog HDL facilitates the accurate modelling of signal contention, bidirectional pass gates, resistive MOS devices, dynamic MOS, charge sharing and other technology dependent network configurations by allowing net signal values to have a wide range of unknown values and different levels of drive strengths.

A declaration begins with the gatetype keyword specifying the required gate or switch primitive. Gatetype keywords include: and, or, not, buf, nmos, pmos, pullup etc. Gate and switch instances include an optional instance name and a required terminal connection list.

The propagation delay from input to output through a logic gate or switch primitive may be specified in a declaration. The drive strengths on the output terminals of a gate declaration instance may also be defined.
'Nets' are a fundamental data type of the language and are used to model an electrical connection. Except for the trireg net, which models a wire as a capacitor that holds electrical charge, nets do not store signal values. Nets only transmit values that are driven on them by structural elements (gate outputs or assign statements) or behavioural models (registers).

### 4.4.1 Design Example: RS Flip-Flop

An RS flip-flop consists of two inputs, SET and RESET, and (normally) two outputs, Q and $\overline{\mathrm{Q}}$. For the purposes of this design example, the $\overline{\mathrm{Q}}$ signal will not be generated as a module output. All the signals, inputs and output, will be active $H I G H$ i.e. positive logic.

When the SET input is asserted (HIGH), the Q output signal goes HIGH and remains HIGH even when the SET input is deactivated. When the RESET input is asserted, the Q output goes LOW and again stays LOW when the RESET signal is deasserted. A conflict will occur if the SET and RESET inputs are both HIGH simultaneously. The logic circuit designer should ensure that this situation never arises.

A circuit implementation consists of two cross-coupled NOR gates [Mano84]:


Figure 8 : Possible Implementation of RS Flip-Flop.

An example module of a structural (gate-level) representation of a RS flip-flop is given overleaf. Each module definition begins with the module keyword and is terminated by
the endmodule statement. The first line of the definition specifies the module name and the names of its ports.

```
module RS_FF(set, reset, Q);
input set, reset;
output Q,
wire Qbar;
nor #10 g1(Q, reset, Qbar);(5)
nor #10 g2(Qbar, set, Q); (6)
endmodule
input set, reset;(3)
endmodule

In lines 2 and 3 in the example, the type of each port is specified: set and reset are input ports, Q is an output port. The module's logic gate primitive components are defined in lines 5 and 6. The first word in the line indicates the component type-name - in this case, nor gates. The \#10 indicates that the propagation delay of the gate from input to output is 10 time units. The nor gates are then instantiated by giving each one an instantiation name ( g 1 and g 2 ) and specifying the gate connections. The output is specified first, followed by any number of inputs (in this example two). There is a net (or wire) which is internal to the module, i.e. it is not an input or output, which connects the output of g2 to an input of g 1 . This internal net is declared and named in line 4.

\subsection*{4.5 Behavioural Modelling}

When a system is modelled as a structural, gate-level representation, very little translation effort is required to convert the HDL model into a correctly functioning physical implementation. However, in many cases the circuit engineer requires the opportunity to derive many design alternatives and consider the merits of each design solution. Behavioural modelling facilitates the architectural refinement of a design. It allows the higher-level functional aspects of the prototype system to be easily evaluated in isolation, without regard to the final implementation of the proposed circuits [Russ89].

The syntax of the Verilog behavioural language is very similar to the high-level programming language ' C ' [Kern88]. It contains a number of procedural constructs which
include the familiar if-then-else conditional execution construct, the conditional assignment (?:) operator and the multi-way branch case statement. Four different statements are provided for iterative sequential behaviour: the for, while, repeat and forever loops. A full range of arithmetic, logical, bit-wise and reduction operators are also incorporated.

\subsection*{4.5.1 Compound Statements}

Two or more statements may be grouped together by means of a block statement so that, syntactically, they act like a single statement. In a SEQUENTIAL block, which is delimited by the keywords begin and end, the statements execute in sequence. Control passes out of the block when the last statement executes. The delay values used in the assignment statements are relative to the execution time of the previous statement:
```

begin
\#10 a = 1;
\#5 b = 0;
\#10 c = a;
end

```

In the example, register \(\mathbf{a}\) is assigned the value 1 ten time units after the execution of the block statement commences. A further five time units later, i.e. fifteen time units from the start of the block statement, register \(\mathbf{b}\) is assigned the value 0 . Register \(\mathbf{c}\) is then assigned the value of \(\mathbf{a}\) (now equal to 1 ) a further ten time units later.

The keywords fork and join surround a CONCURRENT block statement in which the individual statements execute in parallel. Delay values in assignment statements are relative to the simulation time on entry to the block and control passes out of the block when all of the statements have executed:
```

fork
\#10 a = 1;
\#15 b = 0;
\#25 c = a;
join

```

To achieve the equivalent effect to the sequential block statement, the assignment to register \(\mathbf{b}\) in the second line must have a delay value of fifteen time units, since the delay is relative to the start of the block (not relative to the previous statement, as in the sequential block example). In a similar manner, the assignment to register \(\mathbf{c}\) has a delay value of twenty-five time units.

\subsection*{4.5.2 Process Control}

The essence of a Verilog behavioural model is a PROCESS, which can be thought of as an independent flow of activity. The dynamic behaviour of a digital system is then a set of independent, inter-communicating processes. The basic Verilog control construct for describing a process is the always statement: always
<statement> // Continually repeats

The always construct continually repeats the statement following, which may be a block statement (outlined earlier). All of the functionality of a module should be specified within an always statement.

A further Verilog control construct, called the initial statement, describes a process that is executed only once - it provides a means of initialising signals and internal module state variables:
```

initial
begin
busy = `false; // Initialise values
out = 0;
end

```

During simulation of a model, all of the activity flows defined by the initial and always statements start together at simulation time zero.

\subsection*{4.5.3 Timing Control}

Two types of explicit timing control are provided in Verilog to regulate when procedural statements are to occur in simulation time. The first type is a delay control in which
a value expression specifies the time duration between the activity flow reaching a particular statement and the simulation time at which the statement actually executes. The second type of timing control is the event expression, which allows the execution of statements in a particular procedure to wait for the occurrence of some simulation event. The awaited simulation event will be generated by some other, concurrently-executing, procedure. A simulation event can be either the change of a value on a net, or in a register (an IMPLICIT event), or the occurrence of an explicitly named event that is triggered from other procedures (an EXPLICIT event). In many cases, the event control is the positive or negative edge of a clock signal.

Simulation time can only advance by one of the following three methods:

A delay control, which is introduced by the number symbol (\#):
eg. \#100 out = ~in;

After 100 time units, the output is defined to be the inverse of the input signal.
- An event control, which is introduced by the at symbol (@):
```

eg. always @ (negedge clock)

```
\[
\text { out }=\sim \text { in; }
\]

At every clock transition from HIGH to LOW, the output becomes the inverse of the input.

The wait statement, which operates like a combination of a while loop and an event control:
\[
\begin{aligned}
& \text { eg. wait (reset) } \\
& \qquad \text { out }=0 ;
\end{aligned}
\]

Suspend the process until the 'reset' signal is HIGH. When the reset signal does eventually go HIGH, the output signal is forced to zero.

\subsection*{4.5.4 Design Example: Behavioural Representation}
```

module RS_FF(set, reset, Q);
input set, reset;
output Q;
reg Q;
initial
Q = 0;
always @ (set or reset)
case ({set,reset})
2`b10: #10 Q = 1;         2`b01: \#10 Q = 0;
2`b11: begin
\$display("RS_FF: SET and RESET active"); (12)
\#10 Q = x;
end
endcase
endmodule

Again the module definition is enclosed in the module and endmodule keywords and, as in the structural representation, the ports and port types are declared in lines 2 and 3.

Line 4 declares a register with the same name as the output, Q , which will (implicitly) drive the output. Any value assigned to Q will be stored in the register and any value held in the register will be propagated to the output port. Registers are an abstraction of storage devices found in digital systems. Single-bit registers (like Q in the example) are termed scalar; multiple-bit registers are termed vector (eg. addr[31:0] is a 32 bit register).

The initial statement in line 5 is executed only once at the commencement of the simulation. This provides a mechanism for initialising the output value.

The always statement in line 7 is used to provide the dynamic functionality of the module. always @ (set or reset) indicates that the following statements should be executed whenever there is a change to one of the specified signals, i.e. the inputs. The case statement on the following line provides a decision capability based on the values of the SET and RESET inputs. Line 9 means that Q will be set HIGH, if the values of the SET and

RESET inputs, when concatenated together ( $\{$ set, reset \}), match the 2-bit Boolean value 10 (2‘b10). Basically, if the SET input is HIGH and the RESET input is LOW, then the output (register) Q will be set. Similarly, in line 10, if the SET input is LOW and the RESET input is HIGH, the output is driven LOW (reset).

Lines 11 to 14 indicate an important feature of the Verilog behavioural language, namely the ability to report diagnostic messages to the logic circuit designer while the simulation is running. As mentioned in the introduction to the design example, the SET and RESET inputs should never be active simultaneously. If this condition is detected, at line 11, the compound statements (enclosed in the begin and end keywords) on lines 12 and 13 are executed. An appropriate error message is displayed and the output value is set to undefined (x).

### 4.5.5 Programmable Logic Arrays

Verilog allows the modelling of both a synchronous and an asynchronous programmable logic array (PLA). The synchronous form allows the designer to control the simulation time at which the array will evaluate the inputs and update the outputs. For the asynchronous type, evaluation is performed automatically whenever an input term changes value.

PLAs are modelled using 2 orthogonal planes:


The logic equations of the separate planes are defined by loading individual data files containing the associated bit patterns.

### 4.6 Verilog Simulator

The Verilog description of the system may be simulated using a digital logic simulator. This is a software tool that allows many design process tasks to be carried out without the various costs involved in constructing a hardware prototype. These design activities may include [Russ85]:

- Functional Verification.
- Identification of design errors.

D Determination of the feasibility of new design ideas.

- Timing Analysis
- Evaluation of several approaches to a design problem.

The simulator exercises the system model by applying external input signal stimuli. Any generated register or gate output signal changes are then propagated to other gate and module inputs. The main characteristic of the simulator is the ability to manage the concept of time; causing the changed signal values to appear at some specified time in the future. These predicted signal changes are typically stored in a time-ordered event queue.

The RS flip-flop behavioural representation given in Section 4.5.4 is now used as an example to demonstrate the Verilog simulator operation.

In the top-level simulation test file (shown overleaf) the flip-flop module is instantiated (RS_FF) with the instantiation name, $f 1$. The input signal names are set and reset, and the output signal name is $Q$. In the first initial statement block, the input stimulus sequence is specified. In the second initial statement block, the required waveform output display is configured using the \$gr_waves() system task.

```
`timescale 1ps /1ps
module test();
reg set, reset;
wire Q;
RS_FF f1 (set, reset, Q);
initial
begin
    set = 0; reset = 0;
    #50 set = 1; #20 set = 0;
    #50 reset = 1; #20 reset = 0;
    #50 set = 1; reset = 1;
    #20 set = 0; reset = 0;
    #50 reset = 1; #20 reset = 0;
end
initial begin
$gr_waves("set", set
            ,"reset", reset
            "Q", Q );
    $freeze_waves; #340 $stop;
    $unfreeze_waves;
    $ps_waves("waves.ps", "RS_FF simulation example", 0, 330);
        #1 $finish;
end
endmodule
```

The console output text generated during the simulation execution is given below. Note the warning message displayed when the set and reset signals are simultaneously active:

```
VERILOG-XL 1.7 Jan 20, 1995 09:27:16
Compiling source file "test.v"
GRAPHICS 4.2.2 Thu May 27 23:28:23 PDT 1993 (cds2082)
Highest level modules:
test
RS_FF: SET and RESET active @ time=190
```

L29 "test.v": \$stop at simulation time 340
Type ? for help
C1 >
L32 "test.v": \$finish at simulation time 341
114 simulation events
End of VERILOG-XL 1.7 Jan 20, 1995 09:27:37

The graphical display waveforms can also be directed (using the \$ps_waves() system task) to a postscript file, which is shown below:

| Header: RS_FF simulation example |  |  |
| :--- | :--- | :--- |
| User: Robert Kelly | Time Scale From: 0 To: 330 | Page: 1 of 1 |
| Date: Dec 7,1994 09:52:46 |  |  |



## 5. Multi-Level Differential Current Mode Logic

### 5.1 Introduction to Logic Families

Integrated circuit technology has developed dramatically over the past few decades, both in terms of gate switching speed and sophisticated circuit design, as a result of fabrication process enhancements and shrinking minimum geometries.

The nature of the semiconductor product market tends to segment customers into two groups: performance-oriented users who seek leading-edge performance technology at virtually any cost, and cost-sensitive users who need the best performance available at a given price. Since semiconductor economies depend heavily on a volume market, it is the more numerous cost-sensitive users who tend to drive the development of mainstream semiconductor technology [John91].

Early integrated transistors were bipolar, since these were much easier to fabricate. This fact led to the market success of bipolar transistor logic families (DTL, RTL through to TTL) during the early years of IC development. Eventually, the development of the planar process led to the introduction of MOS logic families. Initially, because of the more sophisticated processing requirements of CMOS, NMOS logic dominated. However, as chip sizes increased, power consumption problems emerged and the additional complexity in producing CMOS (the lowest power MOS technology) circuits was justified. CMOS technology has now advanced to become the dominant VLSI technology [West89].

When considering the merits of the various logic families several characteristics are examined:

Transistor switching speed - which translates into logic gate delay.

Noise immunity - a measure of the circuit's resilience to EMI.

Silicon layout size - the degree of integration possible on a given chip size.

Power dissipation - specialist techniques are required for high power circuits.

Fan-Out - the drive capability of the logic gates.

Except for the inability to operate at very high switching speeds, CMOS performs very well when judged against these criteria and as a result currently holds an unassailable advantage in the low and medium frequency ranges of the digital logic market.

At low frequencies, CMOS dissipates considerably less power than bipolar circuits because of its virtually zero static power consumption brought about by its low leakage current.

However, as the operating frequency rises, the dynamic power dissipation of CMOS becomes the dominant factor up to a point where bipolar technologies actually dissipate less power. The power/speed trade-off point between bipolar and CMOS logic families was claimed to be around 50MHz in 1988 [GPS88]. However, with the continuing enhancements of process technologies (particularly with regard to CMOS) the trade-off figure may currently be higher.


Figure 9 : Dynamic Power Dissipation.

Emitter-Coupled Logic (ECL) operates by "steering current" through a differential ("long-tailed") pair of switching bipolar transistors which are coupled through an emitter resistor. ECL is an extremely fast logic family since it is non-saturating and keeps the logic signal swings relatively small (around 0.8 V ).

Differential logic is an enhancement of ECL which still uses the long-tailed pair of switching transistors to steer the gate current to one of the two complementary outputs. However, the input signal and its complement are used as the inputs to the switching transistors.

The noise immunity of the gate is increased by the use of the signal and its inverse as inputs, since any noise is experienced as a common-mode signal. The differential amplifier with complementary inputs possesses a high Common-Mode Rejection Ratio (CMRR). The increased noise immunity of differential logic allows much lower voltage swings to be used resulting in a faster gate switching speed (for the same gate current).

### 5.2 Multi-Level Differential Current Mode Logic

Differential logic (unlike standard ECL or CMOS) can be stacked into a switching "tree" configuration and as a result complex logic functions can be packed into a single gate.

GPS (GEC Plessey Semiconductors) have combined a stacked differential switching tree arrangement with a fabrication process based on Trench-Isolated Bipolar Silicon Technology [Depe89] to produce a logic family known as Multi-Level Differential Current Mode Logic (MDCML) (FAB5 variant).

MDCML has up to 3 levels in the circuit switching tree. This has been chosen as the best compromise between the higher functionality of increasing the number of switching levels and the penalty paid in terms of increased silicon area, the requirement for (voltage) level shifters to transpose signals between levels and the increased power sup-
ply voltage needed to incorporate the many switching levels. GPS estimate that up to $25 \%$ of area and $40 \%$ of current, in the worst case (i.e. random logic), may be required for level shifting [GPS88].


Figure 10 : MDCML 3-Input AND Gate.

A 3-input AND gate structure is shown in Figure 10. There are 3 distinct transistor switching levels. By convention these are known as: LEVEL 3 at the top (inputs A and $\overline{\mathrm{A}}$ ), LEVEL 2 in the middle (inputs B and $\overline{\mathrm{B}}$ ) and LEVEL 1 at the bottom (inputs C and $\overline{\mathrm{C}})$. The voltage difference between the switching levels is defined to be one $\mathrm{V}_{\mathrm{BE}}$ drop, to ensure that the transistors do not saturate, this also simplifies the level shifting circuitry.

The operation of the MDCML 3-input AND gate is as follows:
Assuming all differential input signals are at logic 1, then input A is HIGH and input $\overline{\mathrm{A}}$ is LOW; similarly, inputs B and C are HIGH and inputs $\overline{\mathrm{B}}$ and $\overline{\mathrm{C}}$ are LOW. Transistors $\mathrm{t}_{\mathrm{A} 1}, \mathrm{t}_{\mathrm{B} 1}$ and $\mathrm{t}_{\mathrm{C} 1}$ are ON and transistors $\mathrm{t}_{\mathrm{A} 2}, \mathrm{t}_{\mathrm{B} 2}$ and $\mathrm{t}_{\mathrm{C} 2}$ are OFF . The emitter current flows through $\mathrm{t}_{\mathrm{A} 1}, \mathrm{t}_{\mathrm{B} 1}$ and $\mathrm{t}_{\mathrm{C} 1}$ and causes a voltage drop across the load resistor $\mathrm{R}_{\mathrm{L}}$ connected to the collector of $\mathrm{t}_{\mathrm{A} 1}$. As a result, $\overline{\mathrm{Q}}$ is pulled LOW and since no current flows through $\mathrm{t}_{\mathrm{A} 2}, \mathrm{t}_{\mathrm{B} 2}$ or $\mathrm{t}_{\mathrm{C} 2}, \mathrm{Q}$ is HIGH.

If differential input signal $B$ is then driven to a logic 0 , transistor $\mathrm{t}_{\mathrm{B} 2}$ will turn ON (and $\mathrm{t}_{\mathrm{B} 1}$ will turn OFF) causing the emitter current to flow through $\mathrm{t}_{\mathrm{B} 2}$ and $\mathrm{t}_{\mathrm{C} 1}$, resulting in output Q being pulled LOW. Also, since no current path exists between the $\overline{\mathrm{Q}}$ output and ground, $\overline{\mathrm{Q}}$ is pulled HIGH.

Similarly, if differential input signal C is at logic 0 , while A and B are at logic 1, transistor $\mathrm{t}_{\mathrm{C} 2}$ is ON . A current path exists between the collector of $\mathrm{t}_{\mathrm{A} 2}$ and ground, causing the Q output to be pulled LOW ( $\overline{\mathrm{Q}}$ is again HIGH).

It can be observed that the output Q is HIGH (and its complement $\overline{\mathrm{Q}}$ is LOW) if, and only if, all the differential input signals A, B and C are at logic 1, i.e. the gate performs the AND function.

The logic swing of the gates is defined by the load resistors, $\mathrm{R}_{\mathrm{L}}$, and the gate current, $\mathrm{I}_{\mathrm{E}}$, and is nominally 160 mV . By selecting between alternative gate current-resistor 'pairs' different speed/power options are available.

Due to the differential switching tree arrangement, many complex logic functions can be incorporated into a single gate structure. Two example functions, a 4:1 Multiplexer and a Transparent Latch with Reset are shown in Figures 11 and 12.


The S1 and S0 inputs, at level 1 and level 2 respectively, uniquely select one of the level 3 inputs (A3-A0) for passing to the output.


Figure 12 : Transparent Latch with Reset.

The Reset signal overrides all other inputs and so is at level 1 - when Reset is asserted the Data/Enable switching tree is not active and the output signal is driven to logic $0(\mathrm{OUT}=\mathrm{LOW}, \overline{\mathrm{OUT}}=\mathrm{HIGH})$. When Reset is deasserted, the Data input is passed through to the output (when Enable $=1$ ) or the latch holds the output stable (when Enable $=0$ ). Data storage for the latch is provided by the cross-coupled pair of transistors at level 3 .

In summary, the advantages of MDCML are:
. Non-saturating switching transistors and very small voltage swings allow very high speed operation.

D Differential operation removes the requirement for temperature-compensated voltage reference circuits (needed in ECL).

- Increased noise immunity and enhanced resilience to supply voltage fluctuations, temperature variations and IR drops because of the excellent Com-mon-Mode Rejection Ratio.

The multiple levels of switching transistors results in a high functionality of the standard cells.

The use of differential signals removes the requirement for inverters (a signal is inverted by simply "swapping the wires"), which may represent up to $20 \%$ of all the gates in a system [GPS88].

The high impedance of the long-tailed pair arrangement enhances high fanout operation.

However, disadvantages include:
[ High static power dissipation - although MDCML can be operated at 3 V and the small voltage swings employed result in very small currents when compared with ECL (MDCML - $90 \mu \mathrm{~A}$, ECL $\sim 1 \mathrm{~mA}$ ).

- Extra silicon area and power is required for level shifting circuits.

The routing area needed may be increased by a factor of two, since two wires are needed for each signal. A CAD system may require more sophisticated routing software since, to preserve the common-mode rejection characteristics of differential logic, the two signal wires must be routed as a single entity.

## 6. Verilog Modelling of MDCML

### 6.1 Requirement for Accurate Model of System

The utility of a simulation model of a complex digital system is ultimately determined by the extent to which that model closely reflects reality. A model that is simple and easy to manipulate is of little benefit if it does not mirror the actual switching characteristics of the implementation technology.

Circuit level simulation is normally the lowest level of simulation used in the design of an electronic system and is usually performed on circuits consisting of a few tens of components: transistors, resistors, capacitors etc. [Russ85]. The circuit simulation determines the electrical characteristics of the component group which may form a logic gate primitive, for example an AND gate, and may require a few minutes of CPU time.

Circuit simulation of an entire design consisting of many thousands of transistors may be performed in rare circumstances, but generally, the computing resources required make this approach prohibitive. To produce an accurate design simulation, the standard solution is to model the system at a higher level of abstraction based on information gained from circuit simulation of the primitive components [Hill87].

