DESIGN AND IMPLEMENTATION OF AN ASYNCHRONOUS NOC ROUTER USING A TRANSITION–SIGNALING BUNDLED–DATA PROTOCOL

MATHEUS GIBILUKA
END OF TERM WORK

ADVISOR: PROF. DR. NEY LAERT VILAR CALAZANS
CO–ADVISOR: PROF. MSc. MATHEUS TREVISAN MOREIRA

Porto Alegre
2013
MATHEUS GIBILUKA

DESIGN AND IMPLEMENTATION OF AN ASYNCHRONOUS NOC ROUTER USING A TRANSITION–SIGNALING BUNDLED–DATA PROTOCOL

End of Term work presented as part of the activities to obtain a degree of Computer Engineering at the Faculty of Engineering in the Pontifical Catholic University of Rio Grande do Sul.

ADVISOR: PROF. DR. NEY LAERT VILAR CALAZANS
CO–ADVISOR: PROF. MSc. MATHEUS TREVISAN MOREIRA

Porto Alegre
2013
Current silicon technologies enable the integration of billions of transistors in a single chip, supporting the creation of complex systems on a chip (SoCs). Networks on Chip (NoCs) constitute a suitable alternative for traditional SoC interconnect architectures, as they provide a high level of scalability and parallelism, supporting the ever-increasing number of cores in single chip. Additionally, synchronous design issues that were easily overcome in previous decades - such as clock distribution, skew, and power consumption - are becoming increasingly complex to solve in modern state of the art technology designs. Together, these trends constitute a good motivator for the development of an asynchronous SoC interconnect architecture. This work presents the design and implementation of an asynchronous NoC router using a transition-signaling bundled-data protocol. Additionally, a methodology for synthesis of bundled-data circuits using commercial CAD tools, together with an automated environment for enforcing relative timing constraints, is proposed. The router design was validated through behavioral simulation, and its basic block (a port) was synthesized, validating the implementation through post-synthesis simulation.

Keywords: Networks on chip, asynchronous circuits.
“Stay hungry, stay foolish.”

Steve Jobs
CONTENTS

LIST OF ABBREVIATIONS .......................................................................................................................... XI

LIST OF FIGURES ....................................................................................................................................... XIII

LIST OF TABLES ........................................................................................................................................ XVII

LIST OF EQUATIONS ................................................................................................................................. XVII

1. INTRODUCTION .................................................................................................................................... 1

2. CONCEPTS ............................................................................................................................................. 5
   2.1 Networks on Chip ............................................................................................................................... 5
   2.1.1 Hermes NoC ................................................................................................................................ 5
   2.2 Asynchronous Circuits ...................................................................................................................... 6
   2.2.1 Delay Models ............................................................................................................................... 6
   2.2.2 Handshake Protocols .................................................................................................................... 7
   2.2.3 Bundled-Data ............................................................................................................................... 9
   2.2.4 Asynchronous Logic Components .............................................................................................. 10
   2.2.5 High Performance Pipelines ........................................................................................................ 10

3. STATE OF THE ART ............................................................................................................................. 13
   3.1 The NoC of Ghiribaldi et al. ............................................................................................................ 13
   3.2 The ANoC ....................................................................................................................................... 14
   3.3 The MANGO NoC ............................................................................................................................ 14
   3.4 The Asynchronous QNoC ................................................................................................................ 14
   3.5 The ASPIN NoC .............................................................................................................................. 14
   3.6 The Hermes–A and Hermes–AA NoCs ............................................................................................ 15

4. DESIGN .................................................................................................................................................. 17
   4.1 Specification .................................................................................................................................... 17
   4.1.1 Target Architecture .................................................................................................................... 17
   4.1.2 BaT–Hermes Specification .......................................................................................................... 18
   4.1.3 Timing Constraints ...................................................................................................................... 19
   4.2 Architecture ................................................................................................................................... 19
   4.2.1 Transition Merger and Phase Matcher ...................................................................................... 19
   4.2.2 Input Interface ........................................................................................................................... 21
   4.2.3 FIFO .......................................................................................................................................... 23
   4.2.4 Input Buffer Control ................................................................................................................... 25
   4.2.5 Routing Control .......................................................................................................................... 27
   4.2.6 Output Interface .......................................................................................................................... 28
   4.2.7 Output Control ............................................................................................................................ 30
   4.2.8 Arbiter ....................................................................................................................................... 32
   4.3 Functional Validation ....................................................................................................................... 33
   4.3.1 Simulation Wrappers ................................................................................................................ 33
   4.3.2 Simulation Results ...................................................................................................................... 34

5. IMPLEMENTATION .............................................................................................................................. 37
   5.1 Relative Timing Constraints ............................................................................................................ 37
   5.1.1 Enforcing Relative Timing Constraints on Synopsys Tools ..................................................... 37
   5.1.2 Automated Constraint Enforcement ......................................................................................... 39
   5.2 Synthesis ...................................................................................................................................... 41
   5.2.1 Logic Synthesis Flow ................................................................................................................. 42
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACDC</td>
<td>Asynchronous Constraints for Design Compiler</td>
</tr>
<tr>
<td>DI</td>
<td>Delay Insensitive</td>
</tr>
<tr>
<td>EDA</td>
<td>Electronic Design Automation</td>
</tr>
<tr>
<td>FIFO</td>
<td>First In First Out</td>
</tr>
<tr>
<td>FSM</td>
<td>Finite State Machine</td>
</tr>
<tr>
<td>GALS</td>
<td>Globally Asynchronous Locally Synchronous</td>
</tr>
<tr>
<td>GAPH</td>
<td>Grupo de Apoio ao Projeto de Hardware</td>
</tr>
<tr>
<td>IC</td>
<td>Integrated Circuit</td>
</tr>
<tr>
<td>II</td>
<td>Input Interface</td>
</tr>
<tr>
<td>IP</td>
<td>Intellectual Property</td>
</tr>
<tr>
<td>ITRS</td>
<td>International Technology Roadmap for Semiconductors</td>
</tr>
<tr>
<td>MUTEX</td>
<td>Mutual Exclusion</td>
</tr>
<tr>
<td>NoC</td>
<td>Network-on-Chip</td>
</tr>
<tr>
<td>NRZ</td>
<td>Non Return-to-Zero</td>
</tr>
<tr>
<td>OI</td>
<td>Output Interface</td>
</tr>
<tr>
<td>PVT</td>
<td>Process, Voltage, Temperature</td>
</tr>
<tr>
<td>QDI</td>
<td>Quasi-Delay-Insensitive</td>
</tr>
<tr>
<td>RTZ</td>
<td>Return-to-Zero</td>
</tr>
<tr>
<td>SI</td>
<td>Speed Independent</td>
</tr>
<tr>
<td>SoC</td>
<td>System on a Chip</td>
</tr>
<tr>
<td>ST</td>
<td>Self Timed</td>
</tr>
<tr>
<td>STA</td>
<td>Static Timing Analysis</td>
</tr>
<tr>
<td>YeAH!</td>
<td>Yet Another Hermes</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES**

Figure 1.1 - Example of (a) a synchronous circuit, and (b) an asynchronous circuit. CLi represent combinational logic blocks, R symbolizes registers, and CTRL indicates control logic. Adapted from [SPA01]. .................................................................................................................. 2

Figure 2.1 - 3x3 Hermes NoC, connected in a 2D mesh topology. Adapted from [MOR04]............. 5

Figure 2.2 - A circuit fragment with gate and wire delays [SPA01]. .................................................. 7

Figure 2.3 - Handshake operations using (a) a level–signaling protocol and (b) a transition–signaling protocol. Adapted from [SPA01]........................................................................................................... 8

Figure 2.4 - Handshake channel types: (a) Nonput; (b) Push; (c) Pull; (d) Biput. .............................. 8

Figure 2.5 - An example 8–bit bundled–data channel scheme and its operation: (a) a bundled–data push channel; (b) an example of a level signaling communication; (c) an example transition–signaling communication. Adapted from [SPA01]. ................................................................................................. 9

Figure 2.6 - An abstract view of pipelines: (a) a synchronous pipeline; (b) an asynchronous pipeline. Adapted from [NOW11].................................................................................................................. 11

Figure 2.7 - Two transition–level asynchronous pipelines: (a) Sutherland’s; (b) Mousetrap. Adapted from [NOW11].................................................................................................................. 11

Figure 2.8 - Example of the Mousetrap pipeline operation for Stage 2 signals. ............................... 12

Figure 2.9 - Nonlinear pipelines: (a) Example of a nonlinear pipeline structure; (b) a Mousetrap fork stage; (c) a Mousetrap join stage. Adapted from [SIN07]................................................. 12

Figure 4.1 - Architecture of the YeAH! router .................................................................................. 17

Figure 4.2 - Detailed view of input (a) and output (b) interfaces of the YeAH! router .................. 18

Figure 4.3 - Top view of a) Input Interface and b) Output Interface of BaT–Hermes ..................... 18

Figure 4.4 - Example of a mutually exclusive pipeline fork with a) transition merging and b) phase–matching circuits and pipeline join with c) phase–matching and d) transition merging circuits. 20

Figure 4.5 - Waveform showing the operation of the a) transition merger and the b) phase matcher circuit. ................................................................................................................................. 21

Figure 4.6 - Architecture of the Input Interface ............................................................................. 22

Figure 4.7 - Waveform showing the operation of the Input Interface on the Local port (4) of a BaT–Hermes router with address x11............................................................... 23

Figure 4.8 - Building blocks of the FIFO proposed in [GHI13]: a) write controller; b) read controller; c) 1-hot ring counter. ........................................................................................................... 24

Figure 4.9 - Architecture of the transition–signaling bundled–data FIFO proposed in [GHI13]........ 25

Figure 4.10 - Waveform showing the operation of a 2-place FIFO .............................................. 25

Figure 4.11 - Architecture of the Input Buffer Control .................................................................. 26

Figure 4.12 - Waveform showing the Input Buffer Control Operation .......................................... 27

Figure 4.13 - Routing Control Architecture .................................................................................... 28

Figure 4.14 - Waveform showing the operation of the Routing Control at a local port on a BaT–Hermes router of address x11 ................................................................. 28

Figure 4.15 - Architecture of the Output Interface ...................................................................... 29

Figure 4.16 - Waveform showing the Output Interface communicating with two Input Interfaces .. 30

Figure 4.17 - Outport Control Architecture, with logic blocks highlighted ..................................... 31

Figure 4.18 - Waveform showing the operation of an Output Control circuit ............................. 32

Figure 4.19 - 4–input arbiter design proposed in [GHI13]............................................................... 32
Figure 4.20 - Waveform showing the operation of the Arbiter ................................................................. 33
Figure 4.21 - Fragment showing the control logic of the wrapper used to simulate the Input Buffer Control circuit .......................................................... 34
Figure 4.22 - Operation of BaT–Hermes when sending packets from the East port to the North and Local ports ........................................................................ 35
Figure 4.23 - BaT–Hermes a) operating at peak performance scenario, and b) with contention on the output of the local port ........................................................................... 36
Figure 5.1 – FIFO's write controller circuit a) after logic synthesis and b) after adding extra cells to control the place where the delay line will be inserted ................................................................. 38
Figure 5.2 – Abstract view of wire cell, rotated by 90° ........................................................................... 39
Figure 5.3 – XML file describing the relative timing constraints of the FIFO’s write control circuit, shown in Figure 5.1–b .................................................................................................................. Error! Bookmark not defined.
Figure 5.4 – XML file describing the relative timing constraints of the FIFO Write Control circuit, shown in Figure 5.1(b) ................................................................................................................. 40
Figure 5.5 – Fragment of the TCL scripts generated by ACDC. This function sets minimum delay constraints .......................................................................................... 41
Figure 5.6 – Layout of synthesized a) Input Interface and b) Output Interface circuits ................................................. 43
Figure 5.7 – Post-synthesis simulation of an Input Interface sending packets to the a) west, b) east, c) north, and d) south Output Interfaces ........................................................................ 45
Figure 5.8 – Detailed view of Figure 5.7(a), showing flits propagating through the Input Interface .................... 45
Figure 5.9 – Post-synthesis simulation of an Output Interface receiving four packets simultaneously .................................................................................................................. 46
Figure A.1 – VHDL source code for the settings package (part 1 of 2) ................................................................. 51
Figure A.2 – VHDL source code for the settings package (part 2 of 2) ............................................................... 52
Figure A.3 – VHDL source code for FIFO Write Control module ........................................................................ 52
Figure A.4 – VHDL source code for simulation wrapper of FIFO Write Control (part 1 of 2) .................................. 53
Figure A.5 – VHDL source code for simulation wrapper of FIFO Write Control (part 2 of 2) .................................. 54
Figure A.6 – VHDL source code for FIFO Read Control module (part 1 of 2) .......................................................... 54
Figure A.7 – VHDL source code for FIFO Read Control module (part 2 of 2) ......................................................... 55
Figure A.8 – VHDL source code for simulation wrapper of FIFO Read Control (part 1 of 2) ............................... 55
Figure A.9 – VHDL source code for simulation wrapper of FIFO Read Control (part 2 of 2) ............................... 56
Figure A.10 – VHDL source code for FIFO module (part 1 of 3) ........................................................................... 57
Figure A.11 – VHDL source code for FIFO module (part 2 of 3) ........................................................................... 58
Figure A.12 – VHDL source code for FIFO module (part 3 of 3) ........................................................................... 59
Figure A.13 – VHDL source code for Mousetrap stage ....................................................................................... 60
Figure A.14 – VHDL source code for request splitter circuit ................................................................................ 61
Figure A.15 – VHDL source code for Input Buffer Control module (part 1 of 2) ....................................................... 62
Figure A.16 – VHDL source code for Input Buffer Control module (part 2 of 2) ....................................................... 63
Figure A.17 – VHDL source code for simulation wrapper of Input Buffer Control (part 1 of 2) ............................ 64
Figure A.18 – VHDL source code for simulation wrapper of Input Buffer Control (part 2 of 2) ............................ 65
Figure A.19 – VHDL source code for Input Buffer module ...................................................................................... 66
Figure A.20 – VHDL source code for Routing Unit module .................................................................................... 67
Figure A.21 – VHDL source code for Routing Control module (part 1 of 2) ............................................................ 68
Figure A.22 – VHDL source code for Routing Control module (part 2 of 2) ............................................................ 69
Figure A.23 – VHDL source code for simulation wrapper of Routing Control (part 1 of 2) ........................................ 69
Figure A.24 – VHDL source code for simulation wrapper of Routing Control (part 2 of 2) ................................. 70
Figure A.25 – VHDL source code for Input Interface module (part 1 of 2)...............................71
Figure A.26 – VHDL source code for Input Interface module (part 2 of 2)...............................72
Figure A.27 – VHDL source code for Outport Control module (part 1 of 3)...............................72
Figure A.28 – VHDL source code for Outport Control module (part 2 of 3)...............................73
Figure A.29 – VHDL source code for Arbiter module (part 1 of 3)........................................74
Figure A.30 – VHDL source code for Arbiter module (part 2 of 3)........................................75
Figure A.31 – VHDL source code for Arbiter module (part 3 of 3)........................................76
Figure A.32 – VHDL source code for Output Interface module (part 1 of 2)...............................76
Figure A.33 – VHDL source code for Output Interface module (part 2 of 2)...............................77
Figure A.34 – VHDL source code for router port (part 1 of 2)..............................................78
Figure A.35 – VHDL source code for router port (part 2 of 2)..............................................79
Figura A.36 – VHDL source code for BaT–Hermes router (part 1 of 2).......................................79
Figura A.37 – VHDL source code for BaT–Hermes router (part 1 of 2).......................................80
Figure B.1 – FIFO: constraint ensuring that phase_select_i arrives before req_i in the write control circuit and phase_select_i arrives before ack_i in the read control circuit.................................81
Figure B.2 – FIFO: constraints ensuring that, for each buffer position, data_i arrives before req_wr_i and data_o arrives before req_rd_o.........................................................81
Figure B.3 – FIFO: constraint ensuring that, for each buffer position, reg_en arrives at the register before full_o arrives at the read control circuit.....................................................82
Figure B.4 – FIFO: constraints ensuring that, for each read and write control circuit, phase_select_i arrives before en_i.................................................................82
Figure B.5 – FIFO: constraints ensuring that, for each read control circuit, en_i is disabled before a new full_o request can be made.................................................................83
Figure B.6 – FIFO: constraints ensuring that, for each register, the data is stored before an ack_wr_o is issued...............................................................................................83
Figure B.7 – Input Buffer Control: constraint ensuring that data_i arrives before req_i...............84
Figure B.8 – Input Buffer Control: constraints ensuring that data_o and last_flit_o arrives before req_header_o or req_data_o..............................................................................84
Figure B.9 – Input Buffer Control: constraint ensuring that data_o is properly stored in LT2 and FF1 before ack_o is issued.................................................................84
Figure B.10 – Input Buffer Control: constraint ensuring that, when a request is received, the mux remains stable until FF1 stores the data currently at its input.....................................85
Figure B.11 – Input Buffer Control: constraint ensuring that, when a request is received, the data at the input pin of FF3 only changes after it has been captured..................................85
Figure B.12 – Input Buffer Control: constraint ensuring that, when a request is received, FF5 only captures the data at its input pin after last_flit_lvl has propagated................................85
LIST OF TABLES

Table 1 – Comparisson of 5–port asynchronous NoC routers..........................................................13
Table 2 – Area used by the synthesized Input and Output Interfaces..............................................44

LIST OF EQUATIONS

Equation 1 – (a) The enforced–path’s delay must be equal or grater to the base–path’s delay; (b)
equation to calculate the delay–line’s delay, by isolating it from (a). ...........................................41
1. INTRODUCTION

According to the International Technology Roadmap for Semiconductors (ITRS), increasing transistor density, higher operating frequencies, short time-to-market, and reduced product life cycle characterize today’s semiconductor industry [ITR11]. Smaller transistor feature sizes, now reaching ultra-deep submicron levels, enable the integration of billions of transistors in a single chip, supporting the creation of complex systems on a chip (SoCs). Typically, an SoC is composed by intellectual property (IP) cores, an interconnection architecture, and interfaces to peripheral devices [MOR04]. IP cores are pre-designed functional blocks such as a processor or a memory controller. The reuse of cores helps reducing the system’s time to market.

Traditionally, the interconnect architecture of SoCs was based on dedicated wires or shared buses. The former approach is only applicable to systems containing a small number of cores, since the number of connecting wires increases quite fast as the number of cores grows. Shared buses are more scalable, but only one communication at a time is allowed, limiting its scalability to a few dozen IP cores [MOR04]. Networks on chip (NoCs), on the other hand, provide a higher level of communication parallelism and even higher scalability, when compared to bus-based interconnects [PON10b]. An NoC is an infrastructure that manages communication between IP cores. A NoC-based SoC typically consists of cores connected to routers, which are connected among themselves by communication channels [MOR04]. Each router may handle several simultaneous connections, increasing the available bandwidth. The TILE64™ [BEL08], the Cell processors [KAH05], the Intel Single Chip Cloud Computer [WIJ11] and the Intel Xeon Phi Coprocessor [INT13] are commercial examples of SoCs using an NoC as interconnect architecture.

Most of today’s digital circuits are designed employing the synchronous paradigm: an externally generated global clock signal is used to create a discrete notion of time. This signal controls every storage element, as Figure 1.1(a) illustrates. When the clock transitions, registers sample the data at the input ports, and the stored values are displayed at the output ports – making data flow from one register to the next. The clock period is determined by the computation time of the slowest combinational logic block. This guarantees that data is only sampled when all signals are stable. Therefore, feedback hazards and signal glitches that may occur as combinational circuits stabilize can be ignored, greatly simplifying the design process. However, synchronous design issues that were easily overcome in previous decades - such as clock distribution, skew, and power consumption, are becoming increasingly complex to solve in modern state of the art technology designs. For instance, the clock signal in a high-speed processor represents an average of 45% of the total power [AMD05]. This makes complex synchronous systems less attractive for low power applications.

Asynchronous circuits are an alternative to tackle the problems created in advanced nodes by the use of the synchronous paradigm. As Figure 1.1(b) exemplifies, this class of circuits does not employ a clock signal. Instead, local handshakes between adjacent components perform the necessary synchronization, communication and sequencing of operations [SPA01]. This potentially reduces power consumption, and eliminates problems related to clock distribution and skew [HAU95] [SPA01]. However, the lack of techniques, methodologies and electronic design automation (EDA) tools fully supporting asynchronous
systems prevents traditional circuit designers from taking full advantage of asynchronous circuits [PON10a] [HAU95].

A globally asynchronous locally synchronous (GALS) [CHA84] system is intermediate between a fully synchronous and a fully asynchronous design. In a GALS system, distinct clock signals are used to govern different modules. Internally, each module works as a fully synchronous system; different modules may operate at different clock rates. GALS techniques eliminate the burden of creating a clock distribution network across the whole chip, simplifying the achievement of timing closure in SoC designs [PON10b]. However, synchronization interfaces must be employed to enable reliable communication between distinct clock domains, introducing latency penalties in the system [PON10b] [SHE08].

