An Asynchronous Low-Power Innovative Network-on-Chip including Design-for-Test capabilities

Yvain Thonnart[1], Xuan-Tu Tran[2], Pascal Vivet[1], Edith Beigne[1], Fabien Clermidy[1], Jean Durupt[1]

[1] *CEA-LETI, MINATEC – 17 rue des Martyrs, 38 054 Grenoble, France, {firstname.lastname}@cea.fr*
[2] *VNU-Coltech / SIS laboratory – 144 Xuan Thuy Road, 10 000 Hanoi, Vietnam, tutx@vnu.edu.vn*

Abstract

The demands of scalable, low latency and power efficient System-On-Chip interconnect cannot be satisfied only by point-to-point or shared-bus interconnects. By providing more bandwidth at reasonable power consumption, new communication infrastructures like NoCs seem promising, but are still limited by implementation issues. We present in this paper an Asynchronous Network-on-Chip architecture with two main innovations. Firstly, an automatic power regulation scheme is proposed to dynamically save leakage and dynamic power consumption. Secondly, due to the current lack of testing methodology for asynchronous logic, we propose a novel DfT solution to allow acceptance of the asynchronous NoC. The proposed architecture has been fully implemented in a STMicroelectronics CMOS 65nm technology, integrated in a complex test-chip and fabricated.

1. Introduction

With many hundreds of million transistors available on a single chip, the era of Manycores System-on-Chip (MPSoC) has become a reality. The available synchronous bus architectures cannot meet the increasing demands of communication between the processing elements, because the long-wire loads and resistances result in slow signal propagation. On the other hand, the advantages of Network-on-Chip (NoC) [1] are numerous: high scalability and versatility, high throughput with good power efficiency. A NoC can be implemented using synchronous logic but designers still face the issue of generating a global clock tree over the chip. Multi synchronous design can help but with extra latency due to re-synchronization cost within each router along the path.

The NoC distributed communication architecture is perfectly adapted to the Globally Asynchronous Locally Synchronous (GALS) paradigm [2] where ÏP units are implemented with standard synchronous design methodologies on an independent timing domain, while the NoC itself is implemented in asynchronous logic. The GALS paradigm also offers a natural way to implement Dynamic Voltage and Frequency Scaling (DVFS) thanks to fully decoupled synchronous islands. Several asynchronous NoC architectures have been presented recently [2]. We have proposed recently ANoC [3], an Asynchronous Network-on-Chip architecture, providing Quality-of-Service and fully implemented in Quasi-Delay-Insensitive (QDI) logic [7]. In such architectures, two main issues need to be solved: leakage power and testability.

Regarding power reduction, asynchronous design techniques have shown interesting dynamic power savings, due to their un-clocked nature [6]. Thanks to their local handshaking, asynchronous circuits are automatically in standby state when inactive. The asynchronous logic scheme offers thus the equivalent of both RTL clock gating and architectural clock gating, but without the need of any additional software. Nevertheless, asynchronous circuits, like their synchronous counterparts, also need to reduce their leakage power. In this paper, we present an innovative solution to detect incoming asynchronous activity, which associated to an automatic power regulation, efficiently reduces the supply voltage and thus the leakage power [8].

Regarding testability of a GALS NoC architecture, two issues need to be solved: the test of the (synchronous) IP units and the test of the (asynchronous) NoC itself. For the test of IPs within a NoC architecture, a test wrapper compatible with the IEEE 1500 standard [4] can be used for each embedded IP while the network architecture can be used as a high bandwidth Test Access Mechanism (TAM). Test data of each IP are encapsulated into packets that are transported in the network using the network protocol. There is thus no extra TAM hardware cost and the scalability of the test is improved [5]. However, the GALS NoC architecture should be searched for defects first. It is known that asynchronous logic is difficult to test with standard design techniques. In this paper, we present the design and implementation of an asynchronous DfT architecture, adapted to the ANoC asynchronous router [9].