### 6.2 Determination of Electrical Characteristics of MDCML

The switching characteristics of MDCML logic primitives are determined by circuit simulation of the arrangement of transistors and associated component models required
to produce a particular logic function. The circuit simulation is achieved using HSPICE [Hspi90], a widely-used, commercially-available, development of the original Berkeley SPICE program [Nage73].

For the purposes of demonstrating the simulation procedure, a 2-input AND gate will be used as an example. A circuit diagram of the primitive components used to construct a 2-input AND gate is shown in Figure 13:


Figure 13 : 2-input AND gate.

A textually-based SPICE model of the circuit is produced:

```
*subckt and2 out Nout i3 Ni3 i2 Ni2 Vs
.subckt and2 7 7 6 5
xr1 1 6 rn
xr2 1 7 rn
xq1 6 5 8 t20
xq2 7 4 8 t20
xq3 8 3 9 t20
xq4 7 2 9 t20
xics 9 0 cs90
.lib 'Elibbase' rn
.lib 'Elibbase' t20
.lib 'Elibbase' cs90
.ends and2
```

Figure 14 : SPICE model of AND2.

Each component instance is given an instantiation name. In the AND2 gate model of Figure 14, transistors have been labelled, $\mathrm{xq} 1, \mathrm{xq} 2$, xq 3 and xq 4 , resistors have been labelled xr1 and xr2 and the current source is labelled xics. The circuit connections are
specified in terms of numbered nodes and the name of the model primitive used for each component is indicated. For example, "xq3839t20" specifies a transistor with the instance name xq3, which has its collector, base and emitter connected to nodes 8,3 and 9 respectively and has the circuit behaviour defined by the t 20 transistor model.

The propagation delays from each of the inputs, A and B , to the output are measured for both rising and falling edges. Since both phases of the signal are available in differential logic, delays are measured from the input crossover point to the output crossover point. The A (level 3) input, B (level 2) input and output waveforms are shown in Figure 15.


Figure 15 : 2-input AND gate SPICE waveforms.

The measured propagation delays for an unloaded 2-input AND gate are:
A rising $->$ OUT rising $=178 \mathrm{ps}$
A falling $->$ OUT falling $=178 \mathrm{ps}$
$B$ rising $->$ OUT rising $=241 \mathrm{ps}$
$B$ falling $->$ OUT falling $=193 \mathrm{ps}$

The results shown in Figure 15 indicate that:
i) The level 3 (A input) propagation delay is less than the level2 (B input) propagation delay: the higher levels in the MDCML switching tree switch faster.
ii) The rising and falling delays of level 2 are different. This may be explained by noting that the switching tree is non-symmetrical above the level 2 inputs.

### 6.2.1 Output Loading Effects

The effects on the propagation delay of loading the gate output are now considered. The output load is provided by the successive addition of level 3 buffer circuits. The buffer circuit is chosen for this purpose because no level shifting is required between the AND2 gate output and the input of the buffer. Also, since the buffer circuit has a symmetrical switching tree structure, it should provide an equivalent response to both rising and falling input waveforms. The topology of the test circuit is shown below:


The propagation delay effects of gate output loading were measured for both level 3 and level 2 input signal changes, and for both rising and falling edges. The following results were obtained (all times measured in picoseconds):

| Additional <br> O/P Load | LEVEL 3 |  | LEVEL 2 |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Rising | Falling | Rising | Falling |
| 0 | 178 | 178 | 241 | 193 |
| 1 | 205 | 205 | 262 | 220 |
| 2 | 232 | 233 | 283 | 249 |
| 3 | 262 | 263 | 307 | 281 |
| 4 | 294 | 296 | 333 | 317 |
| 5 | 330 | 332 | 362 | 358 |
| 6 | 370 | 372 | 404 | 404 |
| 7 | 412 | 412 | 440 | 444 |

A graph of additional delay against additional load was plotted (see Figure 16).


Figure 16 : Graphs of Additional Load vs Additional Delay.

The graphs can be approximated to a straight line through the origin. The conclusion drawn from the results is that each additional load applied to the output of a 2 -input AND gate adds around 30ps to the propagation delay for both input levels and for both rising and falling edges.

### 6.2.2 Input Drive Effects

When the preceding gate is heavily loaded, this can have an effect on the propagation delay of the gate in question. This is as a result of the input signal drive capability being ‘shared’ between several successor gate inputs. A level 3 buffer circuit was again used as the (preceding) gate load in the following configuration:


The effects on propagation delay were measured for each level and for rising and falling input signal changes. The following results were obtained:

| Drive <br> Sharing | LEVEL 3 |  | LEVEL 2 |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Rising | Falling | Rising | Falling |
| 0 | 178 | 178 | 241 | 193 |
| 1 | 188 | 189 | 244 | 196 |
| 2 | 199 | 200 | 248 | 199 |
| 3 | 210 | 211 | 251 | 202 |
| 4 | 222 | 222 | 254 | 205 |
| 5 | 232 | 233 | 258 | 207 |
| 6 | 242 | 241 | 262 | 210 |
| 7 | 250 | 248 | 265 | 213 |

A graph of additional delay against extent of drive sharing was plotted (see Figure 17).


Figure 17 : Graphs of Additional Delay vs Extent of Drive Sharing.

The conclusion is that the effect of input drive sharing on the 2-input AND gate is different for each of the input signal levels: for level 3, 10ns is added to the propagation delay for each gate sharing the drive, for level 2 the delay is only increased by 3 ns . This suggests that level 2 signals have a greater drive capability than level 3 signals.

Also, the effect on gate propagation delay of input drive sharing is less significant than the effect of output loading.

### 6.3 Production of Verilog Model

The information obtained from the circuit simulation may then be used to generate a model capable of being simulated at a higher level of abstraction. In this case, a behavioural or gate-level model is produced using the Verilog Hardware Description Language.

On initial examination of the simulation data, two points emerge regarding the switching characteristics of the MDCML 2-input AND gate. Firstly, there are different propagation delay for the rising and falling signal changes of the level 2 input and, secondly, the delays differ for the different input levels.

For gate-level modelling in Verilog, both rising and falling propagation delays may be specified for each logic primitive. For behavioural modelling, both input signal edges may be detected using the "always @ (posedge ...)/always @ (negedge ...)" construct. In this example, however, only a single propagation delay will be specified for each input level (for either signal transition direction) to maintain the model simplicity.

The worst-case delay will be used for each level:
A (level 3) $\quad>$ out $=178 \mathrm{ps}$
B (level 2)

Both a behavioural and a gate-level model of the 2-input AND gate are produced to demonstrate how a logic primitive may be modelled. Also, two approaches to modelling the different input level propagation delays can be shown.

Considering the behavioural module first:

```
module and2 (out, Ain,Bin);
`timescale 1ps/1ps
`define and2A_delay 178
`define and2B_delay 241
input Ain, Bin;
output out;
reg out;
always @ (Ain)
    #(`and2A_delay) out = Ain & Bin;
always @ (Bin)
    #(`and2B_delay) out = Ain & Bin;
```


## endmodule

In the behavioural example, the module is triggered when there is a change in any of the input signal values. Depending on whether the input change is at level 3 (Ain) or level 2 (Bin), the output is specified to change (if an output value change is warranted) after a different time interval, and $2 A \_$delay or and $2 B \_$delay. In this manner, the different propagation delays from input to output of the different levels are modelled.

The gate-level, or structural, module is based on the and and buf logic primitives used in the following configuration:


```
module and2 (out, Ain,Bin);
`timescale 1ps/1ps
`define and2A_delay 178
`define and2B_delay 241
input Ain, Bin;
output out;
wire delb;
and #(`and2A_delay) g1 (out, Ain,delb);
buf #(`and2B_delay - `and2A_delay) g2 (delb, Bin);
endmodule
```

A single propagation delay is specified for the and logic primitive; the minimum of the A and B propagation delays, which is and2A_delay. An additional delay is encountered by a $B$ input signal change and this is modelled by providing a buf (buffer) logic primitive. The buf element has a propagation delay equivalent to the difference between the propagation delays of the two input levels (A and B). In this way, an A input change will propagate to the output after the and 2 A_delay time through the and primitive. Also, a B input signal change will propagate to the output after the 'and2B_delay and 2 A_delay' time through the buf primitive plus the and 2 A_delay time through the and primitive, i.e. a total propagation delay time of and2B_delay.

An example of the waveforms produced by both the behavioural (AND2_Be) and structural (AND2_St) modules when simulated in Verilog using the same input stimuli is shown below:

| Header: MDCML 2-input AND gate |  |  |
| :--- | :--- | :--- |
| User: Robert Kelly |  |  |
| Date: Mar 7,1994 09:52:46 | Time Scale From: 0 To: 4000 | Page: 1 of 1 |



### 6.3.1 Accuracy Comparison

The MDCML 2-input AND gate exhibits differing propagation delays to input changes occurring at the different levels. This suggests that the operation of the Verilog models of the AND2 gate may be sensitive to simultaneous or nearly simultaneous changes in the input signals. The two models of the AND2 gate were simulated under these conditions and a problem was discovered which can be observed in the following
waveforms:


The output waveforms of the two AND2 modules indicates that the behavioural and gate-level models react differently to the specified input stimulus. In particular, the response to a B input change closely followed by an A input change must be examined for each of the models.

For the behavioural module, when the A input change occurs (always @ (Ain)), the output assignment expression, out $=$ Ain \& Bin, will be evaluated. At this point, both the Ain and Bin inputs are HIGH and so the output value will be scheduled to change after the propagation delay time of the Ain input, \#('and2A_delay) out = Ain \& Bin.

For the gate-level (structural) module, when the Ain input change occurs, the previous change of the Bin input has not yet propagated through the buf primitive. So, at this point in time, the A input to the and primitive is HIGH, but the B input to the and primitive is LOW. Therefore, the Ain input change does not directly affect the output of the and primitive in this situation. Some time later, when the effect of the Bin input change has propagated through the buf primitive, both inputs to the and primitive will be HIGH and the output will be scheduled to change.

In summary, for the behavioural model, the Ain propagation delay takes priority, whereas for the gate-level model, the Bin propagation delay takes priority. The question is then: which model most closely reflects reality? To provide the definitive answer, a

HSPICE circuit simulation is performed with the appropriate input stimulus. The results are shown below in Figure 18.


Figure 18 : Ain after Bin SPICE waveforms.

The Ain input change occurs 21.6ps after Bin for the rising transition and 22.3ps after Bin for the falling transition.

The measured propagation delays are:

| A rising | $->$ | OUT rising | $=238 \mathrm{ps}$ |
| :--- | :--- | :--- | :--- |
| B rising | -> | OUT rising | $=260 \mathrm{ps}$ |
| A falling | -> OUT falling | $=113 \mathrm{ps}$ |  |
| B falling | OUT falling | $=135 \mathrm{ps}$ |  |

Considering the rising transitions of the A and B inputs, the actual measured output propagation delay is most closely modelled by the gate-level module. That is, the B input signal change tends to take priority. The output propagation delay of the falling input transitions seems to be anomalous; since the measured delay is much less than either
the normal A or B input falling transition delays. This behaviour may be as a result of both input changes simultaneously tending to force the output LOW.

Since the gate-level, structural, module gives the most accurate modelling behaviour of the 2 -input AND gate, this module is chosen as the basis for the AND2 logic gate in the MDCML Verilog component library. The full component library is constructed by considering each logic gate function in a similar manner.

### 6.3.2 Continuous Assignment

The CPU resources required for the simulation of a large scale digital system can be significant, even when the design is simulated at higher levels of abstraction. The Verilog HDL provides a mechanism for accelerating the performance of the modelling constructs by applying a technique known as continuous assignment. Continuous assignment may be used to increase the simulation performance of models by directly assigning values to outputs of primitives based on the values currently on the inputs.

The gate-level model of the 2-input AND gate can be replaced by a continuous assignment version which increases the simulator performance:

```
module and2 (out, Ain,Bin);
`timescale 1ps/1ps
`define and2A_delay 178
`define and2B_delay 241
input Ain, Bin;
output out;
wire delb;
assign #(`and2A_delay) out = (Ain & delb);
assign #(`and2B_delay-`and2A_delay) delb = Bin;
endmodule
```

Here the assign statement replaces the instantiations of the and and buf primitives in the gate-level model.

The technique is restricted to logic primitives that are:
purely combinatorial, i.e. the primitive does not contain any internal state.

- based on very simple logical operations.


### 6.3.3 Net Delays

The issue of how the output loading and drive sharing information should be incorporated into the Verilog model of the system is now examined.

It can be noted that, essentially, the causes of the propagation delays of an interconnected system of component modules can be divided into two types. The first type, Elemental delays, concern the direct operation of the basic logic function, i.e. the propagation delay through the unloaded gate. The second type, which can be called Topological delays, cause an additional propagation delay to be applied to each basic component depending on the nature of its interconnections, both input and outputs, with the rest of the system, i.e. the fan-in and fan-out of the individual gates.

Verilog provides a facility whereby delay values may be assigned to individual nets (wires) connecting system modules (known as net delays). This would appear to be an ideal method of managing the additional propagation delay effects of the system interconnections. The basic component modules would incorporate the Elemental delays and the interconnecting nets would include the Topological delays.

For completeness, the system models should incorporate the effects of input drive-sharing and output loading. However, for a reasonably large design, calculating the additional gate propagation delay due to these effects by hand would be tedious - some form of automatic netlist generation is required (this is best achieved in conjunction with a schematic design-capture system).

In addition, the circuit delay values due to track capacitance (measured during the physical layout design process) could easily be backannotated into the Verilog simulation model by modifying the net delay values. However, until the actual physical layout of
the design is accomplished, and real track capacitance values can be used to generate real interconnection delay values, the MDCML Asynchronous ARM Verilog model will use only elemental delays.

## 7. MDCML Asynchronous ARM

### 7.1 ARM Architecture

### 7.1.1 Overview

The Advanced RISC Machine (ARM) is a general purpose 32-bit microprocessor architecture based on the Reduced Instruction Set Computer (RISC) principle of a simple, regular instruction set allowing fast and efficient decoding [Furb89,VLSI90]. Together with a three stage (fetch, decode, execute) execution pipeline, this results in a high instruction throughput. The ARM uses a load/store architecture with a register-oriented instruction set.

The ARM6 (the target architecture of the MDCML Asynchronous ARM) has a 32-bit data space and a 32-bit address space [ARM91](see Figure 19). All instructions are one word (32-bits) and all data processing operations are performed on word quantities. Byte quantities (in addition to words) can only be specified for load and store operations.

The ARM6 may be executing instructions in one of six processor modes:

- User - normal program execution.
- Supervisor - protected mode for operating system support.
- IRQ- normal interrupt handling.
- FIQ - 'fast' interrupt handling (for external data I/O).
- Abort - data or instruction prefetch abort.
- Undef - undefined instruction execution.

Most applications programs execute in User mode, the other (privileged) modes are entered to service interrupts or handle processor exceptions.


Figure 19 : ARM6 Block Diagram.

A total of 37 registers are provided, 31 general purpose registers (each 32-bits wide) and 6 status registers. The registers partially overlap such that, depending on the current processor mode, only 15 general purpose registers (R0-R14) and R15 holding the Program Counter (PC) are 'visible'. In all modes the Current Program Status Register (CPSR), which contains the condition code flags and the current mode bits, is visible and in the privileged modes the Saved Program Status Register (SPSR) is also visible. R14 is used as the subroutine link register (receiving a copy of the PC return address on executing a Branch and Link instruction).

The fast interrupt mode (FIQ) has seven 'banked' registers (R8-R14) and all privileged modes have banked R13 (stack pointer) and R14 (subroutine link) registers. There is a SPSR (loaded with the CPSR on exception entry) for each of the privileged modes.

### 7.1.2 Instruction Set

The ARM instruction set consists of ten basic instruction types. All ARM instructions are conditionally executed based on the value of the $\mathrm{N}, \mathrm{Z}, \mathrm{C}$ and V flags in the CPSR. The conditions always (AL) and never (NV) also exist. Conditional execution of all ARM instructions seeks to improve processor performance by removing the need for small-offset forward branches which therefore maintains the execution pipelining.

The data processing instructions can be divided into two groups: those concerned with logical operations (AND, EOR, ORR, BIC, MOV, MVN, TST, TEQ) and those performing arithmetic operations (ADD, ADC, SUB, RSB, CMP, CMN). This class of instruction also contains an $\mathbf{S}$ bit which indicates whether the condition codes should be set based on the result of the specified operation. Since the ARM architecture contains a barrel shifter connected to one of the input operand buses of the Arithmetic Logic Unit (ALU), it is possible to perform various shift functions on one of the input operands before the specified data operation is applied. A subset of the data processing type, the MRS/MSR instructions provide access to the CPSR and the SPSR: the MRS instruction moves the contents of the CPSR or SPSR into a register and the MSR instruction moves a register value into the CPSR or SPSR.

The branch (B) and branch-with-link (BL) instructions allow the PC to be modified by adding a signed offset. A 'jump' instruction can also be achieved by using the MOV instruction to load the PC (R15) directly with an immediate or register value.

A multiply (MUL) or multiply-accumulate (MLA) instruction uses a 2-bit Booth's algorithm to perform integer multiplication, the multiply-accumulate form adds a third operand register value to the result of the basic two input register multiplication.

Two instruction types are concerned with moving data between registers and memory. The LDR/STR data transfer instructions move a single byte or word of data. The LDM/ STM block data transfer instructions are used to move any subset of the currently visible register set.

The software interrupt (SWI) instruction is used to enter supervisor mode in a controlled manner and the single data swap (SWP) instruction is used to swap a byte or word quantity between a register and memory as an 'atomic' (uninterruptable) operation this facility provides the basis for multiprocessing semaphore support.

Three further instruction types are used in the context of coprocessor interaction and will not be discussed further

### 7.2 MDCML Asynchronous ARM

### 7.2.1 Overview

The high-level design of the MDCML asynchronous ARM will closely follow that of the AMULET1 [Furb94b]; an asynchronous ARM microprocessor developed for CMOS technology within the ESPRIT OMI-MAP project involving the AMULET group at Manchester University. The TAM-ARM project will also consider if any of the design enhancements proposed for the CMOS successor to AMULET1, in the light of experience gained while producing the original prototype, will be appropriate for the MDCML implementation.

Much of the architectural design information presented in this chapter is derived from [Pave94], a Ph.D. thesis of one of the principal design team members.

The internal structure of the MDCML asynchronous ARM is shown in Figure 20 overleaf.


Figure 20 : Internal Processor Organisation.

Since the MDCML asynchronous ARM will be a preliminary demonstrator of an asynchronous bipolar implementation of the ARM6 architecture, several features have not been incorporated into the MDCML asynchronous ARM architecture due to system design time constraints. Unimplemented features include the class of instructions used to manage coprocessor interaction, support for 26-bit mode operation (an instruction-set backwards-compatibility issue) and the MLA (Multiply-with-Accumulate) instruction.


The MDCML ARM will employ a transition-signalling bundled-data interface between the asynchronous processor and the external memory subsystem involving the following signals:
a An output bundle containing the requested memory address, associated control signals and, possibly, a write data value (if a write operation is specified).

- An input bundle containing a data value (which may be an instruction) read from memory and the bundled-data protocol control signals.
- A memory abort response. Every data access to memory requires a response signal to indicate whether the access will successfully complete. This allows the processor to support a virtual memory system.

A processor reset and level-sensitive interrupt request signals complete the MDCML asynchronous ARM connections to the external environment.

The structure of the Execution Pipeline, which includes the Register Bank, Execution Unit and Instruction Decode, is outlined in Figure 21 overleaf. In particular, the typical micropipeline structure of Event Registers (shown as shaded boxes) interposed by computational logic can clearly be seen. The pipeline operation is controlled by the transition signalling protocol operating between the event registers and functional blocks, but the details have been omitted from the diagram in the interests of clarity.

The Primary Decode provides the entire decoding for the Register Bank control signals and a partial decode for the Execution Unit function blocks. Note that the decode and control signals are also pipelined, but this does not imply that the datapath and control operate in lockstep. Control signals and data values only synchronise at the appropriate functional unit.


Figure 21 : Micropipelined Structure of the Execution Pipeline.

### 7.2.2 Register Bank

The register bank provides the storage required for the general-purpose registers and program status registers defined by the ARM architecture. For the execution of a typical instruction, one or two operands are read from the register bank onto the $A$ and $B$ buses. They are then subjected to some logical or arithmetic operation to yield a result, which is normally written back to the register bank via the W bus. To improve overall CPU performance, pipeline operation is employed whereby several instructions may be in different phases of execution. At a certain instant in time, the operands of one instruction may be undergoing an ALU operation, while simultaneously, the operands of the next instruction are being read from the register bank and the ALU result of the previous instruction is being written back to the register bank.


Figure 22 : Register Bank Operation.

In an asynchronous design context, pipelined instruction execution with concurrent read and write access to registers presents a number of problems with regard to coherent register operation:

- Due to execution phase pipelining, multiple register write operations may be outstanding. The register bank control logic must maintain a record of the correct sequence of write register addresses.

An operand read action may be requested from a register that has a write operation pending. The register read needs to be suspended until the write actually occurs (the register read should get the new register value).

Asynchronous read and write operations on the same register may interact unpredictably.

These problems may be solved by storing the write (destination) register addresses in a FIFO (First-In, First-Out queue). After the register bank is accessed during the initial stages of instruction execution, to provide the operands, the destination register address is entered into the FIFO. When the instruction result value eventually arrives back at the register bank, the destination (write) address is at the end of the FIFO. The depth of the FIFO determines the number of register write operations that can be outstanding.

On every read access to the register bank to obtain the instruction operands, the Write Address FIFO is examined. If either of the read addresses is found to match any of the write addresses in the FIFO, the read operation for that particular register stalls until after the write operation on the register has completed. Once the operands have been successfully read from the register bank, the destination (write) register address for the instruction result is entered into the Write Address FIFO. The FIFO effectively provides a 'locking' function on the register bank to prevent Read-after-Write register hazards hence the alternative name for the Write Address FIFO is the Lock FIFO [Pave92]. The asynchronous register bank design is shown overleaf in Figure 23.