Networks on chip can be implemented as fully synchronous, GALS, or asynchronous circuits. To avoid a bottleneck at the interconnection, a fully synchronous NoC needs to operate at a frequency high enough to meet bandwidth requirements of the most demanding core, which may result in significant, wasteful power consumption on routers that do not require high throughput [GEB10]. A GALS NoC allows controlling the operating frequency (and bandwidth) of each router, potentially reducing the dynamic power consumption when compared to a single-frequency NoC. However, as mentioned before, synchronization interfaces must be added between routers and cores operating in different clock domains [SHE08].

An asynchronous NoC allows a more flexible integration of components with different timing characteristics. Besides presenting reduced overall power consumption when compared with other implementation styles, zero dynamic power is naturally achieved when the NoC is idle [GIL11] [GEB10] [BJE05]. Also, latency overheads caused by the synchronization interfaces in GALS NoCs reduce the data throughput of the system. An approach to ease this issue is to reduce as much as possible the number of synchronization points throughout the SoC. By employing an asynchronous NoC as communication architecture, only two synchronization interfaces need to be traversed by each packet: one at the sender’s output port, and another at the receiver’s input port [PON10a].

**Figure 1.1 - Example of (a) a synchronous circuit, and (b) an asynchronous circuit. CLi represent combinational logic blocks, R symbolizes registers, and CTRL indicates control logic. Adapted from [SPA01].**
Since the last decade, the research on asynchronous routers and NoCs has gained momentum [PON10b]. Asynchronous NoCs can reduce interconnect overall power consumption, while eliminating increasingly complex clock-related design problems [GIL11] [BJE05]. However, most of the previous works make use of delay insensitive data encodings, usually resulting in large area overheads, or bundled-data level-signaling implementations, which increase the overall latency of the system, when compared to transition-signaling protocols [GHI13].

The main goal of this work was the design of a bundled-data asynchronous network-on-chip router. Along with it, a methodology for synthesis of bundled-data circuits using commercial CAD tools was proposed and an automated environment for enforcing relative timing constraints was developed. In this work, the Author explored several topics not covered during the undergraduate program in Computer Engineering, in particular asynchronous circuits and NoC design, while gaining practical knowledge about the digital system design process. Furthermore, due to the lack of automated design and synthesis tools for asynchronous systems, many challenges had to be overcome to synthesize bundled-data circuits. This presented a unique opportunity to study about how EDA tools work, and learn how to adapt them to provide some degree of automation to the process of designing asynchronous systems.

The remainder of this work is organized as follows. Chapter 2 provides relevant background information. Chapter 3 presents the state of the art in asynchronous NoCs. Chapters 4 and 5 present developed work. The specification and architecture of the designed router is detailed in Chapter 4. The synthesis process is presented in Chapter 5. Final remarks and directions for future work are explored in Chapter 6.
2. Concepts

This Chapter presents some basic concepts about Network-on-Chips (NoCs) and asynchronous circuits, which are needed to support this work.

2.1 Networks on Chip

On-chip communication implemented with dedicated wires is only effective for systems with a small number of cores, as it presents poor reusability and scalability. Shared buses are more scalable and reusable, but only one communication at a time is allowed and the bandwidth is shared among all cores. Moreover, it also lacks scalability, given that all cores share the same communication medium. The NoC approach address all these issues. It consists of a communication infrastructure in which cores are connected to routers and these communicate among themselves through channels. NoCs are reliable, energy efficient, reusable, and present much better bandwidth scalability when compared to traditional bus architectures [MOR04]. In fact, the use of NoCs is already a well-established concept for creating effective intrachip communication infrastructures for contemporary systems. The TILE64™ [BEL08], the Cell processors [KAH05], the Intel Single Chip Cloud Computer [WIJ11] and the Intel Xeon Phi Coprocessor [INT13] are commercial examples of NoC usage.

2.1.1 Hermes NoC

The Hermes NoC, proposed in [MOR04], comprises routers with a set of bidirectional ports that connect to an IP core and to other routers, according to a topology [MOR04]. Figure 2.1 exemplifies a 3x3 Hermes NoC with a 2D mesh topology. Each router has five bi-directional ports: East, West, North, South, and Local. The Local port links to a local IP core, and the other ports link to neighbor routers. Each router’s address is expressed by its XY coordinates.

Figure 2.1 – 3x3 Hermes NoC, connected in a 2D mesh topology. Adapted from [MOR04].
IP cores communicate with each other through message exchange. To forward messages across the network, routers employ a wormhole packet switching approach; therefore, a packet is transmitted between routers in units called *flits*, an abbreviation of *flow control digits*. A 2-flit header and a variable length payload compose each packet: the first flit contains the packet’s destination address and the second indicates the number of flits in the payload. When the header flit goes through a router, it reserves a path from an input port to an output port for the whole packet; the succeeding flits of the packet flow through this reserved path. Wormhole switching provides low latency, requires less memory, and facilitates the multiplexing of one physical channel into more than one logic channel [MOR04].

Each router can handle up to five simultaneous connections. The router’s control logic is centralized in a single block and comprises two basic modules: routing and arbitration. The routing logic analyzes the packet’s header, calculates to which port the packet should be sent, and connects the input port to the correct output port. The arbiter acts as a tiebreaker when more than one input port attempts to connect to the same output port. One of its goals it to prevent starvation, i.e. provide a balanced and fair usage of the output ports between input ports. Each input port has a buffer to help reduce performance degradation caused by a possibly blocked flit [MOR04].

### 2.2 Asynchronous Circuits

Most of today’s digital systems are synchronous, which means they employ a global clock signal to synchronize the operation of all sequential components. This creates the abstraction of discrete time. In an asynchronous circuit, the coordination of modules is performed without a clock signal [MYE01]. Instead, handshake protocols are used to perform the necessary synchronization, communication and sequencing of operations [SPA01]. The discrete-time abstraction of synchronous circuits helps simplifying the design, but removing it can grant several other benefits, like lower power consumption, higher operating speed, lower electromagnetic noise emission, and eliminating clock distribution problems [HAU95] [SPA01]. However, depending on the handshake protocol and data-encoding scheme chosen, asynchronous control logic may introduce significant area, circuit speed, and power consumption overheads [SPA01]. Another drawback of asynchronous circuits is the lack of mature EDA tools supporting their design.

#### 2.2.1 Delay Models

Asynchronous circuits can be classified with respect to the assumptions employed on gate and wire delays during their design. A delay-insensitive (DI) circuit can operate correctly no matter the magnitude of all delays found in the circuit. This means, referring for example to Figure 2.2, that the circuit in that Figure works for any arbitrary values of \(d_A\), \(d_B\), \(d_C\), \(d_1\), \(d_2\), and \(d_3\) [SPA01]. DI is the most robust class of asynchronous circuits, but it is very limited, since systems respecting this assumption can only contain C-elements as multiple-input operators [MAR90].

A circuit that operates correctly without assumptions regarding delays, except on some specific wire forks, called *isochronic forks*, is called quasi-delay-insensitive (QDI) [MYE01]. A fork is isochronic if the delays at all endpoints of the fork are identical (or differ by a very small amount that is controlled during design) [SPA01]. Considering the circuit represented in Figure 2.2, this means that either \(d_2 = d_3\), or \(\delta = |d_2 - d_3|\)
is smaller than some given value \( \kappa \). Even though this assumption may seem of little significance, it does change model properties considerably. By making forks isochronic, only one end of the fork needs to be sensed; therefore, when the sensing gate fires, the signal will have guaranteedly reached all ends of the fork [HAU95]. A further discussion about QDI circuits is outside the scope of this work, and can be found, for example, in references [SPA01] and [HAU95].

A circuit is called speed-independent (SI) when it can operate correctly assuming unknown gate delays and zero-delay wires. Again, considering the circuit shown in Figure 2.2, this means that the circuit works regardless of the values of \( d_a, d_B, \) and \( d_C \), but assuming that \( d_1 = d_2 = d_3 = 0 \). Of course, it is often unrealistic to assume zero-delay wires in today’s semiconductor processes [SPA01].

Delay-insensitivity and speed-independence have mathematically well-defined properties. Circuits whose correct operation relies on more elaborate timing assumptions are called self-timed (ST) [SPA01]. An ST circuit assumes that both gate and wire delays are known, and delay lines are used to satisfy timing assumptions. These are often called matched-delay asynchronous circuits, and are usually simpler than QDI in terms of the required hardware [STE09]. However, some of the drawbacks are its worst-case operation and the sensitiveness to process, voltage, and temperature (PVT) variations [KRZ10]. However, several works available on literature report that these circuits are usually more suited to high speed and low power applications than those designed using QDI techniques and assumptions [GHI13][BHA13][NOW11][STE09].

### 2.2.2 Handshake Protocols

Instead of using a clock signal, asynchronous circuits use handshake protocols between components to perform synchronization [SPA01]. A handshake consists in the exchange of a request and an acknowledge signaling between an active and a passive element, establishing synchronization between both. The active component starts the communication by issuing a request (req); then it waits for the arrival of the corresponding acknowledge (ack), generated by the passive element [PEE96]. This communication mechanism can be implemented using two basic approaches: transition-signaling and level-signaling; also, implementations usually rely on one of four possible channel configurations: nonput, push, pull, and bput [KRZ10].

The level-signaling handshake protocol takes four signal events to complete a communication cycle – this signaling scheme is accordingly called a four-phase protocol, and sometimes named return-to-zero (RTZ) handshake [SPA01]. The events required to complete the handshake, as exemplified in Figure 2.3(a), are: (1) the active component initiates the communication by asserting \( \text{req} \); (2) the passive component senses the request and, in response, asserts \( \text{ack} \); (3) the active element detects the acknowledge and deasserts \( \text{req} \); (4) finally, the passive component notices the deasserted request, and responds by deasserting \( \text{ack} \). At this point

---

**Figure 2.2 - A circuit fragment with gate and wire delays [SPA01].**

A circuit is called speed-independent (SI) when it can operate correctly assuming unknown gate delays and zero-delay wires. Again, considering the circuit shown in Figure 2.2, this means that the circuit works regardless of the values of \( d_a, d_B, \) and \( d_C \), but assuming that \( d_1 = d_2 = d_3 = 0 \). Of course, it is often unrealistic to assume zero-delay wires in today’s semiconductor processes [SPA01].

Delay-insensitivity and speed-independence have mathematically well-defined properties. Circuits whose correct operation relies on more elaborate timing assumptions are called self-timed (ST) [SPA01]. An ST circuit assumes that both gate and wire delays are known, and delay lines are used to satisfy timing assumptions. These are often called matched-delay asynchronous circuits, and are usually simpler than QDI in terms of the required hardware [STE09]. However, some of the drawbacks are its worst-case operation and the sensitiveness to process, voltage, and temperature (PVT) variations [KRZ10]. However, several works available on literature report that these circuits are usually more suited to high speed and low power applications than those designed using QDI techniques and assumptions [GHI13][BHA13][NOW11][STE09].

### 2.2.2 Handshake Protocols

Instead of using a clock signal, asynchronous circuits use handshake protocols between components to perform synchronization [SPA01]. A handshake consists in the exchange of a request and an acknowledge signaling between an active and a passive element, establishing synchronization between both. The active component starts the communication by issuing a request (req); then it waits for the arrival of the corresponding acknowledge (ack), generated by the passive element [PEE96]. This communication mechanism can be implemented using two basic approaches: transition-signaling and level-signaling; also, implementations usually rely on one of four possible channel configurations: nonput, push, pull, and bput [KRZ10].

The level-signaling handshake protocol takes four signal events to complete a communication cycle – this signaling scheme is accordingly called a four-phase protocol, and sometimes named return-to-zero (RTZ) handshake [SPA01]. The events required to complete the handshake, as exemplified in Figure 2.3(a), are: (1) the active component initiates the communication by asserting \( \text{req} \); (2) the passive component senses the request and, in response, asserts \( \text{ack} \); (3) the active element detects the acknowledge and deasserts \( \text{req} \); (4) finally, the passive component notices the deasserted request, and responds by deasserting \( \text{ack} \). At this point
the handshake is completed, and the active component may start another communication cycle. A big disadvantage of level-signaling protocols is its return-to-zero phase (events 3 and 4), as they cost extra time and energy [SPA01]. The transition-signaling protocol eliminates this disadvantage.

On the transition-signaling protocol, often called two-phase protocol or non-return-to-zero (NRZ) signaling, handshake events are encoded as signal transitions – there is no difference between a low-to-high and a high-to-low transition, as both imply the same type of event [SPA01]. This signaling scheme is illustrated in Figure 2.3(b). The active component starts the communication by issuing a request event – that is, switching the logic level of req. Then, the passive component senses the request transition, and ends the communication by sending out an acknowledge event – that is, changing the logic level of ack. At this point, the active component may start a new communication. Transition-signaling protocols can complete a handshake in half the time of a level-signaling one. However, depending on the used data encoding, the transition-signaling control unit may introduce more significant area and power overheads [SPA01].

![Figure 2.3– Handshake operations using (a) a level-signaling protocol and (b) a transition-signaling protocol. Adapted from [SPA01].](image)

In addition to establishing synchronization between components, handshakes can also be used to transfer data between elements – this can be achieved by encoding data in the request, acknowledge, or in both events [PEE96]. Handshake channels that do not convey data are called nonput channels. They only have the request and acknowledge wires, as shown in Figure 2.4(a), and are used for synchronization purposes only. A push channel has, in addition to the req and ack wires, a data bus going from the active component to the passive component, as Figure 2.4(b) shows – it can be said that the active component pushes data through the channel [PEE96]. In this type of channel, data needs to be valid before issuing a request event. If the data flows from the passive to the active component, as depicted in Figure 2.4(c), it is said that the active component pulls the data through the channel – characterizing a pull channel configuration. A channel with such configuration requires that data be valid before the acknowledge event occurs. Data flowing both ways characterizes a biput channel – in this situation, data sent by the active element needs to be valid before the req event, and data coming from the passive component needs to be valid before the ack event. This channel configuration is illustrated in Figure 2.4(d).

![Figure 2.4 – Handshake channel types: (a) Nonput; (b) Push; (c) Pull; (d) Biput.](image)
2.2.3 Bundled-Data

In a synchronous circuit, the role of the clock is to define points in time where data and control signals are stable and valid. Between cycles, signals may present hazards and make multiple transitions as combinational circuits stabilize [SPA01]. Asynchronous circuits, on the other hand, are usually designed to detect the arrival of data. Assuming a push channel, data validity must be inferred either from the data channel, as in QDI, or the explicit request signal, as in ST [KRZ10]. This choice is a trade-off between speed, robustness, area, and power [SPA01]. For instance, dual-rail is a very robust, delay-insensitive, data encoding scheme employed in several QDI designs. It consists of encoding each bit with a pair of wires – one, marked as the true wire, represents the bit value 1 when at the logic level 1, and the other, called false wire, denotes the bit value 0 when at the logic level 1 [SPA01]. When viewed together, each wire pair is a codeword that represents one bit of data or the absence of data (when both wires are at the logic level 0). Due to the DI characteristic of this encoding scheme, the passive element can detect data validity, when all bits have arrived, and, from this, infer the request event and generate an acknowledgement, which can be signaled using a single wire. However, the use of two wires per bit implies silicon area overhead and increases the circuit switching activity, given that every data sent requires the switching of exactly one wire per bit of information [KRZ10].

An alternative is the use of regular Boolean encoding, where one data bit is represented by one wire. This encoding scheme, however, does not convey data validity information. Therefore, an explicit signal (the request mentioned before) is required and the circuit designer needs to guarantee that data is valid before the detection of an event on this signal [SPA01]. The term bundled-data is commonly used to designate this kind of encoding: it suggests that the request and acknowledge wires are bundled with the data signals, hinting on the timing relationship between them. Figure 2.5(a) illustrates a bundled-data push channel. In the example, a delay line is used to match the delay of req to the worst-case delay of the data channel. A circuit using bundled-data scheme falls under the ST category, and either handshake protocol can be used.

The operations of the bundled-data level-signaling and the transition-signaling protocols are illustrated in Figure 2.5(b) and in Figure 2.5(c), respectively. The events required to complete the communication are the same discussed in Section 2.2.2. However, the designer needs to guarantee that, whenever there is a data exchange between an active and a passive component, data signals are valid and stable before the handshake event that notifies the data arrival is detected – that is, the request event on a push channel, the acknowledge event on a pull channel, and both events on a biput channel.

![Figure 2.5](image)

**Figure 2.5** – An example 8-bit bundled-data channel scheme and its operation: (a) a bundled-data push channel; (b) an example of a level signaling communication; (c) an example transition-signaling communication. Adapted from [SPA01].

Therefore, asynchronous circuits' communication protocols are characterized by choices of: a data encoding scheme, a channel configuration, and a handshake protocol. Dual rail encoding is a good choice for robustness and delay-insensitivity, but it introduces significant power and area overheads [KRZ10].
Transition-signaling bundled-data circuits are a suitable choice for systems with high-speed requirements [GHI13] [BHA13] [NOW11] [STE09] [SPA01]. However, some extra design effort is needed to guarantee the fulfillment of delay constraints between data and handshake signals. Since the throughput of the circuit depends on the time it takes to complete a handshake cycle, level-signaling bundled-data has a speed penalty inherent to its handshake protocol [SPA01].

2.2.4 Asynchronous Logic Components

Unlike synchronous systems, asynchronous circuits operate with a continuous notion of time – that is, events, represented by signal transitions, may occur at any moment. To support this characteristic, several asynchronous logic components have been proposed throughout the years. Two of these components are used in this work: the C-element and the mutual exclusion (mutex) gates.

A mutex is used to arbitrate between events that can happen simultaneously. Its task is to forward independently generated request signals at inputs RA and RB to its corresponding outputs GA and GB guaranteeing that at most one output is active at any given time [SPA01]. The C-element is a state-holding gate that can be used for event synchronization: its current output value is held when its inputs are distinct; when all its inputs have the same value, the output changes to reflect that same value. For example, a 2-input C-element changes its output to 0 when both inputs are 0 and to 1 when both inputs are 1, for other combinations the output is kept unchanged [SPA01].

The ASCEnD cell library [MOR11a] [MOR11b], designed at the GAPH, includes more than six hundreds instances of the components mentioned in this Section in several forms, some of which were employed in this work.

2.2.5 High Performance Pipelines

Pipelining is a technique classically used to increase parallelism in a digital system and, consequently, increase the throughput of the latter. In synchronous systems, a pipeline design consists in partitioning complex logic blocks into smaller blocks, where adjacent blocks are separated by registers all of which are controlled by a single clock signal. Since asynchronous circuits do not have such a clock signal, a protocol to control the interaction of neighboring stages must be defined in an asynchronous pipeline, along with the choice of a data encoding and of storage elements [NOW11]. Together, these design choices compose a design template for asynchronous pipelines.

Figure 2.6(a) shows an abstract view of a synchronous pipeline. In such a system, data moves to the next stage at the end of every clock cycle, and this occurs simultaneously in all stages. To work properly, the pipeline clock period needs to be larger than the computation time of the slowest stage. In an asynchronous pipeline, exemplified in Figure 2.6(b), the data flow is controlled by handshakes between stages. Usually, a stage accepts new data if the previous stage is providing new data and the next stage has already stored the previous data item. Unlike what happens in synchronous pipelines, in an asynchronous pipeline different data values can flow at different rates, given that data propagation is not constrained by the worst case delay – which may improve the system’s average latency and throughput [NOW11].
Asynchronous pipelines can be classified based on the logic style of its datapath: static or dynamic. Classically, dynamic logic refers to circuit implementations where the pull-up network is removed, resulting in smaller area usage and higher switching speed [RAB03]. However, this type of implementation increases the design and validation effort [NOW11]. For this reason, this work considers only static logic pipelines.

Figure 2.6 – An abstract view of pipelines: (a) a synchronous pipeline; (b) an asynchronous pipeline. Adapted from [NOW11].

Asynchronous pipelines can be classified based on the logic style of its datapath: static or dynamic. Classically, dynamic logic refers to circuit implementations where the pull-up network is removed, resulting in smaller area usage and higher switching speed [RAB03]. However, this type of implementation increases the design and validation effort [NOW11]. For this reason, this work considers only static logic pipelines.

Figure 2.7 – Two transition-level asynchronous pipelines: (a) Sutherland’s; (b) Mousetrap. Adapted from [NOW11].

Ivan Sutherland pioneered the transition-signaling bundled-data asynchronous pipeline in the late 1980s. Each stage of his proposed pipeline, shown in Figure 2.7(a), has a C-element as the control unit and a special capture-pass latch capable of sensing transitions at the inputs. The Mousetrap pipeline [SIN07], shown
in Figure 2.7 (b), is based on Sutherland’s micropipeline [SUT89], but features less complex signaling and lower overhead [NOW11]. Mousetrap is controlled by a 2-input XNOR gate, and data is stored in a bank of standard active-high level-sensitive D latches. Figure 2.8 exemplifies the operation of a Mousetrap pipeline stage: 1) the incoming request (req₁) traverses the latch and causes the XNOR gate to switch its output to low (en₂), making the latch opaque; in parallel, the latch produces the acknowledge (ack₂) signal; 2) once the acknowledge is received in the previous stage, the latter can start a new handshake; 3) the output request (req₃) flows through the delay line and arrives at the next stage; 4) the next stage acknowledges it, causing the XNOR gate to switch its output to high, making the latch of stage 2 transparent again, which is the original state for all latches. From this moment on, a new handshake cycle can start.

