

An Asynchronous Power Aware and Adaptive NoC Based Circuit

Edith Beigné, Fabien Clermidy, H el ene Lhermet, Sylvain Miermont, Yvain Thonnart, Xuan-Tu Tran, Alexandre Valentian, Didier Varreau, Pascal Vivet, Xavier Popon, and Hugo Lebreton

Abstract—In complex embedded applications, optimisation and adaptation of both dynamic and leakage power have become an issue at SoC grain. A fully power-aware globally-asynchronous locally-synchronous network-on-chip (NoC) circuit is presented in this paper. Network-on-chip architecture combined with a globally-asynchronous locally-synchronous paradigm is a natural enabler for DVFS mechanisms. The circuit is arranged around an asynchronous network-on-chip providing scalable communication and a 17 Gb/s throughput while automatically reducing its power consumption by activity detection. Both dynamic and static power consumptions are globally reduced using adaptive design techniques applied locally for each synchronous NoC units. No fine control software is required during voltage and frequency scaling. Power control is localized and a minimal latency cost is observed.

Index Terms—Network-on-chip, GALS, leakage power, dynamic power, Vdd-hopping, super cut-off, DC-DC converters, DVFS.

I. INTRODUCTION

POWER dissipation has emerged as a major design constraint in today complex system-on-chip (SoC) architectures, limiting performance, battery life and reliability. A lot of dynamic and static power saving techniques exists [1], [2] and most of them are related to the V_{DD} voltage supply level control. It is well known that reducing the voltage affects power consumption quadratically and linearly affects frequency. However, if only one global voltage is controlled or scaled down, there is no real optimization at unit level and the whole system is constrained by the most critical functional unit to meet its timing constraints. As the number of IPs integrated in a SoC is increasing and the power consumption is ever increasing, a fine-grain power management is thus becoming essential. Using both a network-on-chip (NoC) distributed communication scheme and a globally-asynchronous locally-synchronous (GALS) scalable approach offers an easy local power management due to local synchronization and local clock generation [3]. Moreover, it can allow better energy savings since each functional unit can easily have its own independent clock and voltage. Hence, NoC architecture combined with a multi power-domains GALS system appear as natural enablers for distributed power management system as well as for local DVFS and local power gating.

Manuscript received August 25, 2008; revised November 10, 2008. Current version published March 25, 2009. This work was supported by the European Commission in the framework of FP6 with the IST-CLEAN project (Controlling Leakage Power in Nano CMOS SOC's), with the Medea+ LOMOSA 2A708 project (Low-power expertise for Mobile and multimedia system applications) and with the Medea+ NEVA 2A703 project (Network-on-chips designs driven by Video and distributed Application), and also supported in part by STMicroelectronics.

The authors are with CEA-LETI-MINATEC, 38054 Grenoble, France (e-mail: edith.beigne@cea.fr).

Digital Object Identifier 10.1109/JSSC.2009.2014206

Moreover, working with today's deep submicron technologies, designers are facing a severe leakage problem, which is a direct consequence of the scaling down of CMOS technology [4]. On the one hand, because of decreasing supply voltage values (V_{DD}), the threshold voltage of transistors (V_{TH}) is lowered to prevent speed degradation, leading to a huge increase of the subthreshold current. On the other hand, the thinning down of the gate oxide thickness exponentially increases the gate tunnelling current (IG) and the gate-induced drain leakage (IGIDL) current due to higher electrical fields near the drain. The most effective way to cut these leakage currents is to use power switch transistors, in order to switch off the inactive circuits. Among the existing power switch transistors [5], SCCMOS [19] is the most suited to a low- V_{DD} environment since it uses a low- V_{TH} transistor.

We present in this paper ALPIN, an "Asynchronous Low Power Innovative NoC" circuit [22]. This test-chip is designed to demonstrate different adaptive design techniques aiming at reducing both dynamic and static power consumption in a 65 nm CMOS technology. The aim of this work is to propose efficient techniques from system level (presented in Section III) down to the physical level (illustrated in Section VII) proposing also architectural and design levels power optimizations. This GALS architecture presented in Section II is based on a fully asynchronous NoC. The proposed NoC automatic power regulation will be described in Section IV. Due to the GALS paradigm, each ALPIN synchronous island described in Section V is an independent frequency and power domain. The proposed DVFS architecture is based on the association of a local clock generator, its pausable clock system [6] and a local voltage supply unit [10]. The supply unit, presented in Section VI, is based on a V_{DD} -Hopping technique [7] to control active power modes, combined with a SCCMOS power switch [8] to handle standby low leakage mode. Unit voltage and frequency are locally controlled to handle five power consumption modes from high performance to absolutely no leakage. An average voltage and frequency given value is obtained thanks to an automatic hardware mechanism. The main objective is to build an adaptive power-aware system while removing the need for fine control software.

II. POWER AWARE GALS NoC ARCHITECTURE

A. System Architecture

ALPIN (Fig. 1) is a fully power-aware GALS NoC architecture [9]. Each of the synchronous IP units of the dedicated SoC are arranged around a fully asynchronous network-on-chip. As detailed in Fig. 2, synchronization and communication between the NoC router and the synchronous units is done using a pausable clock mechanism called SAS (Synchronous-to-Asynchronous and Asynchronous-to-Synchronous interfaces) [3]. A programmable local clock generator, using a programmable

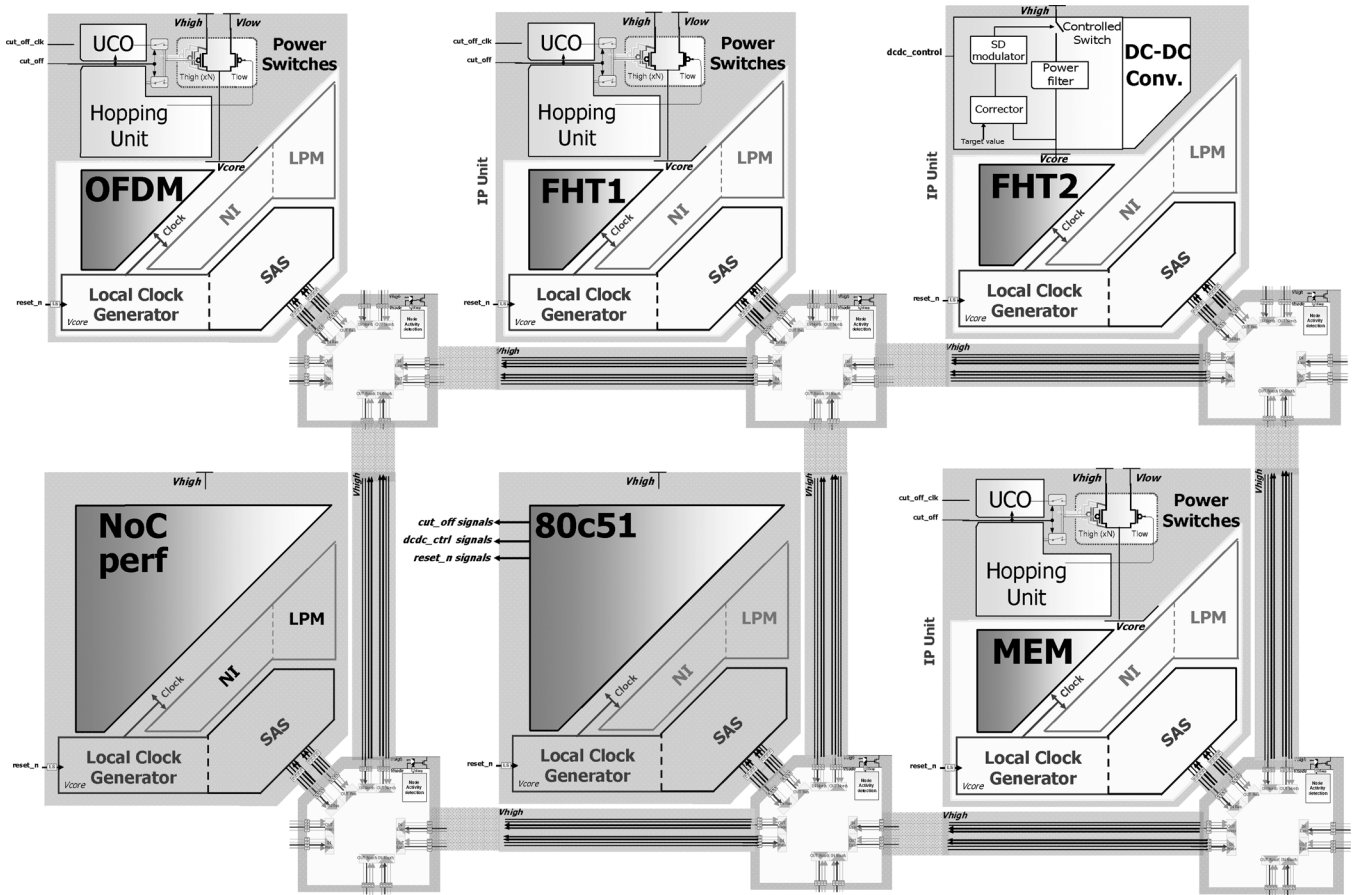


Fig. 1. ALPIN architecture.

