Heterogeneous vs Homogeneous MPSoC Approaches for a Mobile LTE Modem

Camille Jalier, Didier Lattard, A.A. Jerraya, Gilles Sassatelli, Pascal Benoit, Lionel Torres

Heterogeneous vs Homogeneous MPSoC Approaches for a Mobile LTE Modem

Camille Jalier, Didier Lattard and Ahmed Amine Jerraya
CEA, LETI, MINATEC
F38054 Grenoble, France
{firstname.lastname}@cea.fr

Gilles Sassatelli, Pascal Benoit and Lionel Torres
University of Montpellier II, LIRMM
UMR 5506, France
{lastname}@lirmm.fr

***Abstract*— Applications like 4G baseband modem require single-chip implementation to meet the integration and power consumption requirements. These applications demand a high computing performance with real-time constraints, low-power consumption and low cost. With the rapid evolution of telecom standards and the increasing demand for multi-standard products, the need for flexible baseband solutions is growing. The concept of Multi-Processor System-on-Chip (MPSoC) is well adapted to enable hardware reuse between products and between multiple wireless standards in the same device. Heterogeneous architectures are well known solutions but they have limited flexibility.**

Based on the experience of two heterogeneous Software Defined Radio (SDR) telecom chipsets, this paper presents the homoGENEous Processor arraY (GENEPY) platform for 4G applications. This platform is built with Smart ModEm Processors (SMEP) interconnected with a Network-on-Chip. The SMEP, implemented in 65nm low-power CMOS, can perform 3.2 GMAC/s with 77 GBits/s internal bandwidth at 400MHz. Two implementations of homogeneous GENEPY are compared to a heterogeneous platform in terms of silicon area, performance and power consumption. Results show that a homogeneous approach can be more efficient and flexible than a heterogeneous approach in the context of 4G Mobile Terminals.

I. Introduction

A. MPSoCs for LTE applications

The Third Generation Partnership Project (3GPP) has defined the Long Term Evolution (LTE) for 4G radio access. Emerging fourth generation cellular standards like LTE [1] require intensive modem signal processing. This standard involves high data rate, low latency, and relies on OFDMA/MIMO techniques with adaptive modulation. The baseband architecture requires dynamic reconfigurations due to user resource allocation, a high computational demand and low power requirements under real-time constraints. A LTE mobile handset has to meet a 100 Mb/s data rate, processing 100 GOPS with a power budget of 1W. The demodulation stage is characterized by a 10 GOPS workload with a budget around 200mW [2]. One major challenge therefore relies on devising an architecture that meets the flexibility, power and area requirements.

Traditional telecom chipset are designed with dedicated hardwired solutions which are cost-ineffective for multi-standard mobile handset. For more flexibility, MPSoCs (Multiprocessor System-on-Chip) [3] with multiple programmable processors as system components have been introduced in the telecom field. MPSoCs are well suited for systems with concurrent algorithms like telecom applications. The implementation of such algorithms on the heterogeneous FAUST [4] and MAGALI [5] platforms has proven the efficiency of MPSoCs to provide a valuable solution. MAGALI is a heterogeneous architecture based on mostly specialized programmable cores, and specific configurable hardware accelerators.

In the context of SDR platforms, the flexibility and reconfigurability is the key challenge. Homogeneous MPSoCs, which are based on the replication of identical units, can better provide flexibility, fault tolerance and scalability. This attractive solution is challenging in the context of Mobile Terminals where area and power consumption overheads are very costly.

B. Related Works

Previous works have proposed architectures for LTE modem implementation. Multi-core architectures like Picochip [6], Infineon's MuSIC [7] or Sandbridge's SB3011 platform [8] argue high computational performance and high flexibility. They are homogeneous or heterogeneous DSP-centered and accelerator-assisted MPSoCs. But the power consumption of these solutions allows them to mainly target base-stations. Those solutions have an estimated power consumption in the order of magnitude of several Watts.

The MAGALI platform [5] is a heterogeneous NoC-based MPSoC for Mobile Terminal in 65nm low-power CMOS. This solution supports OFDMA/MIMO standards with a reduced power consumption. This platform offers a medium flexibility with a high reconfiguration speed to manage adaptive modulation and coding techniques. This approach is based on a centralized control processor which limits the scalability, the flexibility and the fault tolerance of the architecture.