The paper is organized as follows: in section 2 is presented the proposed ANoC architecture, in section 3 is presented the automatic power regulation scheme adapted to the NoC router to reduce its leakage power, in section 4 is presented the DfT architecture adapted to the NoC router. Finally we present the ALPIN test-chip which has been designed and fabricated in a STMicroelectronics CMOS 65*nm* technology and the associated results.

2. ANoC Architecture

The ANoC framework [3] allows an easy assembly of a complex GALS SoC made of independent synchronous IP units, relying on asynchronous routers and links assembled in a mesh topology as seen on Figure 1.
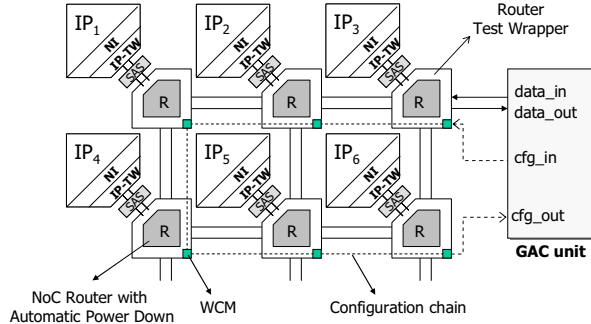

Figure 1: ANoC architecture template

The asynchronous network routers are the basic elements; they have 5 bi-directional ports that connect to four neighboring routers and the local synchronous IP via a network interface (NI) and an asynchronous ⇔ synchronous interface (SAS). Each NoC router includes an automatic power down mechanism in order to save leakage, as presented in detail in section 3.

Regarding test infrastructure, each synchronous IP can be tested using standard scan chains, using the IP test-wrapper (IP-TW) and the NoC as a TAM. For testing the NoC itself, each NoC router is encapsulated with an ad-hoc asynchronous test wrapper, individually controlled through a configuration chain (*dashed-lines*), and a centralized Generator-Analyzer-Controller (GAC) unit performs test pattern generation and analysis. This is presented in detail in section 4.

The NoC protocol between routers is a flit-level *"send/accept"* protocol with two virtual channels in order to provide Quality-of-Service. The NoC channels are 34-bit bidirectional channels, with 32 bits data and 2 bits to encode *Begin-of-Packet* (BoP) and *End-of-Packet* (EoP). Using source routing technique, routing information is encoded in the header flit with a *path-to-target* field shifted in each router, as shown in Figure 2.
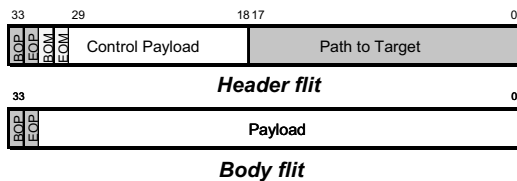

Figure 2: Data flit formats.

At low level, every two bits are implemented in asynchronous QDI logic using a 4-rail / 4-phase protocol in order to reduce the dynamic power.

3. ANoC Router with Automatic Power-Down

The main idea of the proposal is to benefit from the robustness and locality of asynchronous logic to automatically detect activity and use this information to power down and save leakage.

3.1. Asynchronous Activity Detection Scheme [8]

In order to control the leakage of an asynchronous unit (Figure 3), a voltage regulator is used in order to power down the asynchronous logic unit when in standby mode. Activity detection on the incoming and outgoing channels is performed using channel monitors. When no more input and output activity is detected, the voltage regulator powers down the asynchronous logic unit in standby mode for reducing the leakage power. When new incoming activity is detected, the voltage regulator powers up the asynchronous logic in normal mode, without any additional software control and at minimal latency cost.
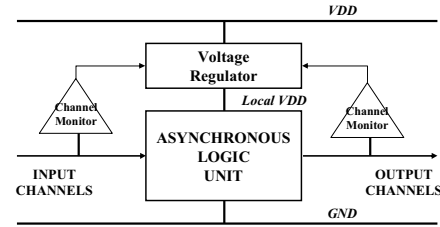

Figure 3: Asynchronous activity detection scheme

Due to their robustness to operating conditions, asynchronous circuits can be easily supplied at low voltage for power reduction. Nevertheless, many issues need to be addressed in this simple proposal. The proper protocol must be chosen in order to detect traffic with a fast and reliable detection. The second constraint is regarding the voltage regulator and the definition of the standby mode.