The operation of the asynchronous register bank is now described - the associated Verilog waveforms in Figures 24, 25 and 26 show a Read cycle, a Read cycle stalled on a register lock and a Write cycle respectively:

The register read request ( $R \_R e q$ ) arrives (Figure 24) and presents two instruction operand register addresses ( $A \_a d d r$ \& $B \_a d d r$ ) and a destination (write result) register address ( $W \_a d d r$ ). Additional addressing information indicates the current processor mode and hence the 'visibility' of a particular register set. The $R \_$Req signal is stalled
(by the Muller-C gate) until the register bank is able to commence another operand read cycle (indicated by the Rgo signal). The A and B bus decoders are then enabled (by $R d e c)$ and at the same time the destination address is latched into the W latch.


Figure 23 : Asynchronous Register Bank Design.

The decoded read enable signals are gated with the Write Address FIFO lock information for the associated register (Lock) - a read will be suspended if the register is locked i.e. a write operation is pending on the register. A read will proceed (A_enb \& B_enb activated) if the register is unlocked, while a read on a locked register must wait until a subsequent write operation clears the lock.

Once both read operations have completed (both $A \_d o n e \& B \_d o n e ~ e v e n t s ~ h a v e ~ o c-~$ curred), the operands (A_bus \& B_bus) are latched (by the $O \_R e q$ signal) and the read decoders are disabled. Loading the read output Event Register causes the $D_{-}$Req event
to be signalled to the Execute Unit, indicating that the instruction operands are now available. The write decoder is then enabled (Wdec) and the decoded destination register address is entered in the Write Address Lock FIFO (LK_Req), thereby locking the register. The write address is stored in the Lock FIFO in decoded form to enable a locked register address to be detected easily (the detail is described elsewhere [Pave91]). Once the Lock FIFO has accepted the new result destination address ( $L K_{-} A c k$ ), the write decoder is disabled and the write address latch is freed ( $L t \_d n$ ). $R \_A c k$ is signalled and the next instruction can now commence its operand read phase.

The second waveform (Figure 25) shows the effect of a locked register on the read cycle. When the read decoders are enabled ( $R d e c$ ), the B bus register enable ( $B \_e n b$ ) is activated and the B bus operand read completes ( $B \_$done, $B_{-}$bus is valid). However, the A bus operand register is locked and therefore the A bus register enables are not activated ( $A \_e n b$ ). Eventually, a subsequent write operation will clear the register lock (Lock), the A bus enables will be activated and the A bus read will complete (A_done). The sequence of events following this point is as outlined previously.

The write request signal ( $W_{\_}$Req), Figure 26, indicates that a result value has arrived on the W bus ( $W \_b u s$ ) for writing into its destination register. When the decoded destination register address ( $W_{-}$reg) is available at the output of the Lock FIFO, the write operation can begin (Wgo). A control signal (valid) also arrives with the data value to provide a facility to clear destination register locks (remove register addresses from the Lock FIFO) without actually writing data into the register bank. This mechanism allows instructions that have failed condition code tests at the ALU to remove write locks from previously 'reserved' destination registers.

If the full register write operation is to proceed ( $W \_o k$ ), the write bus enables are signalled (Wr_reg) and the appropriate register write enable line is activated ( $W \_$_enb) and the data value is written into the register. Once the register write operation has completed ( $W$ _done), the write bus enables are turned off and the write address is removed from the Lock FIFO (Lock) - unlocking the register for subsequent read operations. The
remaining waveforms at the bottom of Figure 26 show a stalled read operation resume once the register lock is cleared.

| Header: Register Bank Read Cycle |  |  |
| :--- | :--- | :--- |
| User: Robert Kelly |  |  |
| Date: Sep 23, $1994 \mathbf{1 8 : 2 2 : 3 8}$ | Time Scale From: $\mathbf{4 6 6 . 5 5}$ ns To: 488.20 ns | Page: 1 of 1 |



Figure 24 : Register Bank Read Cycle Waveform.


Figure 25 : Register Bank Stalled Read Waveform.

| Header: Register Bank Write Cycle |  |  |
| :--- | :--- | :--- |
| User: Robert Kelly |  |  |
| Date: Sep 23, $1994 \mathbf{1 8 : 4 0 : 2 0}$ | Time Scale From: $\mathbf{3 1 3 . 7 3}$ ns To: $\mathbf{3 4 3 . 2 0}$ ns | Page: $\mathbf{1}$ of $\mathbf{1}$ |



Figure 26 : Register Bank Write Cycle Waveform.

### 7.2.3 Memory Interface

The interface between the external memory system and the processor is divided into two parts: the Address Interface issues all address information to memory and the Data Interface is responsible for all data values written to or read from memory.

For a write operation, the address generated by the address interface synchronises with the write data value supplied by the data interface before being passed to the memory subsystem.

For a read operation, the address generated by the address interface is sent directly to the memory subsystem and control information for the access is passed to the data interface. The control information is examined when the memory read value is supplied to the data interface. The read data value is 'routed' to the correct processor function block destination based on the associated control information.

### 7.2.4 Address Interface

One of the primary functions of the address interface is to generate sequential addresses for instruction prefetching. The Program Counter (PC) value circulates around a loop containing the Memory Address Register (MAR), an (address) Incrementer and two PC Holding Latches (see Figure 27). Two holding latches are required because of a potential deadlock situation if only one latch was provided. The deadlock occurs when a data transfer request immediately follows the arrival of a new PC value - this is described in detail elsewhere [Pave94, pp126-127]. In each cycle of the PC loop, the PC value is copied into the Memory Address Register where it initiates an external memory instruction read request. After the processor reset signal is deactivated, the Memory Address Register is forced to all zeros and a memory request event is generated causing instruction prefetching (and therefore instruction execution) to begin at memory address (hex) 00000000.

When the address interface is required to generate a memory address for a data transfer operation (either read or write), the PC prefetching loop must be temporarily suspended. Since the prefetch operation is asynchronous with respect to the rest of the processor operation, arbitration is required to gain exclusive control of the address interface resources.


Figure 27 : Address Interface Structure.

The value of the Program Counter is available to the programmer as register R15 and can be used as a source or destination operand in the same manner as a general-purpose register. Note that writing a new value to R15, changing the PC, has the same effect as a branch instruction. However, because of the 3-stage (fetch, decode, execute) execution pipeline operation of the synchronous ARM 6, the address value read from R15 (the PC) is 8 bytes ( 2 instruction words) ahead of the actual address of the currently executing instruction. In order to ensure that existing ARM instruction code programs have the same functionality, some mechanism must be provided in the asynchronous implementation to mimic the behaviour when register R15 is accessed to provide instruction operands.

The PC Pipe is a 2-stage FIFO into which the instruction prefetch PC address value is copied after it has been used to generate a memory read access. However, after system reset the first two instruction prefetch addresses are not copied into the PC Pipe. When the first prefetched instruction eventually reaches the primary instruction decoder and synchronises with its associated R15 (PC) value, the R15 value will precede the actual memory address location of the decoded instruction by 8 bytes. As a result of the action of the PC Pipe, the asynchronous implementation can emulate the synchronous ARM behaviour of the R15 (PC) register.

The PC Pipe mechanism of maintaining the R15 value 8 bytes ahead of the currently executing instruction temporarily fails after a branch instruction executes. However, the association between R15 values and instructions is only incorrect for those instructions that do not execute, i.e. the instructions prefetched beyond the branch instruction. When the branch target instruction actually begins the decode phase, prior to execution, the PC Pipe mechanism has re-synchronised - further details can be found in [Pave94, p128].

The operation of the address interface is now described - the associated Verilog waveforms in Figures 28, 29, 30 and 31 show the instruction prefetching mechanism, a data transfer address arriving on the W bus, the address interface interaction of a LDM (LoaD Multiple) instruction and the effect of a branch address arriving on the W bus respectively:

The instruction prefetching cycle request (PC_Req), Figure 28, arrives at the address interface control arbiter along with a PC value (PreAddr) as an input to the Memory Address Register (MAR). Eventually, control is granted (PCgo) to the PC loop and the PC value is latched into the MAR by the $M A R_{-} R e q$ signal. A memory read access is then initiated (Mem_Req) with the PC address value contained in the MAR (MemAd$d r)$. The control circuit then triggers the Address Incrementer (Inc_Req) and, once the PC value has been incremented by adding 4 (all ARM instructions are 32 bits wide and are word aligned), a completion signal (Inc_dn) is generated. A control signal, PC/
$L S M$, indicates whether the incrementer has been used to generate a PC value or a Load/ Store Multiple (LSM) address (since a LSM instruction also uses the incrementer functionality to generate sequential addresses). The incrementer output value (Incre) is then latched into the first of the PC holding latches ( $P C_{-} l t l$ ) and, subsequently, the output of the first latch ( $P C x$ ) is copied into the second PC holding latch (on reception of the $P C_{l} l t 2$ event). The output of the second latch (the current $P C$ value) is then entered into the PC Pipe ( $P P_{-} R e q$ ) and, when the PC Pipe indicates that it has accepted the PC value ( $P P_{-} A c k$ ), the instruction prefetch cycle request ( $P C_{-}$Req) is again generated.

A data transfer address can arrive at the address interface directly from the register bank on the A bus or, in this example (Figure 29), on the ALU (write) result bus ( $W_{-}$bus). The data access request ( $W_{-}$Req) is directed to the address interface control arbiter and arbitration takes place between the data transfer request and the PC prefetch loop request. Eventually, the data transfer is given control of the address interface ( Wctl ) and a request grant signal is generated (Wgo). The multiplexer control signals (MuxCtl) are switched to allow the W bus value to pass to the input of the MAR (Pre$A d d r$ ), to be subsequently latched by the $M A R \_R e q$ signal. A memory access request event is then generated (Mem_Req) with the address contained in the MAR (MemAddr). Since this is a single word transfer, the incrementer is not activated (Inc_by) and the data transfer is completed when an acknowledge signal is returned to the source of the W bus value ( $W \_A c k$ ).

The remaining waveforms in Figure 29 indicate a stalled PC prefetch request ( $P C_{-}$Req) which is unable to continue (PCgo) until the W bus acknowledge has occurred ( $W_{-} A c k$ ). The actual PC memory access request must also wait until the previous W bus data transfer memory cycle has completed (indicated by Mem_Ack). The prefetch loop then resumes by incrementing the PC access address.

For the block data transfer instructions (LDM/STM) involving the movement of multiple data values to or from consecutive memory locations (Figure 30), only the base address of the transfer is sent (via the A bus or, in this example, the $W_{\_} b u s$ ) to the ad-
dress interface. The data transfer request ( $W \_$Req ) arrives at the address interface control arbiter and again, eventually, control is given to the data transfer (Wctl) and the grant signal (Wgo) is generated. The MAR input multiplexers are switched (MuxCtl) and the data transfer address (PreAddr) is latched into the MAR (MAR_Req). A memory access is then initiated (Mem_Req) with the block data transfer base address (Me$m A d d r$ ). The address incrementer then operates (Inc_Req) to generate the next sequential LSM address (Incre). The PC/LSM control signal indicates that a LSM instruction triggered the address incrementer and so, when the incrementer operation has completed (Inc_dn), the address value is copied into the LSM (temporary storage) register (LSM_Req). The address interface continues to generate sequential memory addresses until a control signal ( $L D M_{-} d n$ ) indicates that the required number of addresses have been produced. The LSM data transfer then relinquishes control of the address interface arbiter by signalling $W_{-} A c k$. At this point, the PC prefetching loop can again resume.

When the processor executes a branch instruction (Figure 31), the new PC value arrives at the address interface from the ALU via the $W_{-}$bus. The W bus data request ( $W \_$Req ) is directed to the address interface arbiter and eventually exclusive access is indicated ( $W c t l$ ) and a grant signal is generated ( $W g o$ ). The multiplexer is again switched (MuxCtl) and the W bus address value is passed to the input of the MAR (Pre$A d d r$ ) where it is subsequently latched ( $M A R \_R e q$ ). A memory read access is signalled (Mem_Req) with the new PC address contained in the MAR (MemAddr) to fetch the branch target instruction. The first phase of branch instruction interaction with the address interface, namely supplying the target address and initiating an instruction prefetch memory access, is now complete and control of the arbiter is released ( $W \_A c k$ ).

The second phase of the branch interaction involves restarting the PC prefetching loop with the new instruction stream addresses. Once a memory access is activated on the branch target address, the address incrementer is signalled (Inc_Req), the target address is incremented (Incre) and the incrementer completion signal is generated (Inc_dn). The $P C / L S M$ control signal indicates that the incrementer output value is an
instruction prefetch address and it is latched through the PC holding latches (PC_ltl and $P C_{-} l t 2$ ) to become the current $P C$ value. A new instruction prefetching cycle request ( $P C_{-}$Req) is directed to the address interface control arbiter and, when control is granted ( $P C g o$ ), prefetching restarts with the new PC address value. The previous PC_Req* request signal that was stalled at the arbiter, while the branch target address arrived on the W bus, is released $\left(P C g o^{*}\right)$ when the W bus access relinquishes control of the arbiter ( $W \_A c k$ ). Control circuitry in the instruction prefetching loop is able to detect that a new PC value has arrived and so the prefetch request for the old instruction stream is discarded


Figure 28 : Address Interface Instruction Prefetching Waveform.



Figure 29 : Address Interface Data Transfer Waveform.

| Header: Address Interface - $\mathbf{3}$ register LDM |  |  |  |
| :--- | :--- | :--- | :---: |
| User: Robert Kelly |  |  |  |
| Date: Oct 4, 1994 16:52:59 | Time Scale From: $\mathbf{3 7 1 . 6 7}$ ns To: 437.60 ns | Page: $\mathbf{1}$ of $\mathbf{1}$ |  |



Figure 30 : Address Interface Block Data Transfer Waveform.

| Header: Address Interface - Branch => new address into PC loop |  |  |  |
| :--- | :--- | :--- | :---: |
| User: Robert Kelly |  |  |  |
| Date: Oct 4, 1994 17:18:04 | Time Scale From: 73.92 ns To: $\mathbf{1 2 6 . 5 6} \mathbf{n s}$ | Page: $\mathbf{1}$ of $\mathbf{1}$ |  |



Figure 31 : Address Interface Branch Waveform.

### 7.2.5 Data Interface

The data interface controls the interaction between the external data bus and the processor. It handles the values returned from memory after a read access and the data values written out to memory. The overall structure of the data interface is shown below in Figure 32.


Figure 32 : Data Interface Structure.

For memory write operations, in which a byte quantity is specified, the Data Out (DOUT) section has the facility to replicate the least significant byte across all byte positions in the word (to enable byte writes to any byte-aligned address). The memory write data request (indicating that the data value is available) must rendezvous with the Memory Address Register request (indicating that a write address has been generated
by the Address Interface) before the external memory access request is despatched. This rendezvous occurs in the Memory Control (Mem_Control) section.

Once a memory read value arrives at the data interface, it is latched in an Event Register. The Destination Control block will then extract the corresponding control information from the Memory Control Pipe (Mem_Ctl_Pipe) for this read access. Note that the control information was entered into the Memory Control Pipe when the Address Interface generated the read access address. The retrieved control information will indicate whether the memory read value was an instruction or a data value.

Data values read from memory are passed to the Data In (DIN) section, where byte-rotation logic is provided to rotate values read from non-word aligned memory addresses. Also, logic exists for masking the most significant 24 bits of the data word for byte read quantities.

Incoming instructions are buffered before execution in the 5-stage Instruction FIFO Pipeline (I_Pipe). The I_Pipe must be 3 stages longer than the (2-stage) PC Pipe because of a complex deadlock situation - a detailed explanation can be found elsewhere [Pave94, pp130-131]. An instruction emerging from the I_Pipe may also be passed into the Immediate Field Extraction Unit (Imm_Pipe), so that any immediate operand can be retrieved from the appropriate fields of the instruction word prior to full decoding. The output of the Immediate Field Extraction Unit can be multiplexed onto one of the datapath operand buses, if required.

The operation of the data interface is now described - the associated Verilog waveforms in Figures 33, 34, and 35 show a data byte read operation, a data byte write operation and the reception of a prefetched instruction word from which an immediate operand is extracted respectively:

A memory read data byte (or word) operation (Figure 33) begins when the address interface signals (MARr) that a valid memory access address (MemAddr) has been generated. Since this is a read data transfer, the associated control information (Mctl) is
latched into the Memory Control Pipe (MCPir), while the external memory request is generated (MEMr). Some time later, the memory subsystem responds with a data word value ( $M R R$ ) which is latched into the Memory Read (Event) Register by the $M R R r$ request signal. An acknowledge signal is generated (MRRa) once the data is latched and the external memory read cycle is completed when the memory subsystem responds with the MEMa acknowledge event. The read input request (MDr), indicating that a valid read data value is contained in the Event Register (RdData), and the Memory Control Pipe output request (MCPor), signifying the availability of the associated control information for this memory access, synchronise (Sync) in the Destination Control section. The Opcode control signal indicates that the value read from memory is not an instruction and so the DIr request signal latches the value into the Data In (DIN) section. Additional control information is passed to the DIN block indicating that a data byte read has occurred $(B / n W)$ and the position of the required byte within the word (ByteNo). The unwanted bytes are masked out and the byte read value is shifted into the least significant byte position (SelByte). An output request (DI_Req) is then sent to the Execution Pipeline to signal that the output of the Data In section $(D I)$ is now valid and, eventually, an acknowledge event ( $D I \_A c k$ ) will be received when the byte read value has been consumed.

A write data byte (or word) operation (Figure 34) is initiated when a request signal (DOr) is received by the Data Out (DOUT) section of the data interface. It indicates that a write data word value $(D O)$ has been read from the register bank and is available for transfer to memory. A control signal $(B / n W)$ specifies that a byte data transfer is required. The byte-replication logic is then triggered (Rep_Req), which causes the least significant byte position value to be copied into all byte positions in the data word (By$t e R e p)$. The replicated byte value is then latched (Rep_dn) into the Memory Write Register $(M W R)$ contained within the Data Out section. The $M W R r$ request signal indicates to the Memory Control (Mem_Control) section that a write value is now ready for transfer. The $M A R r$ request signal indicates that the address interface has generated the associated memory address (MemAddr) for this write data transfer. When these two re-
quest signals synchronise (Sync) in the Memory Control section, a external memory data transfer is initiated (MEMr). Since this is a write memory access, the associated control information (Mctl) is not entered (no event onMCPir) into the Memory Control Pipe (Mem_Ctl_Pipe). A memory write access is specified (Wen) by the control information and, eventually, the memory subsystem responds with an acknowledge signal (MEMa) indicating that the memory write cycle has completed. The Memory Control section can then clear the Memory Write Register (MWRa) and signal the memory write cycle termination to the address interface (MARa).

An instruction read operation (Figure 35), in a similar manner to a data read operation, commences when an address interface request signal (MARr) arrives indicating that a PC prefetch address has been generated (MemAddr). The associated control information (Mctl) is again latched (MCPir) into the Memory Control Pipe before the external memory read access is requested (MEMr). A request signal (MRRr), generated by the memory subsystem, is used to latch the returned memory value ( $M R R$ ) into the Memory Read Register. When the latch operation has completed, the data interface responds with an acknowledge signal (MRRa) and the external memory cycle is terminated by the $M E M a$ acknowledge signal. The $M D r$ signal indicating the presence of a returned memory value ( $R d$ Data) in the Memory Read Register and the Memory Control Pipe output request (MCPor) synchronise in the Destination Control section. The Opcode control signal indicates that the memory read value is a prefetched instruction and so the value is latched (INr) into the Instruction Pipeline (I_Pipe). Some time later, a request signal to the primary instruction decode ( $I N \_R e q$ ) indicates that the prefetched instruction (IN) has emerged from the Instruction Pipeline. The output of the instruction disassembler (DIS) shows that the instruction does indeed contain an immediate operand value. The full instruction word is subsequently latched (IMr) into the Immediate Field Extraction Unit (Imm_Pipe), where the immediate operand value (IMM) is retrieved. A request event (IM_Req) is sent to the Execute Unit control indicating the validity of the output of the Imm_Pipe. An acknowledge signal (IN_Ack) is received from the primary instruction decode stage when the instruction word has been consumed.


Figure 33 : Data Interface Byte Read Waveform.

| Header: Data Interface - Write data byte |  |  |  |
| :--- | :--- | :--- | :---: |
| User: Robert Kelly |  |  |  |
| Date: Oct 9,1994 13:46:13 | Time Scale From: 614.62 ns To: 677.39 ns | Page: 1 of 1 |  |



Figure 34 : Data Interface Byte Write Waveform.

| Header: Data Interface - Read instruction with immediate value |  |  |
| :--- | :--- | :--- |
| User: Robert Kelly |  |  |
| Date: Oct $\mathbf{7 , 1 9 9 4} \mathbf{1 0 : 3 9 : 5 5}$ | Time Scale From: 762.08 ns To: 842.04 ns | Page: $\mathbf{1}$ of $\mathbf{1}$ |



Figure 35 : Data Interface Instruction Read Waveform.

### 7.2.6 Execution Unit

The execution unit contains the computational logic of the processor. It comprises (see Figure 36) a Multiplier, Shifter, ALU and storage registers for the Current Program Status Register (CPSR). The Multiplier accepts two input operands to produce partial sum and partial carry outputs which are then added together in the ALU to produce a final result. It is based on an iterative shift-and-add operation using carry-save adders and incorporates early-termination detection logic.The Shifter is connected to one of the operand buses in series with the ALU allowing various shift and rotate operations to be performed on one of the ALU input values.


Figure 36 : Execution Unit Structure.

The Arithmetic Logic Unit (ALU) performs all the logical operations and arithmetic functions needed by the ARM architecture. The arithmetic functions requiring the ALU
to carry out an addition are potentially the most time-consuming operations because of the carry propagation between the bit positions of the calculation. A study of ARM instruction execution [Jagg90] indicates that around $20 \%$ of all instructions perform arithmetic data processing. However, since the ALU is also used in calculating addresses for data transfer and branch instructions, the actual percentage of instructions requiring an ALU addition operation is much higher than the above figure.

In a synchronous system, all ALU operations must take place within a fixed clock period and techniques, such as carry-lookahead, have been developed to reduce the time required for an addition. The ARM6 uses a carry-select mechanism. An asynchronous ALU may vary the required computation time, dependent on the actual input data values, and can determine addition operation completion by noting when carry propagation has terminated.