![Figure 2.8 – Example of the Mousetrap pipeline operation for Stage 2 signals.](image)

Many practical applications can be implemented with linear pipelines. However, complex system architectures often require nonlinear pipelines [SIN07]. This implies using pipelines with parallel stages, as Figure 2.9 – Nonlinear pipelines: (a) shows. The Mousetrap template can be easily adapted to support such systems. A pipeline fork stage can only accept a new request when all forking stages have acknowledged the current one. This behavior can be achieved by the use of a C-element to join the ack signals coming from parallel stages, as Figure 2.9(b) shows. Similarly, a join stage must wait until all requests from the parallel stages have arrived to start processing new data, which can be accomplished by employing a C-element at the incoming request signals, as Figure 2.9 – Nonlinear pipelines: (c) illustrates.

![Figure 2.9 – Nonlinear pipelines: (a) Example of a nonlinear pipeline structure; (b) a Mousetrap fork stage; (c) a Mousetrap join stage. Adapted from [SIN07].](image)
3. State of the Art

This Chapter presents an overview of the current literature on asynchronous network on chip routers based on 5-port architectures. As metrics used by each author differ from one another and the designs target different technologies using different delay models, there is not enough data to draw a fair comparison of all implementations. However, most authors provide area usage and throughput information in units of flits/s. Table 1 summarizes the attributes of the NoCs covered in this Section.

<table>
<thead>
<tr>
<th>Router</th>
<th>Asynchronous Style</th>
<th>Technology feature size</th>
<th>Flit Size</th>
<th>Area</th>
<th>Average Latency</th>
<th>Average Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ghiribaldi et al. [GHI13]</td>
<td>Bundled-data, transition-signaling</td>
<td>40nm</td>
<td>32 bits</td>
<td>4,691µm²</td>
<td>1.195ns</td>
<td>35.4 Gbytes/s per port</td>
</tr>
<tr>
<td>Hermes- AA [PON10b]</td>
<td>Dual-rail, level-signaling</td>
<td>65nm</td>
<td>8 bits</td>
<td>114,456µm²</td>
<td>3.709ns for input port, 1.351ns for output port</td>
<td>7.7 Gbits/s per router</td>
</tr>
<tr>
<td>ASPIN [SHE08]</td>
<td>Bundled-data, level-signaling</td>
<td>90nm</td>
<td>32 bits</td>
<td>36,199µm²</td>
<td>1.5ns</td>
<td>4.4 Gbytes/s per router</td>
</tr>
<tr>
<td>ANoC [BEI05]</td>
<td>QDI, level-signaling</td>
<td>160nm</td>
<td>34 bits</td>
<td>0.25mm²</td>
<td>2ns and 2.5ns, depending on traffic priority</td>
<td>5.0 Gbytes/s per router</td>
</tr>
<tr>
<td>MANGO [BJE05]</td>
<td>Bundled-data, level-signaling</td>
<td>120nm</td>
<td>32 bits</td>
<td>0.188mm²</td>
<td>N.A.</td>
<td>15.4 Gbytes/s per router (under typical conditions)</td>
</tr>
<tr>
<td>Hermes–A [PON10a]</td>
<td>Dual-rail, level-signaling</td>
<td>180nm</td>
<td>8 bits</td>
<td>0.33mm²</td>
<td>N.A.</td>
<td>3.6 Gbits/s per router</td>
</tr>
<tr>
<td>Async. QNoC [DOB09]</td>
<td>Bundled-data, level-signaling</td>
<td>180nm</td>
<td>8 bits</td>
<td>N.A.</td>
<td>N.A.</td>
<td>1.1 Gbytes/s per router</td>
</tr>
</tbody>
</table>

3.1 The NoC of Ghiribaldi et al.

Ghiribaldi et al. [GHI13] propose an asynchronous NoC router based on the Mousetrap pipeline template, accordingly using a transition-signaling bundled-data protocol. The router features wormhole switching, 32-bit flits, and algorithmic dimension-order routing. The design was synthesized using a low-power standard-Vt 40nm cell library, with normal process, 1.2V supply voltage, and operating temperature of 300K.

The router presents in average a latency of 1,195ps, and a cycle time of 903ps. Latency is defined as the time it takes for a head flit to traverse from the input port to the output port, assuming an empty router and no congestion. Cycle time is the interval between two successive acknowledgments received at the router output port. The area usage is 4,691µm². Based on the average cycle time and on the flit size it is possible to compute the NoC average throughput: 35.4 Gbits/s per port.
3.2 The ANoC

Proposed by Beigné et al., the ANoC [BEI05] is a QDI router implemented with level-signaling handshake protocol and support for two virtual channels – one for real-time low latency packets, and other for best-effort traffic. It features wormhole routing, odd-even turn model adaptive routing algorithm, and 34-bit flits – 32 bits of data, plus 2 control bits. The router was synthesized using STMicroelectronic’s HCMOS9LL 0.16μm technology. The area used by each router is 0.25mm² and the average cycle time is 4ns, giving the maximum throughput of 5Gbytes/sec, when all inputs and outputs run concurrently. The router latency is 2ns for the high priority virtual channel and 2.5ns for the low priority one.

3.3 The MANGO NoC

The MANGO [BJE05] NoC, proposed by Bjerregaard and Sparso, employs virtual channels to provide connection-oriented guaranteed services and connectionless best-effort routing. Communication is performed using a level-signaling bundled-data protocol. Each router implements wormhole packet switching, 32-bit flit size, and XY routing algorithm. The NoC was synthesized using 0.12μm CMOS standard cell technology. Under worst-case timing parameters (1.08V supply voltage and 125ºC operating temperature), each port presents a performance of 515MHz; under typical conditions, 795MHz was obtained. This corresponds to a worst-case maximum throughput of roughly 10GBytes/sec and a typical case of roughly 15.4GBytes/sec. The router area usage is 0.188mm².

3.4 The Asynchronous QNoC

Dobkin et al. [DOB09] proposed the asynchronous QNoC, a network on chip supporting quality-of-service in four distinct service levels, each with two virtual channels. The routers are implemented using a level-signaling bundled-data protocol, XY routing algorithm, and wormhole packet switching with 8-bit flit size. The NoC was synthesized using a 0.18μm CMOS standard cell library from Tower Semiconductor Ltd. The minimal router data cycle was 4.5ns, resulting in a throughput of 220Mflits/s per port, or 1.1GBytes/sec per router.

3.5 The ASPIN NoC

Proposed by Sheibanyrad et al., the ASPIN NoC [SHE08] uses both bundled-data and dual-rail data encoding: the latter is used only for long wires. Each router features wormhole packet switching, 32-bit flit size, and the XY routing algorithm. Synthesis was targeted at the STMicroelectronics 90nm low-voltage threshold technology. The area used by each router was 36,199μm². The maximum throughput is 131Mflits or 4.423Gbytes/s, and the average packet latency is 1.5ns.
3.6 The Hermes-A and Hermes-AA NoCs

Pontes et al. proposed two asynchronous delay-insensitive NoC router designs: Hermes-A [PON10a] and Hermes-AA [PON10b]. In both cases, routers implement wormhole packet switching, 8-bit flit width, and operate using a QDI dual-rail level-signaling protocol. The first design, Hermes-A, features a distributed XY routing algorithm and uses the XFab 180nm technology – with typical transistor models, 1.8V supply voltage, and operating temperature of 25°C. The second design uses another option of routing algorithm: West-First (WF); Hermes-AA was synthesized using general-purpose standard-Vt 65nm technology from STMicroelectronics; the router operates at 25°C with a supply voltage of 1V.

Hermes-A has a throughput of 727Mbits/s on each router link. In the best case, this performance level can be sustained on all five ports, resulting in a maximum throughput of approximately 3.6Gbits/s for the whole router. In this situation, the total power is 11.14mW. The total area usage is 0.33mm², of which 0.22mm² correspond to standard cell area.

Hermes-AA was able to produce 1.55Gbits/s of throughput on each router link – resulting in a maximum router throughput of 7.75Gbits/s, when all five ports are transmitting in parallel. According to the authors, per-port throughput can reach up to 6.3Gbits/s when asynchronous FIFOs are added to the input ports. However, this value goes back to the saturation throughput (1.55 Gbits/s) when FIFOs are full. Starting from an idle router, the input port latency reaches 3.709ns, and the output port latency is 1.351ns. When using the XY routing algorithm, the area usage is 114456µm² (75144µm² of standard cell area); with the WF algorithm, there was a slight increase of area: 116034µm² (76455µm² of standard cell area). For both routers, asynchronous cells such as C-elements were taken from the ASCEnD library [MOR11a] [MOR11b].
This work focuses on the design of a high-performance asynchronous NoC router based on the Hermes NoC. A transition-signaling bundled-data handshake protocol is a suitable choice for systems with high-speed requirements [SPA01], as they have the potential to lead to low area overheads, when compared to delay-insensitive encoding. Due to its simple control logic, the Mousetrap pipeline template [SIN07] is used for controlling the communication between components.

This Chapter details the specification and architecture of proposed router, and presents the functional validation of the design, obtained through behavioral simulation. The VHDL implementation files are available in Appendix A.

4.1 Specification

4.1.1 Target Architecture

The YeAH! NoC router, designed by the Hardware Design Support Group (GAPH) from the PUCRS, employs a fully distributed control logic to implement functionality similar to that of a Hermes router, introduced in section 2.1.1. In its design, routing and arbitration responsibilities are assigned, respectively, to each input and output interface of the router, instead of being centralized in a control block. This results in a highly modular design that avoids bottlenecks in the control unit, improving performance in congested networks. Figure 4.1 illustrates the architecture of the YeAH! router. Each port comprises two independent components: the input interface and the output interface. Each input interface has a set of independent control channels connected to all output interfaces. Up to five simultaneous internal connections can be handled by the router.

![Architecture of the YeAH! router.](image-url)
Figure 4.2(a) displays the input interface architecture, which consists of an input buffer, a control unit, and the routing algorithm. When the Buffer outputs a header flit, the Control activates the Routing block that issues a request to the chosen output port. Once the request is granted, the packet is transmitted. The Control logic uses the payload size information contained in the second flit of each packet to identify header flits from payload flits. The buffer is implemented as a circular FIFO. Currently, the XY routing algorithm is the only one supported in YeAH!.

Each output interface, as Figure 4.2(b) illustrates, comprises an Arbiter and a (output) Control. The former implements a round-robin algorithm to select which input port should gain access to the output port when multiple simultaneous requests are pending. The (output) Control acts as a multiplexor, selecting which data channel should be sent to the output based on the arbiter’s choice.

![Figure 4.2 - Detailed view of input (a) and output (b) interfaces of the YeAH! router.](image)

### 4.1.2 BaT–Hermes Specification

The architecture of the bundled-data transition-signaling Hermes router (BaT-Hermes) is based on the YeAH! router. Its highly decoupled architecture makes it a good starting point for the design of an asynchronous router as each individual module can be implemented as a handshaking module. In fact, the basic structure shown in Figure 4.1 can be kept, as communication between modules is replaced by handshake protocols. The router can have up to 5 ports, each composed of an Input and an Output Interface. Each Input Interface (II) is connected to all Output Interfaces (OI), but the one located in the same port — that is, there is no loopback connection. Figure 4.3 shows the top view of each interface. IIs are responsible for buffering and routing incoming packets to the correct OI. OIs, in turn, arbitrate between incoming packets from different IIs. Therefore, a 5-port router can handle up to five simultaneous connections, where all OIs receive a packet from a different II. BaT-Hermes is parameterizable with respect to flit size, buffer depth, and which ports are available. At this time, only the XY routing algorithm is implemented. However other routing algorithms can be easily implemented due to the high modularity of the router.

![Figure 4.3 – Top view of a) Input Interface and b) Output Interface of BaT-Hermes.](image)
The BaT-Hermes communication protocol determines the handshakes between Input and Output Interfaces. It defines that the first flit of each packet must be sent through a req_outport request – that is, a request event issued from the II’s req_outport_o port to the OI’s req_outport_i port, while the remaining flits are sent over req_data requests – similarly, sent from II’s req_data_o port to the OI’s req_data_i port. Routing and arbitration tasks take place during the first handshake (through a req_outport request), in a specific II and OI, respectively. In the remaining handshakes, until the last flit is detected, flits flow directly from the II to the OI. The protocol also specifies that the last_flit_i signal needs to switch before the last flit is sent. All communication between interfaces is transition-encoded. The signals req_data_o, last_flit_o and data_o of each II are shared across all OIs, which arbitrate incoming requests. BaT-Hermes uses the same packet layout as the original Hermes NoC: a two-flit header, comprising the packet’s target address and payload size, and a variable-length payload. The implementation details of each module is described in Section 4.2.

The key difference between the architectures of BaT-Hermes and the NoC proposed in [GHI13] is the router’s control unit. In the former, the detection of head and tail flits is based on the payload size, contained in the packet header. The latter uses a 2-bit flit-type field, embedded in each flit, to identify head and tail. The approach taken on the latter greatly simplifies the control units when compared to the Finite State Machines (FSM) required by the former to count the number of flits sent, but reduces the amount of data carried by each flit, since two bits are always employed as control. Additionally, all communications between IIs and OIs in BaT-Hermes is performed with transition-signaling handshakes, while the router proposed in [GHI13] employs level-signaling handshakes to start and finish the transmission of packets – which increases the circuit’s switching activity.

4.1.3 Timing Constraints

Bundled-data circuits rely on carefully designed delay-lines to ensure that handshake events take place only when data is valid. Due to the lack of EDA tools supporting such circuits, the timing relationships between wires and modules must be manually extracted. During the design phase of BaT-Hermes, the timing constraints of each circuit were identified and documented using the following technique: i) for each register, identify all data signal paths connected to the data input pin; ii) identify the control signal path of the register; iii) create a delay line at the control signal in order to guarantee that the path of step 1 is faster than the one of step 2. The list of timing constraints is available in Appendix B.

4.2 Architecture

This Section details the architecture of each component of BaT-Hermes. The techniques described in Section 4.3.1 were employed to obtain the waveforms used to illustrate the operation of the circuits.

4.2.1 Transition Merger and Phase Matcher

Transition-signaling handshake protocols relate changes in the logic value of control signals to actions. Therefore, it is imperative that the request and acknowledge signals remain stable until their related actions
have taken place. Linear transition-signaling pipelines are designed to behave this way, and can be easily modified to support non-linear concurrent stages, as discussed in Section 2.2.5. Furthermore, a pipeline can fork to or join from stages that work in a mutually exclusive manner, that is, only one parallel stage will perform a handshake at a time, while all control signals of the remaining stages are kept stable. This construct, exemplified in Figure 4.4, is widely used in BaT-Hermes to decouple routing handshakes from data transferring handshakes, as the former takes longer to complete due to routing and arbitration overheads. Additional hardware is needed in order to support mutually exclusive joins and forks in transition-signaling pipelines.

The approach taken in BaT-Hermes design to implement forks and joins makes use of two circuits: the transition merger and the phase matcher. The former is used to merge independent signals of the same type to a single wire while keeping the number of transitions at the output consistent – that is, any transition in any of the inputs results in a transition at the output. Note that the inputs do not need to be synchronized and are agnostic of the logic level of the output. The phase-matcher is employed to keep the input control signals of deactivated stages stable while shared control signals switch to perform handshakes with the active stage. As Figure 4.4(a)-(b) show, a pipeline fork requires the use of a transition merger on the acknowledge signals and phase matchers at each parallel stage request signal input; on pipeline joins, as showed in Figure 4.4(c)-(d), phase matchers are used at each parallel stage acknowledge signal input, and a transition merger on the request signals.

![Diagram of mutual exclusive pipeline fork and join with phase matching and transition merging](image)

**Figure 4.4 – Example of a mutually exclusive pipeline fork with a) transition merging and b) phase-matching circuits and pipeline join with c) phase-matching and d) transition merging circuits.**

Both the transition merger and the phase matcher rely on the usage of $n$-input XOR gates, where $n$ is the number of parallel stages in the pipeline fork or join. For instance, a transition merger circuit operating with the request signals, as Figure 4.4(d) shows, takes all output request signals from the parallel stages and generates one signal that switch every time one of its inputs switch, given the logic behavior of the XOR. The waveforms on Figure 4.5(a) show the operation of this circuit. As it can be seen, when $ack_o_{-1}$ switch to ‘1’, the output $ack_o$ also go to ‘1’. Next, when $ack_o_{-2}$ switch to ‘1’, $ack_o$ switch back to ‘0’. Finally, a new transition in $ack_o_{-1}$, to ‘0’, makes $ack_o$ to switch, back to ‘1’. The phase matcher is used to mask transitions directed to the active stages, keeping the logic levels of disabled stages stable. Its operation is similar to the transition merger circuit, but for each handshake cycle its output transitions twice – canceling out the transition that originated the handshake. A latch is used to prevent these two transitions from
propagating to the disabled stage. The XOR gate takes as input the shared control signal, \textit{req\_i} in the example of Figure 4.4(b), and one control signal from each of the other stages. Even though either control signal (\textit{req} or \textit{ack}) can be used, as both will eventually transition during the handshake cycle, it is more efficient to use the signal that switches first, \textit{req\_o} in this example, in order to propagate the transition to the other stages as soon as possible. Figure 4.5(b) shows the operation of the circuit depicted in Figure 4.4(b): 1) a request action on \textit{req\_i} generates a transition at the output of every XOR gate connected to it; 2) the latch on the disabled stage (LT2) blocks the propagation of the transition, while the enabled stage (LT1) propagates it, generating a request event on that stage; 3) the signal flows from the latch LT1 and arrives at the XOR gate at the disabled stage, switching it for the second time and restoring the previous logic level.

![Figure 4.5](image)

**Figure 4.5** – Waveform showing the operation of the a) transition merger and the b) phase matcher circuit.

Although only one request will be received at any given time, as the stages work in a mutually exclusive manner, the pipeline join can be improved to support multiple simultaneous requests being made at the parallel stages. This can be done by the addition of the latch LT3, in Figure 4.4(c), and control logic to choose which stage will be enabled. If it is guaranteed by design that only one request will be made at a time, and during that handshake cycle the other stages will not attempt to start a handshake, the latch LT3 and the control logic can be removed. Also, in pipeline joins, the control of the phase matcher latch (LT4) is performed with an XOR gate. The latch is enabled when a request arrives and automatically disables itself when the acknowledge signal propagates through it. In pipeline forks this task is handled by the control logic used to activate the stage. The logic block that enables the parallels stages may be based on the data input of the stage or on previous handshakes. These circuits are also used in the NoC router proposed by Ghiribaldi et. al. [GHI13].

### 4.2.2 Input Interface

The II, shown in Figure 4.6, combines an Input Buffer and a Routing Control unit. The Input Buffer is composed of a parameterizable depth First-In First-Out (FIFO) queue and a Buffer Control, responsible for identifying header flits from the payload. The Routing Control implements an algorithm to select the target OI and signal a request to it. Transition-signaling handshake protocol is used in all communications between these components. This circuit implements the mutually exclusive parallel stages mentioned in Section 4.2.1: when a handshake is made through the \textit{req/ack\_outport} signals, the \textit{req/ack\_data} signals are not used, and vice versa.
A pair of dedicated handshake signals (req_outport_o and ack_outport_i) is connected to each communicating OI - there is no loopback connection between the IIs and OIs on the same port. The Routing Control uses the req_outport_o wires to request access to the target OI. An ack_outport_i acknowledge means that the target interface has received the flit on data_o (target address flit), and is ready to receive other flits from the packet. Note that, albeit handshake cycles on these signals take a long time to complete due to routing and arbitration overheads, they are performed a single time for each packet, only in the first flit for directing the packet to the correct OI. Subsequent flits flow directly from the Input Buffer Control to the selected OI.

The Input Buffer Control uses the req_data_o signal, shared with all communicating OIs, to send the remaining flits. Each communicating OI has a dedicated ack_data_i wire used to acknowledge data requests destined to it. The shared request signal helps reduce the number of wires and control logic between IIs and OIs. Note that the OI was designed to support out-of-phase req_data_o and ack_data_i signals, in order to support the shared req_data wire. Hence, a transition merger circuit is used at the II to prevent out-of-phase signals in the Input Buffer Control. A transition on the last_flit_o signal indicates that the next data request will deliver the last flit of the packet – freeing the OI to accept requests from another IIs after the last flit is received.