delay line, is implemented within each unit to generate a variable frequency in a predefined and programmable tuning range. Each clock domain is also an independent power domain: each NoC unit owns a local power supply unit to generate and control its internal core voltage supply using V_{high} and V_{low} external supplies through power switches.

ALPIN contains six NoC routers and six NoC units dedicated to Telecom applications: a TRX-OFDM unit, two FHT units, one MEM (DMA-like) unit, a 80c51 microcontroller, and finally a dedicated NoC-perf unit for on-chip NoC traffic generation and NoC performance measurements. Functional details of the computing units are out of the scope of this paper [20]. Within IPs, standard-cell CMOS logic is robust to low-voltage operation with some margins; while embedded SRAMs has been designed to support low-voltage operation (down to 0.7 V) [10].

Regarding dynamic power consumption, two techniques are proposed and compared on-chip. On one hand, an advanced integrated buck-boost inductive DC-DC converter is applied for DVFS onto FHT2 unit through 80c51 programming. On the other hand, a Locally Adaptive Voltage and Frequency Scaling (LAVFS) technique is applied onto TRX-OFDM, MEM, FHT1 units using a fully integrated hopping between V_{high} and V_{low} .

Regarding static power consumption, PMOS power switches controlled by an ultra-cut-off (UCO) technique [8] are inserted on each NoC unit to maintain minimum leakage in standby mode. The power switch gate bias is automatically adapted to PVT conditions, leading to an optimal leakage reduction. As far as leakage is concerned, asynchronous NoC routers are automatically powered down during their inactivity phases.

The power management strategy is programmed by the main CPU, through NoC unit attached Network Interface's Low Power Managers, according to required performance and power constraints. DVFS can be executed during IP computation and communication according to their own activity. The only global signals are regarding units' reset and off control. The main CPU is required to directly disable/enable the units for entering power off mode and executing reset phase. Main principles of the local control and power regulation are described thereafter.

B. Main Principles

Each synchronous IP unit is defined as an independent power domain (using its dedicated local voltage) and an independent frequency domain (using its dedicated local clock based on a delay line described Section VI). The unit handles a set of user-defined power modes (see Section V).

In order to perform efficient local DVS, the main objective is to avoid as much as possible low-level software control to ensure a minimal latency cost of DVS during unit's computing. Within the Power Unit, this leads to implement a hardware controller to automatically switch between V_{high} and V_{low} . By guarantying smooth DVS transitions, the synchronous IP block can continue its own operation. To obtain an average voltage value between V_{high} and V_{low} , the Low Power Manager automatically switches between these two values using a configurable duty-ratio.

Since the power domain and frequency domain of the synchronous unit are identical, it is expected that the local frequency scales approximately linearly with the associated

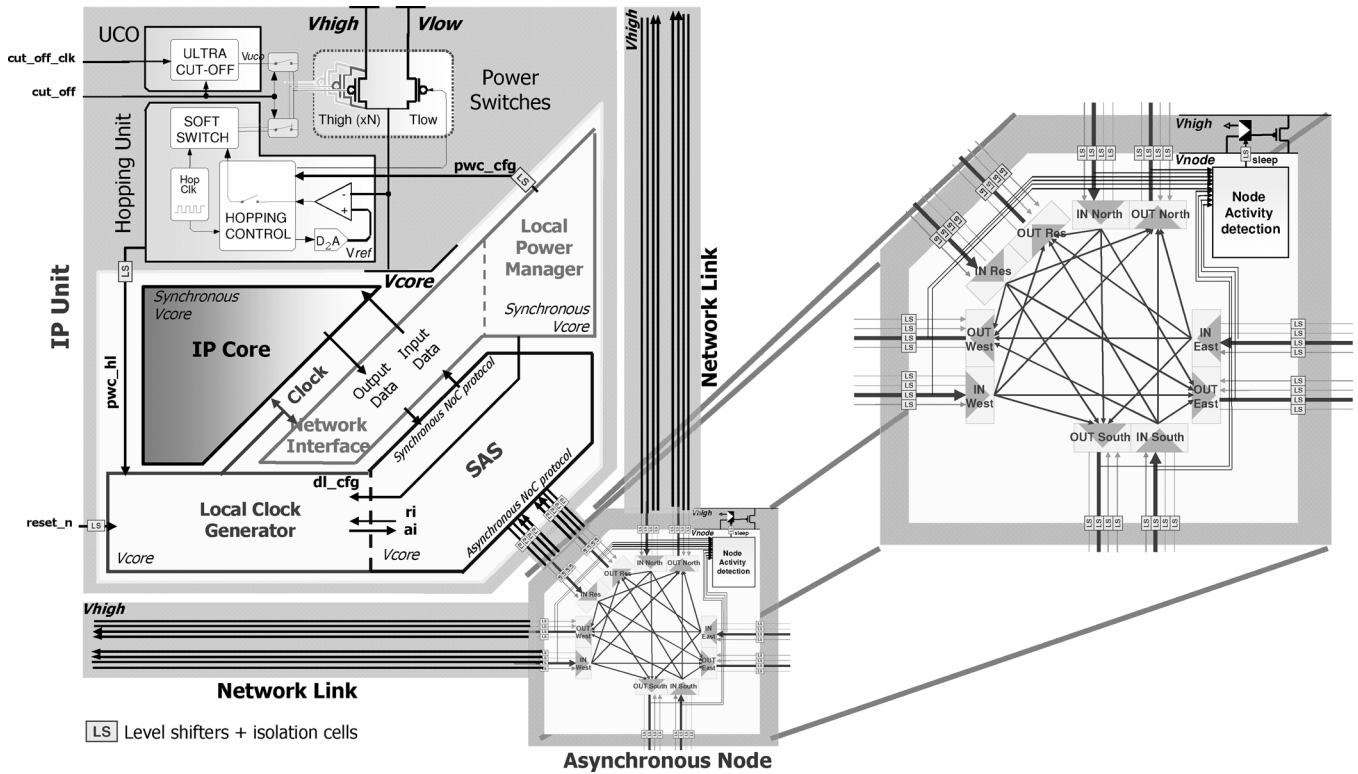


Fig. 2. NoC units architecture. Zoom on a NoC node.

voltage scaling. If the frequency/voltage scaling ratio is not exactly linear, the delay line must be reprogrammed accordingly. The delay value corresponding to V_{low} voltage is adjusted by a correcting factor with respect to the V_{high} delay value, since the delay line is supplied on the *same* power domain.

Finally, using the Pausable Clock technique [9], fast and reliable delay line programming interface is obtained. As a consequence, the synchronous IP unit locally continues its own computations or NoC communications during any DVFS phases.

III. SYSTEM-LEVEL POWER MODELLING

In such architecture, power estimation and profiling at application level is a major concern for proper power optimization. SystemC at the Transaction Level is adapted and largely adopted by the industry [21] to perform platform functional validation, as well as performance estimation. We have developed a generic way to instrument a SystemC Transaction Level Modeling (TLM) platform to model power consumption at a coarse grain [12]. The power model takes into account leakage and dynamic power and generates power traces and power/energy statistics. The proposed approach is applied to generate ALPIN power profile and drive the power management policy considering that low power design techniques are implemented at lower level of abstraction.