The operation of the MDCML Asynchronous ARM ALU has a similar high-level design to that employed in the CMOS AMULET1 [Gars93] in that addition completion is signalled when carry propagation has ceased. The actual implementation of the ALU datapath components in MDCML logic yields a much higher performance than the CMOS counterpart. However, because the circuit design technique of wired logic (wire-AND, wire-OR etc.) is not easily produced in MDCML technology, some aspects of the MDCML ALU control logic are slower than the equivalent CMOS circuit. In particular, the 32-bit AND function used to determine when valid signals have been asserted by all bit positions and the 32 -bit NOR function used to produce the ALU output Z (zero) flag are implemented in (slow) multiple stages of 3-input gates. The average addition time in the MDCML Asynchronous ALU is much faster than the worst-case time and all logical operation are completed in a fixed (short) time period. The exploitation of data-dependent computation time results in a simple ALU design of comparable performance to existing synchronous designs which incorporate carry-lookahead or carryselect techniques.

The operation of the execution unit is now described - the associated Verilog waveforms in Figures 37 and 38 show a multiply operation (followed by the addition of the partial sum and partial carry multiplier outputs in the ALU) and an ALU logical operation involving a shift of one of the ALU input operands respectively:

The execution of a multiply instruction (Figure 37) begins when the datapath input request ( $R D_{\_}$Req) arrives indicating that the two operands specified by the instruction have been read from the register bank (A_bus and B_bus). A control signal (Mult) shows that a multiply operation is required and so the multiplier request signal (Mul_$R e q)$ is generated. When the multiply operation has finished two outputs are produced, the partial sum (Psum) and the partial carry (Pcarry), and a completion signal (Mul_dn). The partial sum and carry are then latched into the ALU input operand event register by the $O p_{-} R e q$ signal. At this point the register bank output register is no longer required to hold the initial instruction operands stable and the execution unit indicates this by generating an acknowledge signal ( $R D_{\_} A c k$ ). The two ALU input operands ( $A_{-} o p$ and $B_{-} o p$ ) are subjected to the required ALU function (Afunc) - in this case, an addition to combine the partial sum and carry - when the ALU is enabled ( $A L U_{-} E n b$ ). When the ALU operation has terminated, an output value $(A L U)$ is produced along with a completion signal $\left(A L U \_d n\right)$. The ALU output latch is then closed ( $\left.A L U_{-} l t\right)$, holding the ALU output result (Result) stable. A request event is generated ( $O \_R e q$ ) to indicate that the result value is available for copying into the execution unit output register (not shown in Figure 36). A W (write) bus request signal ( $W \_$_Req) is forwarded to the Write Control unit, while the execution unit output register value ( $W \_b u s$ ) is placed on the W bus. The Write Control unit will 'steer' the W_Req request signal to the appropriate function unit based on the associated control information (Wctl) for this instruction. Eventually, the specified function unit (in this example, the register bank) will respond with an acknowledge signal ( $W_{-} A c k$ ) when the result value has been received.

For the execution of an instruction involving a (register) shifted operand (Figure 38), the input request ( $R D \_R e q$ ) from the register bank again indicates the validity of
the input operands (A_bus and B_bus). The shifter is enabled (Sh_Enb) with the appropriate function control signals (Sfunc) and eventually the shifter output result (Shift) is produced along with a completion signal (Sh_dn). The A bus instruction operand and the shifted B bus operand are then latched into the ALU input operand event register by the $O p_{-} R e q$ signal and the register bank input request is subsequently acknowledged ( $R D \_A c k$ ). From this point on the execution unit control signals and sequence of events is similar to the ALU operation described for the multiply instruction previously. The ALU input operands ( $A_{-} o p$ and $B_{-} o p$ ) are again subjected to the required function (Afunc - in this example, an AND operation) when the ALU is enabled (ALU_Enb) and, eventually, an output value ( $A L U$ ) is produced followed by a completion signal ( $A L U \_d n$ ). The ALU output value is latched ( $A L U_{-} l t$ ) into the ALU result latch (Result) and the execution unit output register is signalled ( $O \_R e q$ ). The W bus request signal ( $W \_R e q$ ) is generated when the execution unit output value $\left(W_{\_} b u s\right)$ is valid and the appropriate function unit is signalled based on the associated instruction result control information ( $W c t l$ ). An acknowledge signal ( $W_{-} A c k$ ) is received when the result destination function unit has consumed the value.

| Header: Execute Pipeline - Multiply operation |  |  |
| :---: | :---: | :---: |
| User: Robert Kelly |  |  |
| Date: Oct 19, 1994 18:47:23 | Time Scale From: 748.61 ns To: 823.72 ns | Page: 1 of 1 |



Figure 37 : Execution Unit Multiply Waveform.

| Header: Execute Pipeline - Logical operation (shifted operand) |  |  |  |  |
| :--- | :--- | :--- | :---: | :---: |
| User: Robert Kelly |  |  |  |  |
| Date: Oct 19, $1994 \mathbf{1 8 : 5 2 : 1 7}$ | Time Scale From: 926.32 ns To: 981.24 ns | Page: 1 of 1 |  |  |



Figure 38 : Execution Unit Shifted ALU Operand Waveform.

### 7.2.7 Comments on the MDCML Asynchronous ARM Design

The MDCML ARM demonstrates that the implementation of a simple RISC architecture using asynchronous design techniques is attainable. The complex design task is made manageable by employing a modular design methodology, namely micropipelines, with subsystems communicating via a well-defined protocol i.e. the transitionsignalling bundled-data interface. The design of the Register Bank control logic, with the novel, arbiter-free method of allowing concurrent read and write operation interaction, gives an example of how new design problems can be overcome. The data-dependent operation of the ALU shows how an asynchronous system can take advantage of the variable processing rates of a particular functional unit in order to increase overall performance. Also, the autonomous action of the instruction prefetching mechanism in the Address Interface demonstrates the independent operation of the component subsystems.

The MDCML Asynchronous ARM exhibits a very high degree of concurrency which is suggested in many of the Verilog waveforms shown earlier in the chapter. This is as a result of the self-timed constituent function units operation being solely dependent on input data availability. As a consequence of this asynchronous computational parallelism, the total system state at any particular instant is difficult to determine. Similarly, the effects of the interactions between two communicating subsystems, in an overall system context, are difficult to quantify. Developing an understanding of the total system operation is still in the early stages, and the design changes required to increase overall system performance are not immediately obvious. The production of a realistic simulation model of the entire system (described in the following chapter) which has the ability to execute real ARM instruction code programs has proved invaluable in exploring the complex behaviour of the running system.

## 8. Architectural Modelling

### 8.1 Introduction

Verilog is an industry-standard Hardware Description Language which is integrated into the CAD system supplied by the MDCML technology provider, GEC-Plessey Semiconductors (GPS). The entire process, therefore, of architectural modelling through schematic design capture and physical layout to bipolar technology fabrication is more easily accomplished.

Architectural modelling of a system seeks to hide the lowest levels of the implementation complexity from the conceptual design, so that alternative design ideas can be more easily considered and evaluated. The design process iteratively refines the higher levels of abstraction to move towards an implementation of the prototype system. At each stage in the process, the Verilog system model can be simulated to provide an indication of the design correctness and system performance.

### 8.2 Modelling

The initial requirement in developing a model of a prototype system is the production of a library of components that can be used to construct larger functional subsystems. The Verilog HDL has a range of logic primitives incorporated into the language but, because of the switching characteristics of the different signal levels in MDCML, the standard primitives must be combined to produce models of the MDCML gate-level equivalents (see Section 6.3). For example a 3-input OR gate can be modelled in the following manner:

```
'timescale 1ps/1ps
module or3 (Out, Ain,Bin,Cin);
'define A_del 230
'define B_del 300
'define C_del 400
output Out;
input Ain,Bin,Cin;
wire delb,delc;
buf #(`B_del - `A_del) g1 (delb,Bin);
buf #('C_del - 'A_del) g2 (delc, Cin);
or #('A_del) g3 (Out, Ain,delb,delc);
endmodule
```

An 'asynchronous control element' library is also produced using the behavioural modelling language constructs of Verilog. This comprises the micropipeline control circuit elements outlined in Section 3.1. The Verilog behavioural model of the Muller-C element is shown below:

```
‘timescale \(1 \mathrm{ps} / 1 \mathrm{ps}\)
module MullC (Out, Ain,Bin,Rst);
'define A_del 470
'define B_del 640
'define Rst_del 370
output Out;
reg Out;
input Ain,Bin,Rst;
always @ (Ain)
    if ((!Rst) \&\& ((Ain===Bin) || (Ain==='bx)))
    \#('A_del) Out = Ain;
always @ (Bin)
        if \(((!\) Rst \() \& \&((A i n===B i n) \|(\) Bin==='bx) \())\)
            \#('B_del) Out = Bin;
always @ (Rst)
        case (Rst)
            1'b1: \#('Rst_del) Out = 0;
            1 'b0: if (Ain===Bin)
                    \#('B del) Out = Bin;
        1'bx: \#('Rst_del) Out = 'bx;
        endcase
endmodule
```

The dynamic simulation behaviour of the above Muller-C element is provided by the 3 concurrently-executing always @ statements, one for each of the inputs: Ain, Bin and Rst.

As an illustration of the operation of the model, the Ain behavioural code is explained:

At every change of the Ain signal ( always @ (Ain) ), if the Rst (reset) signal is inactive ( $!$ Rst $)$ and both inputs have the same value ( Ain $===$ Bin ) or the Ain input is undefined ( Ain $====$ 'bx ), then the Ain input signal is passed to the output $($ Out = Ain $)$ after the appropriate delay ( \#‘A_del $)$.

Note that any undefined input signal arriving at the Muller-C is propagated to the output. This feature assists in the detection of incorrect operation (see Section 8.5.3).

Larger components, such as 32 -bit Event Registers (Section 3.2.1) can be constructed from their constituent elements: a Muller-C gate and 32 Capture-Pass latches. However, since Event Registers are widely used throughout the MDCML Asynchronous ARM, a behavioural model of an Event Register is produced which improves simulator performance. That is, a single model is invoked for any input data signal change rather than multiple invocations of the constituent models. Also, by producing a single behavioural model for a larger function, additional checking can be incorporated into the model structure to report all occurrences of incorrect circuit operation. For example, the reception of two successive input request events, without an intervening input acknowledge event, results in an error message being displayed during the simulation execution.

The complex computational subsystems of the Asynchronous ARM architecture, including the ALU, shifter and multiplier, are also modelled as behavioural modules. It is much easier to handle the input and output bus values of such components as single data entities (e.g. 32-bit integers) rather than manipulating the individual bit values. For example, consider adding two 32-bit operands in the ALU:

| input | [31:0] A, B; |
| :---: | :---: |
| output | [31:0] out; |
| reg | [31:0] out; |

Once the bit-widths of the input and output buses are specified, the addition result assignment to the output bus is achieved by means of a single arithmetic operator.

The complete MDCML Asynchronous ARM model consists of a single module which instantiates the component subsystems to produce a hierarchical composition of asynchronous modules. The autonomous subsystems communicate using the two-phase bundled-data interface, and the total system is self-starting from reset. Once the global reset signal is deactivated, instruction prefetching commences (from the external memory model) leading to execution of the test program instructions.

The processor diagram shown overleaf in Figure 39 shows the major functional subsystems and the significant control signal and data bundles connected between them. To assist in the clarity of the diagram, some of the signals found in the Verilog processor 'core' model in Figure 40 have not been included in Figure 39. The signal names in the bolder typeface in the processor diagram indicate the connections to the external environment. At the top right-hand corner of the diagram are the bundled-data interface signals used to communicate with the external memory system. These include the Memory Access Control Information, Memory Address, Write Data and Read Data values and the associated protocol control signals. The memory subsystem is modelled using the Verilog behavioural language to generate the required data values and the communication protocol control signal sequences. The two signals names at the bottom of the figure ( $n A b t$ and $\operatorname{Dabt[1:0]\text {)handlethefaultresponsesofthememorysystem.The'nor-}}$ mal' and 'fast' interrupt signals (Nirq and Nfiq) are shown at the top left-hand corner.

The Verilog model in Figure 40, illustrates the top-level components of the MDCML Asynchronous ARM and the connectivity of the processor signals. The full hierarchical model developed by the author is given in Appendix A. For example, the Register Bank (Reg) has the instantiation name $r g$; it produces the $R G a, R W a$ and $R D r$ output signals along with two 32-bit output buses ( $\mathrm{Na}[31: 0]$ and $\mathrm{Nb}[31: 0]$ ). The Register Bank has five input signals ( $R G r, R W r, W c[2], W s e l$ and $R D a$ ), in addition to the global reset signal (Rst), and has three input buses: a 32-bit Write (result) bus, a 30-bit PC (program counter) bus and a 28 -bit control bus (Rs[27:0]).


Figure 39 : MDCML Asynchronous ARM Processor Diagram.

```
MDCML Micropipelined ARM
`timescale 10ps /10ps
module ARMstCore (Add[31:0],Dout[31:0],Ctl[6:0],MEMr,MRRa,
                                    MRR[31:0],Nfiq,Nirq,Dabt[1:0],PAbt,MEMa,MRRr, BigEnd, nAbt,Rst) ;
`include "ARMstCore.inc"
// Ctl[] = Seq, Inc, Ren, Wen, Usr, B/W, Opc
// memory data in blocks and instruction pipeline
DatInt dat (MARa,DOa,MEMr2,MWR[31:0],MRRa,DWr,DW[31:0],DWusr,
                                    DWv, DWpc, INr,IN[31:0], flo[1:0],Nim[31:0], IMa0, IMr2,
                                    MARr, DOr, nMLb [31:0], DObw, MEMa2, MRRr, MRR [31:0] , PAbt, DWa,
                                    INa, {MAR[1:0],MAc[4:0],Mval,MdPC,MpPC},IMr0,IMa2,BigEnd,Rst);
EvtReg #32 lat0 (Dout[31:0],MEMa0,MEMr0, MWR[31:0],MEMr2,MEMa,Rst);
Cgate2 C0 (MEMa2, MEMa0,MEMa1,Rst);
Cgate2 c1 (MEMr, MEMr0,MEMr1,Rst);
// 1st decode stage
Decode1 dec1 (RGr,Rs[27:0],IN2r,IN2[25:0],IN3r,IN3[19:0],PCsel,XPr,
                                    XLa, INa, IMr0,LSMPr, nTRM, r15,NGr0,nGn[1:0],vect [2:0],
                                    INr,IN[31:0], flo[1:0],Nfiq,Nirq,PCpar,RGa,IN2a,IN3a,XPa,
                                    XLr,PPr3, IMa0, LSMPa, ALUgo,ALUok,mode [5:0],NGa0,nAbt,Rst) ;
// 1st execution and 2nd decode stage
Reg rg (RGa,RWa,RDr,Na[31:0],Nb[31:0]
                                    RGr,Rs[27:0], nPC[31:2],RWr,W[31:0],Wc[2],Wsel,RDa,Rst);
NGen nGen (NGa0,NGr2,ng[5:0], NGr0,IN[15:0],nGn[1:0],vect[2:0],NGa2,Rst);
Decode2 dec2 (IN2a,RSa,C2r,{Imd[6:0],SHop[9:0],DObw,c2[7:0]},
                                    IN2r,IN2[25:0],RSr,Na[7:0],C2a,Rst) ;
// IN[] = Xt[1:0],PCpar,cond[3:0],sctls[2:0],I[11:5],
// DObw,toRs,cpCP,~toDO,~toA,nGen,~Mult,NImm
// Imd[] = Xt[1:0],PCpar,cond[3:0]
// c2[] = toRs,cpCP,~toDO,~toA,nGen,~Mult,NImm
// 3rd decode stage
Decode3 dec3 (IN3a,C3r,ALfs[9:0],{vec3[2:0],c3[22:0]}, IN3r,IN3[19:0],C3a,Rst);
// c3[] = UseCP,S,F,C,Wcp[2:0],Ral,Rcnd,~ALUwt,~DabtWt,
// tPCp[1:0],Wreg,Wadd,SP,LSM,Ren,Wen,B/W,Opc,destPC,Rsel
// 3rd control and execution stages
Shift shft (Sh[31:0],ShC,SHd, nMLb[31:0],Nim[31:0],c2[0],SHop[9:0],psrC,SHe);
ExecP excP (RDa,C2a,C3a,NGa2,SHe,IMa2,PCpar,ALUgo,ALUok,mode[5:0],psrC,
        WRr,W[31:0],Wc[9:0],Wq[1:0],APr,DOr, nMLb[31:0],RSr, Dabt0,
                            RDr, Na[31:0],Nb[31:0],C2r, c2[7:0],C3r, c3[22:0],vec3[2:0],
                    ALfs[9:0],NGr2,ng[5:0],Sh[31:0],ShC,SHd, IMr2,Imd[6:0],
                                    DW[31:0],DWv,DWusr,WRa,Wsel,APa,DOa,RSa,Dabt [1:0],Rst);
// write bus control
WrCtl wctl (DWa,WRa,RWr,ADr,Wsel, DWr,DWpc,DWv,WRr,Wq[1:0],RWa,ADa,WLx,Rst);
// the memory address interface
AddInt add (ADa,WLx,APa,PPr3,XPa,XLr,nPC[31:2],LSMPa,
    MARr,MAR[31:0], {MAc[6:0],Mval,MdPC,MpPC }
                                    ADr,W[31:0],Wc[9:0],APr,Na[31:0],LSMPr,nTRM,
                                    r15,INa,XPr,XLa,PCsel,MARa,Dabt[1],Dabt0,Rst);
// MAc[] = Seq, Inc, Ren, Wen, Usr, B/W, Opc
// Mval = valid MdPC = destPC MpPC = PCpar
EvtReg2 # (32,7) lat1 (Add[31:0],Ctl[6:0],MEMa1,MEMr1,
                                    MAR[31:0],MAc[6:0],MEMr2,MEMa,Rst);
endmodule // ARMstCore
```

Figure 40 : MDCML Asynchronous ARM ‘Top-Level’ Verilog Model.

The Verilog comment at line 8 of Figure 40 (// Ctl[] = Seq,Inc,Ren,Wen,Usr,B/W,Opc) indicates the component signals that comprise the $\operatorname{Ctl}[6: 0]$ external memory control bus. Some of the other MDCML Asynchronous ARM control buses are also expanded in the comment lines.

### 8.3 Features

### 8.3.1 Instantiation Parameters

One useful feature of the Verilog modelling environment is the use of parameters when instantiating components. These parameters may be used, for example, to specify different propagation delay times for different instances of the same module (to reflect particular gate loading effects) or to specify multiple bit-widths for certain components. To illustrate this point, a register can be modelled by specifying a multiple bit-width parameter for the data input and output nets of a latch.

```
'timescale 1ps/1ps
module T_latch (out, in, enable);
parameter width=1; // default data width = 1
parameter Data_delay=330;
parameter Enb_delay=490;
output [width-1:0] out;
reg [width-1:0] out;
input [width-1:0] in;
input enable;
always @ (enable)
    case (enable)
        1'b1: #'Enb_delay out = in;
        1'bx: out = 'bx;
    endcase
always @ (in)
    if (enable)
        #`Data_delay out = in;
endmodule
```

When instantiating components, the required parameters must be specified in the same order as they are given in the particular component definition - in the case of the T_latch above: width,Data_delay,Enb_delay. If no parameters are specified, the default values are used, i.e width $=1$, Data_delay $=330$ ps, Enb_delay $=490$ ps.

The modules are then instantiated in the following manner:

Two single-bit latches with differing data propagation delays

```
T_latch #(1,300,500) t1 (out1, in1, enb1);
T_latch #(1,400,500) t2 (out2, in2, enb2);
```

t 1 is a single-bit T_latch, with a 300 ps input-output data propagation delay and a 500ps enable-output propagation delay, where the data input signal is called in1, the output is called outl and the enable signal is called enbl.
t 2 is also a single-bit T_latch, this time with a 400 ps input-output data propagation delay and again an enable-output propagation delay of 500ps.

A 3-stage pipeline for 32-bit data values:

```
T_latch #(32,300,500) p1 (o1[31:0], pin[31:0], enb);
T_latch #(32,300,500) p2 (o2[31:0], o1[31:0], Nenb);
T_latch #(32,300,500) p3 (pout[31:0], o2[31:0], enb);
```

The pipeline is constructed by instantiating T_latch components with 32-bit data widths. The input of the pipeline, $\operatorname{pin}[31: 0]$, is fed into the input of the first latch, p1. The output of p 1, o1[31:0], is fed into the input of the second latch, p 2 , and so on.

The enable signals of the successive stages of the pipeline operate in antiphase, causing data values to move one stage along the pipeline for every two transitions of the enable signal.

### 8.3.2 Test Vector Generation

A standard technique of generating test patterns for validating a fabricated chip is to apply stimuli to the simulation model of the design and then dump the values of the significant control signals and data buses at suitable time intervals to an activity file. In a synchronous system, this normally occurs at the clock edge, when all signals are usually stable. For an asynchronous system, however, given that subsystems operate concurrently at their own rate, the sequential ordering of changes in logic level of two independent signals internal to two separate subsystems cannot be specified. Therefore, the total system state at any given instant cannot be known.

One method used to automatically generate test patterns for a micropipelined system is to locally delay the acknowledge signal into each subsystem until sufficient time has elapsed so that all internal signals have reached a stable state. Effectively, the module is deadlocked awaiting the acknowledge input. The subsystem state is then recorded at the instant of the acknowledge input event using the $\$$ fstrobe() Verilog function. Test patterns for the entire chip can be produced by delaying the external memory access acknowledge input for each memory access and dumping the control and bus values of interest.

The \$strobe() Verilog system task allows the value, at the end of the current timestep, of any signal wire or register to be displayed on the standard output device. The \$fstrobe() function allows the values to be written to a file via an output channel identifier. For example:

## always @ (input)

\$fstrobe(chan_id, " \%b \%b \%h", input, output, state);
The $\$ \mathbf{f s t r o b e}()$ task is triggered on every input signal change ( always @ (input) ). The signal values are written to the file which was bound to the chan_id channel identifier when it was initially opened. The signal values are written on the same line, for each input signal change, in the following order: input, output, state. The format of the signal values ( " \%b \%b \%h") is Binary for the input and output, and Hexadecimal for the state.