Figure 4.7 shows the operation of the II: 1) data received in the II is buffered in the FIFO and delivered to the Input Buffer Control; 2) the Input Buffer Control receives data from the FIFO, detects that it is a target address flit, and sends a request to the Routing Control; 3) the Routing Control choses the target output port based on the routing algorithm and requests its use; 4) the OI of the selected port acknowledges the request, indicating that it is ready to receive the packet; 5) the other flits are transmitted by the Input Buffer Control using the shared req_data_o signal; 6) only the OI requested on step 3 acknowledges the request made on step 5; 7) before the last flit of the packet is sent, the Input Buffer Control notifies the OI switching the last_flit_o signal. The FIFO works in parallel with the Input Buffer Control, buffering the flits sent to the input interface while the handshakes take place.
of each handshake cycle, when 
by the XNOR of a request and an acknowledge signal (e.g. matcher circuit when employed in pipeline joins, as explained in Section 4.2.1. The ring counter is controlled making it increment at the end of each handshake cycle, when ack_i transitions.

4.2.3 FIFO

Buffering is a technique commonly used in NoC design to help reducing performance degradation caused by a possibly blocked flit [MOR04]. The buffer used in BaT-Hermes is a transition-signaling bundled-data First-In First-Out (FIFO) circular queue proposed by Ghiribaldi et al. in [GH13]. Figure 4.8 details the three building blocks of the FIFO: (a) write controller (wr_ctrl); (b) read controller (rd_ctrl); (c) 1-hot ring counter (read/write_counter). Write and read controllers work in pairs, as they access the same data register. The wr_ctrl circuit employs a phase matcher controlled by an XNOR gate and an enable signal (en_i) – very similar to the Mousetrap stage, introduced in Section 2.2.5. The signal reg_en_o controls the active-high latch that stores data. full_o is a request signal employed to inform the pairing rd_ctrl that new data is available to be read. empty_o is an acknowledge signal indicating that the data has been read. New data can only be written when the wr_ctrl is active (i.e. en_i is high) and the currently stored data has been read – which can be detected by the XNOR of full_o and empty_i, resembling a Mousetrap stage. The rd_ctrl works as the phase matcher circuit when employed in pipeline joins, as explained in Section 4.2.1. The ring counter is controlled by the XNOR of a request and an acknowledge signal (e.g. req_i and ack_i) – making it increment at the end of each handshake cycle, when ack_i transitions.
Figure 4.8 – Building blocks of the FIFO proposed in [GHI13]: a) write controller; b) read controller; c) 1-hot ring counter.

Figure 4.9 shows the FIFO’s architecture. The pair of signals \( \text{req}_\text{wr}_i \) and \( \text{ack}_\text{wr}_o \) and \( \text{req}_\text{rd}_o \) are used to write to and read from the buffer, respectively. Note that a transition merger is used join the \( \text{ack}_o \) signals from the multiple \( \text{wr}_\text{ctrl} \) into a single \( \text{ack}_\text{wr}_o \) signal. Similarly, the \( \text{req}_o \) signal from the \( \text{rd}_\text{ctrl} \) employ the same circuit to generate \( \text{req}_\text{rd}_o \). The enable signal of each \( \text{wr}_\text{ctrl} \) and \( \text{rd}_\text{ctrl} \) is generated by the \( \text{write}_\text{counter} \) and \( \text{read}_\text{counter} \), respectively. The latter is also used as the mux selector for choosing the data output signals.

Since the order of activation of each controller is known, it is possible to reduce the number of signals used in the phase matcher to two: the shared control signal (\( \text{req}_i \) or \( \text{ack}_i \)) and a \( \text{phase}_\text{select}_i \) signal that is a function of the FIFO’s depth and the controller’s index.

- For even-depth FIFOs:
  - The \( \text{phase}_\text{select}_i \) of even-index controllers is the next \( \text{full}_i/\text{empty}_i \) signal (\( \text{phase}_\text{select}_i[i] \leftarrow \text{full}_i/\text{empty}_i[i + 1] \), when \( i \mod 2 = 0 \));
  - The \( \text{phase}_\text{select}_i \) of odd-index controllers is the previous \( \text{full}_i/\text{empty}_i \) signal (\( \text{phase}_\text{select}_i[i] \leftarrow \text{full}_i/\text{empty}_i[i - 1] \), when \( i \mod 2 \neq 0 \)).

- For odd-depth FIFOs:
  - The \( \text{phase}_\text{select}_i \) of even-index controllers is 0, which means that the \( \text{req}_i/\text{ack}_i \) signal can be connected directly to the latch;
  - The \( \text{phase}_\text{select}_i \) of odd-index controllers is 1, which means that the XOR gate can be replaced by an inverter.

The waveform presented in Figure 4.10 illustrates the operation of a 2-place FIFO: 1) a request to write in the fifo is made (\( \text{req}_\text{wr}_i \) switches) and directed to the active write controller; 2) the data latch controlled by the active \( \text{wr}_\text{ctrl} \) is disabled, storing the data, while both \( \text{full}_o \) and \( \text{ack}_o \) signals switch; 3) \( \text{ack}_o \) propagates, switching \( \text{ack}_\text{wr}_o \) and incrementing the \( \text{write}_\text{counter} \); 4) simultaneously with step 3, \( \text{full}_o \) propagates to \( \text{rd}_\text{ctrl} \) switching \( \text{req}_\text{rd}_o \), indicating that there is data in the queue; 5) once this data is read, \( \text{ack}_\text{rd}_i \) switches, incrementing \( \text{read}_\text{counter} \) and freeing the data register to receive new data. These steps repeat for each position of the queue.
Figure 4.9 – Architecture of the transition-signaling bundled-data FIFO proposed in [GHI13].

Figure 4.10 – Waveform showing the operation of a 2-place FIFO.

4.2.4 Input Buffer Control

The Input Buffer Control implements the BaT-Hermes communication protocol and is responsible for interacting with the OI. The protocol defines that the first flit of each packet must be sent through a `req_outport_o` request, while the remaining flits are sent over `req_data_o` requests. It is also stated that the
signal last_flit_o must switch its logic value before the request sending the last flit of the packet is made. The Input Buffer Control circuit, detailed in Figure 4.11, implements a Finite State Machine capable of fulfilling the requirements of the protocol. A Mousetrap stage (LT1 and LT2) was inserted to reduce the average latency of the router. It allows the FIFO to fetch new data while the Input Buffer Control is busy. All latches and flip-flops of the circuit are reset to zero, except FF2 and FF5.

The FSM circuit can be split in two blocks: a counter and the control logic for selecting which signal will be used to perform the handshake (req_outport_o or req_data_o), depending if the current flit is the first or the remaining flits of the packet. Flip-flops FF1, FF2, and FF3 compose the counter circuit. The signal flit_counter, stored in FF1 denotes the number of flits that remain to be sent. Its value is initialized with the payload size, read from the second flit, and decremented as each flit is sent. last_flit_lvl is a level-encoded signal stored in FF2 that indicates when the handshake of the last flit is occurring. The output of FF3, called size_flit, controls the mux that selects which value will be stored in FF1: the flit in the data_i bus, or the current value of flit_counter decremented by 1. The data stored in these three registers are updated every time the Mousetrap stage accepts a request – therefore, the sampled values correspond to what was taking place on the previous handshake. For example, when the last_flit_lvl register (FF2) detects that the value of flit_counter is 1, it means that when the previous handshake took place there was still one flit remaining to be sent; consequently, the current flit is the last flit. This allows the computation of the next values to start as soon as the request is made, which means that the circuit doesn’t need to be delayed if the computation time is smaller than the handshake round-trip time – that is the sum of: propagation delay of request signal, propagation delay of acknowledge signal, and the minimum amount of time it takes for the OI to issue an acknowledge.

![Figure 4.11 – Architecture of the Input Buffer Control.](image)

The second block of the FSM, composed by LT3, LT4, FF5 and the adjoining logic gates implements a mutually exclusive pipeline fork to control which output signal will perform the next request: req_header_o or req_data_o. A transition merger connects ack_header_i and ack_data_i to the Mousetrap stage. The Phase Matcher circuits are implemented with latches LT3 and LT4, plus connecting XOR gates. The signal data_hs, generated by FF5, controls the phase matcher’s enable signal: when data_hs is asserted, the request is signaled by req_data_o; when deasserted, req_header_o performs the request. The value stored in FF5 is updated as soon as the request signal propagates through the active phase matcher latch – the XNOR gate performs this detection: its output switches to low when the inputs have a different logic value. The mux prevents temporary transitions on the deactivated stage from reaching FF5. The new value of FF5, calculated by the NAND gate, is determined as follows: i) If data_hs is deasserted (req_header_o handshake), the next
data_hs will be asserted (data_hs handshake); ii) If data_hs is asserted and last_flit_lvl is also asserted, the next data_hs will be deasserted; iii) otherwise, data_hs is kept asserted.

In order to reduce latency, the Input Buffer Control was designed to avoid stricture timing constraints on the request signal path. Even though the last_flit_lvl signal is shared between both blocks of the FSM, its propagation delay only affects FF5, as the signal at the input data pin needs to arrive before the one at the clock pin. As long as this delay is smaller than the round-trip time of the request signal, it does not affect the performance of the circuit.

Figure 4.12 shows the step-by-step operation of the Input Buffer Control circuit: 1) req_i request is made; 2) as soon as the data is stored in the Mousetrap stage, an acknowledge is issued by switching ack_o; 3) in parallel with step 2, the flip-flops FF1, FF2 and FF3 update the stored values – note the size_flit signal rising to select data_i as the input of FF1 for the next handshake. Also in parallel, the request propagates to req_header_o, and the data_hs signal is updated; 4) when the acknowledge signal ack_header_i arrives, the payload size is stored in flit_counter and size_flit is disabled – changing the mux to decrement the counter. In parallel, the request is made using the signal req_data_o; 5) the following flits are sent through req_data_o, while the flit_counter is decremented; 6) when flit_counter reaches 1, it asserts the last_flit_lvl signal, indicating that the next flit is the last of the packet. As a consequence, the positive edge-triggered flip-flop FF4 updates its output with the inverse of the current value, causing last_flit_o to switch. As soon as the req_data_o propagates through LT4, data_hs is deasserted, restoring the initial conditions of the circuit.

![Figure 4.12 – Waveform showing the Input Buffer Control Operation.](image)

The simulation technique used for functional validation doesn’t simulate gate delays. For this reason, the signals seem to switch at the same time – for example, the en signal appears as a spike instead of a pulse long enough to fulfill the timing restrictions of the connected gates. Details about how the behavioral simulations were performed can be found in Section 4.3.1.

### 4.2.5 Routing Control

The Routing Control is responsible for requesting use of OIs. The circuit, as shown in Figure 4.13, is composed of a Routing Unit and a mutually exclusive fork with a parallel stage for each communicating OI. The routing algorithm is a combinational circuit that compares the value of signal target_address_i with the hardwired router address and port. Once the computation is ready, the target OI phase matcher is enabled, generating a request through the respective req_outport_o. An XOR gate is used to guarantee that the Routing...
Unit is deactivated when routing is not taking place. The VHDL implementation of this module, through the use of “if generate” statements, can optimize the circuit by removing logic related to inexistent OI. This situation may happen, for example, in routers placed at the corner of a mesh network.

![Routing Control Architecture](image)

Figure 4.13 – Routing Control Architecture.

Figure 4.14 shows the operation of this circuit: 1) the arrival of a req_route_i activates the Routing Unit. A request on the correct req_port_o is made as soon as routing algorithm completes its computation; 2) the acknowledge sent by the OI through the respective ack_out_port_i signal is forwarded to ack_route_o.

![Waveform showing the operation of the Routing Control at a local port on a BaT-Hermes router of address x11.](image)

Figure 4.14 – Waveform showing the operation of the Routing Control at a local port on a BaT-Hermes router of address x11.

4.2.6 Output Interface

The OI of BaT-Hermes is responsible for arbitration tasks. Figure 4.15 shows the architecture of this circuit, which combines a set of Output Control blocks with Phase Matchers at the ack_i inputs, an Arbiter, a multiplexor, and a Transition Merger on the request signals. Each Output Control is connected to an II and is capable of handling out-of-phase req_data_i/ack_data_o signals events, in order to support the shared req_data_i signal. The Arbiter mediates simultaneous requests, guaranteeing that only one Output Control can access the output channel at any time. The grant signal from the Arbiter (arbiter_grant_i) is also used to control the data multiplexor. When a req_outport_i request arrives, the Output Control issues a level-encoded request signal to the Arbiter. Once this request is granted, and as long as the level-encoded request
signal remains asserted, the *Outport Control* circuit can interact with the output channel through *req_o*, *ack_i*, and *data_o* signals.

Unlike the handshakes between the other components of BaT-Hermes, the *Arbiter* communicates using a level-signaling protocol. This was mandatory, since no transition-sensitive mutex is available in the cell library used.

Figure 4.15 – Architecture of the Output Interface

Figure 4.16 illustrates the operation of the circuit: 1) two simultaneous *req_outport_i* requests are made (0 and 1); 2) the *Arbiter* grants access to the output channel to one of the *Outport Control* blocks that issued a request on step 1; 3) Once the initial handshake is completed, the next flits are sent using the shared *req_data_i* signal; 4) the *Outport Control* circuit handles the out-of-phase handshake signals gracefully; 5) the last flit detector, built inside the *Outport Control*, senses the transition on the *last_flit_i* wire and deasserts the arbiter request signal; 6) the arbiter grants access to the output channel to another *Outport Control* block. The transition of *req_data_i*[0] before the first handshake is generated by the test bench to simulate an out-of-phase *req_data_i* signal.
4.2.7 Outport Control

The Outport Control implements the BaT-Hermes communication protocol and is responsible for controlling the interactions between II, Arbiter and output channel. The circuit, detailed in Figure 4.17, can be split into six blocks: i) last flit detector, formed by FF2, FF3 and the adjoining XOR gate; ii) programmable phase matcher for the signal req_data_i, implemented with FF1, LT2 and the connected XOR gates; iii) phase matcher circuit to control ack_outport_o, composed by LT1 and the adjoining XOR gates; iv) phase matcher to control ack_data_o, comprised of LT3 and the connected XOR gates; v) transition merger circuit, connecting the input request signals to LT5; vi) arbiter handshake control, implemented with LT4 and the remaining gates.

The interaction between Arbiter and Outport Control is similar to a level-signaling handshake: first the Outport Control asserts the signal arbiter_request_i to request use of the output channel. Once the request is granted, the Arbiter signals the Outport Control by asserting arbiter_grant_i. From this moment on, the Outport Control has exclusive use of the output channel. This use can be relinquished by deasserting the signal arbiter_request_o. A new request can only be made after the deasserted arbiter_grant_i propagates to the Outport Control. The four-phase handshake prevents a new req_outport_i request, received soon after the end of the previous packet, from being sent to the output channel due to a delayed deassertion of arbiter_grant_i. The handshake behavior, implemented by the AND gate connected to the set pin of LT4, blocks new requests while the signal arbiter_grant_i is asserted. The grant signal, generated by the AND gate connecting arbiter_request_o and arbiter_grant_i, enables the latches LT1, LT2, and LT5, allowing handshake signals to propagate. This signal is disabled as soon as arbiter_request_o is deasserted in order to prevent incoming handshake signals from propagating. An example of this situation is given at the end of this Section.
The programmable phase matcher circuit is needed in order to support out-of-phase data requests. An XOR gate tests if the signal stored in LT2 is equal to req_data_i. The output of this gate is stored in FF1 on the low-to-high transition of the grant signal – that is, when the arbiter grant is given, through arbiter_grant_i. The stored value is fed to another XOR gate, which acts as a programmable inverter of req_data_i. When the signals are out-of-phase, logic ‘1’ is stored in FF1, inverting req_data_i in order to match the value stored in LT2.

Similar to FF1, FF3 stores the value of last_flit_i when the arbiter grant is given. The last flit of a packet can be detected when the last_flit_i becomes different from the value stored in FF3. This test is performed by the negative edge-sensitive flip-flop FF2 at the end of each data handshake cycle. Logic ‘1’ is stored in FF2 when the arrival of the last flit is identified, which causes LT4 to reset – deasserting arbiter_request_o. FF2 is reset when the grant signal goes to zero.

BaT-Hermes has timing constraints to guarantee that the signals req_data_i and last_flit_i are stable before a req_outport_i request can be issued.

---

**Figure 4.17 – Outport Control Architecture, with logic blocks highlighted.**

Figure 4.18 exemplifies the Outport Control operation: 1) the arrival of a req_outport_i request causes LT1 to be set, requesting use of the output channel to the Arbiter; 2) once the Arbiter grant is given, the last flit detector and programmable phase matcher are initialized in parallel, while the req_outport_i propagates to the output req_o; 3) the ack_i signal propagates to ack_outport_o, making LT1 opaque and LT2 transparent; 4) a sequence of req_data_i requests and ack_data_o acknowledges takes place while the packet is transferred; 5) at the end of the data handshake cycle the last flit detector notices the transition on last_flit_i, which makes the output of FF3 (last_flit) to switch to ‘1’, resetting LT4, and, as a consequence, deasserting arbiter_request_o. The internal grant signal falls as soon as the signal arbiter_request_o becomes a logic ‘0’, resetting FF3 – this is shown in the waveform as a spike on the last_flit signal. 6) A new req_outport_i request arrives, but is blocked until the deasserted arbiter_grant_i signal propagates to the Outport Control. If the four-phase handshake with the Arbiter was not implemented, the request that was blocked in step 6 would
have been propagated to the output channel, as the \textit{arbiter\_grant\_i} signal was still high when the \textit{req\_outport\_i} request arrived.

Figure 4.18 – Waveform showing the operation of an Output Control circuit.

4.2.8 Arbiter

The \textit{Arbiter} is responsible for mediating simultaneous output channel usage requests, granting access to only one \textit{Output Control} at a time. As discussed in Section 2.2.4, arbitration tasks in asynchronous circuits are usually performed with mutual exclusion gates. Bat-Hermes requires a 4-input arbiter, as each OI can communicate with up to 4 IIs. Figure 4.19 shows the 4-input arbiter proposed by Ghiribaldi et. al., and used in BaT-Hermes. Details about its design can be found in [GHI13]. The C-element and mutex gates needed to implement this circuit are available in the ASCEnD Cell Library.

Figure 4.19 – 4-input arbiter design proposed in [GHI13].

Figure 4.20 illustrates the operation of the \textit{Arbiter}: 1) in the absence of contention the grant is immediately given; 2) in the presence of contention, similar to the 2-input mutex, only one grant is given at a time and is retained for as long as the request signal is asserted; once the request is deasserted, grant is given to next request. This design provides a certain degree of fairness when choosing the next request to be granted: two requests signals that share the same mutex will not be given consecutive access if a request has
been made on the other mutex. This is similar to a round robin policy implemented between the requests on the left and right sides of the arbiter [GHI13]. This behavior can be seen on the waveform below.

![Waveform showing the operation of the Arbiter.](image)

**Figure 4.20 – Waveform showing the operation of the Arbiter.**

### 4.3 Functional Validation

The function validation of BaT-Hermes was performed through behavioral simulation using Mentor Graphics’ ModelSim. This section describes the techniques used to enable the simulation of an asynchronous circuit using a commercial tool designed for synchronous systems. Simulation results are also presented.

#### 4.3.1 Simulation Wrappers

The behavioral simulation of asynchronous circuits using tools designed for synchronous system simulation can be problematic, since there is no clock signal giving a timing reference to the circuit. Nevertheless, there are several ways to overcome this problem. One simple technique is to use the VHDL statement “after” on signal assignments to simulate propagation delays. This method, however, requires careful planning of the amount of delay applied to each signal in order to fulfill all timing constrains of the circuit. Another approach is to treat each module as an independent self-contained block with internal delay long enough to fulfill its timing requirements. For the simulation of BaT-Hermes, this technique was implemented in the form of circuit wrappers.

A wrapper is a VHDL entity with interface identical to the circuit under test (CUT). It employs external latches between the control ports of the instantiated CUT and the ports of the wrapper to control signal propagation. The order in which latches are enabled and disabled depends on the functionality of the block, and must emulate the circuit’s desired sequence of operations – during synthesis, this is achieved by the use of relative timing constraints. Only modules capable of generating requests and acknowledges need to be wrapped – for example, a Mousetrap stage, that issues an acknowledge when a request is received. However, other blocks can be wrapped to simulate propagation delay. Extra latches may be needed in order to fulfill dependencies related to signals coming from other blocks.