A. Power Modelling Framework

We do not intend to estimate power consumption from a pure functional model. Our estimator must be fed with given power consumption data. Hence, the power modelling framework works as a calculator supplied with external data, instrumenting the system level model (Fig. 3). Each power model is specific to the estimated IP model. A unique generic monitor watches all power models, in order to report and trace results. The power

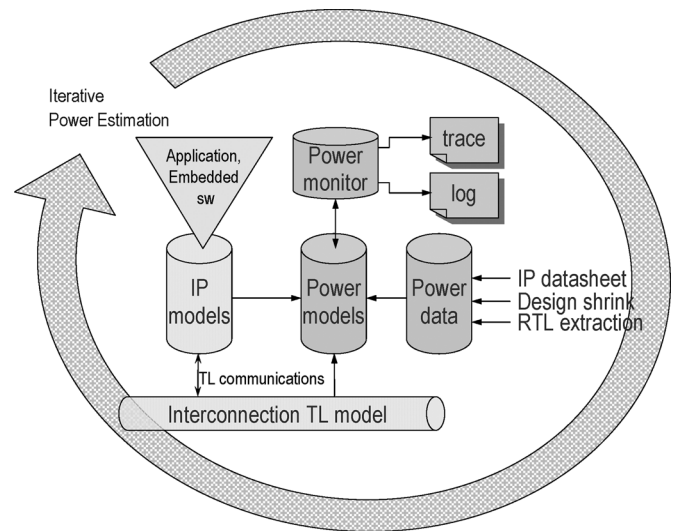


Fig. 3. Power estimation framework.

model allows iterative power estimations and optimization, by modifying embedded software, IP models, power model refinement or power data. A top-down approach enables to perform power-aware design space exploration. It also gives earlier estimates of the power consumption which are useful for designing power related mechanisms (hardware or software). The goal of the bottom-up approach is more specific. It implies that a hardware description is already available. Thus, it assumes HW/SW partitioning and HW design is already performed and does not aim at improving the design. The bottom-up approach is useful when application parameters are expected to widely affect power consumption of the SoC. This is basically the case

when the architecture implements dynamic power management techniques [16].

B. IP Unit Modelling

We have validated our approach on ALPIN units. The existing TLM/SystemC models of ALPIN units have been instrumented to model their power behavior. We supply our power model with data extracted from PrimePower simulations (gate level, after place and route, and with back-annotated gates). The power model takes into account leakage and dynamic power and generates power traces and power/energy statistics. The high level model provides the IP unit power profile using DVFS mechanisms. It also gives the NoC power profile including its automatic power regulation [12]. The obtained Transaction Level simulations are fast and exhibit a relative error of less than 10% compared to PrimePower simulations.

The proposed power modelling has been applied to investigate system level power optimization: IP scheduling, DVFS parameters, and NoC traffic analysis.

IV. LOW-POWER ASYNCHRONOUS NoC DESIGN

In this section, we first recall the topology of the asynchronous NoC already published in [11] and a brief explanation of the activity control detection implemented to reduce leakage power within the NoC when inactive. Finally, we give silicon measurement results on ALPIN chip.

A. NoC Overview

ALPIN NoC Architecture (Fig. 1) is a 2D-mesh based topology. The asynchronous node (Fig. 2) is a communication crossbar that implements the physical layer of the NoC. It is composed of five input and five output controllers, which respectively route and store incoming flits, and arbitrate and transmit outgoing flits. At link level, a handshake protocol using specific “send” and “accept” signals enables the flits composing the packets to be transmitted from node to node. Quality-of-service is provided using two virtual channels, enabling priority management. One virtual channel is reserved to guaranteed-service, while the second virtual channel is dedicated to best-effort traffic. NoC nodes and links are implemented using quasi-delay-insensitive (QDI) logic with a four-phase protocol and four-rail encoding scheme [13]. Due to local handshaking, asynchronous circuits are automatically in standby state when inactive. Nevertheless, asynchronous circuits, like synchronous ones, also need to reduce their leakage power. A power analysis of the asynchronous nodes revealed that leakage power represents about a quarter of the dissipated power and most of the static power consumption is dissipated while the node was inactive. Furthermore, to reach low latency specifications, the nodes are implemented using a specific asynchronous cell library designed with low voltage threshold (LVT) transistors, which leads to higher leakage currents for those specific cells.

The design objective is to propose a low leakage mode within NoC nodes, during inactivity phases, by reducing the supply voltage, while keeping the node in a functional state, even though at low speed because of asynchronous logic robustness due to its delay insensitivity [13]. A global power management at NoC level is not feasible because of the unpredictability of the traffic in the network: real-time software to control the nodes activity is unrealistic. This needs to be done node per node using a local control based on data traffic within each node.

TABLE I
NoC NODE POWER AND SPEED PERFORMANCES

Node State	Power On @ 1,2V	Power Down @ 0,6V
Throughput	1,8 ns (550 Mflits/s)	7,2 ns (140 Mflits/s)
Latency	2,5 ns	5,8 ns
Leakage	210 μ A	80 μ A

B. Node Activity Detection and Leakage Control

In the NoC topology, each node is in charge of its own power regulation independently from the others. NoC links between NoC nodes always remain at a high supply voltage to be able to transmit flits from a high power node to a low power one.

The node is considered to be inactive whenever there is no data stored inside any input or output controller of the node. It is only an interconnection medium, and has no incidence on the quantity of data: each data flit is admitted in the node, and is then transferred out of the node. It is sufficient to count data flowing in and out of the node to determine whether it can be put in low-power mode. An activity control block is defined [11], which listens to every input or output port and increments a counter whenever a begin-of-message (BOM) gets in, and respectively decrements this counter whenever an end-of-message (EOM) gets out. When the message counter is equal to zero, the node can be powered down. When a new data gets in the node, it must be powered up as fast as possible.

Using a *sleep* signal generated by the activity control, a Power Regulation Unit made of a gate-controlled PMOS transistor is in charge of automatically reducing the internal power supply down to 0.6 V in order to save leakage in low power mode. As explained before, the node has to stay functional during this phase either for protocol encoding reasons or reset constraints at power up. Thanks to asynchronous logic robustness, the node remains functional while reducing its leakage power (Table I). This adaptive low-power mechanism is performed without any latency cost and any specific software for leakage control.

V. POWER ADAPTIVE NoC UNIT ARCHITECTURE

Each synchronous IP unit of the ALPIN chip is defined as an independent power domain (using its dedicated local voltage V_{core}) and an independent frequency domain (using its dedicated Local Clock Generator). The IP architecture is described here below followed by the set of user-defined power modes description.

A. NoC Unit Integration for DVFS

In order to integrate a synchronous unit within the proposed NOC DVFS scheme (Fig. 2), each synchronous IP core is encapsulated within the following.

- Its own Network Interface (NI) and Local Power Manager (LPM). The Network Interface provides HW primitives for NoC packet generation and reception, and task synchronization. The LPM implements the local power mode defined for that given unit and controls respectively the power supply unit and the Local Clock Generator.
- Its own NoC Test Wrapper (NTW) for handling test-mode (not represented in Fig. 2 for clarity of illustration). For test mode, the NoC topology is used to carry the test patterns which are used to feed the synchronous scan chains through the NTW [9].

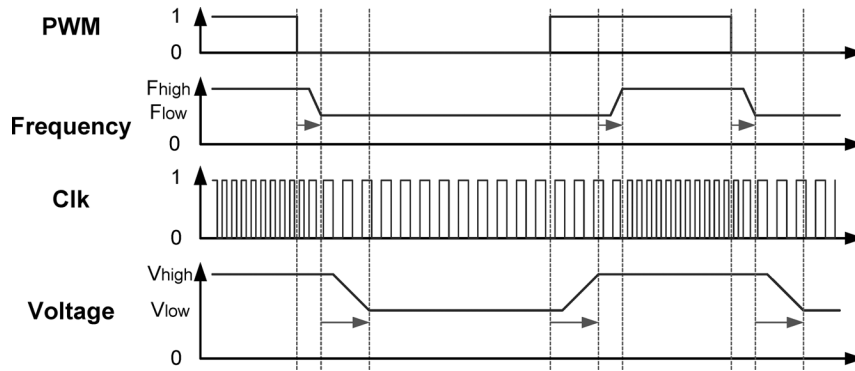


Fig. 4. LPM control, sequence example.