The example illustrated overleaf is of the MDCML Asynchronous ARM Chip model (ARMst), which consists of the processor core and the bond pad driver circuits. In order to reduce the pin count, the input data bus (MRR[31:0]) and the output data bus (Dout[31:0]) use the same external data bus ( $\mathrm{Xd}[31: 0]$ ) by means of tristate driver circuitry.

The activity file is opened, in a Verilog initial timing control block, using the following file operation system task:

```
dump_chan = $fopen("ARMst_vecs");
if (dump_chan == 0) $finish; // quit simulation if $fopen() fails.
```

The interface signals for the ARMst module are then recorded in the activity file by the \$fstrobe() system task whenever the Reset (XRst), Memory Read Request (XRr) or Memory Access Acknowledge (XMa) signals change:
always @(XRst or XRr or XMa)
\$fstrobe(dump_chan,
" \%b \%b \%b \%b \%b \%b \%b \%b \%b \%b \%h \| \%h \%b \%b \%b - \%0d", XBigEnd, XnAbt, XPAbt, XDabt, XNfiq, XNirq, Xdbe, XRr, XMa, XRst, Xd, Xa, Xc, XMr, XRa, \$time);

The format of the resulting activity file is shown below:


The prototype silicon can then be tested by subjecting the test specimen to the input stimulus given on the left-hand side of each line in the activity file. Eventually, the specimen outputs should (for a fully-functioning device) assume the associated test vector file output values for each stimulus line.

Since the signal values are only 'sampled' when the system state is stable, there is a risk that timing errors may be overlooked. However, only timing errors on the external interface signals may be missed, since any internal data-bundling timing errors will propagate incorrect data values to the outputs - which will then be detected. Design effort must be directed to the external interface control elements to ensure data-bundling errors are eliminated.

### 8.4 Code Execution

### 8.4.1 Compilation Method

The executable binary is generated from the actual program and an ARM assembler file which contains various initialisation and library functions. The files are compiled, assembled and linked using the ARM Cross-Development Toolkit. This allows code to be generated by a SPARC-based workstation for execution on an ARM processor. A binary executable is produced, which is then converted to a text format suitable for loading into the asynchronous ARM Verilog model.

### 8.4.2 Validation Suite

Since the MDCML Asynchronous ARM is binary code-compatible with the existing synchronous ARM devices, the test program suite used by Advanced RISC Machines (ARM) Ltd. to test prototype devices can also be used to test the design of the asynchronous implementation.

The ARM Validation Suite consists of over a dozen test programs written in ARM assembler [Cock87]. The suite includes programs to exercise the data processing subsystems of the ARM architecture, involving the Arithmetic Logic Unit, Shifter and Multiplier. Further validation programs test the operation of the Register Bank, including the reading and writing of the Current and Saved Processor Status Registers (CPSR and SPSRs) and the interaction of the processor with the external memory system via the Load/ Store Register (LDR/STR) and Load/Store Multiple (LDM/STM) instructions. The branch (and branch-and-link) mechanism of the processor is also fully tested.

As mentioned previously, the MDCML Asynchronous ARM has no support for coprocessor interaction and the Multiply-with-Accumulate (MLA) instruction is not implemented, therefore these aspects of the ARM Validation Suite are not considered during the design test phase.

The simulated execution of the ARM Validation programs revealed a number of errors in the Asynchronous ARM model. In particular, running the Multiplier function test program contained in the Validation Suite exposed an error in the Verilog Multiplier module. The cause of the problem was traced to a specific feature of the Verilog modelling environment. When the assignment of a new value to a bus or control signal occurs, but the newly assigned value is the same as the previous value, then no Verilog event is generated for the assignment. This means that any event control statement dependent on the signal value (eg. always @ (signal)) is not triggered. The Multiplier behavioural model had to be modified and the addition of an extra control signal was required.

Complete verification of the MDCML Asynchronous ARM architectural model, by running the ARM Validation Suite, gives a significant degree of confidence in the overall asynchronous design and the component subsystems.

### 8.4.3 Dhrystone Benchmark

As a high-level language platform, a computer architecture should efficiently execute those features of a programming language that are most frequently used in actual programs. This ability can be measured by a program known as a benchmark. A benchmark can be a real application program supplied with specific input data chosen to provide a representative task or a specially-written (synthetic) program incorporating a wide range of high-level language statements and constructs.

The original Dhrystone synthetic benchmark program (written in Ada) was published in the CACM in October 1984 [Weic84]. A 'C' version was produced in 1988. The program contains statements of a high-level programming language in a distribution which is considered representative of a general-purpose, integer-computational processor workload. The program statement statistics used to develop the Dhrystone benchmark are based on the execution of over 700 programs written in several languages.

The actual benchmark statement distribution is as follows (' C ' version):

| assignments | $51.0 \%$ |
| :--- | :--- |
| control statements | $32.3 \%$ |
| procedure, function calls | $16.7 \%$ |

The distribution of statements is also balanced with respect to operators (arithmetic, logical, comparison etc.), operand type (integer, character, pointer, Boolean etc.) and operand locality (global, local, procedure parameters, function results etc.). The program does not compute anything meaningful, but is syntactically and semantically correct.

There are several areas where the execution details (compiler influence, timing measurement method, cache interaction etc.) have to be checked very carefully whenever a synthetic benchmark program is used for comparison of different processors or different systems. However, for evaluation of design alternatives of the functional components of a prototype microprocessor, the Dhrystone benchmark, with its representative mix of program statement types, provides a useful metric.

The executable binary is generated from three files: two ' C ' source files (dhry_1.c and dhry_2.c) which contain the actual benchmark program and an ARM assembler file which contains initialisation and library functions. A 16Kbyte binary executable is produced, which is then converted to the text format suitable for loading into the Verilog external memory model.

The model executes 1 Dhrystone loop in approximately 344 seconds and indicates a simulated time of 22.9 microseconds, the ratio of the actual running time to the simulated time is $15,000,000: 1$. This translates to a Dhrystone benchmark figure of around 43,500 Dhrystones per second. For the purposes of the benchmark execution, an external memory access time figure of 5 ns is assumed. Also, the result is based on typical parameters for the underlying $1.2 \mu \mathrm{~m}$ bipolar technology.

For comparison, the $1 \mu \mathrm{~m}$ CMOS AMULET1 device yields a figure of 20,500 Dhrystones per second [Furb94b].

### 8.5 Usage

### 8.5.1 Instrumentation

Since the Verilog language has a rich and powerful behavioural modelling capability, custom design tools and system modelling instrumentation functions can also be quickly and easily produced. The following tools, which assist the digital engineer in exploring the dynamic behaviour of the prototype system, have been written by the author.

The data bundling constraint (see Section 2.2.4) is an integral and necessary part of the interface protocol. A Bundle Checker module has been written in the Verilog behavioural modelling language and attached to each of the data "bundles" of interest (data bus + request signal) to determine the validity of the data value change and request signal event sequencing. This has enabled modules with an insufficient bundling tolerance to be identified and modified. The bundle checking code is also incorporated into the behavioural representation of the Event Register module, since these components are widely used throughout the Asynchronous ARM design.

Usually, a 1 nanosecond bundling margin is considered safe, i.e. the data arrives at least 1ns before the request event. However, an Event Register has a 'built-in' bundling margin of around 1.2 ns because of the circuit topology (see Figure 6). The ReqIN request signal must pass through the Muller-C element and then through a power driver circuit (not shown in Figure 6) before the Capture-Pass element begins to latch the incoming data, Din. Therefore, even if the data and request signals arrive simultaneously at the Event Register external inputs, the 'built-in' bundling margin results in a safe transfer of data.

A sample output of the Bundle Checker module and Event Register during a simulation run is shown below (all times are given in picoseconds):

```
Bund_Chk test.sys.cpu.core.rg.LkF.ChkA: Margin = 90 @ 14573
Bund_Chk test.sys.cpu.core.rg.LkF.ChkM: Margin = 90 @ 26342
EvtReg test.sys.cpu.core.dat.memCP.l0: 2730 @ 9671
EvtReg test.sys.cpu.core.dec1.l3: 4230 @ 35185
Bund_Chk test.sys.cpu.core.rg.ChkI: 2450 @ 6206
Bund_Chk test.sys.cpu.core.rg.chki: 
Bund_Chk test.sys.cpu.core.rg.LkF.ChkA: 90 @ 14573
Bund_Chk test.sys.cpu.core.rg.LkF.ChkM: 90 @ 26342
EvtReg test.sys.cpu.core.add.xpipe.e1: 21280 @ 46342
```

The first block shows where (and when) the Bundle Checker has detected a bundling margin below 1000 ps (i.e. 1 ns ) while the simulation is running. The second block involves each bundle checking component (including Event Registers) reporting its minimum bundling values at the end of the simulation run. Note that some of the modules indicate a bundling margin well in excess of 1 ns . This suggests areas where the control circuit performance may be increased.

Another, behaviourally-modelled, design tool which has been implemented is the Pipeline Occupancy Monitor. This is used to collect information regarding the effectiveness of each of the FIFO buffering pipelines used throughout the design, and can clearly be used to influence the pipeline depth in the design. The effect that the number of pipeline stages has on performance is examined in greater detail in Section 8.6.4.

A further useful tool when attempting to understand the operation of a microprocessor is a disassembler, since it is often useful to know the specific instruction that a particular functional unit is processing. This can be achieved by disassembling the 32 -bit value representing the instruction (as in most RISC architectures, the ARM instructions are a fixed width). A Verilog Disassembler module can be connected to the input stage of the instruction (buffering) pipeline in the Data Interface, to note when a particular instruction of interest is (pre)fetched from external memory. Alternatively, it could be connected to the input of the Instruction Decoder to determine when the instruction actually begins decoding. Usually, the latter option is chosen because it represents the commencement of the actual instruction execution.

Writing the disassembler module in the Verilog behavioural language was relatively straightforward for the author because of two factors: The Verilog language syntax is very similar to an existing high-level procedural language ( ' C ') of which the author has programming experience; and the instruction set generally follows the RISC principle of having few instruction formats with regular bit-field positions. The output of the Ver-ilog-ARM Disassembler is in the form of a text string which is suitable for display in conjunction with other signal and bus values using the Verilog waveform display described in the following section.

A Disassembler output example can be seen towards the bottom of Figure 35 (labelled DIS) in the previous chapter. In this case, the disassembler module is connected to the input stage of the Primary Instruction Decoder.

### 8.5.2 Graphical Output

The Verilog waveform output mechanism is implemented by the gr_waves() system task. The user can continuously monitor the waveforms via the interactive graphics interface as the simulation progresses. Two different screens are provided: the Waves screen, on which the signal waveforms are displayed as timing diagrams, and the Select screen, which displays the list of signals from which the user can choose a subset for current display. The unknown (or X) state of a signal or bus is displayed as a solid filled box. The high impedance ( or Z ) state is displayed as a horizontal line which is vertically centred between the ' 0 ' and ' 1 ' levels. The $g r$ r_waves system task was used to produce the waveform diagrams illustrated in the previous chapter.

Verilog provides an interactive graphics interface to display data as a screen of text along with the formatted values of system model nets and registers. The gr_regs() system task defines the layout of the screen and specifies the text and variables to be displayed and the appropriate formats. The graphics screen is updated whenever a value changes for any of the variables defined in the gr_regs task during the simulation execution

A time bar is displayed at the top of the $g r_{-} r e g s$ window representing the total time period of the simulation execution. The interactive mode allows the user to select a particular instant in time by positioning the cursor at a particular point on the time bar - the required data values for the corresponding time are then displayed in the gr_regs window.


Figure 41 : Register Bank Display using Verilog gr_regs() system task.

This graphical output feature is particularly useful for displaying information about the internal state of the prototype system. The Asynchronous ARM Register Bank is displayed using this feature in Figure 41. The Register Bank consists of 31 general-purpose registers (including the Program Counter (PC) - R30) and 6 SPSRs (Saved Processor Status Registers). Only a subset of the entire Register Bank is 'visible' to the programmer in any one of the processor execution modes. The gr_regs Register Bank
display shows the current simulation time in the top right-hand corner ( $\mathrm{t}=20893.49 \mathrm{~ns}$ ). The value of each of the 37 registers, at this particular time, is shown in the left-hand column. The previous value of each of the registers, and the (simulation) time at which each register was modified is shown in the middle and right-hand columns respectively.

Verilog also provides the capability to view data as dynamically changing bar graphs. The gr_bars() system task allows the user to set up charts with multiple bar graphs and update the bars as the simulation proceeds.


Figure 42 : Pipeline Occupancy using Verilog gr_bars() system task.

The $g r$ _bars facility can be used in conjunction with the Pipeline Occupancy module (see Section 8.5.1) to display information regarding the occupancy of all buffering structures used throughout the Asynchronous ARM design. Figure 42 shows the occupancy of the pipelines and buffer structures used in the processor core at a particular instant in (simulated) time. These include: the Immediate Field Extraction Unit, Instruction Pipe, Write Data Buffer and Memory Control Pipe in the Data Interface; the PC Pipe and Exception Pipe in the Address Interface and the Memory Lock FIFO and ALU Lock FIFO in the Register Bank.

### 8.5.3 Detecting Incorrect Operation

One of the primary requirements in exercising a simulation model of a prototype system is to determine when, and where, the system functions incorrectly. For an asynchronous design, erroneous operation may be easier to detect and locate than for a synchronous design, since, in many cases, the asynchronous system will deadlock. A control element may have generated a request signal and not received an acknowledge due to a design fault in the control circuit. Location of the fault is usually achieved by determining which request-acknowledge signalling pairs have not yet completed their communication actions and examining the control circuits responsible for generating these signals.

In some circumstances, the control signal events going to a particular control element may not appear in the correct sequence. For example, an Arbiter may receive a request signal on input R1 and then receive a second request event on R1 before a done signal is received (D1), releasing the Arbiter after the first R1 request. Also, for a Call element, the common (subroutine) done signal may be received before any of the request input channels has actually received a request event. Generally, the cause of incorrect sequencing of the control signals is (as above) design faults in the control circuits. The Verilog behavioural models of many of the asynchronous control elements contain extra checks to detect incorrect interface signal sequencing and report errors (including the module instance concerned and the time) while the simulation is running. Also, when incorrect sequencing is detected, the outputs of the particular control element are forced into the undefined state, since in the real system the output values would not be valid if the control element functions incorrectly.

In an asynchronous system composed of functional units communicating using transition signalling, an event occurs when the logic value on any signal wire changes between the logic 0 and logic 1 levels - in either direction. An undefined value on any of the control signal wires in such a system could prove catastrophic, particularly if the undefined state remains undetected. Usually, an undefined control signal causes the re-quest-acknowledge communication protocol to fail and the system will deadlock. In or-
der to ensure that the system deadlocks as quickly as possible in such circumstances, the Verilog models of the control blocks of the MDCML Asynchronous ARM have been written so that undefined control signals are rapidly propagated throughout the system. Any undefined input signal arriving at a control element is immediately propagated to all outputs of that control element. Total system deadlock results very quickly, enabling the source of the original undefined signal to be easily detected.

The case of incorrect operation which is most difficult to detect is where the bundling constraint is not met when a data value is passed between two asynchronous modules using the bundled-data interface. If the transfer request event arrives at the receiving module before the actual data value, the receiver may latch (capture) an incorrect data value. The request and acknowledge control signals are correctly generated and received by the sender and receiver, respectively, and in the correct sequence. As a result, the system will continue to operate, but with the 'wrong' data value. The effects of propagating an incorrect data value may be significant, particularly if the value is subsequently used to generate system control signals. It is in consideration of this factor that a great deal of design effort must be directed towards eliminating 'data bundling' errors. The Bundle Checker module (section 8.5.1) assists the asynchronous logic designer appreciably.

### 8.6 Performance analysis

### 8.6.1 Subsystem Processing Performance

The Dhrystone benchmark program has been used as a general test program to evaluate alternative design decisions and to provide a performance measure. In particular, it allows the effect of a change in processing rate of a given datapath component, in the context of overall system performance, to be assessed in order to pinpoint computational bottlenecks. The effect on the execution time of varying a module's processing rate by
altering the delay between its Request_In and Request_Out control signals is shown in Figure 43 for the ALU, Register Bank and Primary Decode PLA.

It might be expected that at a high processing rate (for a given subsystem block) the graphs would be almost horizontal until a point was reached, as the processing rate decreased, when the time delay through the block would move it onto the 'critical path' causing system performance to be severely impacted.


Figure 43 : Graph of Block Processing Time vs Dhrystone performance.

The results however do not show this. Instead they seem to suggest that in an asynchronous system of inter-communicating modules, when considered over a number of executed instructions, every subsystem on the datapath is on the 'critical path', i.e. a change in processing rate of any subsystem has an effect on overall system performance. The Dhrystone performance graph for the Primary Decode PLA is approximately constant over the processing delay range shown. This tends to indicate almost complete overlap with concurrent, slower datapath operation. By considering the gradient of the graphs for each subsystem, it can be noted that the degree of linkage between subsystem performance and overall system performance is different. Design effort, to increase system performance, should therefore be concentrated on those subsystems which produce the
steepest gradient processing rate graphs, since this will have the most beneficial effect on overall instruction throughput.

### 8.6.2 Non-symmetrical Propagation Delays

Due to the characteristics of the underlying bipolar technology, with inputs to the basic circuit elements being at different voltage levels, the propagation delay from input to output for many of the primitive logic functions is non-symmetrical i.e. it is different for each of the inputs. To determine if the effect of the non-symmetrical propagation delays significantly affects the performance of the MDCML Asynchronous ARM, the switching characteristics of two of the most heavily used primitive asynchronous control elements, the XOR gate and the Muller-C element, are examined.

The XOR gate acts as a MERGE element for events (see Section 3.1.1). An output event (transition) is generated for every input event. Initially, the most active input of each XOR gate instance is determined i.e. the input that switches most during a complete run of the benchmark program. The most active input signal is then assigned to the fastest switching (level 3) input terminal of each of the XOR gates and the benchmark is again run. For comparison, the XOR gate inputs are reversed (with the most active input signal assigned to the slower switching input terminal, at level 2) and the benchmark program is again executed.

The Muller-C gate acts as a JOIN element for events (see Section 3.1.2). An output event is generated only after an event has been received on both inputs. In contrast to the 'most active input' technique for the XOR gate, the 'later switching' input must be determined for each Muller-C element i.e. the last event to arrive for each input event 'pair'. The later switching input signal is assigned to the fastest switching (again level 3) input terminal of the Muller-C and the benchmark program executed. Again, for comparison, the benchmark is executed with the Muller-C elements inputs reversed.

The results for the XOR gate and Muller-C element are shown in the table below (all figures are expressed in Dhrystones per second):

|  | Fastest | Slowest | Difference |
| :---: | :---: | :---: | :---: |
| XOR | 43535 | 43137 | 398 |
| MULLER-C | 43950 | 42644 | 1306 |

Figure 44 : Effect of Non-Symmetrical Propagation Delays.

The figures indicate that, in the case of the XOR gate, the non-symmetrical propagation delays have little effect on overall performance. However, in the case of the Muller-C element, a 3\% improvement in system performance can be achieved simply by connecting the gate the "optimum way round".

### 8.6.3 Processor-Memory Interaction

The MDCML Asynchronous ARM processor core is contained within an external simulation environment which includes a simple MMU and memory model capable of supporting the bundled-data communication protocol. In order to determine if the processor performance is limited by the external memory access time, for prefetching instructions or reading and writing data values, several simulation runs of the Dhrystone benchmark program were performed with different access time values in the memory model for each run. The results are shown in the Figure 45 overleaf.

The graph shows that the processor performance is, to some extent, limited by the external memory speed. Although a doubling of memory speed does not result in a doubling of processor speed, any increased memory performance is reflected in increased processor performance. Also, an indication of the peak performance of the processor can be obtained by extrapolating the graph backwards to the zero point on the x -axis, i.e. memory access time is 0 ns (an infinitely fast memory). This gives a theoretical peak performance of around 46,500 Dhrystones per second.

Memory Access


Figure 45 : Effect of Memory Speed on Processor Performance.

### 8.6.4 Internal Pipeline Efficiency

The Asynchronous ARM processor contains several pipeline structures which act as buffers to even out the flow of data between functional units of differing speed in the design. Some Event Registers between datapath stages are necessary to support concurrent operation since a previous result can be held while a unit computes its next result. In an attempt to improve the performance of the overall system, the efficiency of these pipelines must be examined. The lengths of some of the internal processor pipelines are fixed, since they perform a particular function or are used to prevent potential deadlock situations. For example, the PC Pipe in the Address Interface must be 2 stages long (see Section 7.2.4) and the 5 -stage Instruction FIFO Pipeline in the Data Interface must be 3 stages longer than the PC Pipe to prevent a complex deadlock state (see Section 7.2.5). Also, the Memory Control Pipe in the Data Interface must be the same length as the Instruction FIFO Pipeline.

The operation of 4 particular pipelines will be examined in detail. These are the ALU and Memory Lock FIFOs in the Register Bank, the Immediate Field Extraction Unit and the Write Data buffering structure in the DOUT section of the Data Interface.

The Pipeline Occupancy Monitor module was connected to the external request and acknowledge signals of the pipelines under investigation and the Dhrystone benchmark program was executed. The results are displayed, using the $\$ g r_{-}$bars() system task, in Figure 46 below:


Figure 46 : Pipeline Occupancy during Benchmark Execution.

For each of the pipelines, the fraction of the total simulation time that the pipeline occupancy was a particular value is shown. For example, for $89 \%$ of the total time, the ALU Lock FIFO was empty and for $10 \%$ of the time, the ALU Lock FIFO contained only one item.

The results seem to suggest that the ALU Lock FIFO, Memory Lock FIFO and Write Data Buffering pipelines are too long and could be reduced to contain only 1 stage (or
possibly removed altogether). The Immediate Field Extraction Unit is probably the correct length.

The investigation was carried further by modifying the length of each of the pipelines, in isolation, and noting the effect on the processor performance when executing the Dhrystone benchmark. Performance may be improved by shortening pipelines, which reduces the latency of the pipeline, i.e. the time taken for a single item to pass through an empty pipeline. The results are shown in the graphs below in Figure 47.