Wrappers for the following circuits were created to simulate BaT-Hermes: Input Buffer Control, FIFO Write Control, FIFO Read Control, and Routing Control. The first two actively perform handshakes, while the remaining were wrapped to simulate the propagation delay. Implementation files for these wrappers are available in Appendix A. Figure 4.21 shows the control logic of the wrapper used to simulate the Input Buffer Control circuit. Signals whose name starts with “hold” are used to control the extra latches – by asserting
them, the latches become opaque. Signals starting with “aux” are the outputs of the CUT that are connected to
the inputs of their respective latches. In this wrapper, latches to hold the following control signals are used:
ack_o, req_header_o, and req_data_o. Initially, after reset, the propagation of all these signals is blocked by
making the latches opaque. Next, the wrapper waits for a request signal to arrive and for the respective
acknowledge generated by the circuit under test – which, happens immediately, since there is no gate or wire
delays. After 1 ns, the acknowledge signals is released, by deasserting hold_ack_o. This simulates the time
needed by the data latch inside Input Buffer Control to store data. After 4 ns, the signals req_header_o and
req_data_o are released. This delay simulates the propagation time of the logic inside the circuit under test.
After that, all latches become opaque again, and the wrapper waits for a new request signal.

```vhdl
-- Control Logic
control: process
begin
  -- Initializing control signals
  hold_ack_o <= '0';
  hold_req_header_o <= '0';
  hold_req_data_o <= '0';
  wait until reset_i = '0';
  loop
    -- Hold all signals
    hold_ack_o <= '1';
    hold_req_header_o <= '1';
    hold_req_data_o <= '1';
    -- Wait for req_i request
    if (req_i = aux_ack_o) then
      wait until (req_i /= aux_ack_o);
    end if;
    -- Wait for ack_o acknowledge. (Mousetrap stage generates ACK right away)
    if (req_i /= aux_ack_o) then
      wait until (req_i = aux_ack_o);
    end if;
    -- Release #1: ack_o, after 1 ns (latch setup constraint)
    wait for 1 ns;
    hold_ack_o <= '0';
    -- Release #2, req_header_o and req_data_o, after 4 ns (ctrl logic delay)
    wait for 4 ns;
    hold_req_header_o <= '0';
    hold_req_data_o <= '0';
    -- Wait 1 ns before holding it again
    wait for 1 ns;
  end loop;
end process;
```

Figure 4.21 – Fragment showing the control logic of the wrapper used to simulate the Input Buffer
Control circuit.

4.3.2 Simulation Results

Three test cases were used to perform functional validation of BaT-Hermes. In the first, three packets
were sent from the east port to the north and local ports – the first and last packets to the latter. This test shows
packets flowing through the router and illustrates the operation of the programmable phase matcher
implemented in the Outport Control. Figure 4.22 shows the waveforms generated when the test was
performed: 1) the first flit was sent to the local port using the req_outport signal, causing the initialization
of the programmable phase matcher associated with the Outport Control interacting with the east port; 2) the
next flits are sent through the shared req_data signal; 3) the second packet is sent to the north port; by the end
of the transmission, the phase of the shared req_data signal will be different than its initial state; 4) similarly
to step 1, the programmable phase matcher is initialized to allow the out-of-phase \textit{req\_data} signal interact with the \textit{Outport Control}.

![Figure 4.22 – Operation of BaT-Hermes when sending packets from the East port to the North and Local ports.](image)

The second test simulates the router operating at its maximum throughput, which happens when there are five connections simultaneously active. In the situation shown in Figure 4.23(a), the connections happen as follows: \textit{\textit{i})} from the east port to the local port; \textit{\textit{ii})} from the local port to the north port; \textit{\textit{iii})} from the north port to the west port; \textit{\textit{iv})} from the west port to the south port; \textit{\textit{v})} from the south port to the east port.

The last scenario, shown in Figure 4.23(b), tests the router with presence of contention: all input ports, with the exception of the local port, try to send packets to the local output port. It can be seen that the \textit{Arbiter} circuit works as expected, allowing only one Input Interface access the Output Interface at any given time.
Figure 4.23 – BaT-Hermes a) operating at peak performance scenario, and b) with contention on the output of the local port.
5. IMPLEMENTATION

After functional validation, the Input and Output Interfaces of BaT-Hermes were synthesized to layout level using the STMicroelectronics 65nm CMOS technology. The ASCEnD cell library, designed at GAPH, provided the mutex and C-element cells required by the Arbiter circuit. The timing constraints identified during design were applied in the synthesis process to ensure the correct operation of the circuit. This Chapter details how the timing constraints were enforced and describes the synthesis flow used to generate the layouts. It also describes post-synthesis simulation with back-annotated delays performed to validate the circuit’s functionality.

5.1 Relative Timing Constraints

Bundled-data circuits fall under the category of self-timed asynchronous circuits, as they rely on carefully designed delay lines to ensure that handshake events take place only when data is valid – that is, the request signal must arrive only after the data signal is stable. Relative timing constraints define signal-arrival order, and can be used to fulfill these requirements. These constraints relate a base path, an enforced path, and a delay line: the enforced-path delay must be greater than the base-path delay; if it is not, a delay line must be inserted in the enforced path to fulfill the requirement. Commercial EDA tools, however, are not adapted to deal with such constraints in a natural manner [GHI13].

5.1.1 Enforcing Relative Timing Constraints on Synopsys Tools

Even though Synopsys tools do not natively support relative timing constraints, an approach for enforcing them is proposed in [GHI13], but no EDA support is made available in that reference. The methodology consists in iteratively extracting delays from base paths, using the get_timing_path command, and applying to enforced paths, using the set_min_delay command, until all constraints are met. However, several issues, explained below, may arise when setting minimum delay values. In order to avoid such problems, the following guidelines were defined during the development of this work for creating or enforcing constraints:

\( G_{i} \) A delay-line cannot be inserted in a path that is common to both base and enforced paths of the same constraint;

\( G_{ii} \) Delays cannot be set on paths with forks;

\( G_{iii} \) Given that \( d(x,y) \) calculates the path delay from point \( x \) to point \( y \), the delay of a path that goes through a delay line should be calculated as: \( d(a,b) = d(a,m) + d(m,n) + d(n,b) \), where \( a \) and \( b \) are, respectively, the start and endpoints of the path; and \( m \) and \( n \) are, respectively, the start and endpoints of the delay line;

\( G_{iv} \) A delay-line start and endpoint must be referenced by pin, not by net name.
The FIFO’s write control circuit, introduced in Section 4.2.3, can be used to illustrate the issues related to the set_min_delay command. Figure 5.1(a) shows this circuit after logic synthesis. The detailed part has one relative timing constraint: signal phase_select_i must arrive at pin $D$ of full_reg (full_reg/$D$) before req_i. If the delay is set on the path from req_i to full_reg/$D$, it is not possible to infer where the tool will insert the delay line, which can be: 

- **a)** between $U10/Z$ and $U9/D$ or $U9/Z$ and full_reg/$D$ the delay is applied to both req_i and phase_select_i, preventing the constraint from ever being met, and justifying the first guideline (Gi); 
- **b)** from the fork of req_i to $U9/B$ or $U10/B$ the delay is not applied to all paths from req_i to full_reg/$D$, therefore, more iterations may be needed in order to fix delays of remaining paths, which creates redundant delay lines, and justifies the second guideline (Gii); 
- **c)** from req_i to its first wire fork, which is the right place to add the delay-line as it affects the whole path of req_i without interfering with phase_select_i. A timing path, defined by a timing start and endpoint, is used by Synopsys’ Static Timing Analysis (STA) tool to calculate the delays across wires and gates. The timing at these points is broken by the STA tool, therefore delays can only be calculated to and from these points – the tool cannot compute the delay across a timing points. When the start and endpoints of a delay line are not timing start and endpoints, they become one, breaking the timing path through them. For this reason, the delay of a path that goes through a delay line must be measured as stated on the third guideline (Giii). Also for this reason, nets should not be used as start or endpoints, as it is not possible to infer in which cell the timing path will be broken, justifying the guideline number four (Giv).

Figure 5.1 – FIFO’s write controller circuit a) after logic synthesis and b) after adding extra cells to control the place where the delay line will be inserted.
As stated above, for the circuit shown in Figure 5.1(a) it is desired to insert a delay line between \textit{req\_i} and its first wire fork. However, wire forks are not referable in Synopsys’ tools, and, according to guideline number four, a delay-line start and endpoints must not be referenced by a net name. The approach taken to overcome these issues is the insertion of extra cells in order to create a path where delay lines can be appropriately placed. These extra cells cannot change the circuit’s logic - therefore, only buffers and pairs of inverters can be used. In an attempt to avoid inserting extra delay where interconnect delay alone are enough to fulfill timing requirements, a custom “wire cell”, composed of a single wire with input and output pins, was created. This cell was called \textit{HS65\_GS\_BFX0} to maintain the naming pattern used by the standard-cell library, and the property \textit{size\_only} was set, preventing it from being removed in logical optimizations, but allowing it to be replaced by actual buffers when needed. The extra cells were manually inserted, based on the timing constraints defined during design. Figure 5.1(b) shows the FIFO’s write control circuit with a pair of wire cells inserted. The delay-line can be created from the startpoint of the wire cell connected to \textit{req\_i} to the endpoint of the wire cell connected to the wire fork, ensuring that (Giv) is respected.

Situations where, due to how the circuit was synthesised, it is not possible to find a place meeting the guidelines to insert the extra cells can be solved by creating a new entity wrapping the problematic circuit. Since the synthesis flow keeps the hierarchy of the circuit, the additional entity will have a port for each signal, creating a place to insert extra cells.

The wire cell was modeled as a single wire with an input and an output pin, where the output is directly connected to the input pins. Parasitics were extracted using Mentor Graphic’s Calibre PEX tool. Electrical characterization was performed using Cadence’s Encounter Library Characterization tool employing the same non-linear table stimuli used in the characterization of a minimum-size buffer (output capacitance vector and input slope vector). The abstract view, shown in Figure 5.2, was created using Cadence’s Abstract Editor.Nota that the wire had to be split to generate the .LEF files, since the tool does not allow input and output pins to share the same net. However, this does not compromise the functionality of the circuit, given that this view is only required for place and route, and power and timing models employed the RC model of a full wire. Additionally, the wire cell had the same height of a standard-cell of the library, in order for it to be compatible during the placement; VDD and GND rails are also in the default position of the library. For this design, after the physical synthesis, all wire cells used in the IIs and OIs were substituted by buffers, suggesting that interconnect delay was not enough to fulfill the timing constraints on these circuits.

![Figure 5.2 – Abstract view of wire cell, rotated by 90°.](image)

### 5.1.2 Automated Constraint Enforcement

Scripts and functions were created to generate an automated environment for the iterative process of constraint enforcement described in the previous Section. This environment was called Asynchronous Constraints for Design Compiler (ACDC) and comprises a set of scripts written in TCL and Python that take as input an XML file describing the relative timing constraints of the circuit after logic synthesis, as exemplified in Figure 5.3. From this file, it generates TCL scripts capable of checking, setting and reporting
the defined constraints inside Synopsys’ tools. ACDC performs consistency checks in order to find mistakes in the manually generated XML file. At first, it checks if the file is well formatted and semantically correct – it verifies, for example, if there is exactly one delay target set on the enforced path, among many other verifications. After that, the validity of path names is check with Synopsys’ tools, and constraints containing non-existent paths are disabled. If a delay target is shared between constraints, the tool calculates the maximum target delay value of all shared constraints, and applies it as the path’s minimum delay, ensuring that all constraints are fulfilled.

The XML structure used by ACDC, shown in Figure 5.3, is very flexible and allows the representation of complex paths. Sets, which can be nested, are used to represent a group of paths. Each set has an action property that defines how its delay is calculated. The “sum” action returns the sum of all enclosed path- and set-delays, which is useful to represent linear paths. The “max” action returns the maximum delay among all enclosed paths and sets, and is helpful to represent path branches.

```
<design name="input_interface">

<constraint type="relative" name="wr_ctrl:: req_i vs. phase_select ">
  <description>data_o vs. wr/full_o to fifo/req_rd for all registers</description>
  <base>
    <path>
      <startpoint>phase_select_i</startpoint>
      <endpoint>full_reg/D*</endpoint>
    </path>
  </base>

  <enforced>
    <set action="sum">
      <path>
        <startpoint>req_i</startpoint>
        <endpoint>eco_req_i_1/A</endpoint>
      </path>
      <path delayTarget="true">
        <startpoint>eco_req_i_1/A</startpoint>
        <endpoint>eco_req_i/Z</endpoint>
      </path>
      <path>
        <startpoint>eco_req_i/Z</startpoint>
        <endpoint>full_reg/D*</endpoint>
      </path>
    </set>
  </enforced>
</constraint>
</design>
```

**Figure 5.3 – XML file describing the relative timing constraints of the FIFO Write Control circuit, shown in Figure 5.1(b).**

Additionally, two TCL functions were created to simplify delay extraction and enforcement. The `custom_get_delay` function returns the delay between the start and endpoints used as, respectively, the first and second parameters of the function. A third optional parameter is the type of delay, which can be maximum (max, the default option), or minimum (min). `Custom_set_delay` wraps the `set_min_delay` command to check if the delay constraint was properly set – stopping the execution if an error occurs. It takes three parameters: the start and endpoints of the delay line, and the target delay value.
Figure 5.4 – Fragment of the TCL scripts generated by ACDC. This function sets minimum delay constraints.

Figure 5.4 shows the ACDC-generated function used to set the constraints described in the XML file shown in Figure 5.3. Lines 4 through 11 extract the minimum and maximum delays of each path referenced in the XML file and assign them to custom variables. Line 19 calculates the maximum delay of the base path (AC_path1), while the next line calculates the minimum delay of the enforced path – composed by the delay from the path’s startpoint to the delay-line’s startpoint (AC_path0_min), delay-line’s delay (AC_path2_min), and delay from the delay-line’s endpoint to the path’s endpoint (AC_path3_min). Line 22 calculates the delay-target’s delay, as shown in Equation 1(b). Lines 25 and 26 apply the previously calculated delay to the delay line. If the delay line was shared between constraints, the maximum delay target value of all constraints would be calculated in line 25.

**Equation 1** – (a) The enforced-path’s delay must be equal or grater to the base-path’s delay; (b) equation to calculate the delay-line’s delay, by isolating it from (a).

\[ AC_{\text{path0}}_{\text{min}} + AC_{\text{path2}}_{\text{min}} + AC_{\text{path3}}_{\text{min}} \geq AC_{\text{path1}} \] \hspace{1cm} (a)

\[ AC_{\text{path2}}_{\text{min}} \geq AC_{\text{path1}} - (AC_{\text{path0}}_{\text{min}} + AC_{\text{path3}}_{\text{min}}) \] \hspace{1cm} (b)

### 5.2 Synthesis

Synthesis was the most challenging step in the development of this work. Synthesis flows for bundled-data asynchronous circuits using commercial tools are scarce. In fact, the only reference found was [GHI13], where very little details are given about the synthesis process. During the many synthesis attempts throughout the development of this work, it was observed that, as systems become larger, more relative timing constraints are needed in order to ensure proper operation of the circuit. Also, as the number of constraints grows, more overlaps may happen, which can cause the addition of redundant delay lines, degrading the circuit’s performance. Additionally, the number of iterations needed to fulfill the constraints increases.
Based on those observations, a synthesis flow was created in order to avoid conflicts between constraints. It consists of synthesizing each module of the system individually, in a bottom up manner. The post physical synthesis netlist of each circuit, generated after the respective timing constraints were met, is used, instead of its behavioral VHDL implementation, to instantiate the module. Even though it is not guaranteed that the constraints will still be fulfilled once the netlist is instantiated, as the placement and routing of the new circuit will be different from the one when the netlist was generated, it is fair to assume that, if needed, the amount of extra delay to meet the constraints will be small. Therefore, when synthesizing a circuit that instantiates other blocks, the relative timing constraints calculations can take into account the delays of these blocks, which are known at this stage (post-synthesis), possibly creating smaller delay lines. Note that the XML constraints file has to include the internal timing constraints of the instantiated modules as well, in order to guarantee that all constraints are met.

The following logic and physical flows were used to perform the synthesis of every component of BaT-Hermes. Synopsys' Design Compiler and IC Compiler were used to perform, respectively, logic and physical synthesis.

5.2.1 Logic Synthesis Flow

The logic synthesis flow can be divided in three phases: settings, initial synthesis, and optimization. In the first phase, tool configurations and design-specific settings are applied. Next, an initial synthesis is performed in order to map the design using the standard cell library. Finally, the design is optimized to a given target: low area, high performance, or low power.

In asynchronous systems, not only the logic function, but also the structure of the circuit defines the system’s behavior [GHI13]. To guarantee that Design Compiler does not make structural changes during synthesis, logic optimizations can be disabled with the command set_structure. Next, timing loops of the instantiated modules need to be disabled to avoid the insertion of loopbreakers. A timing loop is a combinational circuit with feedback signal – for example, the Mousetrap stage control latch, where its output signal is fed back as the enable signal, after passing through an XNOR gate. As such loops prevent Design Compiler from performing static timing analysis, the tool inserts extra buffers in the feedback path and disables the timing through them, breaking the feedback path – these buffers are called loopbreakers. If loopbreakers are inserted in a path with a relative timing constraint, ACDC may not be able to calculate the delay of the path, and may fail. Also, the dont_touch flag is set for instantiated modules to preserve the delay lines and prevent logic changes.

After these initial settings, logic synthesis is performed with the command compile_ultra. The flag no_autoungroup is used to keep the hierarchical structure of the circuit. The optimization phase consists in setting constraints to achieve a desired target and running compile_ultra again to optimize the circuit. BaT-Hermes aims for high performance, therefore maximum delay constraints for critical paths were set with the set_max_delay command. Additionally, timing loops related to the newly synthesized circuit also have to be disabled. The insertion of extra buffers to allow the creation of delay lines can be performed after the optimizations. Before storing the design, the size_only flag is applied to the circuit to allow cell resize and buffer insertion while preventing logic changes.
5.2.2 Physical Synthesis Flow

Similarly to the logic synthesis, the physical synthesis flow can also be divided in three phases: settings, initial synthesis and constraint enforcement. The input to this flow is the mapped netlist generated in the logic synthesis. Initially, the timing loops of the circuit to be synthesized are disabled and the *dont_touch* property is applied to the circuit. The *set_cost_priority* command is used to increase the priority of minimum delay constraints, since they are used by ACDC to enforce the relative timing constraints. Next, the initial physical synthesis is performed using a standard synthesis flow.

After initial synthesis, the *dont_touch* property is replaced by *size_only* in order enable cell resize and buffer insertion. The buffers and inverters from the core library present asymmetric rise and fall times, which can greatly degrade the performance of the circuit if used to create the delay lines. This is because the circuit is based on transition signaling and, therefore, the worst case between rising and falling propagation delays is employed for dimensioning the delay lines. The bigger the difference between these delays, the bigger is the impact in the average latency, as the handshake protocol relies on alternated falling and rising edges of the request signal. The solution found was to build the delay lines with standard cells designed for clock tree synthesis, as they are symmetric with respect to rise and fall times. The ACDC environment is then employed to enforce the relative timing constraints.

5.2.3 BaT–Hermes Synthesis

The BaT-Hermes router is composed by a set of ports, each with IIs and OIs. Each interface can be synthesized individually and saved as a hard macro. The same OI macro can be instantiated across all ports, whereas a different II macro must be generated for each port, since the Routing Control logic depends on the address of the router and the in which port it will be used. The router is assembled by interconnecting IIs to OIs, as discussed in Section 4.1.2.

Unfortunately, due to the problems that had to be overcome in order to create the synthesis flow, a full router could not be synthesized in time to be included in this work. However, the local port of a router, featuring 16-bit flit size and an 8-flit input buffer, was successfully synthesized and validated.

![Figure 5.5 – Layout of synthesized a) Input Interface and b) Output Interface circuits.](image-url)
The synthesis targeted the STMicroelectronics 65nm general-purpose standard-Vt CMOS technology. The ASCEnD cell library, available for this technology, provided the required asynchronous cells. Figure 5.5 shows the layout of the synthesized interfaces. Table 2 compares the total cell area of each circuit with the area occupied by buffers and inverters, most of which were used in delay lines. Post-synthesis functional validation is shown in Section 4.3.

<table>
<thead>
<tr>
<th></th>
<th>Total Cell Area</th>
<th>Buffer and Inverter Area</th>
<th>% of Buffers and Inverters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input Interface</td>
<td>4357µm²</td>
<td>1489µm²</td>
<td>34.1%</td>
</tr>
<tr>
<td>Output Interface</td>
<td>1428µm²</td>
<td>600µm²</td>
<td>42.0%</td>
</tr>
</tbody>
</table>

### 5.3 Post-Synthesis Validation

After physical synthesis, the netlist and .SDF file of the Input and Output Interfaces were exported and used to perform the post-synthesis functional validation with back-annotated delays. Cadence’s SimVision simulator was employed in this task.

#### 5.3.1 Input Interface

The test case applied to the II simulates an IP Core connected to the local port sending packets to all other cores of a 3x3 mesh network. Figure 5.6 shows an overview of the waveforms generated by this test case. Note the stalled `ack_o` signal (1) generated when the input buffer becomes full. The signals that communicate with the Output Interface were grouped to help identifying the destination of each request: Outport 0 groups signals sent to the east port; Outport 1 combine signals sent to the west port; signals intended for the north port are grouped under Outport 2; Outport 3 combines signals sent to the south port. Shared signals, like `data_o`, `req_data_o`, and `last_flit_o` were replicated in each of these groups to help to comprehend which handshakes are taking place at each instant. The signals from Outport 4 refer to the local port and are not shown, since loopback connections are not supported by BaT-Hermes.
Figure 5.6 – Post-synthesis simulation of an Input Interface sending packets to the a) west, b) east, c) north, and d) south Output Interfaces.