TABLE II
UNIT POWER MODES

INIT	At reset, the unit is at V_{high} with no clock
HIGH	The unit is supplied by V_{high} voltage
LOW	The unit is supplied by V_{low} voltage
HOPPING	The unit is automatically switched between V_{high} and V_{low} voltages, <i>for DVFS</i>
IDLE	The unit is idle, with maintain of its current state at V_{low} voltage, <i>for reduced leakage power</i>
OFF	The unit is switched OFF, with no maintain of its current state, <i>for minimal leakage power</i>

- Its own Pausable Clock interface, which contains the SAS interface, the Local pausable Clock Generator and a delay line programming interface.
- Its own power supply unit (PSU), which contains an ultra-cut-off (UCO) mechanism for leakage reduction and a hopping unit for DVFS, generates the V_{core} voltage. Lastly, some isolation cells are inserted on the external asynchronous NoC link for proper voltage conversion between the local unit and the NoC topology.

In the proposed scheme, the complete Synchronous IP core is clocked by the locally generated clock, and supplied by the V_{core} voltage, locally generated by the PSU. The synchronous unit has thus its own frequency and power domain.

B. Power Mode Definition

In the proposed architecture, each unit can be set in one of the six available power modes (Table II):

- in INIT mode, supply voltage is V_{high} , and no clock is sent to the core. This is the “post-reset” mode.
- in HIGH mode, supply voltage is V_{high} and core clock is on. This is the “nominal” working high performance mode.
- in LOW mode, core clock is still on, but supply is switched to V_{low} . Clock frequency is lower than nominal, and energy per cycle decreases. This is the “low power” mode.
- in HOPPING mode, core clock is on and supply voltage automatically hops between V_{high} and V_{low} . Frequency and duty-ratio of hopping transitions is configurable. The obtained performance is an average value between V_{high} and V_{low} modes, depending on the given duty-ratio. This is the “DVFS” mode.
- in IDLE mode, core clock is off and leakage power is reduced due to the V_{low} supply voltage. This is the “low-power dormant” mode.

- in OFF mode, the unit is switched off either by an MTCMOS classical approach [18] or by the UCO device (see Section VI) to further reduce the leakage power. Since the NI is inactive in this mode, the OFF mode is enabled/disabled through an external “cut_off” signal controlled by another unit in the NoC (for instance the main host processor). This is the “low-leakage” mode.

All power modes, beside the OFF mode, can be selected per each unit through programming of the unit Network Interface.

VI. NoC UNIT DESIGN

In this section is presented the design of the various elements (from Fig. 2) to encapsulate one synchronous IP core within the proposed low power mechanisms. We describe respectively the Local Power Manager, the Local Pausable Clock Generator [9] and the power supply unit.

A. Local Power Manager

The Local Power Manager (LPM) is in charge of handling the unit’s power modes. This manager contains a set of programmable registers, which can be programmed through the NoC, to define the unit power mode, to configure the programmable delay line, and to configure and control the power supply unit. The Power Unit manager contains a “mode” register to define the unit power mode (Section V). In INIT and IDLE mode, the Power Manager set an idle signal in order to gate the clock of the core unit. The power supply unit is then required to be in V_{high} supply in INIT mode while in V_{low} supply in IDLE mode.

In order to control the HOPPING mode, a Pulse Width Modulation (PWM) is used. Two dedicated registers define the “PWM frequency” and the “PWM duty-ratio”. As an illustration, Fig. 4 shows a sequence between the HIGH and LOW states using the PWM.

Finally, the LPM also contains a set of registers to precisely control the necessary Hopping-unit signals: the hopping-unit clock control (through a dedicated programmable delay-line) and transition slopes control between V_{high} and V_{low} .

B. Local Pausable Clock Generator

The Local Pausable Clock Generator (Fig. 5) is in charge of IP core *Clk* generation and in charge of temporary pausing the clock during synchronous-to-asynchronous (S-A), asynchronous-to-synchronous (A-S) data communications and during voltage transitions. It handles the concurrent requests: from either the A-S, S-A or LPM interfaces, respectively: *ri_as*, *ri_sa*, *ri_dlp*. An asynchronous request is arbitrated

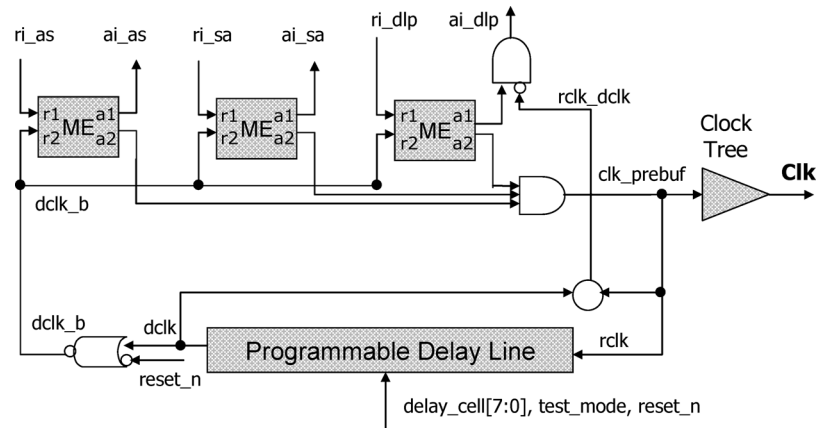


Fig. 5. Local pausable clock generator.

using a mutual-exclusion element which causes the clock to be momentarily paused. Once the request is released, the clock can restart again. If no asynchronous request is received, the clock (half-) period is determined by the value programmed in the delay line. The generated clock is then applied to the synchronous unit through a classical clock-tree.

The main difficulty is to obtain a precise, small and low-power delay line, while using if possible only standard-cells so that it can be placed-and-routed as a hard-macro and then reused for various units. The delay line is composed of delay elements built with either available delay-cells or inverter-cells (according to the required delays), and of latches and multiplexers. The latches offer a good compromise in terms of delay and energy in the targeted technology (STMicroelectronics 65 nm), while filtering out all unnecessary pulses within the delay line. Only one path through the delay line is activated according to the programmed binary value.

C. Power Supply Unit

The power supply unit (PSU) manages the unit supply voltage V_{core} according to the selected power modes using supply voltages provided by off-chip DC-DC converters. The power supply unit (Fig. 6) is composed of three main devices: the power switches (T_{high} and T_{low} power transistors), the Ultra Cut-Off voltage generator (UCO), and the V_{DD} -Hopping unit. The UCO is used during 'OFF' mode to reduce the unit leakage current. The V_{DD} -Hopping unit ensures smooth transitions between V_{high} and V_{low} without stopping the unit clock and computations. In the case of FHT2 unit, the PSU is replaced by a fully integrated buck-boost DC-DC converter.

1) *Dynamic Power Control*: To do fine-grain voltage scaling and reduce dynamic power consumption of SoC and MPSoC containing more than ten functional cores, traditional DC-DC converters have reached their limit. The simplest ones are linear converters, small and easily integrated, but their efficiency is lowered at low output voltage. More efficient converters like capacitive or inductive ones are widely used in the industry, but they are using capacitors and inductors that cannot be easily integrated. A fully integrated inductive buck-boost DC-DC converter has been implemented to compare on-chip a classical Local Dynamic Voltage and Frequency Scaling (LDVFS) approach with an innovative Local Adaptive Voltage and Frequency Scaling approach using a hopping technique.

a) *Integrated Buck-Boost DC-DC Converter*: A micropower up-and-down converter switching power supply is used to

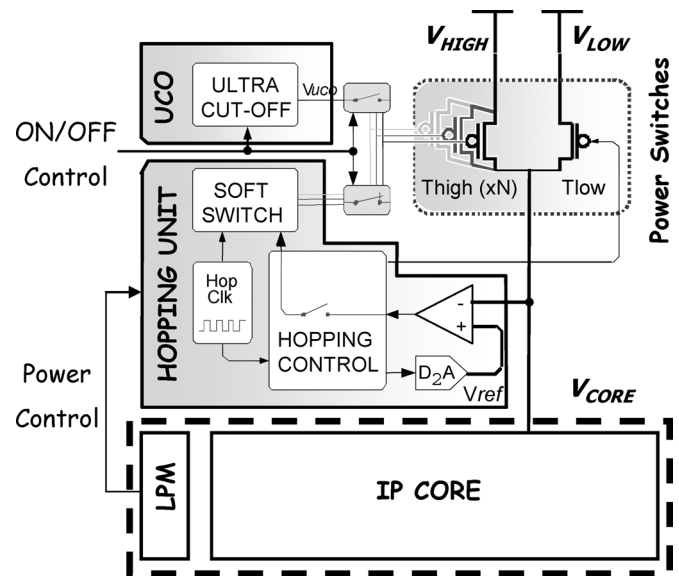


Fig. 6. Power supply unit.

convert the available power from a battery into a regulated and controllable power supply. Ten discrete set point values between 0.6 V and 1.2 V are available for the power supply voltage to achieve dynamic voltage scaling on the supplied FHT2 digital block.