Figure 47 : Effect of Pipeline Length on Processor Performance.

The *'s in each of the graphs shows the length of that particular structure in the current MDCML Asynchronous ARM design. Note that the Immediate Field Extraction Pipe must contain at least one stage for correct system operation.

The results indicate that the ALU Lock FIFO should be shortened by 2 stages (to 1 stage), the Memory Lock FIFO should be shortened by 1 stage (to 3 stages), the Immediate Field Extraction Pipe should be shortened by 1 stage (to 1 stage) and the Write Data Buffer should contain only 1 stage (shortened by 2 stages). These 'recommended' pipeline modifications were simultaneously incorporated into the processor design and the benchmark was again executed. The resulting Dhrystone performance was measured at 44,045 Dhrystones per second - the increase in performance was approximately equal to the sum of the performance increases when the best case of each individual pipeline graph is considered separately. Of greater importance in the bipolar design, than the performance gain, is the decrease in silicon area used for active cells and routing in addition to the power reduction.

The conclusion from this pipeline efficiency study is that the lengths of the pipelines in the current MDCML Asynchronous ARM design should be reduced, in some cases by as much as 2 stages. However, the results only apply to the execution performance of a particular program (the Dhrystone benchmark). There is a requirement to consider a range of general-purpose applications, where individual pipeline structures may be more heavily stressed and, unless silicon area is at a premium, it is better to provide extra buffering to smooth out processing 'hot spots'.

### 8.6.5 Comments on the Performance Analysis.

As mentioned in the concluding comments of the previous chapter, developing an understanding of the total system operation of the MDCML Asynchronous ARM, with its complex integration of inter-communicating, self-timed subsystems is still in its early stages. However, the ability to develop user-instrumentation for a wide variety of monitoring tasks using the Verilog behavioural modelling language and to present the resulting information in its most appropriate form using the graphical and text output Verilog system tasks assists the asynchronous logic designer appreciably in designing working (i.e. correct) systems and exploring the dynamic behaviour of those systems.

## 9. Conclusions

The principal aim of this project was to build an architectural model of the MDCML (bipolar) Asynchronous ARM processor capable of supporting the simulated execution of real ARM instruction code programs. This has been achieved. Furthermore, the model has then been used to explore the dynamic behaviour of the system. Various forms of user-instrumentation were written by the author to enable detailed examination of particular function units, and to present the resulting information in a wide variety of forms. Design enhancements were then proposed and tested by the execution of a wide-ly-used benchmark program.

When designing systems incorporating new ideas, whether these are implementation technology developments or new architectural features, the risks of encountering difficulties are increased over a more mature foundry process or circuit design style. Simulation offers the opportunity to exhaustively test the prototype system before it is committed to the integrated circuit manufacture, where design changes are not possible.

### 9.1 Production of the System Model

Initially, circuit simulation of the basic bipolar logic primitives was carried out to provide information regarding the switching characteristics of the target implementation technology. The knowledge gained was then employed to construct structural and behavioural models of the standard logic primitives (AND, OR, etc.) and asynchronous control elements in the Verilog modelling environment. By producing gate-level and
functional models of the system building blocks, the simulation of a large-scale processor design becomes computationally feasible.

The functional subsystems of the Asynchronous ARM, including the Register Bank, ALU, Memory Interface and Decode logic were then developed. They were constructed, either from the combination of the logic primitives and asynchronous control elements or from representations involving a behavioural description. The complete system consists of the functional subsystems supported by a transition-signalling communication protocol.

### 9.2 Current State of the Project

The architectural model of the MDCML Asynchronous ARM processor has been completed and, at the time of writing, the major part of the datapath is near submission for fabrication. This has only been possible through the use of simulation since it involves a novel design methodology and a new target implementation technology. A simulation environment, consisting of a simple Memory Management Unit (MMU) and an external memory model, has also been produced. The Asynchronous ARM model successfully executes all the programs in the ARM Validation Suite, except for those instructions requiring specific hardware resources which will not be implemented in the target bipolar technology. A number of design and monitoring aids have been written by the author which expose significant parts of the internal operation of the asynchronous system. Information gained during the processor pipeline length investigation enabled the length of the ALU and Memory Lock FIFOs to be reduced in the original design to improve performance and reduce silicon area.

### 9.3 Comments on the Verilog Modelling Environment

Verilog provides an ideal environment for modelling a micropipelined asynchronous microprocessor architecture. Its modular, hierarchical structure is in harmony with a system composed of inter-communicating asynchronous functional units, and asyn-
chronous operation maps well onto its control constructs. The flexibility and suitability of Verilog is further demonstrated by the production of custom tools and test vectors specifically for our prototype design.

The bottom-up, incremental design style and verification of individual primitive components is easily accommodated into a high-level, behavioural view of the overall system, which is largely technology independent. The production of a full execution codecompatible architectural model results in a valuable aid in analysing the dynamic behaviour of the system and gives a degree of confidence in the design approach. Alternative design decisions have been more easily evaluated and an indication of expected performance has been gained.

In common with many other digital logic modelling environments, a Verilog design description is exercised by means of an event-driven simulator. This simulation paradigm fits particularly well with the event-driven computational model of asynchronous logic. Furthermore, the timing control mechanisms incorporated into the Verilog behavioural language, especially the event control constructs, would be ideal for modelling a selftimed system developed using any asynchronous design methodology.

A high degree of concurrency is supported in the Verilog system model through the use of the fork and join compound statements (see Section 4.5.1), allowing a non-deterministic ordering of the notionally parallel execution of the individual statements. Also, multiple always @ event control blocks across the entire design result in many 'threads of execution' being simultaneously active throughout the prototype system.

The modular approach to designing with asynchronous inter-communicating subsystems afforded by the micropipeline approach is closely reflected in the architectural modelling environment of Verilog with its hierarchical module structure. All these features make Verilog sympathetic to an asynchronous design style.

### 9.4 Future Research

### 9.4.1 Technology Migration

The architectural modelling of the MDCML Asynchronous ARM design has been achieved in a hierarchical, modular fashion at a relatively high level of abstraction. The figures used for the propagation delays of the standard logic primitives and asynchronous control elements are based on the circuit simulation of their realisation in the target bipolar technology. By re-designing the required low-level components in a different implementation technology and determining the respective propagation delay times, the characteristics of the new target technology can be incorporated into the basic Verilog system models. This results in the Asynchronous ARM processor design being easily migrated to a new fabrication technology and would, for example, enable a comparison between MDCML and CMOS on the basis of performance.

Of course, detailed design of the functional units will consider if any circuit optimisations exist in the new implementation technology to increase the performance, reduce the gate count, or improve the power efficiency of the system. For example, the lack of a Wire-OR circuit design technique in the MDCML bipolar technology significantly increased the amount of logic required and the propagation delay times for the ALU Completion logic and the Zero-Detect function in the current ALU design. Although the basic switching speed of the MDCML technology is superior to that of CMOS, the circuit design flexibility afforded by CMOS can produce faster and smaller component designs in certain circumstances.

### 9.4.2 Architectural Design Alternatives

In producing the MDCML Asynchronous ARM design, various datapath functional units and control circuit components have been developed using the behavioural modelling language of the Verilog environment. The vast amount of simulation and valida-
tion performed on the system containing these components should convince the logic designer of the integrity of these components.

The system designer is now free to compose these datapath and control elements to produce and explore novel asynchronous computational structures. Multiple functional units can be combined to produce an asynchronous superscalar design or more radical architectures, such as dataflow, may be considered.

## References

[ARM91] ARM6 Macrocell datasheet. ARM Ltd., Cambridge, England, September, 1991.
[Brow91] Brown A., VLSI Circuits and Systems in Silicon. McGraw-Hill, ISBN 0-07-707221-9, 1991.
[Brun91] Brunvand E., Translating Concurrent Communicating Programs into Asynchronous Circuits. PhD thesis, Carnegie Mellon University, 1991.
[Cock87] Cockerell P., ARM Assembly Language Programming. MTC, England, 1987.
[Depe89] Depey M.P. et al., A 10K-Gate 950-MHz CML Demonstrator Circuit Made with a 1-mm Trench-Isolated Bipolar Silicon Technology. IEEE Journal of Solid-State Circuits, 24(3):552-557, June, 1989.
[Dobb92] Dobberpuhl D. et al., A 200 MHz 64b Dual-Issue CMOS Microprocessor. IEEE Journal of Solid-State Circuits, 27(11):155-1565, November, 1992.
[Furb89] Furber S.B., VLSI RISC Architecture and Organization. Marcel Dekker, New York, 1989.
[Furb94a] Furber S.B., Day P., Garside J.D., Paver N.C., Woods J.V., AMULET1: A Micropipelined ARM. Proceedings of IEEE Computer Conference (CompCon'94), San Francisco, USA, March, 1994.
[Furb94b] Furber S.B., Day P., Garside J.D., Paver N.C., Temple S., Woods J.V., The Design and Evaluation of an Asynchronous Microprocessor. IEEE International Conference on Computer Design (ICCD ‘94), October, 1994.
[Gars93] Garside J.D. A CMOS VLSI Implementation of an Asynchronous ALU. Proceedings of the IFIP working conference on Asynchronous Design Methodologies, Manchester, England, 1993.
[Gopa90] Gopalakrishnan G., Jain P., Some Recent Asynchronous System Design Methodologies. Tech. Rep. UU-CS-TR-90-016, Dept. of Computer Science, University of Utah, October, 1990.
[GPS88] GEC - Plessey Semiconductors, Differential Logic Design Manual (FAB 4) 1.0 Edition. July, 1988.
[Hart87] Hartenstein R., Hardware Description Languages. Advances in CAD for VLSI Series, Vol. 7, Elsevier Science Publishers B.V., ISBN 0--707221-9, 1987.
[Hauc93] Hauck S., Asynchronous Design Methodologies: An Overview. Tech. Rep. 93-05-07, Dept. of Computer Science and Engineering, University of Washington, U.S.A. 1993.
[Henn90] Hennessy J.L., Patterson D.A., Computer Architecture: A Quantitative Approach. Morgan Kaufmann Publishers Inc., San Mateo. CA. 1990.
[Hill87] Hill D.D., Coelho D.R., Multi-level Simulation for VLSI Design. Kluwer Academic Publishers, ISBN 0-89838-184-3, 1987.
[Hoar85] Hoare C.A.R., Communicating Sequential Processes. Prentice-Hall, 1985.
[Hspi90] HSPICE User Manual, Meta-Software Inc., CA.
[Jagg90] Jagger D.V., A Performance Study of the Acorn RISC Machine. M.Sc. Thesis, University of Canterbury, New Zealand, 1990.
[John91] Johnson M., Superscalar Microprocessor Design. Prentice-Hall International, ISBN 0-13-875634-1, 1991.
[Kern88] Kernighan B.W., Ritchie D.M., The C Programming Language. Prentice-Hall, Englewood Cliffs, New Jersey, 1988.
[Kogg82] Kogge P.M., The Architecture of Pipelined Computers. Hemisphere, 1982.
[Mano84] Mano M.M., Digital Design. Prentice-Hall International, ISBN 0-13-212325-8, 1984
[Mart89] Martin A.J., Burns S.M., Lee T.K., Borkovic D., Hazewindus P.J., Design of an Asynchronous Microprocessor. Advanced Research in VLSI 1989: Proceedings of the Decennial Caltech Conference on VLSI, ed. C. L. Seitz, MIT Press, pp 351-373, 1989.
[Mart90] Martin A.J. Synthesis of Asynchronous VLSI Circuits. Formal Methods for VLSI Design, editor J. Staunstrup, North-Holland, 1990.
[Moln92] Molnar C.E., Jones I., Sutherland I.E., A Way to Compose Petri Nets. Tech. Rep. SMLI \#92:0354, Sun Microsystems Inc., October, 1992.
[Nage73] Nagel L.W., Pederson D.O., Simulation Program with Integrated Circuit Emphasis (SPICE). Report ERL-M383, University of California, Berkeley, Electronics Research Lab., 1973.
[Pave91] Paver N.C., Condition Detection in Asynchronous Pipelines. UK Patent no 9114513, October, 1991.
[Pave92] Paver N.C., Day P., Furber S.B., Garside J.D., Woods J.V., Register Locking in an Asynchronous Microprocessor. Proceedings of ICCD ‘92, pp 351-355, October, 1992.
[Pave94] Paver N.C., The Design and Implementation of an Asynchronous Microprocessor. PhD thesis, University of Manchester, June, 1994.
[Pete81] Peterson J., Petri net theory and modelling of systems. Prentice-Hall, 1981.
[Russ85] Russell G., Kinniment D.J., Chester E.G., McLauchlan M.R., CAD for VLSI. Van Nostrand Reinhold (UK), ISBN 0-442-30618-0, 1985.
[Russ89] Russell G., Sayers I.L., Advanced Simulation and Test Methodologies for VLSI Design. Van Nostrand Reinhold (Int.), ISBN 0-7476-0001-5, 1989.
[Seit80] Seitz C.L., System Timing. In Introduction to VLSI Systems, editors Mead C.A., Conway L.A., Chapter 7, Addison-Wesley, 1980.
[Suth86] Sutherland I.E., Sproull R.F., Asynchronous Systems. Sutherland, Sproull \& Associates, Palo Alto, California, September, 1986.
[Suth89] Sutherland I.E., Micropipelines. Communications of the ACM. 32(6):720-738, January, 1989.
[Thom92] Thomas D.E., Moorby P., The Verilog Hardware Description Language. Kluwer Academic Publishers, ISBN 0-7923-9126-8, 1992.
[Veri92] Verilog-XL Reference Manual, Volumes 1\&2. Cadence Design Systems Inc., 1992
[VLSI90] Acorn Risc Machine (ARM) Family Data Manual. VLSI Technology Inc., Prentice-Hall, Englewood Cliffs, New Jersey, 1990.
[Weic84] Weicker R.P., Dhrystone: A Synthetic Systems Programming Benchmark. Communications of the ACM. 27(10):1013-1030, October, 1984.
[West89] Weste N., Eshraghian K., Principles of CMOS Design - A Systems Perspective. Addison-Wesley, Wokingham, England, 1989.

## Appendix A: Verilog Model

## The following complete listing of the MDCML (bipolar) Asynchronous ARM Verilog Model contains:

Top-level processor core ..... 143
ARM functions behavioural library ..... 152
Asynchronous component library ..... 159
Standard gate functions library ..... 168
Example of PLA structure modelling ..... 173
External environment model - MMU and memory ..... 174
ALfs [9:0], NGr2, ng [5:0], Sh [31:0], ShC, SHd, IMr2, Imd [6:0],
DW[31:0], DWv, DWusr, WRa, Wsel, APa,DOa,RSa,Dabt [1:0],Rst);

EvtReg2 \# (32,7) lat1 (Add [31:0],Ctl[6:0], MEMa1, MEMr1, MAR[31:0], MAc[6:0], MEMr2, MEMa, Rst); endmodule // ARMstCore
// Primary Instruction decode logic
module Decode1 (RGr, Rs[27:0], I2r, I2[25:0], I3r, I3[19:0], Dabt, XPr,
Ir, IO [31:0], flo[1:0],Nfiq, Nirq, PCpr, RGa, I2a, I3a, XPa,
XLr, PCr, IMa, LSMPa, ALUgo,ALUok, mode[5:0],NGa, nAbt,Rst);
// Instruction Disassembler
//ARM_Dis dis (DIS, I0[31:0], Ir2, Rst);
// arbitrate the exceptions and prioritize
DLor2 sgo (nFiq, Nfiq, mode[4])
$\begin{array}{llll}\text { DLinv } & \text { inv0 } & \text { (Fiq, } & \text { nFiq); } \\ \text { DLor2 } & \text { sg1 } & \text { (nIiq, } & \text { Nirq, mode[5]); }\end{array}$
// arbitrate the exceptions and prioritize (NB mode[4] is omitted!)
DLor2 sg0 (nFiq, Nfiq, mode[4]);

 // check PC parity, and reject if wrong
DLxor2 x0
(NpOK, PAR, flo[0]);
$\begin{array}{lll}\text { Select } & \text { slo } & \text { (Ir2, Irj, } \\ \text { DLx1, Npor2 } & \text { Irst); } \\ \text { x1 } & \text { (Ia, } & \text { Irj,Ia4); }\end{array}$
request until ready
until ready
flo [0], Ir, Rst);
NAR, Ipar);
XLr, NpOK2,Rst);
DABr, XLa);
DABr,Ir2);
$\begin{array}{lll}\text { // force } \mathrm{I}[] \text { into a known state during Dabts } \\ \text { DLinv } & \text { sg4 } & (\mathrm{Z}, \\ \text { Band \#32 } & \mathrm{Iz} & (\mathrm{Iz}[31: 0], \\ \text { Dabt }) ;\end{array}$
// sync with the PC and exceptions
Cgate3 c0 (Ir1, Ir0, Ir, PCr, Rst) ;
Cgate3 c0 (Ir1,

// Micropipelined ARM
'timescale 10ps /10ps
'timescale $10 \mathrm{ps} / 10 \mathrm{ps}$
module ARMstCore (Add[31:0], Dout [31:0], Ctl[6:0],MEMr, MRRa, $\quad$ MRR[31:0], Nfiq,Nirq, Dabt[1:0], PAbt,MEMa, MRRr, BigEnd, nAbt,Rst); `include "ARMstCore.inc"
// Ctl[] = Seq, Inc, Ren, Wen, Usr, B/w, Opc
// memory data in blocks and instruction pipeline
Dat Int dat (MARa, DOa, MEMr2, MWR [31:0], MRRa, DWr, DW [31:0], DWusr,

H32 lat0 (Dout [31:0] MEMaO, MEM
Cgate2 c0 (MEMa2, MEMa0,MEMa1,Rst);
Cgate2 c1 (MEMr,
// 1st decode stage
Decode1 dec1 (RGr, Rs[27:0], IN2r, IN2[25:0], IN3r, IN3[19:0], PCsel, XPr,
XLa, INa,IMr0, LSMPr, nTRM, r15, NGr0, nGn[1:0], vect[2:0],
INr, IN [31:0], flo[1:0],Nfiq, Nirq, PCpar, RGa, IN2a, IN3a, XPa,
XLr, $\operatorname{PPr} 3$, IMa0, LSMPa, ALUgo, ALUok, mode $[5: 0]$, NGa0, nAbt,Rst);
// 1st execution and 2nd decode stage
Reg rg (RGa, RWa, RDr, $\mathrm{Na}[31: 0], \mathrm{Nb}[31: 0]$,
Reg rg (RGa, RWa, RDr, $\mathrm{Na}[31: 0], \mathrm{Nb}[31: 0]$, $\quad$ RGr, $\operatorname{Rs}[27: 0], \mathrm{nPC}[31: 2], \operatorname{RWr}, \mathrm{W}[31: 0], \mathrm{Wc}[2], \mathrm{Wsel}, \mathrm{RDa}$, Rst) ; NGen nGen (NGa0,NGr2,ng[5:0], NGr0,IN[15:0], nGn[1:0], vect[2:0],NGa2,Rst);

Decode2 dec2 (IN2a,RSa, C2r, \{Imd[6:0], SHop [9:0], DObw, C2[7:0]\}, $\operatorname{IN2r,\operatorname {IN}2[25:0],\operatorname {RS}r,\operatorname {Na}[7:0],C2a,\operatorname {Rst})\text {,}}$
IN[] $=X t[1: 0]$, PCPar, $\operatorname{cond}[3: 0], \operatorname{sct} 1 \mathrm{~s}[2: 0], I[11: 5]$, DObw,
toRs, cpCP, $\sim$ toDo, $\sim$ toA, nGen, $\sim$ Mult, NImm
$\begin{array}{ll}\text { Imd [] }=x t[1: 0], \text { PCpar, cond[3:0] } & \text { toRs, cpCP, } \sim \text { toDO, } \sim \text { toA, nGen, } \sim \text { Mult, NImm } \\ c 2[] & =\text { toRs, cpCP, } \sim \text { toDO, } \sim \text { toA, nGen, } \sim \text { Mult, NImm }\end{array}$
// 3rd decode stage
Decode3 dec3 (IN3a, C3r, ALfs [9:0], \{vec3[2:0],c3[22:0]\}, IN3r, IN3[19:0],C3a,Rst);
// c3[] = UseCP,S,F,C,Wcp[2:0],Ral,Rcnd, $\sim$ ALUwt, $\sim$ DabtWt, tPCp [1:0],
Wreg, Wadd,SP, LSM, Ren, Wen, B/W, Opc, destPC, Rsel // 3rd control and execution stages

ExecP excP (RDa,C2a,C3a,NGa2, SHe, IMa2, PCpar,ALUgo, ALUok, mode[5:0],psrC,

[^3]| // kick | off the | immediate pipeline | e if needed |
| :---: | :---: | :---: | :---: |
| Select | s16 | ( IMr, noimm, | Pri, NImm,Rst); |
| DLxor2 | $\times 9$ | ( IMx, | IMa, noImm) ; |
| // copy PC to X pipe if a load or store |  |  |  |
| Select | s17 | ( XPr , nox, | Pr1,NtoX,Rst) ; |
| DLxor2 | $\times 10$ | ( PPx , | XPa, noX) ; |
| // kick off nGen when needed |  |  |  |
| Select | s18 | ( $\mathrm{noN}, \mathrm{NGr}$, | Pri,ton, Rst) ; |
| DLxor2 | $\times 11$ | ( NGx , | NGa, noN) ; |
| // wait for ALU to complete mode change \& interrupt disable? |  |  |  |
| Select | s19 | (ALr, noA, | Pr1, nALw1,Rst) ; |
| Cgate2 | c2 | (ALg, | ALr, ALUgo, Rst) ; |
| DLxor2 | x12 | (ALx, | ALg, noA) ; |
| // wait for all outputs to complete |  |  |  |
| Cgate2 | c3 | (xpal, | ALx, $\mathrm{XPx} \times \mathrm{Rst})$; |
| Cgate2 | c4 | (imng, | NGx, IMx, Rst) ; |
| Cgate2 | c5 | (I32x, | I2x, I3x, Rst) ; |
| Cgate2 | c6 | (Ia0, | xpal,imng,Rst) ; |
| Cgate2 | c7 | (rg32, | I32x,RGx,Rst) ; |
| Cgate2 | c8 | (Ia1, | Ia0, rg32,Rst) ; |
| endmodule // Dec1 |  |  |  |