Figure 5.7 shows a detailed view of Figure 5.6(a) to illustrate the flits flowing through the II. Arrows show the propagation of the first flit of each packet sent to the west OI.

Figure 5.7 - Detailed view of Figure 5.6(a), showing flits propagating through the Input Interface.

The forward latency, measured as the time it takes for an incoming request to propagate to the output, is 2.255ns. The average time between data handshakes is 1.08ns. At the input, the average delay between successive acknowledges is 0.745ns.

### 5.3.2 Output Interface

The test case applied to the Output Interface simulates four IIs trying to send a packet simultaneously. Figure 5.8 shows the simulation waveforms. Note that the req_data_i (1) signals, connected to Outputs 0 and 2, are out-of-phase with the respective ack_data_o. This allows simulating the programmable phase matcher circuit, for the sake of validation. All req_outport_i (2) signals switch at the same time, competing for access to the output channel. Arbitration works as expected, granting access to only one Outport Control at any given time.
Figure 5.8 - Post-synthesis simulation of an Output Interface receiving four packets simultaneously.

The average forward delay, computed from the $req\_data\_i$ to $req\_o$, is 0.677ns. The average cycle time at the output channel is 0.95ps.
6. **Conclusions and Future Work**

This work proposed an asynchronous transition-signaling bundled-data NoC router, called BaT-Hermes. The design, based on the highly decoupled architecture of the YeAH! router, is composed of several independent modules, that can be connected to assemble a router. A full BaT-Hermes was validated through behavioral simulation.

Unfortunately, the synthesis process was far more complex than anticipated and a full router could not be synthesized in time to be included in this work, and is left as a future work. However, one port, composed by an II and OI was synthesized to the STMicroelectronics 65nm CMOS technology and validated through post-synthesis simulation, enabling the validation of the proposed synthesis methodology and the ACDC environment. In this way, apart from the designed NoC router, another contribution from this work is the methodology for synthesis of bundled-data circuits using commercial CAD tools, along with automated environment for enforcing relative timing constraints, namely ACDC.

### 6.1 Future Work

This work presents many topics for further research. The most immediate is the synthesis of a full BaT-Hermes router and its comparison with the fully synchronous YeAH! router. Due to the many structural similarities between them, this is a great opportunity to assess the differences of each implementation style with respect to metrics like latency, throughput, area, and power consumption.

Another interesting topic for research is alternative ways to implement the router, aiming for a simpler synthesis process. One idea is to only use Mousetrap stages as storage elements, implementing the router as a “pure pipeline”. This approach can greatly simplify the synthesis because relative timing constraints are restricted to each pair of stages. An interesting study could be conducted comparing the final circuit performance with the synthesis effort for various design techniques.

Other possibility for future work is on the improvement of the ACDC environment to support automatic constraint detection. Additionally, related work could be conducted about synchronization interfaces that need to be used when interfacing asynchronous and synchronous circuits.
REFERENCES


A. BaT–Hermes VHDL Description

This appendix contains the VHDL implementation code and simulation wrappers for each module of BaT-Hermes.

A.1 Settings Package

```vhdl
-- VHDL source code for the settings package

library ieee;
use ieee.std_logic_1164.all;
use ieee.std_logic_unsigned.all;

package hermes_bd_package is
  -- The flit size should be modified with scripts
  -- since was not possible to change its value
  -- with generics yet
  constant FLIT_SIZE : integer range 1 to 64 := 16;
  constant PORTS_NUMBER: integer := 5;
  -- Port IDs
  constant EAST : integer := 0;
  constant WEST : integer := 1;
  constant NORTH : integer := 2;
  constant SOUTH : integer := 3;
  constant LOCAL : integer := 4;

  -- Types
  type flit_size_array is array(natural range 0 to FLIT_SIZE-1 downto 0) of std_logic_vector(FLIT_SIZE-1 downto 0);

  -- Functions
  function xor_all_bits_considering_comm_ports (X : std_logic_vector; COMM_PORTS : std_logic_vector; THIS_PORT : integer) return std_logic;
  function xor_all_bits_considering_comm_ports (X : std_logic_vector; COMM_PORTS : std_logic_vector; N : integer; COMM_PORTS : std_logic_vector; THIS_PORT : integer) return std_logic;
  function number_of_used_ports (COMM_PORTS : std_logic_vector; THIS_PORT : integer) return integer;

end package;

package body hermes_bd_package is

  -- Helper function to XOR all bits, considering COMM_PORTS and THIS_PORT values.
  function xor_all_bits_considering_comm_ports (X : std_logic_vector; COMM_PORTS : std_logic_vector; THIS_PORT : integer) return std_logic is
    begin
      variable aux : std_logic := '0';
      begin
        for i in X'range loop
          if ((i /= THIS_PORT) and (COMM_PORTS(i) = '1')) then
            aux := aux xor X(i);
          end if;
        end loop;
        return aux;
      end xor_all_bits_considering_comm_ports;

  -- Helper function to XOR all bits but one, considering COMM_PORTS and THIS_PORT values.
  function xor_all_bits_but_one_considering_comm_ports (X : std_logic_vector; COMM_PORTS : std_logic_vector; THIS_PORT : integer) return std_logic is
    begin
      variable aux : std_logic := '0';
      begin
        for i in X'range loop
          if ((i /= N) and (i /= THIS_PORT) and (COMM_PORTS(i) = '1')) then
            aux := aux xor X(i);
          end if;
        end loop;
        return aux;
      end xor_all_bits_but_one_considering_comm_ports;

end package;
```

Figure A.1 – VHDL source code for the settings package (part 1 of 2).
-- Helper function to calculate the actual number of outports being used
function number_of_used_ports (COMM_PORTS : std_logic_vector; THIS_PORT : integer) return integer is
    variable cnt : integer := 0;
begin
    for i in COMM_PORTS'range loop
        if ((i /= THIS_PORT) and (COMM_PORTS(i) = '1')) then
            cnt := cnt + 1;
        end if;
    end loop;
    return cnt;
end number_of_used_ports;

Figure A.2 – VHDL source code for the settings package (part 2 of 2).

A.2 FIFO

library ieee;
use ieee.std_logic_1164.all;

entity fifo_wr_ctrl is
    port(
        reset_i : in std_logic; -- Active-high reset signal.
        en_i    : in std_logic; -- Signal to enable this controller.
        phase_select_i : in std_logic; -- Indicates that the data has been stored.
        req_i   : in std_logic; -- Request signal indicating that new data is available to be written.
        ack_o   : out std_logic; -- Acknowledge signal relative to req_i. Indicates that the data has been read.
        full_o  : out std_logic; -- Signal that the register is full. This, latches the output of the register.
        empty_i : in std_logic; -- Signal that controls the register latch. High makes latch transparent, low
        reg_en_o: out std_logic; -- Signal to control the register(latch). High makes latch transparent(store); low
        -- turns it opaque(hold).
    );
end fifo_wr_ctrl;

architecture fifo_wr_ctrl of fifo_wr_ctrl is
    signal full : std_logic; -- Same as full_o
    signal req : std_logic; -- req_i signal phase-matched to full_o
    signal reg_en : std_logic; -- Signal that controls the full_o latch; Same as reg_en_o
    full_o <= full;
    ack_o <= full;
    reg_en_o <= reg_en;

begin
    full_o <= full;
    ack_o <= full;
    reg_en_o <= reg_en;

    -- Matching req_i signal to the phase of full_o
    req <= req_i XOR phase_select_i;

    -- Enables next write if the register's data has been read
    -- [It's this register's turn to write AND the register is empty (full/empty handshake is complete)]
    reg_en <= en_i AND (empty_i XOR full);

    -- Latch that controls requests (full_o)
    process(reset_i, reg_en, req)
    begin
        if (reset_i = '1') then
            full <= '0';
        elsif (reg_en = '1') then
            full <= req;
        end if;
    end process;

end fifo_wr_ctrl;

Figure A.3 – VHDL source code for FIFO Write Control module.
Figure A.4 – VHDL source code for simulation wrapper of FIFO Write Control (part 1 of 2).
loop
    begin
        -- Wait request (full_o transitions when a valid req_i transition is made)
        if (en_i /= '1') OR (aux_full_o = empty_i) then
            wait until (en_i = '1') AND (aux_full_o /= empty_i);
        end if;
        -- Release #1: full_o, after 1 ns
        wait for 1 ns;
        hold_full_o <= '0';
    end loop;
end process;

loop
    begin
        -- Release #2: full_o, after 1 ns
        wait for 1 ns;
        hold_full_o <= '0';
        -- Hold handshake signals
        hold_ack_o <= '1';
        hold_full_o <= '1';
    end loop;
end process;

-- Wait acknowledge from full/empty handshake
if (empty_i /= aux_full_o) then
    wait until empty_i = aux_full_o;
end if;

end loop;
end process;

-- Latches to hold signals
latch_ack_o: process(reset_i, hold_ack_o, aux_ack_o)
begin
    if ((reset_i = '1') OR (hold_ack_o = '0')) then
        ack_o <= aux_ack_o;
    end if;
end process;

latch_full_o: process(reset_i, hold_full_o, aux_full_o)
begin
    if ((reset_i = '1') OR (hold_full_o = '0')) then
        full_o <= aux_full_o;
    end if;
end process;

end sim;

Figure A.5 - VHDL source code for simulation wrapper of FIFO Write Control (part 2 of 2).

library ieee;
use ieee.std_logic_1164.all;

entity fifo_rd_ctrl is
    port(
        reset_i : in std_logic;  -- Active-high reset signal.
        en_i : in std_logic;     -- Signal to enable this controller.
        full_i : in std_logic;   -- Request signal indicating that new data has been written on the register.
        empty_o : out std_logic; -- Acknowledge signal relative to full_i. Indicates that the data has been read.
        phase_select_i : in std_logic; -- Signal used for empty_o phase matching.
        req_o : out std_logic;    -- Request signal indicating that new data is available to be read.
        ack_i : in std_logic;     -- Acknowledge signal relative to req_o. Indicates that the data has been read.
    );
end fifo_rd_ctrl;

Figure A.6 - VHDL source code for FIFO Read Control module (part 1 of 2).
Figure A.7 - VHDL source code for FIFO Read Control module (part 2 of 2).

Figure A.8 - VHDL source code for simulation wrapper of FIFO Read Control (part 1 of 2).
architecture sim of fifo_rd_ctrl_wrapper is

--- Control signals
signal hold_en_i : std_logic;
signal hold_empty_o : std_logic;

--- Aux signals
signal aux_empty_o : std_logic;
signal aux_en_i : std_logic;

begin

--- Instance of fifo_rd_ctrl
fifo_rd_ctrl_i: entity work.fifo_rd_ctrl
port map (
  reset_i => reset_i,
  en_i => aux_en_i,
  full_i => full_i,
  empty_o => aux_empty_o,
  phase_select_i => phase_select_i,
  req_o => req_o,
  ack_i => ack_i
);

--- Control logic
control: process
begin
  -- Initializing control signals
  hold_en_i <= '0';
  hold_empty_o <= '0';
  wait until reset_i = '0';
  loop
    hold_en_i <= '1';
    hold_empty_o <= '1';
    -- Wait for valid handshake (rd_ctrl enabled and full_i request)
    if (en_i /= '1') OR (full_i = aux_empty_o) then
      wait until ((en_i = '1') AND (full_i /= aux_empty_o));
    end if;
    -- Release #1: en_i, after 1 ns
    wait for 1 ns;
    hold_en_i <= '0';
    -- Wait for ack_i acknowledge
    if (aux_empty_o = full_i) then
      wait until (aux_empty_o = full_i);
    end if;
    -- Release #2: empty_o, after 1 ns
    wait for 1 ns;
    hold_empty_o <= '0';
    -- Wait 1 ns before holding it again
    wait for 1 ns;
  end loop;
end process;

-- Latches to hold signals
latch_en_i: process(reset_i, hold_en_i, en_i)
begin
  if ((reset_i = '1') OR (hold_en_i = '0')) then
    aux_en_i <= en_i;
  end if;
end process;

latch_empty_o: process(reset_i, hold_empty_o, aux_empty_o)
begin
  if ((reset_i = '1') OR (hold_empty_o = '0')) then
    empty_o <= aux_empty_o;
  end if;
end process;
end sim;

Figure A.9 - VHDL source code for simulation wrapper of FIFO Read Control (part 2 of 2).
--- @file fifo.vhd
--- @brief Handshake-based FIFO using transition signaling protocol.
--- @author Matheus Gibiluka, matheus.gibiluka@acad.pucrs.br
--- @date 2013-07-14

library ieee;
use ieee.std_logic_1164.all;
use ieee.std_logic_misc.all;
use work.hermes_bd_package.all;

entity fifo is
  generic
  -- BUFFER_DEPTH: Defines the depth of the buffer.
  BUFFER_DEPTH : integer := 8 ;
  port
  reset_i  : in std_logic ; --! Active-high reset signal.
  req_wr_i : in std_logic;  --! Request signal indicating that new data is available to be written.
  ack_wr_o : out std_logic; --! Acknowledge signal relative to req_wr_i. Indicates that the data has been stored.
  req_rd_i : in std_logic;  --! Request signal indicating that new data is available to be read.
  ack_rd_o : out std_logic; --! Acknowledge signal relative to req_rd_o. Indicates that the data has been read.
  data_i   : in std_logic_vector(BUFFER_DEPTH-1 downto 0) ; --! Data input.
  data_o   : out std_logic_vector(BUFFER_DEPTH-1 downto 0) ; --! Data output.
  end port;

architecture circular of fifo is
  signal write_ptr : std_logic_vector(BUFFER_DEPTH-1 downto 0); -- Write Pointer (1-hot)
  signal read_ptr  : std_logic_vector(BUFFER_DEPTH-1 downto 0); -- Read Pointer (1-hot)
  signal ack_wr_ctrl : std_logic_vector(BUFFER_DEPTH-1 downto 0); -- Acknowledge signals generated by wr_ctrl (data has been written)
  signal req_rd_ctrl : std_logic_vector(BUFFER_DEPTH-1 downto 0); -- Request signals generated by rd_ctrl (data is available to be read)
  signal full      : std_logic_vector(BUFFER_DEPTH-1 downto 0); -- Request signals generated by wr_ctrl (data is available to be read)
  signal empty    : std_logic_vector(BUFFER_DEPTH-1 downto 0); -- Request signals generated by rd_ctrl (data has been read)
  signal reg_en   : std_logic_vector(BUFFER_DEPTH-1 downto 0); -- Register enable signals generated by wr_ctrl.
  signal ack_wr    : std_logic; --! Same as ack_wr_o
  signal req_rd    : std_logic; --! Same as req_rd_o
  signal next_wr_ptr : std_logic; -- Control signal for Write Counter (increment on rising edge)
  signal next_rd_ptr : std_logic; -- Control signal for Read Counter (increment on rising edge)
  type reg_t is array (0 to BUFFER_DEPTH-1) of std_logic_vector(FLIT_SIZE-1 downto 0);
  signal reg : reg_t; -- Register bank (latches)
begin
  -- Write Interface
  process(reset_i, next_wr_ptr)
  begin
    if (reset_i = '1') then
      write_ptr <= (others=>'0');
    elsif next_wr_ptr = '1' then
      -- Ring counter
      write_ptr <= write_ptr(BUFFER_DEPTH-2 downto 0) & write_ptr(BUFFER_DEPTH-1);
    end if;
  end process;
  -- Write Control
  for i in 0 to BUFFER_DEPTH-1 generate
    -- Generate for even-depth FIFO
    wr_ctrl_even : if ((BUFFER_DEPTH rem 2) = 0) then
  end for;
end fifo;

Figure A.10 - VHDL source code for FIFO module (part 1 of 3).
-- Generating control for even-depth FIFO and even control index
wr_full_gen_1: if ((i rem 2) = 0) generate
  wr_ctrl_i: entity work.fifo_wr_ctrl_wrapper
  port map (  
    reset_i,  
    en_i,  
    req_i,  
    ack_o,  
    empty_i,  
    full_o,  
    phase_select_i => full(i+1),  
    reg_en_o => reg_en(i)  
  );
end generate wr_full_gen_1;

-- Generating control for even-depth FIFO and odd control index
wr_full_gen_2: if ((i rem 2) /= 0) generate
  wr_ctrl_i: entity work.fifo_wr_ctrl_wrapper
  port map (  
    reset_i,  
    en_i,  
    req_i,  
    ack_o,  
    empty_i,  
    full_o,  
    phase_select_i => full(i),  
    reg_en_o => reg_en(i)  
  );
end generate wr_full_gen_2;
end generate wr_ctrl_even;

-- Generate for odd-depth FIFO
wr_ctrl_odd: if ((BUFFER_DEPTH rem 2) /= 0) generate
  -- Generating control for odd-depth FIFO and even control index
  wr_full_gen_3: if ((i rem 2) = 0) generate
    wr_ctrl_i: entity work.fifo_wr_ctrl_wrapper
    port map (  
      reset_i,  
      en_i,  
      req_i,  
      ack_o,  
      empty_i,  
      full_o,  
      phase_select_i => '1',  
      reg_en_o => reg_en(i)  
    );
  end generate wr_full_gen_3;
end generate wr_ctrl_odd;
end generate wr_ctrl;

-- Read Interface
-----------------------------------------------------------------------------------------------------------------------------------
req_rd <= xor_reduce(req_rd_ctrl); -- Generates req_rd with the correct phase
req_rd_o <= req_rd;

-- Read Counter
next_rd_ptr <= ack_rd_i AND req_rd; -- Rising edge means that handshake was completed
read_counter: process (reset_i, next_rd_ptr)
begin
if (reset_i = '1') then
  read_ptr <= (others=>'0');
  read_ptr(0) <= '1';
elsif next_rd_ptr='1' then  
  -- Ring counter
  read_ptr <= read_ptr(BUFFER_DEPTH-2 downto 0) & read_ptr(BUFFER_DEPTH-1) ;
end if;
end process;

-- Read Control
rd_ctrl: for i in 0 to BUFFER_DEPTH-1 generate
  -- Generate for even-depth FIFO
  rd_ctrl_even: if ((BUFFER_DEPTH rem 2) = 0) generate
    rd_ctrl_gen_1: if ((i rem 2) = 0) generate
      rd_ctrl_i: entity work.fifo_rd_ctrl_wrapper
      port map (  
        reset_i,  
        en_i,  
        req_i,  
        ack_o,  
        empty_i,  
        full_o,  
        phase_select_i => '0',  
        reg_en_o => reg_en(i)  
      );
    end generate rd_ctrl_gen_1;
  end generate rd_ctrl_even;
end generate rd_ctrl;

Figure A.11 - VHDL source code for FIFO module (part 2 of 3).
en_i          => read_ptr(i),
full_i        => full(i),
empty_o       => empty(i),
phase_select_i => empty(i+1),
req_o         => req_rd_ctrl(i),
ack_i         => ack_rd_i
);
end generate rd_ctrl_gen_1;

-- Generating control for even-depth FIFO and odd control index
rd_ctrl_even: if ((i rem 2) /= 0) generate
  rd_ctrl_i: entity work.fifo_rd_ctrl_wrapper
  port map (
    reset_i     => reset_i,
en_i         => read_ptr(i),
full_i       => full(i),
empty_o      => empty(i),
phase_select_i => empty(i-1),
req_o        => req_rd_ctrl(i),
ack_i        => ack_rd_i
);
end generate rd_ctrl_gen_2;
end generate rd_ctrl_even;

-- Generate for odd depth FIFO
rd_ctrl_odd: if ((BUFFER_DEPTH rem 2) /= 0) generate
  rd_ctrl_i: entity work.fifo_rd_ctrl_wrapper
  port map (
    reset_i     => reset_i,
en_i         => read_ptr(i),
full_i       => full(i),
empty_o      => empty(i),
phase_select_i => '0',
req_o        => req_rd_ctrl(i),
ack_i        => ack_rd_i
);
end generate rd_ctrl_gen_3;

-- Generating control for odd-depth FIFO and even control index
rd_ctrl_gen_3: if ((i rem 2) = 0) generate
  rd_ctrl_i: entity work.fifo_rd_ctrl_wrapper
  port map (
    reset_i     => reset_i,
en_i         => read_ptr(i),
full_i       => full(i),
empty_o      => '0',
phase_select_i => '1',
req_o        => req_rd_ctrl(i),
ack_i        => ack_rd_i
);
end generate rd_ctrl_gen_3;
end generate rd_ctrl_odd;
end generate rd_ctrl;

-- Data Path
--=========================================================================
-- Data registers (latches)
reg_gen: for i in 0 to BUFFER_DEPTH-1 generate
  data_reg: process(reg_en(i), data_i)
  begin
    if (reg_en(i) = '1') then
      reg(i) <= data_i;
    end if;
  end process;
end generate reg_gen;