The input power is processed by power switches and an L-C power filter, yielding the conditioned output power. The passive devices values are chosen to be compatible with above-IC technologies to avoid any external devices: consequently, the inductor and capacitor values are drastically constrained. The system controls the switch duty cycle so that the output voltage follows a given reference. Therefore, the difference between one part of the power filter output voltage ($\alpha \cdot V_{out}$) and a voltage setpoint is amplified and modulated into pulse density information. The obtained clock is used to gate a drive circuit (non-overlapping clocks generator and buffer) controlling the power filter MOS switches. The innovative pulse density modulation (PDM) is based on an asynchronous passive $\Sigma\Delta$ modulator instead of the traditional PWM controller for simplicity of implementation (two RC filters, a comparator and two inverters), low power consumption and spectral spread of the switch noise.

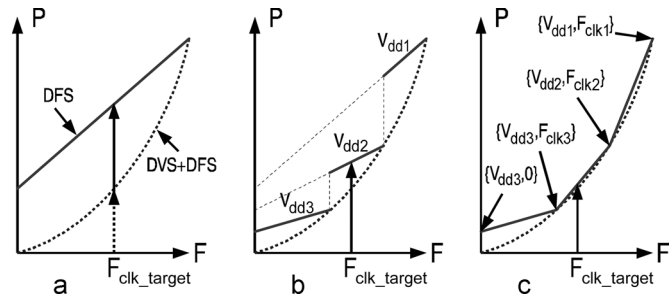


Fig. 7. Voltage dithering principle.

b) V_{DD} -Hopping Technique: The second method, called V_{DD} -Hopping with dithering [7], [15], uses two voltages V_{high} and V_{low} provided by external DC-DC converters to control the local voltage of the functional core. Fig. 7 illustrates the difference between a traditional DVFS and V_{DD} -Hopping. Fig. 7(a) illustrates a classical continuous DVFS approach obtained using an integrated local DC-DC converter. Fig. 7(b) represents a discrete DVS approach using three voltage supply levels but without voltage dithering. As a comparison, Fig. 7(c) illustrates the same system with voltage dithering. By using a dithering method between two power modes (voltage/frequency pairs such as $[V_{high}, F_{high}]$ and $[V_{low}, F_{low}]$ for the ALPIN proposal) V_{DD} -Hopping allows to control the core average frequency F_{avg} and to reduce the power nearly as efficiently as a continuous voltage converter. The obtained F_{avg} depends on the duty ration between T_{high} , time spent in HIGH mode and T_{low} , time spent in LOW mode:

$$F_{avg} = \frac{(F_{low} \cdot T_{low}) + (F_{high} \cdot T_{high})}{T_{low} + T_{high}}$$

To hop between two supply voltages, a device called a power supply selector (PSS) is necessary. The simplest PSS is composed of two power switch transistors (each connecting a different source to core power supply network) and a delay element [15]. To hop from one source to the other, one power transistor is switched on while the other is switched off as illustrated Fig. 8. Because the change of supply voltage and source current is very fast, there will be unpredictable voltage variations (under-shoot and over-shoot) caused by the complex impedance of the power supply network. Those variations can cause delay faults in active nearby digital circuits and are unacceptable. Moreover, the two transistors are partially ‘on’ at the same time, causing a current surge from the V_{high} to the V_{low} source. By stopping the clock during transitions and using a multiple-stage staggered switching for both V_{high} and V_{low} power switches, current injection from one source to an other is avoided [16], [17]. But voltage transients are still present in the power supply network. They will not affect the local functional core, whose clock is inactive, but may cause errors in active nearby functional cores. To solve these problems, without stopping the clock, we have proposed a more efficient transition principle illustrated in Fig. 9 [7]. During transitions, the core supply voltage is provided by the PSS acting as a linear regulator with a voltage set-point given by a DAC. This precise control allows changing

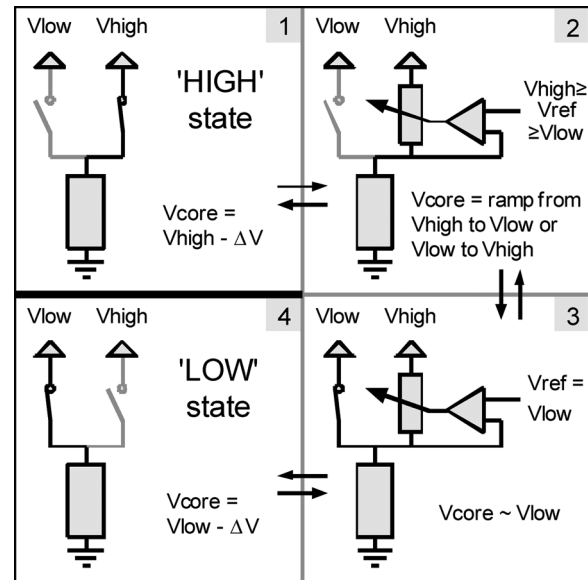


Fig. 8. Hopping transition principle.

the supply voltage following a controlled ramp (V_{ref}), limiting wide current variations, avoiding any supply voltage under- or over-shoot and current flowing from one source to another. Because of smooth transitions, with no undershoot or overshoot, the IP does not need to be stopped. Hopping occurs without latency cost at application level. The transition duration can be adjusted from ~ 40 ns to ~ 500 ns.

c) Compared Results: The Hopping Unit and its power switches are fully integrated and are 20 times smaller than the integrated buck-boost DC-DC converter, both designed for the same FHT block. Hopping Unit efficiency has been evaluated to be close to 95% on the full voltage range. Fig. 10 is giving measurement results of the FHT1 (Hopping Mode) and FHT2 (Classical DVFS Mode). On this curve, hopping technique is compared to an optimal DVS with 100% efficiency and to a classical DC-DC approach with 75% efficiency. The curves are fitting the voltage dithering principle illustrated Fig. 7(c). With given V_{high} and V_{low} voltages set to 1.2 V and 0.9 V for the hopping mode, according to applicative constraints, we can obtain any dynamic power consumption gain from 1 up to 3 without any latency cost in the unit.