// synchronise exception requests with the instruction stream
DLinv invR (NRst,
inst);



endmodule // ArbitX
// Secondary (shift) decode logic
module Decode2 (IN2a, RSa, C2r, d2 [25:0], IN2r, IN2 [25:0], RSr, Rs [7:0], C2a, Rst);
 // Dobw, toI3, $\sim$ Rs, cpCP, $\sim$ toDo, $\sim$ toA, nGen, $\sim$ Mult, NImm EvtReg \#8 Rs1 (Rs2[7:0],RSa,RSr2, Rs[7:0],RSr,RSa2,Rst); IN2 [15], IN2[14], IN2[13], IN2[12]) ; shR[9:0]);
`include "Decode2.inc"
$\qquad$ // IN2[] = xt[1:0],PCpar, cond[3:0], sctls[2:0],I[11:5], $/ / / \mathrm{d} 2[]=\operatorname{Xt}[1: 0]$, PCpar, cond[3:0],SHop[9:0],
DObw, toI3, $\sim$ Rs, cpCP, $\sim$ toDo, $\sim$ toA, nGen, $\sim$ Mult, NImm $]$

Shfc_PLA \#100 shc (IN2[10:9],Rs2[7:0],
// the first stage decoders
// ip [] = ALct $[3: 0]$, sctls[2:0], BW
RegC_DPLA \#400 dec1 (\{nXR, Rs[4], I


// the multicycle sequencer
seqC_PLA \#200 seq (sq1[2:0],ALUok, I[24],I[21:20], P[7],r15A, Rs[4], sq2[1:0],multi);
EvtLch \#2 ell (sq0[1:0], sq2[1:0], Ia2,Rst);
$\begin{array}{llll}\text { Select } & \text { s12 } & \text { (Ia3, Irm, } & \text { Ia2, multi,Rst); ; } \\ \text { DLxor2 } & \text { x5 } & \text { (Ir4, Ir3, } & \text { Irm); }\end{array}$
// the register address muxes
$\operatorname{pc}[3: 0], \mathrm{I}[19: 16], \mathrm{I}[11: 8], \mathrm{Pc}[3: 0], \mathrm{P}[1: 0]) ;$
$\operatorname{Rd}[3: 0], \mathrm{I}[19: 16], \mathrm{I}[15: 12], \mathrm{I}[3: 0], \mathrm{P}[3: 2]) ;$
$\operatorname{Pc}[3: 0], \operatorname{Rd}[3: 0], I[15: 12], \operatorname{Pc}[3: 0], P[5: 4]) ;$
$\operatorname{pc}[3: 0], 1 \mathrm{k}[3: 0], \mathrm{I}[19: 16], \mathrm{I}[15: 12], \mathrm{P}[7: 6]) ;$

$\begin{array}{lll}\text { mux4 } & \# 4 & m \times 1 \\ \operatorname{mux} 4 & \# 4 & m \times 2 \\ \operatorname{mux} 4 \# 4 & m \times 3\end{array}$ // detect PC targets
$\operatorname{Rs}[10], \operatorname{Rs}[9], \operatorname{Rs}[8], \operatorname{Rs}[7]) ;$
$\operatorname{Rs}[14], \operatorname{Rs}[13], \operatorname{Rs}[12], \operatorname{Rs}[11]) ;$

// load \& store multiple
Select
sl3
(RdGr,

// propagate the instruction to decode 3
// I3[] $=$ vect $[2: 0]$, r15M, r15A, I[24:19,16:15], ALct [3:0], $\sim$ ALUwt, $\sim$ Dabtwt, B/W


request the register operation
$\operatorname{Rs}[]=\operatorname{RdEn}, \operatorname{ArNp}, \operatorname{Ra}[3: 0], \operatorname{Rb}[3: 0], \operatorname{Psel}[1: 0], \operatorname{LrNp}, \operatorname{Rm}[3: 0], \operatorname{Rw[3:0],} \operatorname{NBS}[6: 0]$



// the read completion and read data latches
Cgate2 abC (ABdn, Adn, Bdn, Rst);
EvtReg2 \# $(32,32)$ Dlat (Na [31:0], Nb [31:0], Rdn, Dra endmodule // Reg
module LkFIFO (LaA, LaM, Wgo, Wa $[29: 0], \operatorname{WPa}[4: 0], \operatorname{Psel}[1: 0], \operatorname{Lkd}[29: 0], \operatorname{LkP}[4: 0]$,
$\operatorname{LrA}, \operatorname{LrM}, \operatorname{lkA}[29: 0], \operatorname{lkM}[29: 0], 1$ PA $[4: 0], \operatorname{lPM}[4: 0], \operatorname{Ps} 0[1: 0]$, Wrq. Wma, Wdn, Rst) $;$ ‘include "LkFIFO.inc"
 LkReg stg2 2 $29: 0]$, P2[4:0], Ps2 [1:0], k2[29:0], Pk2[4:0], rq3,Aak, Rst) ; // the external (memory) sourced result address lock FIFO
LkReg stg
$(14[29: 0], P 4[4: 0], \operatorname{Ps} 4[1: 0], \operatorname{lk} 4[29: 0], \operatorname{Pk} 4[4: 0], \operatorname{LaM}, r q 5$,
 LkReg $\quad$ stg $6[29: 0], P 4[4: 0], P s 4[1: 0], 1 k 4[29: 0], P k 4[4: 0], r q 5, a k 6, R s t) ; ~$
 $16[29: 0], P 6[4: 0], \operatorname{Ps} 6[1: 0], 1 k 6[29: 0], \operatorname{Pk} 6[4: 0]$, rq7,Mak,Rst);
// the locked output inverters and write address multiplexers

$\begin{array}{lll}\text { mux2 \#30 mux1 } & \text { (Wa[29:0], } & 17[29: 0], 13[29: 0], \text { Wma) ; } \\ \text { m } & \text { (WPa[4:0], } & \text { P7[4:0], P3[4:0], Wma); }\end{array}$ $\begin{array}{lll}\operatorname{mux} 2 \# 5 & \operatorname{mux} 2 & (\text { WPa[4:0], } \\ \operatorname{mux} 2 \# 2 & \operatorname{mux} 3 & \text { (Psel[1:0], } \\ \text { P7[4:0],P3[4:0],Wma); } \\ \text { Ps } 7[1: 0], \operatorname{Ps} 3[1: 0], \text { Wma) }\end{array}$ // the write request steering logic
Select sel2 (Wram,Wrqa, Wra, Wma, Rst); $\begin{array}{llll}\text { Cgate2 } & \text { C1 } & \text { (Mgo, } & \text { Wrqm, Mrq, Rst); } \\ \text { Cgate2 } & \text { C2 } & \text { (Ago, } & \text { Wrqa, Arq,Rst); }\end{array}$
(Mak, Aak, Wgo, Mgo, Ago, Wdn, Rst) ; endmodule // LkFIFO
module AddInt (Wa0, Wa, APa0, PPr3,XPa0, XLr1,R15[31:2],LSMPa, MAr, MA [31:0],MAc[9:0],
Wr0, W0 [31:0],Wc0 [9:0], APr0,A $[31: 0]$, LSMPr, Ntrm,
dPC, PPa3, XPr0, XLa1, PCsel, MAa, Dabt1, Dabt0,Rst);
// Wc[] = SP, LSM, Ren, Wen, Usr, B/W, Opc, valid, destPC, PCpar
// MAC[] = Seq, Inc, Ren, Wen, Usr, B/W, Opc, valid, destPC, PCpar
Sink \#3 sk1 (\{nc0,nc1,nc2\})
// Data Abort control of X pipe

| DecWait $2 \times 1$ dw |
| :--- |$\quad$ (nDbt, XLr0, endmodule

module AddC
module AddC (Wa, PCa, APa, LSMa, LSMPa, MARr, PCPar, Usr, Adsl[1:0],
Wr, SP, LSM, PCpin, Usr0, PCr, APr, LSMr, LSMPr, Nt rm, MARa, Rst) ;
input Wr, SP, LSM, PCpin, Usr0, PCr, APr, LSMr, LSMPr, Ntrm, MARa, Rst; output [1:0] Adsl; $\begin{aligned} & \text { output Wa, PCa, APa, LSMa, LSMPa, MARr, PCpar, Usr; }\end{aligned}$
// Wr/a bundles SP (special load = LSRA or LSM), LSM, PCpin(parity in), Usr0 // LSMPr/a bundles Ntrm (= not last LSM cycle), dPC (= dest is
// MARr/a bundles PCpar, Usr, Adsl[1:0],
// arbitrate into the PC loop
ArbitA arb (Wgo, PCgo,Wmux, Wr, PCr,Wa, PCa, Rst);
// latch W sourced booleans
EvtLch el0 (PCpar, PCpin,Wgo,Rst); // latch W sourced booleans
EvtLch elo
EvtLch ell
// sort out the $W$ sourced operations // sort out the W sourced operations
Select slo
(r1, LSPr,
Select
Sle
sll $\begin{array}{llll}\text { Cgate2 } & \text { C0 } & \text { (r0, } & \text { LAr, APr, Rs } \\ \text { DLxor2 } & \text { xor0 } & \text { (x0, } & \text { APa, d1); }\end{array}$ $\begin{array}{ll}\text { // the PC processing bit } & \\ \text { Select sl2 } & \text { (r2,PCx, }\end{array}$ $\begin{array}{lll}\text { // the PC processing bit } \\ \text { Select } & \text { sl2 } & \text { (r2, PCx, } \\ \text { DLxor2 } & \times 3 & \text { (PCk, } \\ \text { DLxor2 } & \text { x } 4 & \text { (PCa, }\end{array}$

$$
\begin{aligned}
& \text { DLxor2 }{ }^{\text {x4 }} \text { (PCa, } \\
& \text { // the LSM processing bit }
\end{aligned}
$$


(LSSMa,
(LSMPa,
(r0

[^4]module MemCtl (Da, MAa,MCr, MMr, Dr, MAr, Wen, Ren, Val, Opc, dPC, MCa, MMa,Rst); input Dr, MAr, Wen, Ren, Val, OpC, dPC, MCa, MMa, Rst;

| Select | so | ( n W, w, | MAr, Wen, Rst) ; |
| :---: | :---: | :---: | :---: |
| Cgate2 | mullo | (co, | W, Dr, Rst) ; |
| Call | call | (Da, MaO, ca, | c0, now, c1,Rst) ; |
| DLxor2 | $\times 0$ | (MAa, | Ma0, Da); |
| DLor2 | go | (NReg, | Opc, dPC) ; |
| DLinv | inv | (Reg, | NReg) ; |
| DLor2 | g1 | (DoIt, | Reg, Val); |
| DLand2 | g2 | (MCen, | DoIt,Ren); |
| Select | s1 | ( $\mathrm{noR}, \mathrm{MCr}$, | ca, MCen, Rst) ; |
| Select | s2 | ( $\mathrm{nom}, \mathrm{MMr}$, | ca,Val,Rst); |
| DLxor2 | xor1 | (x1, | noR, MCa) ; |
| DLxor2 | xor2 | (x2, | nom, MMa) ; |
| Cgate2 | mullı | (c1, | x1, x2,Rst) ; |

endmodule // MemCtl

## module IMMpipe (i2[31:0],a0,r2, i0[31:0],r0, a2,Rst); input [31:0] i0; input r0,a2, Rst; output [31:0] i2; output a0,r2; wire [31:0] i1, j1;

// the immediate field extraction and pipeline:
 EvtReg \#32 lat1 (i2[31:0],a1,r2, j1[31:0],r1,a2,Rst); endmodule // Ipipe
// Micropipel ined ARM Execution Pipe
module ExecP (RDa, C2a, C3a, NGa2, SHe, IMa2, PCpar, ALUgo, ALUok, mode [5:0], Psrc,

ALfs [9:0], NGr2, ng [5:0], Shis1:0], ShC, SHa, IMr2, 1 (1:0], Rst)
(C2a, APr, DOr, RDa, IMa2, OPr0, MLe, SHe, CPr0,NGa2, RSr,

C3r, C3[22:0],CPr1,ALd,OPr1,WLa, Dabt [1:0], Pass,Usr0, Usr, Rst);
Ct12
$\stackrel{n}{\Psi}$
// the data input block, including byte selection and alignment logic
// b0[5:0] contains: byte[1:0]; Usr; Byte/Word; Valid; PCdest; // b0 [5:0] contains: byte[1:0]; Usr; Byte/Word; Valid; PCdest;
 EvtReg2 \# $(32,3)$
endmodule // Din
module Dout (D3 [31:0],a0,r3, D0[31:0],bw,r0,a3,Rst); input $[31: 0]$ D0; input bw,r0, a3,Rst
output $[31: 0]$ D3; output a0,r3; wire $[31: 0]$ D1,D1a,D2;
// the data out block, including byte replication logic
EvtReg2 \# (32,1) lat0 (D1 [31:0],b1,a0,r1, D0 [31:0],bw, r0, a1, Rst); $\begin{array}{llll}\text { Dout_blk } & \text { dout } & \text { (D1a[31:0], } & \text { D1 [31:0],b1); } \\ \text { EvtReg \#32 } & \text { lat1 } & \text { (D2[31:0], a1,r2, } & \text { D1a[31:0],r1,a2,Rst); }\end{array}$ endmodule // Dout
module DstCtl (MMa,MCa,DIr, IPr, MMr, MCr, Val,Opc, DIa, IPa, Rst); input MMr, MCr,Val,
out put MMa, MCa, DIr, IPr;
// read data destination control: sends the incoming data to Din or Ipipe
// - also sets Din up early to receive incoming data Select so (Dat, Op, MCr,Opc,Rst); $\begin{array}{lrll}\text { Select } & \text { s0 } & \text { (Dat,Op, } & \text { MCr, Opc, Rst); } \\ \text { Select } & \text { s } & \text { (nVd, Vd, } & \text { Dat,Val,Rst); }\end{array}$ $\begin{array}{ll}\text { Decwait2x1 dw } & \text { (IPr,Vd1, MMr, Op,Vd,Rst); } \\ \text { Call (b0,b1,DIr, nVd,Vd1,DIa,Rst); }\end{array}$ (MMa
endmodule // DstCtl
module MemCP (d5[7:0],a0,r5, do[7:0],r0,a5,Rst);
module MemCP (d5[7:0],a0, r5,
input [7:0] d0; input r0, a5, output $[7: 0] \mathrm{d} 5$; output $\mathrm{a} 0, \mathrm{r} 5$;
wire $[7: 0] \mathrm{d} 1, \mathrm{~d} 2, \mathrm{~d} 3, \mathrm{~d} 4$;
// the memory control pipeline
// MCp[]
(d1[7:0],a0,r1, d0[7:0],r0,a1,Rst);

endmodule // MemCP

| DLxor2 | $\times 6$ | (DOx, | DOa, noDO) ; |
| :---: | :---: | :---: | :---: |
| // the CPSR copy latch( SHr used to capture ShC ) |  |  |  |
| Select | sl6 | ( $\mathrm{noCP}, \mathrm{CPr}$, | SHr, c2 [5],Rst) ; |
| DLxor2 | $\times 7$ | ( CPx , | CPa, noCP) ; |
| Cgate2 | c5 | (CPx1, | CPx, OPa, Rst) ; |
| // the | Rs latch |  |  |
| Select | s17 | (RSr, noRS, | RDgo, c2[6],Rst); |
| DLxor2 | $\times 8$ | (RSx, | RSa, noRS) ; |
| // call operand latch |  |  |  |
| Cgate2 | c6 | (OPr, | SHx, NGgo,Rst); |
| // wait for all outputs to complete |  |  |  |
| Cgate2 | c7 | (apdo, | APx, DOx, Rst) ; |
| Cgate2 | c8 | (rsng, | NGx1,RSx,Rst); |
| Cgate 2 | c9 | (C2a, | apdo,rsng, Rst); |
| DLbuff | $\times 9$ | (RDa, | C2a); |

 input $[22: 0]$ c3;
input $[1: 0]$ Dabt;
input C3r, CPr1, ALd, OPr, WLa, Pass, Usr0, Usr, Rst;
output [12:0] alp;
output $33: 0]$ CPmx;
output C3a, CPa1,ALe
output $[3: 0]$ CPmx;
output C3a, CPa1,ALe, ALel, OPa, WLr, PCp1, PCpar, ALUgo, Dabt0, ALx;
 // alP[] = Rsel,Wreg, Wadd, SP, LSM, Ren, Wen, Usr, B/W, OpC, valid, destPC, PCpar
// the CPSR mux control
CPmxC_PLA $\# 100$ dec0 (Rst, Pass, Usr0, c3[21:16], CPmx[3:0]);
C3r, OPr, Rst) ;
C3r, OPr, Rst);
Pass);
Fail, $\mathrm{c} 3(22]) ;$
Lr0, UseCP,Rst,
CPr1, Ucp,Rst);
and


ExecDP
(MLd, nMLb [31:0],OPa0, OPr1,CPa0, Pass, CPr1, PsrC,
mode [5:0], Usr0, Usr, ALd,W[31:0],Wc[9:0],Wq[1:0],WLa, WRr,
mode [5:0], Usr0, Usr, $\operatorname{ALd}, \mathrm{W}[31: 0], \mathrm{Wc}[9: 0], \mathrm{Wq}[1: 0], \mathrm{WLa}, \mathrm{WRr}$,
$\mathrm{Na}[31: 0], \mathrm{Nb}[31: 0], \mathrm{MLe}, \mathrm{C2}[1], \mathrm{C2}[2], \mathrm{ng}[5: 0], \mathrm{Sh}[31: 0], \mathrm{OPr} 0, \mathrm{OPa} 1$,
 (ALUok, alp[2]);
excDP
DLbuff bf0
endmodule // ExecP
module Ct12 (C2a, APr, DOr, RDa, IMa, OPr, MLe, SHe, CPr, NGa, RSr,
C2r, C2[7:0],APa, DOa, RDr, IMr, OPa, MLd, SHd, CPa, NGr, RSa, ALX, Rst) ;
// Control 2 decode logic
module Ctl2 (C2a, APr, DOr, RDa,
input [7:0] c2;
input C2r, APa, DOa, RDr, IMr, OPa, MLd, SHd $, C P a, N G r, R S a, ~ A L X, R s t ; ~$
 // c2[] = toI3, $\sim$ Rs, cpCP, $\sim$ todo, $\sim$ toA, nGen, $\sim$ Mult, NImm // sync with appropriate shifter
Select input
s 10 $(\mathrm{ShIm}, \mathrm{C} 2 \mathrm{r} 1, \quad \mathrm{C} 2 \mathrm{r}, \mathrm{c} 2[0], \mathrm{Rst})$;
 // bypass for cycles which
 $\begin{array}{llll}\text { // activate multiplier if required } \\ \text { Select } & \text { s12 } & \text { (MLe, nML, } & \text { SHg0, c2[1], Rst); } \\ \text { DLxor2 } & \text { x1 } & \text { (SHg2, } & \text { nML, MLd); }\end{array}$
merge calls and activate shifter SHa) ;
ALx) ;
SHgo, nSHa, nALx, Rst);
SHst, CPx1);
SHst, (Pxa1);
SHd, Rst);
RDx, SHx0);
SHgO, Shim);
NGO, c2 22$],$ Rst);

NG1, noNG, CPx1,Rst);
NG2,NGa);
C2r, RDr,Rst) ;
RDgo, c2[3],Rst);
APa, noA);
RDgo, c2[4],Rst); (nSHa,
(nALx,
SHst,
SHe,
SHr, SHa,
(SHx,

(RDgo,
(APr,
(APx,
n
// the data out pipe
Select
sl5 5 (Dor, noDo,



input [1:0] Wc;
input Dr, destPC, Val, Wr, RWa, ADa, WLx, Rst;
output Da,Wa, RWr, ADr, Wmux;
// Wc[] = Wreg, Wadd
Arbita arb

$\begin{array}{ll}\text { //W bus deadlock avoidance } \\ \text { Select s4 } & \text { (nWA, WAr, } \\ \text { Cgate2 co } & \text { (WAok, } \\ \text { DLinv go } & \text { (Lok, } \\ \text { DLxor2 xr3 } & \text { (Wr1, } \\ \text { DecWait } 2 \times 1 \text { dwo } & \text { (Dd1, Wd1, } \\ \text { DLxor2 xr4 } & \text { (x4, } \\ \text { Cgate2 c2 } & \text { (dwf, } \\ \text { DLinv inv1 } & \text { (Nx4, } \\ \text { DLxor2 xr6 } & \text { (Da, } \\ \text { endmodule // WrCtl } & \end{array}$
$\mathrm{v} 1=(\mathrm{A}<0 \& \& \mathrm{~B}<0 \& \&$ out $>=0) \quad| | \quad(\mathrm{A}>=0 \& \& \mathrm{~B}\rangle=0 \& \&$ out $<0)$
delay $=$ ALU_Add;
default: begin
Sdisplay("Illegal ALU function code\n");
\$display("Current simulation time is \%d\n", \$time); out $=\sim 0$; nzcv $=(($ out $<0) \ll 3)+(($ out $==0) \ll 2)+(c 1 \ll 1)+\mathrm{v} 1 ;$
\#(delay) alu = out;
flags = nzcv;
\#100 rout = 1;
end s @(negedge rin)
fork
\#100 alu = `bx; \#100 flags = `bx;
\#(`ALU_Recover) r join \#(`ALU_Recover) rout $=0$; endmodule
 -define MultDin_Dout 27
-define MultMux_Dout 43
'define MultRst Rout 47 'define MultRst_Rout 47
'define MultCompute 2500 output [31:0] outA, outB;
reg [31:0] outA, outB; reg [31:0] outA, outB;
output rout; reg rout;
input $[31: 0]$ inA, inB;
input rin, Mmux, rst; input $[31: 0]$ inA, $1 n B ;$
input rin, Mmux, rst;
reg [31:0] sum, carry;
'define Mult_reset \#(`MultRst_Rout) rout=0; 'define Mult_undef begin outA='bx; outB='bx; rout=1'bx; end always @(rst) case (rst) \(\qquad\) 1'bx: 'Mult_undef 1: 'Mult_reset b0: begin \#('MultDin (atB \(=\) inB \(\left.\quad \begin{array}{l}\text { \#( MultDin_Dout) outA }=\text { inA; } \\ \text { outB }=\text { inB; }\end{array}\right]\) endcase \(\qquad\)    module ALU(alu[31:0], flags [3:0], rout, \(\operatorname{ain}[31: 0], \operatorname{bin}[31: 0], f u n c[9: 0], \operatorname{psrC}, \mathrm{psrV}, \mathrm{shC}, \mathrm{rin})\); define ALU_Add 800 -define ALUL Logic 400 def fine ALU_Recover 300 output [31:0] alu; reg [31:0] alu; output :3 reg put:0] alu; output \([3: 0]\) flags; reg \([3: 0]\) flags; reg rout; bin; input \([31: 0]\) ain, bin; input \([9: 0]\) func; input psrC, \(p s r V, s h C, r i n ;\) input psrC, psrV,shC,rin; integer A, B, out; eg \([31: 0]\) delay; eg \([9: 0] \mathrm{f}\); eg \([3: 0] \mathrm{nzcv}\); reg \([3: 0]\) nzcv; reg \(c 0, c 1, v 1 ;\) always @(posedge rin) delay \(=`\) ALU_Logic;
$\mathrm{A}=0 ; \mathrm{B}=0 ; \mathrm{c} 1=\operatorname{shC} ; \mathrm{v} 1=\mathrm{psrV} ; \mathrm{f}=\sim$ func; $\mathrm{nzcv}=0 ;$