-- MUX that selects which register will be read
reg_mux_gen: for i in 0 to BUFFER_DEPTH-1 generate
  data_o <= (reg(i) when read_ptr(i) = '1' else
             others => 'Z');
end generate reg_mux_gen;
end circular;

Figure A.12 - VHDL source code for FIFO module (part 3 of 3).
A.3 Input Buffer

```vhdl
-- @file mousetrap_ctrl.vhd
-- @brief Control stage of the Mousetrap pipeline template.
-- @author Matheus Gibiluka, matheus.gibiluka@acad.pucrs.br
-- @date 2013-11-10

------------------------------------------------------------------------

-- Interface description:
-- -------------
-- |             |
-- reset_i---------->|             |
-- -------------
-- |             |
-- req_i---------->|             |---------->req_o
-- -------------
-- |             |
-- |             |---------->ack_i
-- -------------
-- |             |
-- |             |---------->en_o
-- -------------

library ieee;
use ieee.std_logic_1164.all;

entity mousetrap_ctrl is
  port(
    reset_i : in std_logic; --! Active-high reset signal.
    req_i : in std_logic; --! Incoming Request signal.
    ack_o   : out std_logic; --! Acknowledge relative to req_i.
    req_o   : out std_logic; --! Outgoing request signal. Same as req_i, after passing through the latch.
    ack_i   : in std_logic; --! Acknowledge relative to req_o.
    en_o    : out std_logic --! Active-high enable signal used to control the latch. May be used to control a data latch.
  );
end mousetrap_ctrl;

architecture mousetrap_ctrl of mousetrap_ctrl is
  signal req : std_logic; -- req_i after passing through latch.
  signal en  : std_logic; -- Signal to control the latch.
begin
  en <= req XNOR ack_i;
  en_o <= en;
  req_o <= req;
  ack_o <= req;
  -- Latch
  process(reset_i, en, req_i)
  begin
    if (reset_i = '1') then
      req <= '0';
    elsif (en = '1') then
      req <= req_i;
    end if;
  end process;
end mousetrap_ctrl;
```

Figure A.13 - VHDL source code for Mousetrap stage.
Library ieee;
use ieee.std_logic_1164.all;

entity request_splitter is
port(
  reset_i : in std_logic;  -- Active-high reset signal.
  req_i : in std_logic;  -- Request signal.
  ack_o : out std_logic;  -- Acknowledge relative to req_i (ack_header_i xor with ack_data_i).
  req_header_o : out std_logic;  -- Header request signal. The next req will be redirected to req_data_o.
  ack_header_i : in std_logic;  -- Acknowledge relative to req_header_o.
  req_data_o : out std_logic;  -- Request signal.
  data_hs : in std_logic;  -- Signal that indicates when a data handshake will be performed.
  hs_complete : in std_logic;  -- Signal indicating that the next request to be made is a header request.
  last_flit_lvl_i : in std_logic);  -- Signal that indicates when a handshake is completed.
end entity request_splitter;

architecture request_splitter of request_splitter is
begin
  process(reset_i, hs_complete)
  begin
    if (reset_i = '1') then
      ack_o <= '0';
      data_hs <= '0';
      req_header <= '0';
      req_data <= '0';
    elsif (hs_complete'event and (hs_complete = '1')) then
      data_hs <= not(data_hs and last_flit_lvl_i);
    end if;
  end process;

  -- req_i after being phase-matched to req_header.
  signal req_header : std_logic;
  signal req_data : std_logic;
  signal req_data_phasematched : std_logic;
  signal req_data_o : std_logic;
  signal req_header_o : std_logic;
  signal data_hs : std_logic;
  signal hs_complete : std_logic;

  req_header_phasematched <= req_i XOR req_header_o;
  req_data_phasematched <= req_i XOR req_data;

  process(reset_i, data_hs)
  begin
    if (reset_i = '1') then
      req_header_phasematched <= '0';
      req_data_phasematched <= '0';
    elsif (data_hs = '1') then
      req_header_phasematched <= req_header_phasematched XOR req_data_phasematched;
      req_data_phasematched <= req_data_phasematched XOR req_data;
    end if;
  end process;

  process(reset_i, data_hs, req_header_phasematched)
  begin
    if (reset_i = '1') then
      req_header_phasematched <= '0';
      req_data_phasematched <= '0';
    elsif (data_hs = '1') then
      req_header_phasematched <= req_header_phasematched XOR req_data_phasematched;
      req_data_phasematched <= req_data_phasematched XOR req_data;
    end if;
  end process;
end architecture request_splitter;

Figure A.14 - VHDL source code for request splitter circuit.
library ieee;
use ieee.std_logic_1164.all;
use ieee.std_logic_unsigned.all;
use work.hermes_bd_package.all;

entity input_buffer_ctrl is
port(
  reset_i : in std_logic; -- Active-high reset signal.
  req_o   : out std_logic; -- Request signal indicating that new data is available on data.i.
  ack_o   : out std_logic; -- Acknowledge relative to req_i. Indicates that the data was consumed.
  data_i  : in std_logic_vector(FLIT_SIZE-1 downto 0); -- Data input.
  data_o  : out std_logic_vector(FLIT_SIZE-1 downto 0); -- Data output.
  last_flit_o : out std_logic -- Transition-encoded signal indicating that the flit on data_o is the last flit of the current packet.
  ack_data_i : in std_logic; -- Acknowledge relative to req_data_o. Indicates that the payload was consumed.
  req_data_o : out std_logic; -- Request signal indicating that the flit on data_o is the packet header.
  ack_header_i : in std_logic; -- Acknowledge signal relative to req_header_o. Indicates that header was consumed.
  req_header_o : out std_logic; -- Request signal indicating that the flit on data_o is the packet header.
  en_o    => en
); 

architecture input_buffer_ctrl of input_buffer_ctrl is

begin

data_o <= data;

-- Mousetrap stage at the input

-- Control Latch
mousetrap_ctrl_i: entity work.mousetrap_control
port (reset_i, req_i, ack_o, req_o, ack_i, en_o
); 

begin

-- Data Latch
process(reset_i, en, data_i)
begin
if (reset_i = '1') then
  data <= (others => '0');
elsif (en = '1') then
  data <= data_i;
end if;
end process;

-- Request Splitter, to generate req_header_o and req_data_o

Figure A.15 - VHDL source code for Input Buffer Control module (part 1 of 2).
--=========================================================================
request_splitter_i: entity work.request_splitter
port map (
  reset,           => reset_i,
  req_i            => req_i,
  ack_o            => ack_o,
  req_header_o     => req_header_o,
  ack_header_i     => ack_header_i,
  req_data_o       => req_data_o,
  ack_data_i       => ack_data_i,
  last_flit_lvl    => last_flit_lvl
);

-- Control State Machine
--=========================================================================
-- flit_counter Flip-Flop (Sampled at the request transition, reset by last_flit_lvl)
process(last_flit_lvl, en)
begin
  if (last_flit_lvl = '1') then
    flit_counter <= (others => '0');
  elsif (en'event and (en = '0')) then
    if (size_flit = '1') then
      flit_counter <= data_i;
    else
      flit_counter <= flit_counter - 1;
    end if;
  end if;
end process;

-- last_flit_lvl Flip-Flop (Sampled at request transition)
process(reset_i, en)
begin
  if (reset_i = '1') then
    last_flit_lvl <= '1';
  elsif (en'event and (en = '0')) then
    if (flit_counter = x'1') then
      last_flit_lvl <= '1';
    else
      last_flit_lvl <= '0';
    end if;
  end if;
end process;

-- size_flit Flip-Flop (Sampled at request transition)
process(reset_i, en)
begin
  if (reset_i = '1') then
    size_flit <= '0';
  elsif (en'event and (en = '0')) then
    size_flit <= last_flit_lvl;
  end if;
end process;

-- Transition-encoded last_flit_o Generator
--=========================================================================
-- last_flit_o Flip-Flop (Sampled at low-to-high transition of last_flit_lvl)
last_flit_o <= last_flit;

process(reset_i, last_flit_lvl)
begin
  if (reset_i = '1') then
    last_flit <= '0';
  elsif (last_flit_lvl'event and (last_flit_lvl = '1')) then
    last_flit <= not(last_flit);
  end if;
end process;

end input_buffer_ctrl;

Figure A.16 - VHDL source code for Input Buffer Control module (part 2 of 2).
```
-- File: input_buffer_ctrl_wrapper.vhd
-- @brief Wrapper for Input Buffer Control
-- @author Mathieu Gibiluka, mathieu.gibiluka@acad.pucrs.br
-- @date 2013-08-06

interface input_buffer_ctrl_wrapper is
  port (reset_i : in std_logic; -- Active-high reset signal.
        req_i : in std_logic; -- Request signal indicating that new data is available on data_i.
        ack_o : out std_logic; -- Acknowledge relative to req_i. Indicates that the data was consumed.
        req_header_o : out std_logic; -- Request header to data_o. Indicates that header was consumed.
        data_i : in std_logic_vector(FLIT_SIZE-1 downto 0); -- Data input.
        req_data_o : out std_logic; -- Request signal indicating that the flit on data_o is payload.
        ack_header_i : in std_logic; -- Acknowledge signal relative to req_header_o. Indicates that header was consumed.
        ack_data_i : in std_logic; -- Acknowledge relative to req_data_o. Indicates that payload was consumed.
        data_o : out std_logic_vector(FLIT_SIZE-1 downto 0); -- Data output.
        last_flit_o : out std_logic; -- Active-high signal indicating that the flit on data_o is the last flit of the current packet.
  );
end input_buffer_ctrl_wrapper;

architecture sim of input_buffer_ctrl_wrapper is
  -- Control signals
  signal hold_ack_o : std_logic;
  signal hold_req_header_o : std_logic;
  signal hold_req_data_o : std_logic;
  -- Aux signals
  signal aux_ack_o : std_logic;
  signal aux_req_header_o : std_logic;
  signal aux_req_data_o : std_logic;
begin
  -- Instance of input_buffer_ctrl
  input_buffer_ctrl_i : entity work.input_buffer_ctrl
    port map (reset_i => reset_i,
              req_i => req_i,
              ack_o => aux_ack_o,
              data_i => data_i,
              req_header_o => aux_req_header_o,
              ack_header_i => aux_ack_header_o,
              req_data_o => aux_req_data_o,
              ack_data_i => aux_ack_data_i,
              data_o => data_o,
              last_flit_o => last_flit_o);
  -- Control Logic
  control: process
  begin
    -- ...
end;```

Figure A.17 - VHDL source code for simulation wrapper of Input Buffer Control (part 1 of 2).
-- Initializing control signals
hold_ack_o <= '0';
hold_req_header_o <= '0';
hold_req_data_o <= '0';
wait until reset_i = '0';
loop
  -- Hold all signals
  hold_ack_o <= '1';
  hold_req_header_o <= '1';
  hold_req_data_o <= '1';
  -- Wait for req_i request
  if (req_i = aux_ack_o) then
    wait until (req_i /= aux_ack_o);
  end if;
  -- Wait for ack_o acknowledge. (Mousetrap stage generates ACK right away)
  if (req_i /= aux_ack_o) then
    wait until (req_i = aux_ack_o);
  end if;
  -- Release #1: ack_o, after 1 ns (latch setup constraint)
  wait for 1 ns;
  hold_ack_o <= '0';
  -- Latches to hold signals
  latch_ack_o: process(reset_i, hold_ack_o, aux_ack_o)
  begin
    if ((reset_i = '1') OR (hold_ack_o = '0')) then
      ack_o <= aux_ack_o;
    end if;
  end process;
  latch_req_header_o: process(reset_i, hold_req_header_o, aux_req_header_o)
  begin
    if ((reset_i = '1') OR (hold_req_header_o = '0')) then
      req_header_o <= aux_req_header_o;
    end if;
  end process;
  latch_req_data_o: process(reset_i, hold_req_data_o, aux_req_data_o)
  begin
    if ((reset_i = '1') OR (hold_req_data_o = '0')) then
      req_data_o <= aux_req_data_o;
    end if;
  end process;
end loop;
end process;
end sim;

Figure A.18 - VHDL source code for simulation wrapper of Input Buffer Control (part 2 of 2).
Figure A.19 - VHDL source code for Input Buffer module.
A.4 Routing Control

```vhdl
library ieee;
use ieee.std_logic_1164.all;
use ieee.std_logic_unsigned.all;
use work.hermes_bd_package.all;

entity routing_ctrl is
  generic
  ROUTING_ALGORITHM : string := "XY"; -- Choice of routing algorithm.
  ROUTER_ADDRESS : std_logic_vector((FLIT_SIZE/2)-1 downto 0) := "x'11'"; -- Address of the router.
  THIS_PORT : integer := LOCAL; -- Defines which port is this interface representing (East, West, North, South, Local).
  COMM_PORTS : std_logic_vector(PORTS_NUMBER-1 downto 0) := "x'1111'" -- Defines to which ports this interface connects (index THIS_PORT is ignored). Ports with '0' are disconnected.

  port
  reset_i : in std_logic; -- Active-high reset signal.
  req_route_i : in std_logic; -- Signal requesting a route towards the target_address_i router.
  ack_route_o : out std_logic; -- Acknowledge relative to req_route_i. Indicates that the output towards the destination is ready to receive data.
  target_address_i : in std_logic_vector((FLIT_SIZE/2)-1 downto 0); -- Address of the destination router.
  req_outport_o : out std_logic_vector(PORTS_NUMBER-1 downto 0); -- Signal requesting the use of an output port.
  ack_outport_i : in std_logic_vector(PORTS_NUMBER-1 downto 0) -- Acknowledge signal relative to req_outport_. Indicates that the port use has been granted.

begin
end routing_ctrl;
```

architecture routing_ctrl of routing_ctrl is
  signal outport_en : std_logic_vector(PORTS_NUMBER-1 downto 0); -- Signal to enable a given port's handshake
  signal req_outport : std_logic_vector(PORTS_NUMBER-1 downto 0); -- Used to phase-match req_route_i with each
  signal ack_route : std_logic; -- Same as ack_route_o
  signal routing_en : std_logic; -- Signal to enable all handshakes (Request has been made, but hasn't received
  acknowledge yet)

begin
  routing_en <= req_route_i XOR ack_route; -- Enables the computation of the output port; High when handshake hasn't
  been completed yet.
  ack_route <= xor_all_bits_considering_comm_ports(ack_outport_i, COMM_PORTS, THIS_PORT); -- Generates ack_route with
  the correct phase

  -- Routing algorithm
  xy_routing_i : if (ROUTING_ALGORITHM = "XY") generate
    routing_unit_i : entity work.routing_unit(XY)
    generic map
      ROUTER_ADDRESS => ROUTER_ADDRESS,
      THIS_PORT => THIS_PORT,
      COMM_PORTS => COMM_PORTS
    port map
      en_i => routing_en,
      target_address_i => target_address_i,
      req_outport_o => outport_en
    );
  end generate xy_routing_i;

  -- Latches and phase-matches for req_outport_o
  req_outports_ctrl: for i in # to PORTS_NUMBER-1 generate
```
VHDL source code for Routing Control module (part 1 of 2).
Figure A.22 - VHDL source code for Routing Control module (part 2 of 2).

Figure A.23 - VHDL source code for simulation wrapper of Routing Control (part 1 of 2).
architecture sin of routing_ctrl_wrapper is
  -- Control signals
  signal hold_ack_route_o : std_logic;
  signal hold_req_outport_o : std_logic;
  -- Aux signals
  signal aux_ack_route_o : std_logic;
  signal aux_req_outport_o : std_logic_vector(PORTS_NUMBER downto 0);
begin
  -- Instance of routing_ctrl
  routing_ctrl_i : entity work.routing_ctrl
  generic map(
    ROUTING_ALGORITHM => ROUTING_ALGORITHM,
    ROUTER_ADDRESS => ROUTER_ADDRESS,
    THIS_PORT => THIS_PORT,
    COMM_PORTS => COMM_PORTS
  )
  port map (
    reset_i => reset_i,
    req_route_i => req_route_i,
    ack_route_o => aux_ack_route_o,
    target_address_i => target_address_i,
    req_outport_o => aux_req_outport_o,
    ack_outport_i => aux_ack_route_o
  );
  -- Control Logic
  control: process
  begin
    -- Initializing control signals
    hold_ack_route_o <= '0';
    hold_req_outport_o <= '0';
    wait until reset_i = '0';
    loop
      -- Hold all signals
      hold_ack_route_o <= '1';
      hold_req_outport_o <= '1';
      -- Wait for req_route_i request
      if (req_route_i = aux_ack_route_o) then
        wait until (req_route_i /= aux_ack_route_o);
      end if;
      -- Release #1: req_outport_o, after 4 ns (routing logic delay)
      wait for 4 ns;
      hold_req_outport_o <= '0';
      -- Wait for ack_outport_i acknowledge.
      if (req_route_i /= aux_ack_route_o) then
        wait until (req_route_i = aux_ack_route_o);
      end if;
      -- Release #2: ack_route_o, after 1 ns (phase-matching delay)
      wait for 1 ns;
      hold_ack_route_o <= '0';
      -- Wait 1 ns before holding it again
      wait for 1 ns;
      end loop;
    end process;
    -- Latches to hold signals
    latch_ack_route_o: process(reset_i, hold_ack_route_o, aux_ack_route_o)
    begin
      if ((reset_i = '1') OR (hold_ack_route_o = '0')) then
        ack_route_o <= aux_ack_route_o;
      end if;
    end process;
    latch_req_outport_o: process(reset_i, hold_req_outport_o, aux_req_outport_o)
    begin
      if ((reset_i = '1') OR (hold_req_outport_o = '0')) then
        req_outport_o <= aux_req_outport_o;
      end if;
    end process;
  end process;
sim;
A.5 Input Interface

```vhdl
library ieee;
use ieee.std_logic_1164.all;
use ieee.std_logic_misc.all;
use work.hermes_bd_package.all;

entity input_interface is
  generic
  ROUTING_ALGORITHM : string := "XY"; -- Choice of routing algorithm.
  BUFFER_DEPTH : integer := 8; -- Defines the depth of the buffer.
  ROUTER_ADDRESS : std_logic_vector((FLIT_SIZE/2)-1 downto 0) := x"11"; -- Address of the router.
  THIS_PORT : integer := LOCAL; -- Defines which port is this interface representing (East, West, North, South, Local).
  COMM_PORTS : std_logic_vector(PORTS_NUMBER-1 downto 0) := "11111" -- Defines to which ports this interface connects (index THIS_PORT is ignored). Ports with '0' are disconnected.

  port( 
    reset_i : in std_logic; -- Active-high reset signal.
    req_i : in std_logic; -- Request signal indicating that new data is available on data_i.
    ack_o : out std_logic; -- Acknowledge signal relative to req_i. Indicates that the data has been stored.
    data_i : in std_logic_vector(FLIT_SIZE-1 downto 0); -- Data input.
    req_outport_o : out std_logic_vector(PORTS_NUMBER-1 downto 0); -- Signal requesting use of an output port.
    ack_outport_i : in std_logic_vector(PORTS_NUMBER-1 downto 0); -- Acknowledge signal relative to req_outport_o. Indicates that the port use has been granted, and the flit on data_o has been consumed.
    data_o : out std_logic_vector(FLIT_SIZE-1 downto 0); -- Signal requesting a route towards the address on data_i.
    req_outport : out std_logic; -- Active-high signal indicating that the flit on data_o is the last flit of the current packet.
  );
end input_interface;

architecture input_interface of input_interface is
  signal data : std_logic_vector(FLIT_SIZE-1 downto 0); -- Same as data_o
  signal req_header : std_logic; -- Signal requesting a route towards the address on data_i.
  signal ack_header : std_logic; -- Acknowledge from req_header. The routing is completed, and data_o has been stored
  signal ack_data : std_logic; -- Acknowledge from req_data (same as all ack_data_i XORed)

begin
  data_o <= data;
  ack_data <= xor_all_bits_considering_comm_ports(ack_data_i, COMM_PORTS, THIS_PORT);

  -- Buffer Instantiation
  input_buffer_i : entity work.input_buffer
  generic map( 
    BUFFER_DEPTH => BUFFER_DEPTH 
  )
  port map(
    reset_i => reset_i,
    req_i => req_i,
    ack_o => ack_o,
    data_i => data_i,
  );
end architecture;
```

Figure A.25 - VHDL source code for Input Interface module (part 1 of 2).
Figure A.26 - VHDL source code for Input Interface module (part 2 of 2).