3.2. ANoC Router with Automatic Power-Down

Instead of doing asynchronous detection at handshake level (i.e. power up and power down a router between each individual NoC flit), it has been chosen to use a high level encoding to properly detect NoC traffic. We propose to detect activity at NoC message level: this is encoded within the packet header; by using 2 bits encoding *Begin of Message (BOM)* and *End of Message* (*EOM*) (see Figure 2). This encoding is performed in the Network Interface by software configuration (computed offline or online): only the required data packets are tagged with *BOM* and *EOM* flags.

The activity detection logic in a NoC router consists in counting the number of active messages traveling through it (Figure 4). When *BOM* rises, a counter is incremented; when *EOM* rises, the counter is decremented. When the counter is zero, no more message is active through that router: it can be switched down to save energy.
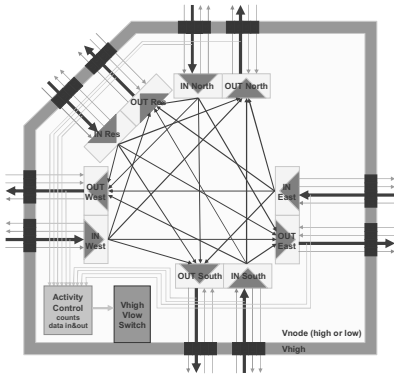
Figure 4: ANoC router with its activity detection

For standby mode, we need to trade-off between leakage reduction, power regulation logic and wake up time. In order to avoid complex power on reset and long wake-up time, we define a standby mode where the router is still functional (0.7V). The power regulation is implemented with a simple PMOS switch. The asynchronous detection logic is then powered with the same supply voltage: the router is yet slower but always functional. Lastly, to properly isolate the router when in power down mode, we implement level shifters on all NoC channel signals.

4. ANoC Router with Test Wrapper

In the proposed DfT architecture (Figure 1), the main idea is to encapsulate each ANoC router by an asynchronous test wrapper in order to improve its controllability and observability. The test wrappers are used: (a) to insert test vectors to the elements-under-test (routers and network links) and to get out the test results; (b) with network links, to establish high bandwidth asynchronous TAMs to transport test data.

All operations of the test wrapper are controlled by its WCM control module (Wrapper Control Module). To establish a test flow within the NoC, the 2-bit configuration chains of all wrappers' control modules are connected in sequence. Test flows are controlled by the GAC unit (*Generator-Analyzer-Controller*), which generates and analyzes the test patterns through the whole architecture.

4.1. Test Wrapper Architecture [9]

Like the ANOC router which is composed of 5 input and 5 output ports, the test wrapper (Figure 5) is composed of 5 input test cells (ITC), and 5 output test cells (OTC), plus the additional wrapper control module (WCM) to control all test cells. The ITCs and OTCs are alternatively interconnected to establish a boundary-scan path around the router. Because the test wrapper interface is similar to the network router interface, there is no change to ANoC external interface except the additional 2-bit test configuration chain.
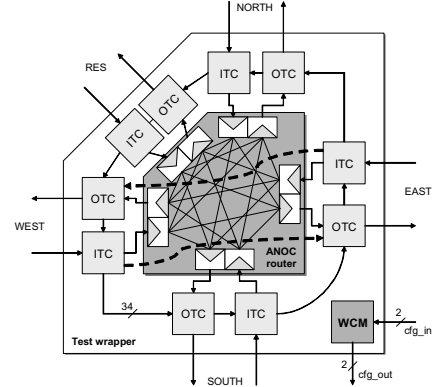

Figure 5: ANoC router with its test wrapper

In order to reduce test time and minimize test complexity, a bypass function is proposed, as classically used in DfT architectures. It allows reducing the length of test paths by giving short-circuits between the inputs and the outputs of test wrappers. In Figure 5 is illustrated a bidirectional bypass (*dashed bold lines*) between the EAST and WEST input/output ports. With bypass function, only the router-under-test is in test mode, which seems directly connected to GAC unit; other routers being inactive.