2) Leakage Power Control: The use of power switch transistors to reduce leakage currents in digital and memory circuits is now a quite old field of research [14], [18]. Three main types of power switches have been proposed. The older one is the MTCMOS, standing for Multiple Threshold CMOS [18]. It consists in using low- V_T transistors for the logic and a high- V_T one for the power switch. The power switch is thus inserted in-between the logical circuit and the supply/ground lines. This technique has been implemented on-chip using a standard- V_T transistor on the power line of the MEM unit. The drawback of the MTCMOS technique resides in the high- V_T or standard- V_T transistor which provides poor performances in a low supply voltage V_{DD} environment. To cope with this issue, the super cut-off CMOS (SCCMOS) has been introduced [19]. It is a low- V_T transistor whose leakage current is exponentially reduced by reverse-biasing its gate, rendering the gate-to-source

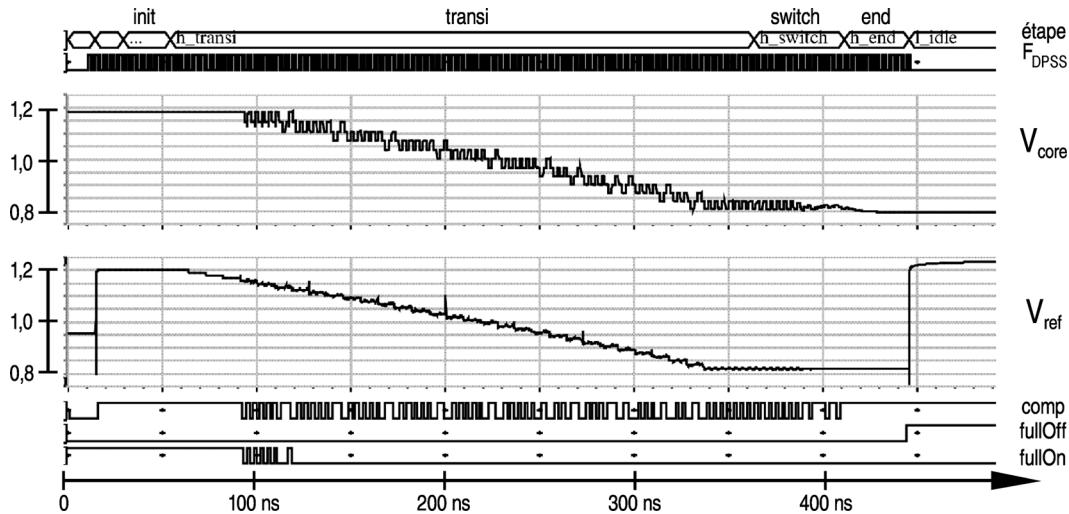


Fig. 9. Controlled transition in the hopping mode.

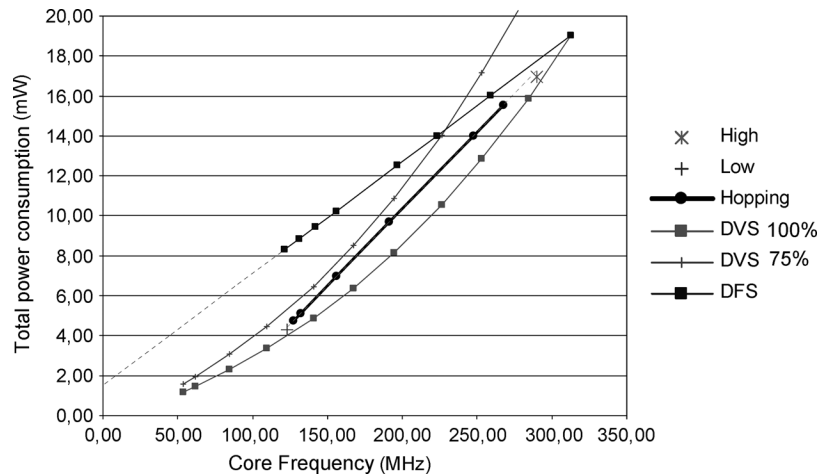


Fig. 10. Power measurement results on FHT1 and FHT2 units.

voltage negative in the case of an NMOS transistor and positive in the case of a PMOS one. For our application, a multiple-voltage islands circuit with a V_{DD} -hopping scheme, we chose a P-type SCCMOS power switch on the TRX-OFDM unit since the supply voltages can go as low as 0.8 V. It has been shown in [8] that a compromise has to be found between the three main leakage currents to minimize the total leakage current of the transistor. However, this optimal value cannot be defined once and for all, since it depends on the varying operating conditions, namely process, voltage and temperature. That is the reason why we have proposed a UCO circuit that automatically biases the gate of the SCCMOS power switch transistor to its point of minimum leakage. This circuit [8] optimally reduces the total leakage current of the power switch transistor whatever the environmental conditions: it efficiently compensates for temperature and corners variations. Although the accuracy of the bias voltage is affected by mismatch, the proposed circuit is nevertheless able to reduce the variation of the total leakage current by biasing the power switch transistor in a region where its current does not vary exponentially.

Leakage measurements (Table III) were performed on the two IP units, namely TRX-OFDM and MEM units. The TRX-OFDM unit, with the higher power dissipation, is driven by the

TABLE III
UCO (TRX-OFDM UNIT) AND MTCMOS (MEM UNIT) LEAKAGE REDUCTION RESULTS

TRX-OFDM unit			UCO	
Mode	V _{CORE}	I(V _{CORE})		R _{on} (k Ω /μm)
HIGH	1,166 V	30,5 mA		6,13
INIT	1,199 V	195 μA	loff gain	loff (pA/μm)
OFF	1,7 mV	0,27 μA	706	49

MEM unit			MTCMOS	
Mode	V _{CORE}	I(V _{CORE})		R _{on} (k Ω /μm)
HIGH	1,162 V	13,9 mA		7,11
INIT	1,198 V	284 μA	loff gain	loff (pA/μm)
OFF	10,2 mV	2,4 μA	118	923

UCO-type LVT power switch, while the MEM unit is driven by a classical MTCMOS-type SVT power switch. In HIGH mode, the voltage drop across the power switches is equal to 30 mV in both cases. The difference between the INIT and OFF modes illustrates the leakage gain provided by the power switches: leakage current of the UCO transistor is 8 times lower than the MTCMOS one, while its I_{ON} current is 2.5 times higher.

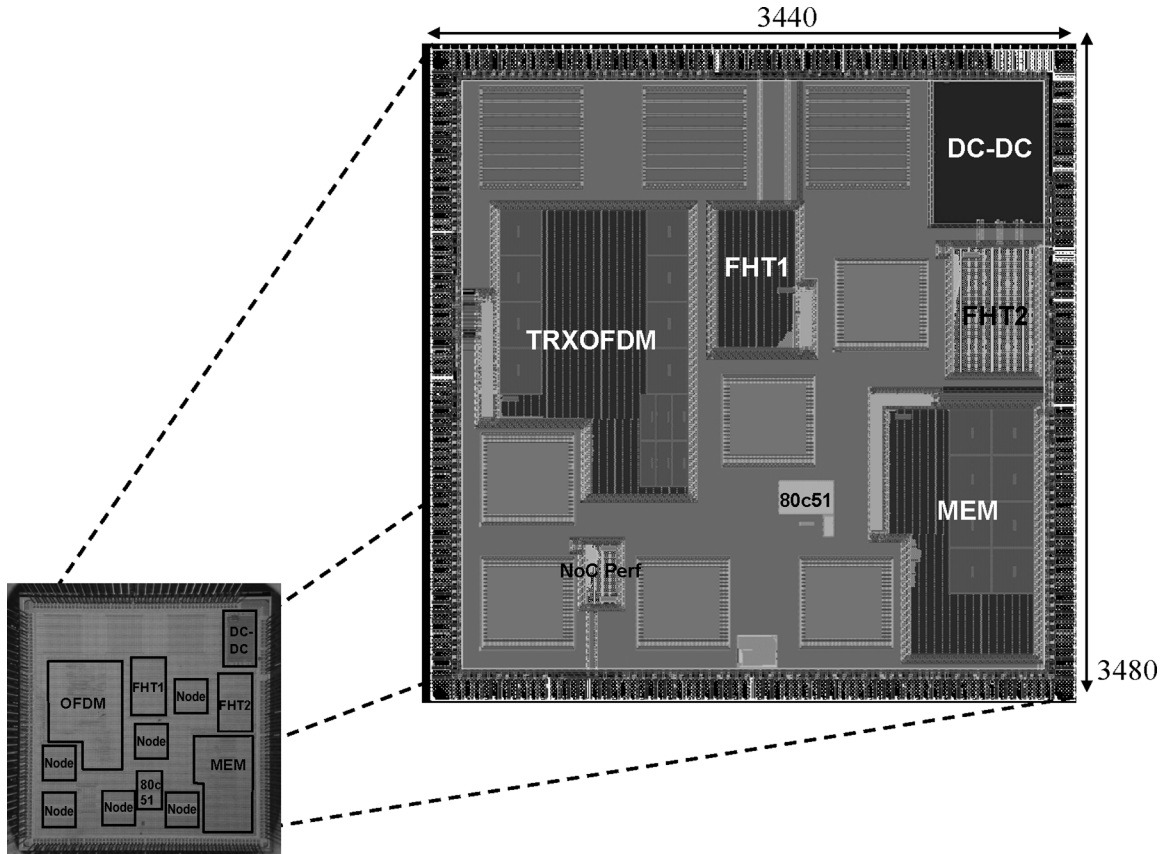


Fig. 11. ALPIN physical implementation.