A.6 Outport Control

```vhdl
library ieee;
use ieee.std_logic_1164.all;
use work.hermes_bd_package.all;
entity outport_ctrl is
  port(
    reset_i    : in std_logic; -- Active-high reset signal.
    req_outport_i : in std_logic; -- Signal requesting use of the outport.
    ack_outport_o : out std_logic; -- Acknowledge relative to req_outport_i. Indicates that port use has been granted and the data at data out port was received.
    arbiter_request_o : out std_logic; -- Level-encoded grant relative to arbiter_request_o. This signal is kept high as long as the arbiter_request_o is asserted.
    req_data_i : in std_logic; -- Signal indicating that a new data flit is ready to be sent.
    req_data_o : in std_logic; -- Signal indicating that the current flit is the last flit of the packet.
    ack_i     : in std_logic; -- Acknowledge relative to req_o. Indicates that the flit has been sent.
  );
architecture outport_ctrl of outport_ctrl is
signal outport_request : std_logic; -- Level-encoded signal that indicates when a request at [req/ack]_outport_i is detected
signal ack_outport : std_logic; -- Same as ack_outport_o
signal req_data : std_logic; -- Phase-matched version of req_data_i
signal req_data_phase : std_logic; -- Signal to indicate when the phase of req_data_i needs to be inverted
signal ack_data : std_logic; -- Same as ack_data_o
end outport_ctrl;
-- Part of a race-condition avoidance mechanism
if (reset_i = '1') then
  req_header_o <= req_header,
  ack_header_i <= ack_header,
  req_data_o <= req_data_o,
  ack_data_i <= ack_data,
  data_o <= data,
  last_flit_o <= last_flit_o
else
end if;
-- Routing control
routing_ctrl_i entity work.routing_ctrl_wrapper
generic map( ROUTING_ALGORITHM => ROUTING_ALGORITHM,
ROUTER_ADDRESS => ROUTER_ADDRESS,
THIS_PORT => THIS_PORT,
COMM_PORTS => COMM_PORTS )
port map( reset_i => reset_i,
req_route_i => req_route_i,
ack_route_o => ack_route_o,
target_address_i => data((FLIT_SIZE/2)-1 downto 0),
req_outport_o => req_outport_o,
ack_outport_i => ack_outport_i
);
end input_interface;
```

Figure A.27 - VHDL source code for Outport Control module (part 1 of 3).
VHDL source code for Outport Control module (part 2 of 3).
A.7 Arbiter

```vhdl
library ieee;
use ieee.std_logic_1164.all;
use work.hermes_bd_package.all;

entity arbiter is
  generic(
    SIZE : integer := 4  -- Number of requests supported (max. 4),
  );
  port(
    request_i : in std_logic_vector(SIZE-1 downto 0);  -- Vector of level-encoded requests,
    grant_o : out std_logic_vector(SIZE-1 downto 0)  -- Vector of grants, based of the requests made in request_i,
  );
end arbiter;

architecture arbiter of arbiter is
  -- Declaration of MUTEX implemented in ascend_behavioral.v and ascend_behavioral_udps.v
  component MUTEX2
    port (  
      RA : in std_logic;  
      RB : in std_logic;  
      AB : out std_logic  
    );
  end component;
  -- Declaration of C-Element implemented in ascend_behavioral.v and ascend_behavioral_udps.v
  component GPSVT_BP_CSVM2X1
    port (  
      Q : out std_logic;  
      A : in std_logic;  
      B : in std_logic  
    );
  end component;
begin
  -- Arbiter with 1 input (Just to make it very generic)
  arb1: if (SIZE = 1) generate
    grant_o <= request_i;
  end generate arb1;
  
  -- Arbiter with 2 inputs
  arb2: if (SIZE = 2) generate
    -- Instantiation of Mutex
    mutex_1: MUTEX2
      port map (  
        RA => request_i(0),  
        RB => request_i(1),  
        AB => grant_o(0)  
      );
  end generate arb2;
  
  -- Arbiter with 3 inputs
  arb3: if (SIZE = 3) generate
    signal grant    : std_logic_vector(1 downto 0);  
    signal request : std_logic_vector(1 downto 0);  
    signal grant_mutex : std_logic_vector(3 downto 0);  
    signal mutex_3_ra : std_logic;
    begin
      grant_o(1 downto 0) <= grant(1 downto 0);  
      grant_o(2) <= grant_mutex(3);
      request(0) <= request_i(0) AND NOT(grant(1));  
      request(1) <= request_i(1) AND NOT(grant(0));
      mutex_3_ra <= request(0) OR request(1);
      begin
        grant_o(1 downto 0) <= grant(1 downto 0);  
        grant_o(2) <= grant_mutex(3);
        request(0) <= request_i(0) AND NOT(grant(1));  
        request(1) <= request_i(1) AND NOT(grant(0));
        mutex_3_ra <= request(0) OR request(1);
      end;
    end;
  end generate arb3;

begin
  -- Arbiter with 4 inputs
  arb4: if (SIZE = 4) generate
    signal grant    : std_logic_vector(3 downto 0);  
    signal request : std_logic_vector(3 downto 0);  
    signal grant_mutex : std_logic_vector(4 downto 0);  
    signal mutex_4_ra : std_logic;
    begin
      grant_o(1 downto 0) <= grant(1 downto 0);  
      grant_o(2) <= grant_mutex(3);
      request(0) <= request_i(0) AND NOT(grant(1));  
      request(1) <= request_i(1) AND NOT(grant(0));
      mutex_4_ra <= request(0) OR request(1);
    end;
  end generate arb4;

end arbiter;
```

Figure A.29 - VHDL source code for Arbiter module (part 1 of 3).
VHDL source code for Arbiter module (part 2 of 3).
A.8 Output Interface

```vhdl
library ieee;
use ieee.std_logic_1164.all;
use work.hermes_bd_package.all;

entity output_interface is
    generic(
        THIS_PORT : integer := LOCAL; --! Defines which port is this interface representing (East, West, North, South, Local).
        COMM_PORTS : std_logic_vector(PORTS_NUMBER-1 downto 0) := "11111" --! Defines to which ports this interface connects (index THIS_PORT is ignored). Ports with '0' are disconnected.
    );
    port(
        reset_i : in std_logic; --! Active-high reset signal.
        req_outport_i : in std_logic_vector(PORTS_NUMBER-1 downto 0); --! Signal requesting use of the output port.
        ack_outport_o : out std_logic_vector(PORTS_NUMBER-1 downto 0); --! Auxiliary signal to connect arbiter_grant to aux. that the port use has been granted, and the flit on data_i has been consumed.
        req_data_i : in std_logic_vector(PORTS_NUMBER-1 downto 0); --! Signal requesting use of the output port.
        ack_data_o : out std_logic_vector(PORTS_NUMBER-1 downto 0); --! Acknowledge signal relative to req_data_i. Indicates that the data has been stored.
        data_i : in flit_size_array(PORTS_NUMBER-1 downto 0); --! Data input.
        last_flit_i : in std_logic; --! Active-high signal indicating that the last flit of the current packet.
        req_o : in std_logic; --! Request signal indicating that new data is available on data_o.
        ack_i : in std_logic; --! Acknowledge signal relative to req_o. Indicates that the payload was consumed.
        data_o : out std_logic_vector(FLIT_SIZE-1 downto 0) --! Data output.
    );
end output_interface;
```

Figure A.31 - VHDL source code for Arbiter module (part 3 of 3).

```vhdl
library ieee;
use ieee.std_logic_1164.all;
use ieee.std_logic_misc.all;
use work.hermes_bd_package.all;

entity output_interface is
    generic(
        THIS_PORT : integer := LOCAL; --! Defines which port is this interface representing (East, West, North, South, Local).
        COMM_PORTS : std_logic_vector(PORTS_NUMBER-1 downto 0) := "11111" --! Defines to which ports this interface connects (index THIS_PORT is ignored). Ports with '0' are disconnected.
    );
    port(
        reset_i : in std_logic; --! Active-high reset signal.
        req_outport_i : in std_logic_vector(PORTS_NUMBER-1 downto 0); --! Signal requesting use of the output port.
        ack_outport_o : out std_logic_vector(PORTS_NUMBER-1 downto 0); --! Auxiliary signal to connect arbiter_grant to aux. that the port use has been granted, and the flit on data_i has been consumed.
        req_data_i : in std_logic_vector(PORTS_NUMBER-1 downto 0); --! Signal requesting use of the output port.
        ack_data_o : out std_logic_vector(PORTS_NUMBER-1 downto 0); --! Acknowledge signal relative to req_data_i. Indicates that the data has been stored.
        data_i : in flit_size_array(PORTS_NUMBER-1 downto 0); --! Data input.
        last_flit_i : in std_logic; --! Active-high signal indicating that the last flit of the current packet.
        req_o : in std_logic; --! Request signal indicating that new data is available on data_o.
        ack_i : in std_logic; --! Acknowledge signal relative to req_o. Indicates that the payload was consumed.
        data_o : out std_logic_vector(FLIT_SIZE-1 downto 0) --! Data output.
    );
end output_interface;
```

Figure A.32 - VHDL source code for Output Interface module (part 1 of 2).
Figure A.33 - VHDL source code for Output Interface module (part 2 of 2).
A.9 BaT–Hermes

---

```vhdl
library ieee;
use ieee.std_logic_1164.all;
use work.hermes_bd_package.all;

entity hermes_bd_port is
  generic(
    BUFFER_DEPTH : integer := 8; -- Defines the depth of the buffer.
    ROUTER_ADDRESS : std_logic_vector((FLIT_SIZE/2)-1 downto 0) := x"11": -- Address of the router.
    THIS_PORT : integer := LOCAL; -- Defines which port is this interface representing (East, West, North, South, Local).
    COMM_PORTS : std_logic_vector(PORTS_NUMBER-1 downto 0) := "11111": -- Defines to which ports this interface connects (index THIS_PORT is ignored). Ports with '0' are left disconnected.
  );
  port(
    reset_i : in std_logic; -- Active-high reset signal.
    import_reg_i : in std_logic; -- Request signal indicating that new data is available on import_data_i.
    import_ack_o : out std_logic; -- Acknowledge signal relative to req_i. Indicates that the data on import_data_i has been stored.
    import_data_i : in std_logic_vector(FLIT_SIZE-1 downto 0); -- Data input of the input interface.
    import_ack_outport_o : out std_logic_vector(PORTS_NUMBER-1 downto 0); -- Signal requesting use of an output port.
    import_reg_outport_i : in std_logic_vector(PORTS_NUMBER-1 downto 0); -- Data input of the output interface.
    import_data_o : out std_logic_vector(FLIT_SIZE-1 downto 0); -- Indicates that the port use has been granted, and the flit on input_data_o has been consumed.
    import_last_flit_o : out std_logic; -- Signal indicating that new payload is available on input_data_o.
    import_ack_data_i : in std_logic; -- Acknowledge relative to req_data_o. Indicates that the payload was consumed.
    import_data_o : out std_logic_vector(FLIT_SIZE-1 downto 0); -- Indicates that the flit on data_o is the last flit of the current packet.
    import_last_flit_i : in std_logic_vector(PORTS_NUMBER-1 downto 0); -- Signal requesting use of the output port.
    import_ack_outport_o : out std_logic_vector(PORTS_NUMBER-1 downto 0); -- Signal indicating that new payload is available on output_data_o.
    input_data_o : out std_logic_vector(PORTS_NUMBER-1 downto 0); -- Request signal indicating that new

Figure A.34 - VHDL source code for router port (part 1 of 2).
payload is available on outport_data_o.
outport_ack_data_o : out std_logic_vector(PORTS_NUMBER-1 downto 0); /* Acknowledge relative to req_data_o.
  Indicates that the payload was consumed.
outport_last_flit_o : in std_logic_vector(PORTS_NUMBER-1 downto 0); /* Output interface data input.
  Flit on outport_o is the last flit of the current packet.
outport_req_o     : out std_logic; /* Request signal indicating that new data is available on outport_data_o.
outport_ack_o     : in std_logic; /*! Acknowledge signal relative to req_o. Indicates that the data has been
  received.
outport_data_o    : out std_logic_vector(FLIT_SIZE-1 downto 0); /*! Output interface data output.
}
end hermes_bd_port;
architecture hermes_bd_port of hermes_bd_port is
begin
  -- Input Interface Instance
  input_interface_i: entity work.input_interface
  generic map (  
    ROUTING_ALGORITHM => ROUTING_ALGORITHM,
    BUFFER_DEPTH => BUFFER_DEPTH,
    ROUTER_ADDRESS => ROUTER_ADDRESS,
    THIS_PORT => THIS_PORT,
    COMM_PORTS => COMM_PORTS )  
  port map (  
    reset_i   => reset_i,
    req_i     => import_req_i,
    ack_o     => import_ack_o,
    data_i    => import_data_i,
    reqoutport_o => import_reqoutport_o,
    ackoutport_i => import_ackoutport_i,
    reqdata_o => import_reqdata_o,
    ackdata_i => import_ackdata_i,
    data_o   => import_data_o,
    last_flit_o => import_last_flit_o  
  );
  -- Output Interface Instance
  output_interface_i: entity work.output_interface
  generic map (  
    THIS_PORT => THIS_PORT,
    COMM_PORTS => COMM_PORTS )  
  port map (  
    reset_i   => reset_i,
    reqoutport_i => output_reqoutport_i,
    ackoutport_o => output_ackoutport_o,
    reqdata_i => output_reqdata_i,
    ackdata_o => output_ackdata_o,
    data_i    => output_data_i,
    last_flit_i => output_last_flit_i,
    req_o     => output_req_o,
    ack_i     => output_ack_i,
    data_o   => output_data_o  
  );
end hermes_bd_port;

Figure A.35 - VHDL source code for router port (part 2 of 2).

/* File: hermes_bd_router.vhd */ /* Brief: A bundled-data transition-signaling handshake-based parameterizable NoC router. */ /* Author: Matheus Gibiluka, matheus.gibiluka@acad.pucrs.br */ /* @date 2013-10-14 */

-- Dependencies:  
  => hermes_bd_package.vhd

-- Generics for configuration:  
  ROUTING_ALGORITHM: Choice of routing algorithm.  
  BUFFER_DEPTH: Defines the depth of each input buffer,  
  ROUTER_ADDRESS: Address of the router.  
  PORTS: Defines which ports are implemented in the router. Ports with '0' are left unconnected.

-- Interface description:  
  reset_i: in std_logic;  
  req_i: in std_logic;  
  req_o: out std_logic;  
  reqdata_i: in std_logic_vector(PORTS_NUMBER-1 downto 0);  
  reqdata_o: out std_logic_vector(PORTS_NUMBER-1 downto 0);  
  ack_i: in std_logic;  
  ack_o: out std_logic;  
  ackdata_i: in std_logic_vector(PORTS_NUMBER-1 downto 0);  
  ackdata_o: out std_logic_vector(PORTS_NUMBER-1 downto 0);  
  data_i: in std_logic_vector(PORTS_NUMBER-1 downto 0);  
  data_o: out std_logic_vector(PORTS_NUMBER-1 downto 0);

Figure A.36 - VHDL source code for BaT-Hermes router (part 1 of 2).
entity hermes_bd_router is
  generic(
    ROUTING_ALGORITHM : string := "XY"; -- Choice of routing algorithm.
    BUFFERDEPTH : integer := 8; -- Defines the depth of the buffer.
    ROUTERADDRESS : std_logic_vector((FLIT_SIZE/2)-1 downto 0) := x"1"; -- Address of the router.
    PORTS : std_logic_vector(PORTS_NUMBER-1 downto 0) := x"111"; -- Defines which ports are implemented.
  );
  port(
    reset_i : in std_logic; -- Active-high reset signal.
    req_i : in std_logic_vector(PORTS_NUMBER-1 downto 0); -- Request signal indicating that new data is available.
    ack_o : out std_logic_vector(PORTS_NUMBER-1 downto 0); -- Acknowledge signal relative to req_i. Indicates that
    the data has been stored.
    data_i : in flit_size_array(PORTS_NUMBER-1 downto 0); -- Data input for the input ports.
    req_outport_inv : ports_array_t; -- Generate ports
    port_gen := use work.hermes_bd_port generic map:
    ROUTING_ALGORITHM => ROUTING_ALGORITHM,
    BUFFERDEPTH => BUFFERDEPTH,
    ROUTERADDRESS => ROUTERADDRESS,
    THIS_PORT => 1,
    COMM_PORTS => PORTS
  );
  begin
    used_port_gen;
    end hermes_bd_router;
architecture hermes_bd_router of hermes_bd_router is
type ports_array_t is array(0 to PORTS_NUMBER-1) of std_logic_vector(PORTS_NUMBER-1 downto 0);

  signal req_outport : ports_array_t; -- Signal to hold each import_req_outport_o. Use:
  req_outport(map<port/interfaces>)(output_port);-- Signal to hold each import_req_outport_o. Use:
  reset_i : std_logic_vector(PORTS_NUMBER-1 downto 0);
  ack_outport : std_logic_vector(PORTS_NUMBER-1 downto 0);
  signal data_o : std_logic_vector(PORTS_NUMBER-1 downto 0); -- Output interface for the output ports.
  std_logic_vector(PORTS_NUMBER-1 downto 0) := x"111";
  last_flit : std_logic_vector(PORTS_NUMBER-1 downto 0); -- Signal to hold each import_last_flit_o Use:
  last_flit(map<port/interfaces>);
  end architecture;

library ieee;
use ieee.std_logic_1164.all;
use work.hermes_bd_package.all;

Figure A.37 - VHDL source code for BaT-Hermes router (part 1 of 2).
B. Relative Timing Constraints

This appendix contains graphical representations of the relative timing constraints found in BaT-Hermes. Wires are color-coded green and red, representing, respectively, base and enforced paths. Delay lines are inserted in the latter. Constraints between blocks are accounted in the input and output signals of each circuit – for example, the timing restrictions between the FIFO and the Input Buffer Control are accounted in the req_rd_o, ack_rd_i, data_o, req_i, ack_o and data_i signals of these modules.

B.1 FIFO

![Figure B.1 – FIFO: constraint ensuring that phase_select_i arrives before req_i in the write control circuit and phase_select_i arrives before ack_i in the read control circuit.](image)

![Figure B.2 – FIFO: constraints ensuring that, for each buffer position, data_i arrives before req_wr_i and data_o arrives before req_rd_o.](image)
Figure B.3 – FIFO: constraint ensuring that, for each buffer position, \textit{reg\_en}\textsubscript{i} arrives at the register before \textit{full\_o}\textsubscript{i} arrives at the read control circuit.

Figure B.4 – FIFO: constraints ensuring that, for each read and write control circuit, \textit{phase\_select\_i}\textsubscript{o} arrives before \textit{en\_i}.\textsubscript{o}
Figure B.5 – FIFO: constraints ensuring that, for each read control circuit, \( en_i \) is disabled before a new \( full_o \) request can be made.

Figure B.6 – FIFO: constraints ensuring that, for each register, the data is stored before an \( ack_{wr.o} \) is issued.
### B.2 Input Buffer Control

![Input Buffer Control Diagram](image)

**Figure B.7** – Input Buffer Control: constraint ensuring that `data_i` arrives before `req_i`.

![Input Buffer Control Diagram](image)

**Figure B.8** – Input Buffer Control: constraints ensuring that `data_o` and `last_flit_o` arrives before `req_header_o` or `req_data_o`.

![Input Buffer Control Diagram](image)

**Figure B.9** – Input Buffer Control: constraint ensuring that `data_o` is properly stored in `LT2` and `FF1` before `ack_o` is issued.
Figure B.10 – Input Buffer Control: constraint ensuring that, when a request is received, the mux remains stable until FF1 stores the data currently at its input.

Figure B.11 – Input Buffer Control: constraint ensuring that, when a request is received, the data at the input pin of FF3 only changes after it has been captured.

Figure B.12 – Input Buffer Control: constraint ensuring that, when a request is received, FF5 only captures the data at its input pin after last_flit_lvl has propagated.
Figure B.13 – Input Buffer Control: constraint ensuring that LT3 is enabled only after all its input signals have propagated.

Figure B.14 – Input Buffer Control: constraint ensuring that LT4 is enabled only after all its input signals have propagated.

Figure B.15 – Input Buffer Control: constraint ensuring that all FSM data signals propagate during a handshake cycle.
Figure B.16 – Input Buffer Control: constraints ensuring the minimum period of the signal used to clock FF5 is large enough to fulfill the flip-flop’s timing requirements.

Figure B.17 – Input Buffer Control: constraints ensuring that the control signal of the mux only changes after the signals at the clock input of FF5 have propagated.
B.3 Routing Control

Figure B.18 – Routing Control: constraints ensuring that, for each latch connected to a *req_outport_o* signal, the signals at the input propagate before the enable signal.

Figure B.19 – Routing Control: constraints ensuring that, for each latch connected to a *req_outport_o* signal, the enable signal disabling the latch propagates before *ack_route_o*.
**B.4 Outport Control**

Figure B.20 – Outport Control: constraint ensuring the last flit detector is activated before the connected Input Interface receives the `ack_data_o` acknowledge.

Figure B.21 – Outport Control: constraint ensuring the programmable phase matcher and the last flit detector are initialized before an `ack_outport_o` acknowledge is issued.
Figure B.22 – Outport Control: constraint ensuring that a new \textit{req\_outport\_i} request is received after LT5 has been disabled.

Figure B.23 – Outport Control: constraint ensuring that the last flit detector only tests a valid \textit{last\_flit\_i} signal.
Figure B.24 – Outport Control: constraints ensuring that the signals to be stored in FF1 and FF3 arrive before the arbiter grant is given.
Figure B.25 – Output Interface: constraint ensuring that data_o signal is stable before a req_o request is issued.