The proposed DfT architecture has been designed and implemented in QDI asynchronous logic, because it is compatible with the existing design (no dedicated test clock, no async. ⇔ sync. interface overhead). Most of all, QDI circuits are well adapted to test since they will stall in case of single stuck-at output faults: every transition is necessary to the good operation of the circuit [6].

4.2. Test Wrapper Protocol

For controlling the proposed DfT architecture, a test protocol has been defined. The Test Configuration Frames (TCFs) are sent serially by the GAC unit through the WCMs to control properly all ITCs and OTCs. Each TCF includes 25 symbols of 2 bits.
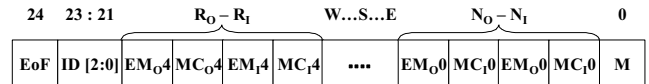
24	23 : 21	$R_O - R_I$			W...S...E	$N_O - N_I$				0
EoF	ID [2:0]	EM_O4	MC_O4	EM_I4	MC_I4	**••••**	EM_O0	MC_I0	EM_O0	MC_I0	M

Figure 6: Test Configuration Frame (TCF)

The TCF (Figure 6) contains an ID field to identify each router along the configuration chain, then it encodes the various configuration bits (Enable and Control values for each successive OTC and ITC test cells of the 5 input/output ports R/W/S/E/N). Finally, the TCF contains the test wrapper mode (Normal/Test/Bypass) and the end of frame information (EoF).

Using this protocol, it is then possible to test iteratively each ANoC link and each ANoC router within the NoC topology. By using successive data values for each NoC port, thanks to QDI logic stuck-at fault stall property, a minimal number of test vectors (320) is finally required.

5. ALPIN circuit results

The Automatic Power Regulation and the Design-for-Test architecture have been fully designed in a STMicro-electronics 65nm LP CMOS technology. These two mechanisms have been integrated in the "Asynchronous Low-Power Innovative NoC" test-chip, the so-called ALPIN circuit (Figure 7).
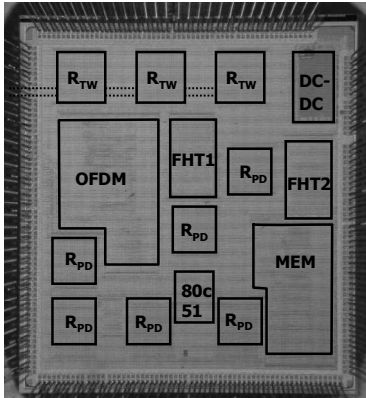

Figure 7 : ALPIN circuit picture [10]

The ALPIN circuit implements the ANoC architecture with 9 routers: 3 routers integrate the Test-Wrapper (noted R_{TW}) and 6 routers integrate the Automatic Power Down mechanism (noted R_{PD}). The ALPIN circuit also contains an innovative Dynamic and Voltage Scaling mechanism applied to each synchronous IP [10]. There are 5 IPs: an OFDM unit, a duplicated FHT unit, a MEMORY unit and finally a small 80C51 micro-controller. All these synchronous IPs are tested using an IP test-wrapper (IP-TW) as presented section 2. For router testing using the TCF protocol, the configuration chains of the 3 routers are connected in sequence, while the GAC unit is implemented off-chip (actually within an FPGA).