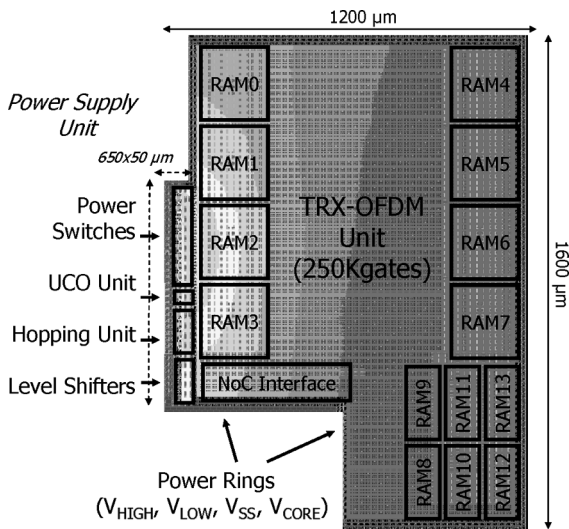


Fig. 12. TRX-OFDM Unit physical implementation and IR-drop distribution.

VII. PHYSICAL IMPLEMENTATION

ALPIN circuit has been implemented in a 65 nm STMicroelectronics technology and covers 11.56 mm² including 224 pads (Fig. 11). The chip contains the six previously mentioned IP units and the six NoC nodes all implementing their local PSU. We present in this section the layout architecture for the proposed low power features.

Fig. 12 gives the zoom on the TRX-OFDM unit with its attached PSU, and the corresponding IR-drop distribution [10].

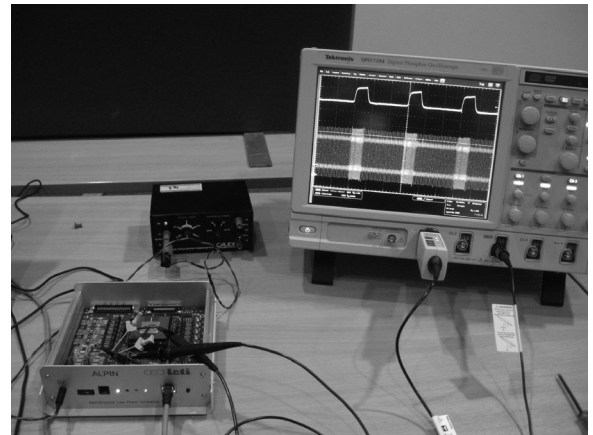


Fig. 13. ALPIN test board and hopping demonstration.

The TRX-OFDM unit is a complex unit providing FFT computations for both RX and TX OFDM protocols, this unit integrates 14 low-power low-voltage SRAMs (290 Kb total). The unit is supplied through four power rings: the external V_{high} , V_{low} , V_{ss} , and the generated V_{core} . The power supply unit represents only 2.5% of the total OFDM unit area.

VIII. RESULTS AND CONCLUSION

The ALPIN chip has been fully measured and validated using a test board (Fig. 13) including a FPGA implementing an extended NoC for traffic generation. The power supply unit area

TABLE IV
ALPIN MEASUREMENT RESULTS

	FHT	MEM	OFDM
Complexity (inc. SRAMs)	56 k gates	220 k gates (SRAMs : 149 k gates)	280 k gates (SRAMs : 187 k gates)
HIGH@1.2V	20.12mW@285MHz	32.63mW@220MHz	81.07mW@220MHz
LOW@0.8V	2.66mW@85,5MHz	4.31mW@66MHz	10.07mW@66MHz
HOPPING@50%	12.3mW@185MHz	19.4mW@145MHz	46.4mW@145MHz
IDLE@1.2V	682 μ W	883 μ W	632 μ W
OFF	720 nA	7.8 μ A	930 nA

cost for each power domain is less than 4% for a delay penalty of 5%. In Table IV are presented the power results of the three units in various power modes. The static power consumption is reduced by 2 decades in OFF mode using UCO, while the dynamic power consumption can be reduced up to a factor of 8 from HIGH to LOW mode. In HOPPING mode, with a duty ratio of 50% between V_{high} and V_{low} , as an example, power consumption is divided by 2 while respecting the speed constraints of the NoC units. Lastly, the power efficiency of the HOPPING mode is 97%, taking into account Hopping transitions. In terms of throughput, the NoC node and NoC pipelined links provide a raw 17 Gb/s onto 32-bit flits.

The proposed ALPIN architecture provides efficient and adaptive power reduction techniques. These innovations will be implemented in a reconfigurable chip for Telecom multi-applications and will be used as a first basis to improve robustness to variability in deep submicron technologies.

ACKNOWLEDGMENT

The authors would like to thank STMicroelectronics for their support and contribution to this work.

REFERENCES

- [1] M. Weiser, B. Welch, A. J. Demers, and S. Shenker, "Scheduling for reduced CPU energy," *Operating Systems Design and Implementation*, p. 1323, Apr. 1994.
- [2] P. Royannez *et al.*, "90 nm low leakage SoC design techniques for wireless applications," in *IEEE ISSCC'2005 Dig.*, San Francisco, CA, Feb. 2005, pp. 138–139.
- [3] M. Krstic, E. Grass, F. K. Gurkaynak, and P. Vivet, "Globally asynchronous, locally synchronous circuits: Overview and outlook," *IEEE Design & Test of Computers*, vol. 24, no. 5, pp. 430–441, Sep.–Oct. 2007.
- [4] N. S. Kim, T. Austin, D. Blauuw, T. Mudge, K. Flautner, J. Hu, M. J. Irwin, M. Kandemir, and V. Narayanan, "Leakage current: Moore's law meets static power," *IEEE Comput.*, vol. 36, Dec. 2003.
- [5] H. Kawaguchi, K. Nose, and T. Sakurai, "A super cut-off CMOS (SC-CMOS) scheme for 0.5-V supply voltage with picoampere stand-by current," *IEEE J. Solid-State Circuits*, vol. 35, no. 10, pp. 1498–1501, Oct. 2000.
- [6] R. Mullins and S. Moore, "Demystifying data-driven and pausable clocking schemes," presented at the IEEE Int. Symp. Asynchronous Circuits and Systems (ASYNC'07), Berkeley, CA, Mar. 2007.
- [7] S. Miermont, P. Vivet, and M. Renaudin, "A power supply selector for energy- and area-efficient local dynamic voltage scaling," presented at the PATMOS'2007, Göteborg, Sweden, Sep. 3–5, 2007.
- [8] E. B. Valentian, "Gate bias circuit for an scmos power switch achieving maximum leakage reduction," presented at the ESSCIRC 2007, Munich, Germany, Sep. 11–13, 2007.
- [9] E. Beigné *et al.*, "Dynamic voltage and frequency scaling architecture for units integration within a GALS NoC," presented at the NOCS'2008, Newcastle, UK, Apr. 2008.
- [10] E. Beigné, S. Miermont, A. Valentian, P. Vivet, S. Barasinski, F. Blisson, N. Kohli, and S. Kumar, "A fully integrated power supply unit for fine grain DVFS and leakage control validated on low-voltage SRAMs," presented at the ESSCIRC'2008, Edinburg, UK, Sep. 2008.
- [11] Y. Thonnart *et al.*, "Automatic power regulation based on an asynchronous activity detection and its application to ANOC node leakage reduction," presented at the ASYNC'2008, Newcastle, UK, Apr. 2008.
- [12] H. Lebreton and P. Vivet, "Power modeling in SystemC at transaction level, application to a DVFS architecture," presented at the Int. Symp. VLSI (ISVLSI'08), Montpellier, France, Apr. 2008.
- [13] M. Renaudin, P. Vivet, and F. Robin, "ASPRO: An asynchronous 16-bit RISC microprocessor with DSP capabilities," in *Proc. 25th European Solid-State Circuits Conf (ESSCIRC'99)*, Sep. 1999, pp. 428–431.
- [14] K. Itoh, "Reviews and prospects of deep sub-micron DRAM technology," in *Extended Abstracts, 1991 Int. Conf. Solid-State Devices and Materials*, Aug. 1991, pp. 468–471.
- [15] H. Kawaguchi, G. Zhang, S. Lee, Y. Shin, and T. Sakurai, "A controller lsi for realizing vdd-hopping scheme with off-the-shelf processors and its application to mpeg4 system," *IEICE Trans. Electronics*, vol. E85-C, no. 2, pp. 263–271, Feb. 2002.
- [16] D. Truong, W. H. Cheng, T. Mohsenin, Z. Yu, T. Jacobson, G. Landge, M. Meeuwssen, C. Watnik, P. Mejia, A. Tran, J. Webb, E. Work, Z. Xiaoh, and B. M. Baas, "A 167-processor 65 nm computational platform with per-processor dynamic supply voltage and dynamic clock frequency scaling," presented at the Symp. VLSI Circuits, Jun. 17–20, 2008.
- [17] W. H. Cheng and B. M. Baas, "Dynamic voltage and frequency scaling circuits with two supply voltages," presented at the IEEE Int. Symp. Circuits and Systems (ISCAS'08), May 18–21, 2008.
- [18] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada, "1-V power supply high-speed digital circuit technology with multithreshold-voltage CMOS," *IEEE J. Solid-State Circuits*, vol. 30, no. 8, pp. 847–854, Aug. 1995.
- [19] H. Kawaguchi, K. Nose, and T. Sakurai, "A CMOS scheme for 0.5 V supply voltage with pico-ampere standby current," presented at the IEEE ISSCC, Feb. 1998.
- [20] D. Lattard, E. Beigne, F. Clermidy, Y. Durand, R. Lemaire, P. Vivet, and F. Berens, "A reconfigurable baseband platform based on an asynchronous network-on-chip," *IEEE J. Solid State Circuits*, vol. 43, no. 1, pp. 223–235, Jan. 2008.
- [21] The Open SystemC Initiative. [Online]. Available: http://www.systemc.org/community/about_systemc/
- [22] E. Beigne, F. Clermidy, J. Durupt, H. Lhermet, S. Miermont, Y. Thonnart, T. Tran-Xuan, A. Valentian, D. Varreau, and P. Vivet, "An asynchronous power-aware and adaptive NoC based circuit," presented at the Symp. VLSI Circuits, Honolulu, HI, Jun. 2008.