$\begin{aligned} \text { if (f \& `h20) } & B=b i n ; \\ \text { else } & \text { if (f \& `h10) } \\ \text { B } & =\sim \operatorname{bin} ;\end{aligned}$
$\mathrm{B} ;$
$\mathrm{B} ;$
$\mathrm{B} ;$
1/A xor $B$
//A or $B$
// $A$ and $B$
// A + B
in logic

\$display("Illegal carry select in ALU\n");
\$stop (1);


2000 ®
end
 'define ShCompute 450
'define ShRecover 100


end
always @(nPC)
begin
(`REGdn_delay) $A d n=0 ; B d n=0$; end $A d n=$



©
always @(enA)
begin
// register read A

## dA $=$ un $2 \mathrm{bin}(\mathrm{en} A) ;$ $\left(r d A!=6^{\prime} h 3 F\right)$ <br> \# (`REGrd_delay) $\mathrm{Na}=\operatorname{Rbnk}[\mathrm{rdA}]$;

end
end

rdPS
if (
end

$r d B=u n 2 b i n(e n B) ;$
if (rdB!=6'h3F)
\# ( VREGrd
Bdn $=1 ;$
end
always @(negedge Ren)

" 1111 always @ (enW)
begin
reg [31:0] Rbnk [0:36]; module REG(Na,Nb, Adn, Bdn, W, nPC, enA, psA, enB, enW, psWm, psWf,Ren);

[^5]f (wrw!=6'h3F)

end
endmodule // REG
 module WriteEn (enW, psWm, psW
'define WriteEn_delay 20
output [29:0] enW;
output $[4: 0]$ psWm,psWf;
reg $[4: 0]$ psWm, psWf;
reg [4:0] psWm,psWf;
input [29:0] Wa;
input $[4: 0]$ WPa;
input $[1: 0]$ Psel;
input Wen;
always @(Wa or WPa or Psel or Wen)
if (Wen)
\# ( 'WriteEn_delay) enW $=$ Wa;
\#('WriteEn_delay) enW = 0;
if (Wen \&\& (Psel\&2))
\#('WriteEn_delay) psWf = 0;

-define Din_delay 73
module Din_blk (out, in, byte, bw);
define Din_delay 73
output [31:0] out;
reg $[31: 0]$ out;
reg [31:0] out;
input [31:0] in;
input bw;
reg [31:0] val, sh;
$\underset{\substack{\text { always } @(i n \\ \text { begin }}}{\text { or byte or bw) }}$
sh $=$ byte * 8 ,
val $=($ sh>0 $) ?(($ in $\gg \operatorname{sh}) \mid($ in<< $(32-s h))):$ in;
\#( Din_delay $)$ out $=b w ?(($ in>>sh $) \&$ hFF $):$ val;
 //*********************, in, bw); `define Dout_delay 35 output [31:0] out; reg [31:0] out; input [31:0] in; reg [31:0] out; input [31:0] in; input bw; always \(@(i n\) or bw) begin end \#(`Dout_delay) out = bw ? ((in \& 'hFF) * 'h01010101) : in;
endmodule

module Immext (out, in);
`define Imm_delay 73 output \([31: 0]\) out; reg \([31: 0]\) out; input \([31: 0]\) in; reg \([31: 0]\) val;   end  // NOTE: Implement using Carry-Save adders ?? `define NGenRst_Rout 50
©define NGenRin_Rout 100
output [3:0] out;
reg [3:0] out;
reg [3:0] out;
output rout;
reg rout;
input [15:0] in;
input rin, rst;
input rin, rst;
reg [4:0] cnt, index;
define NGen_reset begin out='bx; \# (`NGenRst_Rout) rout=0; end ©define NGen_undef begin out='bx; rout='bx; end always @(rst) always \(@(\) rst) case (rst) 1'b1: 1'bx: endcase 'NGen_reset 'NGen_undef \(\stackrel{\text { begin }}{\text { if }} \underset{\text { (rin }===1 ' b x)}{\text { 'NGen_undef }}\) always \(@(\) in \()\) case \(((\) in>>26) \& 3\()\) // LDR/STR 12-bit offset // Sign-extend branch offset / 8-bit immediate  module Incre(out[31:0], in[31:0], rin, rout, aout, rst); 'define IncDin_Dout 27 'define IncDin_Dout 27 'define IncRst_Rout 47 define IncRst_Rout 47 'define IncRin_Rout 400 'define IncAout_Rout 750 output \([31: 0]\) out; reg [31:0] out; output rout; reg rout, ack, req; input [31:0] in; input rin, aout, rst; 'define Inc_reset begin ack=1; req=0; out=in; \#('IncRst_Rout) rout=0; end 'define Inc_undef begin ack=1; req=0; out='bx; rout='bx; end output \([31: 0]\) out; reg [31:0] out; output rout; reg rout, ack, req; input [31:0] in; input rin, aout, rst; 'define Inc_reset begin ack=1; req=0; out=in; \#('IncRst_Rout) rout=0; end 'define Inc_undef begin ack=1; req=0; out='bx; rout='bx; end output \([31: 0]\) out; reg [31:0] out; output rout; reg rout, ack, req; input [31:0] in; input rin, aout, rst; 'define Inc_reset begin ack=1; req=0; out=in; \#('IncRst_Rout) rout=0; end 'define Inc_undef begin ack=1; req=0; out='bx; rout='bx; end output \([31: 0]\) out; reg [31:0] out; output rout; reg rout, ack, req; input [31:0] in; input rin, aout, rst; 'define Inc_reset begin ack=1; req=0; out=in; \#('IncRst_Rout) rout=0; end 'define Inc_undef begin ack=1; req=0; out='bx; rout='bx; end output [31:0] out; reg [31:0] out; output rout; reg rout, ack, req; input [31:0] in; input rin, aout, rst; 'define Inc_reset begin ack=1; req=0; out=in; \# (`IncRst_Rout) rout=0; end
'define Inc_undef begin ack=1; req=0; out='bx; rout='bx; end define Inc_und






val
\[

$$
\begin{aligned}
& \text { if (in \& `h800000) } \\
& \text { val = val | 'hFF000000; } \\
& \text { \#(`mm_delay) out = val; }
\end{aligned}
$$
\]

default: \#(`Imm_delay) out = (in \& 'hFF); // 8-bit immediate endmodule
2'b01:
2'b10: begi
mm_delay) out $=($ in \& 'hFFF $) ;$
n
val $=($ in \& 'hFFFFFF);
val $=$ (in \& 'hFFFFFF);
if (in \& 'h800000)
endcase
endmodule





always @(req1 or req2)
if (!rst)

'define arbit_reset fork

$$
\begin{aligned}
& \text { \# ( 'rst_ID) active=0; } \\
& \text { caller=0; } \\
& \text { \# (`rst_G1) grant1 }=0 ; \\
& \text { \# ( `rst_G2) grant2=0; }
\end{aligned}
$$

'define arbit_undef fork active='bx;caller=0; grant1='bx; grant2='bx; join
always @(rst)
$\begin{aligned} \text { case (rst) } & \\ \text { 1'b1: } & \text { `arbit_reset } \\ \text { 1'bx: } & \text { `arbit_undef } \\ \text { endcase } & \end{aligned}$
\$display("\%m: Arbiter done1 clash!!");
\$display("Time=\%t", \$time);

## 

 //*************************************
module CP_Reg (out, in, capture, pass);
parameter width $=1$;
`define data_delay 32 output [width-1:0] out; output [width-1:0] out; reg [width-1:0] out; input [width-1:0] in; \(\begin{aligned} & \text { always @ (in) } \\ & \text { if (capture }==\text { pass) } \\ & \text { \# (data_delay) out }=\text { in; }\end{aligned}\) always @(capture or pass) if (capture \(==\) pass) \#(pass_delay) out \(=\) in;  module DecWait 2 x (out1, out 2 , req, c1, c2, rst); 'define DWrst_delay 29 'define DWreq_delay 85 'define DWcx_delay 47    output out1, out2;  g reg req_in, c1_in, c2_in, block; DW2x1_reset fork req_in \(=0 ;\) define DW \(2 \times 1\) _undef begin req_in \(=0 ; c 1 \_\)in \(=0 ; c 2 \_i n=0 ; b l o c k=0 ;\) out1 \(=~ ` b x ; o u t 2=\) 'bx; end
initial
$\begin{aligned} & \text { nitial } \\ & \text { if (rst) } \\ & \text { 'DW2x1_reset }\end{aligned}$
always @(rst)
case (rst)
'DW2x1_reset
s @(req)
if (req $==1$ 'bx)
'DW $2 \times 1$ _undef
else if (!rst)

else if (!rst)
begin
fork i
エ)
for
$\stackrel{+}{4}$

$$
\begin{aligned}
& \text { join } \\
& \text { end }
\end{aligned}
$$

ways $@(c 1)$
if $(c 1===1$ 'bx)
'DW $2 \times 1$ _undef

reg pass, inRdy;
time Datat, ReqT, BundT, MinT;
`define Ereg_reset fork pass=1; inRdy=0; \#('EregRst_delay) out=in; Ain=0; Rout=0; join -define Ereg_reset fork pass=1; inRdy=0; \#(`EregRst_delay) out=in; Ain=0; Rout=0; join
`define Ereg_undef begin pass=1; inRdy=0; out='bx; Ain='bx; Rout='bx; end initial begin ReqT \(=0 ;\) Datat \(=0 ;\) BundT \(=0 ;\) MinT \(=99999999 ;\)     ReqT = \$time; if ((ReqT - DataT) < MinT) Mint \(=(\) ReqT - DataT \() ;\) Bund \(=\) ReqT; end BundT \(=\) ReqT if (Rin === \(\begin{aligned} & \text { 'bx) } \\ & \text { 'Ereg_undef } \\ & \text { inRdy }=1 \text {; }\end{aligned}\) ind \#( (VregDin_Dout) out = in; \#(`EregRin_Rout) Rout $=\sim$ Rout;
\#(`EregRin_Ain) Ain \(=\sim\) Ain; pass \(=0 ;\) inRdy \(=0 ;\) end initial MinT = 99999999; \(\quad / /\) some arbitrary large value !! (in) if (!rst) Datat \(=\) \$time; if (inRdy) \$display ("Time= \(\%\)  end \#(`EregDin_Dout) out = in;
always @(Rin)
always ${ }^{@(R i n)}$
if (!rst)
begin
always @(Aout)
if (!rst)
begin


end

| ```//***************************************************************************// // 2-stage FIFO Pipeline //******************************************************************************/ module Pipe2(out, Ain, Rout, in, Rin, Aout, rst);``` |  |
| :---: | :---: |
|  |  |
|  |  |

> $\begin{array}{llll}\text { EvtReg \#32 } & \text { e1 } & \text { (out1, Ain, R1, in, Rin, A1, rst); } \\ \text { EvtReg \#32 } & \text { e2 } & \text { (out, A1, Rout, } & \text { out1, R1, Aout, rst); }\end{array}$
> output [31:0] out
output Ain, Rout;
> input [31:0] in;
input Rin, Aout, rst
> input Rin, Aout, rst;
wire [31:0] out1;
wire A1, R1; output [31:0] out; output
output Ain
input [31:
input Rin,

endmodule
// 3-stage FIFO Pipeline
//************************************************************************************)
module Pipe3(out, Ain, Rout, in, Rin, Aout, rst); module Pipe3(out, Ain, Rout, in, Rin, Aout, rst);
output [31:0] out;
output Ain, Rout;
input [31:0] in;
input Rin, Aout, rst;
wire [31:0] out1, out2;
wire A1, A2, R1, R2;
in, Rin, A1, rst);
out1, R1, A2, rst);
out2, R2, Aout, rst);
e1 (out1, Ain, R1,
e2 (out2, A1, R2,
e3 (out, A2, Rout,
\# (`LregAout_Dout) outP = inP; \#(`LregOrB_delay) $1 P O=($ outP $\mid$ lPI $) ;$
\# \#(`LregOrB_delay) lPO = (outP | lPI)
join
endmodule
join // 2-stage FIFO Pipeline
//******************************************************)
module Pipe2 (out, Ain, Rout, in, Rin, Aout, rst);

EvtReg \#32
EvtReg \#32
EvtReg \#32 endmodule



| output out; input a, b, c; wire delb, delc; |  |  |
| :---: | :---: | :---: |
| assign \#('xor3C_delay-'xor3A_delay) | delc $=\mathrm{c}$; |  |
| assign \#('xor3B_delay-'xor3A_delay) | delb $=$ b; |  |
| assign \#('xor3A_delay) | out = (a ^ delb ^ delc); |  |
| endmodule |  |  |
| //******************************************************************// |  |  |
| // Differential logic 4-input And gate |  |  |
| module DLand4 (out, a, b, c, d); |  |  |
| `define and4A_delay 29} \\ \hline \multicolumn{2}{\|l|}{'define and4B_delay 48} \\ \hline \multicolumn{2}{|l|}{`define and4C_delay 58 |  |  |
| {`define and4D_delay 70} \\ \hline \multicolumn{2}{\|l|}{output out;} \\ \hline \multicolumn{2}{|l|}{input a, b, c, d;} \\ \hline \multicolumn{2}{|l|}{wire delb, delc, deld;} \\ \hline assign \#(`and4D_delay-`and4A_delay) & deld \(=\) d; \\ \hline assign \#(`and4C_delay-`and4A_delay) & delc \(=\mathrm{c}\); \\ \hline assign \#(`and4B_delay-`and4A_delay) & delb \(=\) b; \\ \hline assign \#(`and4A_delay) |  | out $=(\mathrm{a} \& \mathrm{delb} \& \mathrm{delc} \& \mathrm{deld})$; |
| endmodule |  |  |
| //*****************************************************************/ |  |  |
|  |  |  |
| module DLor4 (out, a, b, c, d); |  |  |
| `define or4A_delay 29} \\ \hline \multicolumn{2}{\|l|}{`define or4B_delay 48 |  |  |
| `define or4C_delay 58} \\ \hline \multicolumn{2}{\|l|}{`define or4D_delay 70 |  |  |
| output out; |  |  |
| input a, b, c, d; wire delb, delc, deld; |  |  |
|  |  |  |
| assign \#('or4D_delay-'or4A_delay) | deld $=$ d; |  |
| assign \#('or4C_delay-'or4A_delay) | delc $=\mathrm{c}$; |  |
| assign \#('or4B_delay- 'or4A_delay) | delb $=\mathrm{b}$; |  |
| assign \#(`or4A_delay) | out $=(\mathrm{a} \mid$ delb $\mid$ delc $\mid$ deld); |  |
| endmodule |  |  |
| //*******************************************************************/ |  |  |
| // 2:1 Multip | r (any width) // |  |
|  |  |  |

output out;
input a, b;
wire delb;
endmodule
'define and3A_delay 28
-define and3B_delay 40
'define and3C_delay 52
input $a, b, c ;$
wire delb, delc;
$\begin{array}{ll}\text { assign \#(`and3C_delay-'and3A_delay) } & \text { delc }=\mathrm{c} ; \\ \text { assign \#(`and3B_delay-'and3A_delay) } & \text { delb }=\mathrm{b} ; \\ \text { assign \#(`and3A_delay) } & \text { out }=(\mathrm{a} \& \text { delb \& delc); }\end{array}$
endmodule
 //************************ b (out,
module DLor3 (out,
-define or3A_delay 27
'define or3B_delay 44
input a, b, c;
wire delb, delc
assign \#(`or3C_delay- assign \#(`or3B_delay-
assign \#(`or3A_delay)
endmodule
 module DLxor3 (out, $\mathrm{a}, \mathrm{b}, \mathrm{c}$ );
'define xor3A_delay 26
'define xor3B_delay 44
'define xor3C_delay 59
parameter width $=1$ ；

always＠（a）
if（！sel）
（ ${ }^{\left.m u x 4 s e l 1 \_d e l a y\right) ~ o u t ~}=a$
ï $\ddot{0}$ ö
䓂萻萻萻
always＠（a）
if $\left(\mathrm{sel}===2^{\prime} \mathrm{b} 00\right)$
\＃（｀mux4data＿delay）out $=\mathrm{a} ;$
always＠（b）
if（sel $===2$＇b01）
\＃（｀mux4data＿delay）out $=\mathrm{b} ;$ always＠（c）
if（sel $===2$＇b10）
\＃（＇mux4data＿delay）out $=c ;$
always＠（d）
if（sel $===2^{\prime}$＇b11）
\＃（ $\mathbf{M u x}^{2}$ data＿delay）out $=d ;$


## \＃（ $\operatorname{mux} 4$ sel0＿delay） \＃（ $\operatorname{mux} 4 \operatorname{sel} 10 \_$delay）  default：out $=~ ' b x ; ~$ endcase

endmodule

 module LatchR（out，in，enable，rst）；
parameter width＝1； ＇define LatchRrst＿delay 47 ne LatchR output［width－1：0］out；
reg［width－1：0］out；
input［width－1：0］in；
input enable，rst； input［width－1：0］in
input enable，rst；

[^6]
\[

$$
\begin{aligned}
& \text { initial } \\
& \text { out }=(a \& b) \text {; } \\
& \text { always @(a) }
\end{aligned}
$$
\]

$$
\begin{aligned}
& \text { input [width-1:0] } a, b ; \\
& \text { initial } \\
& \quad \text { out }=(a \& b) ;
\end{aligned}
$$

$$
\begin{gathered}
\text { \# ( `Band } \\
\text { always @ (b) } \\
\text { \# ( 'Band }
\end{gathered}
$$

endmodule


## endmodule

(enable)
\#( 'LatchRenb_delay) out $=$ in;
1'b0:
1'b1:
1'bx:
endcase
s @(enable)
if (!rst)
case(en
1'b
l'bx
endcase
$\begin{aligned} & \text { always @(in) } \\ & \text { if (!rst \&\& enable) } \\ & \text { \#(`LatchRdata_delay) out }=\text { in; }\end{aligned}$
endmodule
 module Latch(out, in, enable);
parameter width $=1$;
-define Latchdata_delay 33
output [width-1:0] out;
reg [width-1:0] out;
output [width-1:0]
reg [width-1:0] out;
input [width-1:0] in;
input enable;
initial (enable)
if
out $=$ in;
always @(enable)
always @(enable)
case (enable)
1'b1:
1'b1:
1'bx:
endcase
always @(in)
endmodule
 module Band (out, a, b);
'define BandA_delay 26
'define BandB_delay 42
output [width-1:0] out;
reg [width-1:0] out;

// Mode decode PLA model - OR plane
11111111111
?1?1?1?1?11
1111111111
11????????
????1111???
$1111111111 ?$
$1111111 ? ? 1$



${ }^{\text {begin }}$ Sdisplay (" ${ }^{\star \star \star}$ NIQ interrupt in \%0d ns ***", Nval);

end
$\stackrel{\square}{\circ}$
营
әт npoupuә
input Wrq, Rst;
reg [7:0] char;
reg [31:0] Fval, Nval;
define Tube_reset
reg [31:0] Fval, Nval;
'define Tube_reset begin Niq=1; Fiq=1; end
'define Tube_undef begin Niq='bx; Fiq='bx; end define END_OF_RUN_CHAR
define SIM_TIME_CHAR
define NS
always @(Rst)

endcase
always @(Wrq)
if (Addr $[15: 13]==3^{\prime} \mathrm{b} 110$ )



[^0]:    1. Fundamental mode operation requires that a circuit achieve a stable internal state after every individual input signal
[^1]:    1. Since Micropipeline control circuits are usually concerned with the control of datapath elements using the bundleddata convention, this example circuit is not typical of those found in an asynchronous Micropipelined microprocessor.
[^2]:    1. Verilog is a trademark of Cadence Design Systems, Inc.
[^3]:    $\mathrm{WRr}, \mathrm{W}[31: 0], \mathrm{Wc}[9: 0], \mathrm{Wq}[1: 0], \operatorname{APr}, \operatorname{DOr}, \operatorname{NMLb}[31: 0], \operatorname{RSr}, \operatorname{Dabt} 0$,
    $\operatorname{RDr}, \mathrm{Na}[31: 0], \operatorname{Nb}[31: 0], \mathrm{C} 2 r, \operatorname{c2}[7: 0], \mathrm{C} 3 r, \mathrm{c} 3[22: 0], \operatorname{vec} 3[2: 0]$,

[^4]:    // produce the address mux controls
    AddMxCtl_PLA \#100 mxctl (Wmux,LSMmx,SP,LSM, Adsl[1:0]);
    endmodule // AddC

[^5]:    function [5:0] un2bin;
    input [31:0] unary;
    reg [5:0] bin, index;
    begin
    // convert unary representation to binary
    (index=0; index<32; index=index+1)
    if (unary \& $(1 \ll$ index))
    bin $=$ index; bin $=$ index
    un2bin $=$ bin;

[^6]:    ial
    if
    else else if（enable）
    out $=$ in；
    always＠（rst）
    case（rst）