Table 1 : ALPIN results

ANoC Router	Original Version	Router with Power Down		Router with Test Wrapper
		Power On	Power Down	
Supply Voltage	1.2 V	1.2 V	0.6 - 0.8 V	1.2 V
Flit Cycle Time	1.8 ns	1.8 ns	7.2 ns	1.8 ns
Flit Throughput	550 Mflit/s	550 Mflit/s	140 Mflit/s	550 Mflit/s
Flit Latency	2.3 ns	2.5 ns	5.8 ns	2.64 ns
Leakage Cur. (Static Power)	200 µA (240 µW)	210 µA (250 µW)	80 µA (100µW)	264 µA (316 µW)
Energy	30 pJ/flit	30 pJ/flit	14 pJ/Flit	37.5 pJ/flit
Router Area	0.17 mm²	0.20 mm²		0.23 mm²

Compared with the original version of the ANoC router (Table 1), the area cost of automatic power down is 17% while the area cost of the test-wrapper is 32%. In terms of speed, throughput is always preserved (550 Mflit/s) with a minor increase in latency due to either level shifters or to the additional ITC/OTC cells.

For power reduction, the Automatic Power Down offers 60% leakage reduction in idle mode and up to 50%

dynamic power reduction if traffic bandwidth in power-down mode is sufficient. Switching between the two power modes is immediate and reliable without any real-time software.

For testability, the DfT architecture offers a complete coverage of each ANoC router (99.86% with single-stuck-at faults on both inputs and outputs). The DfT architecture and TCF protocol are efficient: 2*ns* per TCF symbol, which provide a test time per router of *32µs*. This gives a test-time of the ALPIN routers of about *0.3ms*.

6. Conclusion

We have presented in this paper two advanced innovations to improve an asynchronous NoC architecture: an automatic power regulation scheme in order to control automatically the leakage and dynamic power, and a complete DfT architecture of the NoC framework to allow fast and reliable test of its asynchronous logic.

7. References

[1] A. Jantsch and H. Tenhunen (Eds). *Networks on Chip*. Kluwer Academic Publisher, Feb. 2003.
[2] Krstic, M.; Grass, E.; Gurkaynak, F.K.; Vivet, P., "Globally Asynchronous, Locally Synchronous Circuits: Overview and Outlook", *IEEE Design & Test of Computers*, Volume 24, Issue 5, Sept.-Oct. 2007, pp. 430 – 441.
[3] E. Beigné, F. Clemidy, P. Vivet, A. Clouard, and M. Renaudin, "An Asynchronous NoC Architecture Providing Low Latency Service and its Multi-level Design Framework", In *Proceedings of ASYNC'05*, pp. 44–53, March 2005.
[4] IEEE 1500 working group. IEEE 1500 Standard for Embedded Core Test. *http://grouper.ieee.org/groups/1500*.
[5] A.M. Amory, K. Goossens, E.J. Marinissen, M. Lubaszewski, and F. Moraes, "Wrapper Design for the Reuse of Networks-on-Chip as Test Access Mechanism". In *Proceedings of the IEEE European Test Symposium (ETS)*, pp. 213–218, Southampton, UK, May 2006.
[6] J. Sparso, and S. Furber. *Principles of Asynchronous Circuit Design – A System Perspective*. Kluwer Academic Publishers, Dec. 2001.
[7] A.J. Martin, "Programming in VLSI: From Communicating Processes to Delay-Insensitive Circuits". In *Developments in Concurrency and Communication*, C.A.R. Hoare, editor, pp. 1–64, Addison-Wesley, 1990.
[8] Y. Thonnart, E. Beigné, A. Valentian, P. Vivet, "Power Reduction of Asynchronous Logic Circuits using Activity Detection", *IEEE Transactions on VLSI Systems*, July 2009, Vol. 17, n° 7, pp. 893-906.
[9] X.-T. Tran, J. Durupt, Y. Thonnart, V. Beroulle, C. Robach, "Design-for-Test Approach of an Asynchronous Network-on-Chip Architecture and its Associated Test Pattern Generation and Application", *IET Journal on Computers and Digital Techniques,* September 2009, Volume 3, Issue 5, pp. 487-500.
[10] E. Beigné, et all., "An Asynchronous Power Aware and Adaptive NoC based Circuit", *IEEE Journal Of Solid State Circuits,* April 2009, Vol.44, pp.1167-1177.