Edith Beigné was born in Lamastre, France, in 1975. She received the Electronic Engineering Diploma from the National Polytechnic Institute of Grenoble, France, in 1998.

In 1998, she joined the CEA-LETI laboratory in the Center for Innovation in micro & nanotechnology (MINATEC), Grenoble. She was first involved in contactless RFID mixed signal systems. In 2001, she began the asynchronous logic design activity in cryptographic and contactless systems. As regards the development of the FAUST project, she designed a part of the asynchronous Network-On-Chip. Since 2006, she has been in charge of ALPIN, a power-aware GALS SoC implementing dynamic and static low power techniques based on an asynchronous NoC. She is now working on variability issues and power-performance-yield improvement in advanced CMOS circuits.



Fabien Clermidy was born in Bourg-en-Bresse, France, in 1971. He received the Electronic Engineering Diploma from ENSIMEV in 1994 and the Ph.D. degree in microelectronics from the National Polytechnic Institute of Grenoble, France, in 1999.

In 2000, he joined the CEA-LIST Laboratory in Paris. He was involved in the design of an application-specific parallel computer as a designer. In 2003, he moved to the CEA-LETI Laboratory in the Centre for Innovation in Micro and Nanotechnology (MINATEC), Grenoble. From 2003 to 2006, he was the

architect of the FAUST NoC structure. In 2006, he took the lead of the MAGALI project for software and cognitive radio.



Hélène Lhermet was born in Le Mans, France, in 1974. She received the Electronic Engineering Diploma from Institut Supérieur d'Electronique du Nord (ISEN), France, in 1997, and the Ph.D. degree in physics from the University of Grenoble, France, in 2000.

Currently, she is a Design Engineer for CEA/LETI in the Center for Innovation in micro & nanotechnology (MINATEC) and focuses on electronics for MEMS and power management.



Sylvain Miermont received the engineering diploma in electronics from ENSSAT (Lannion), in 2004, and the Ph.D. degree in microelectronics from Grenoble Institute of Technology, in 2008.

In 2005, he joined CEA-LETI Laboratory in the Center for Innovation in Micro and Nanotechnology (MINATEC), Grenoble, as a Ph.D. student. For three years, he worked on a fine-grain Dynamic Voltage Scaling architecture for GALS-SoC. He has developed a device allowing practical implementation of such architecture, and has patented an architecture for

fine-grain Adaptive Voltage Scaling addressing present and future variability problems.



Yvain Thonnart was born in Paris, France, in 1980. He graduated and received the M.S. degree from the Ecole Polytechnique, France, in 2003. He received the Engineering Diploma from Telecom ParisTech, France, in 2005, specializing in electrical engineering.

In 2003, he worked in STMicroelectronics, Crolles, France, where he focused on the development of digital SOI architectures. In 2005, he joined the CEA-LETI Laboratory in the Center for Innovation in Micro and Nanotechnology (MINATEC), Grenoble, France. His main research interests are asynchronous

logic and low-power design, network-on-chip architectures, formal verification of asynchronous systems, and distributed systems communication protocols. He is currently involved in the development of a NoC-based platform for next-generation wireless telecommunication standards.



Xuan-Tu Tran (M'08) received the Ph.D. degree in Micro and Nano Electronics from the ASIC Design Department of the CEA-LETI, MINATEC and the National Polytechnic Institute of Grenoble (INPG), France, in February 2008. He received the M.Sc. degree from Vietnam National University, Hanoi, Vietnam, in 2003 and the B.Sc. degree from Hanoi University of Science in 1999, all in electronics engineering and communication.

He is currently a Lecturer in the College of Technology (Coltech), Vietnam National University, and

Vice-Head of The Key Laboratory on Smart Integrated Systems. His research interests include design and test of system-on-chip, networks-on-chip, design-for-testability, and asynchronous/synchronous VLSI designs.



Alexandre Valentian was born in Mantes-La-Jolie, France, on August 5, 1977. He received the M.S. degree from ISEP (Institut Supérieur d'Electronique de Paris) in 2001 and the Ph.D. degree from ENST (Ecole Nationale Supérieure des Télécommunications), Paris, in 2005, both in electronic engineering.

He joined CEA-LETI laboratory in the Center for Innovation in micro & nanotechnology (MINATEC), Grenoble, France, in 2005, where he was involved in the development of low-power and low-leakage design solutions for a telecommunications network-on-

chip processor. His current research topics include standard cells library development and leakage and variability control in FDSOI technology.



Didier Varreau was born in Dôle, France, in 1954. He received the Electronic higher technical diploma from Grenoble University, France, in 1975.

In 1976, he joined CEA-LETI to develop instrumental electronic boards for medical and nuclear purpose. From 2003 to 2006 he worked on the FAUST project developing integrated synchronous IPs. Since 2006, he was in charge of physical implementation of the ALPIN chip. He is currently working on multiprocessor system-on-chip.



Pascal Vivet received the Master of Electronics from UJF, Grenoble, France, in 1994. He received the Ph.D. degree in 2001 in the France Telecom Lab, Grenoble, designing a quasi-delay-insensitive microprocessor.

After four years with STMicroelectronics, he joined CEA-LETI in 2003 in the Advanced Design Department. His topics of interest are focused on network-on-chip, globally-asynchronous-locally-synchronous architecture, and low power design.



Xavier Popon received the Engineer degree in electronics from CPE-Lyon, France, in 1990. After working in R&D in two small companies, where he developed electronic boards and FPGA designs for simulators and virtual studios, he received the Masters degree in microelectronics from INPG Grenoble in 2003.

He joined CEA-Leti in 2003 and is currently working on design and programming of electronic boards for demonstrators in wireless applications.



Hugo Lebreton was born in Redon, France, in 1984. He received the Electronics and Computer Sciences degree from the ENSSAT Engineering School at the University of Rennes 1, France, in 2007.

He joined CEA-LETI in spring 2007. His research activities and interests are currently focused on architecture level power modeling and application level power optimization in GALS SoC.