C. Homogeneous approach for LTE applications

Earlier papers have presented the heterogeneous MAGALI platform [5] which is a state-of-the-art reference implementation in this study. In this paper, we present and benchmark two versions of the homogeneous processor array GENEPY, one with a host processor (v0) and one fully distributed (v1). We perform a rigorous comparison of the three different designs MAGALI, GENEPY v0 and GENEPY v1 in terms of silicon

978-3-9810801-6-2/DATE10 © 2010 EDAA

area, performance and power consumption. We show that homogeneous MPSoCs are not only competitive for base-stations but also for mobile terminals.

In this work, we address two issues throughout the two implementations. The first issue is to design a reconfigurable processing unit capable of supporting all algorithms of our LTE applications. The second issue is to implement fully distributed and flexible control mechanisms to support synchronisation and dynamic reconfiguration. For a rigorous comparison and accurate results, each solution has been placed and routed in 65nm low-power CMOS technology.

The rest of this paper is organized as follows. The reference heterogeneous design and the two proposed homogeneous MPSoC are discussed and compared in terms of silicon area in section II. In section III, we present a LTE demodulation application and compare performance and power consumption of our solutions as an application case study.

II. From Heterogeneous to Homogeneous Architectures

A. Heterogeneous reference architecture

The heterogeneous MAGALI platform supports OFDMA/MIMO TX/RX baseband algorithms. In this work, we focus more precisely on the RX modem, the most demanding in terms of computing power. We consider this heterogeneous design as our reference silicon prototype. Indeed it is a state-of-the-art 65nm telecom platform.

The architecture is composed of several cores running in parallel, interconnected by a network-on-chip (NoC) as shown in figure 1 . To perform demodulation algorithms, two different cores have been designed. The DSP unit is a VLIW 32-bit low-power DSP, optimized to handle complex numbers (16-bit I + 16-bit Q), including Complex MAC and Cordix operators. The MMC unit [9] is a Microprogrammable Memory Controller for intensive data manipulation involving synchronization, buffering, duplication and reordering. The platform is designed with DSP or MMC units as needed . The NoC, called ANoC [10], is a fully asynchronous system to exploit the Globally Asynchronous Locally Synchronous concept. All units are plugged on the NoC using a Network Interface (NI). The NI is in charge of packetization, depacketization and flow control using credits, handled by Input/Output Communication Controllers (ICCs and OCCs). The platform control (scheduling and configuration) is semi-distributed. The global control is performed by a host processor, using direct addressing and interrrupt mechanisms. A local Configuration and Communication Controller (CCC) [11] handles the configuration data transferts, the storage of micro-programs for the input communications, output communications and the core processing, and the local scheduling of the NI and the core. Figure 2 shows the resulting unit architecture.

Depending on the application requirements, the platform comprises mostly Mephisto and MMC units. This architecture is not fully scalable, the host processor has to keep a global view over the control. This situation induces a bottleneck and can interfere with real-time constraints.

B. Homogeneous Processor Array with a host processor

To go towards a homogeneous design, we need to support data manipulation and data processing on the same unit. We propose a new unit, called SMEP (v0), integrating a Smart Memory Engine (SME) and a processing cluster with two DSPs. The SME handles four logical buffers mapped on a same 32KB local memory (RAM data). The buffers (size, base pointer) are dynamically configurable to fit applicative needs and are managed as circular buffers for data-flow operations. Data manipulation on the four buffers are performed by four attached Read Processes (RP). A Read Process executes microinstructions to read data from a buffer and therefore generates read addresses, writes data to a specific target and handles synchronizations between Read Processes. The write target of RPs can be either the Network Interface to access other units, another buffer in the local memory, or one of the two MEPHISTOs in the processing cluster. Communications inside the SMEP are supported by a 6x6 crossbar. This interconnect can handle 6 parallel 32-bits transfers at 400MHz, i.e. a 77GBits/s bandwidth. This bandwidth is achieved using a local memory divided into 4 physical RAM banks with interleaved access based on the least significant address bits. The crossbar is configured at each clock cycle based on RPs requests.

In the processing cluster, each DSP reads incoming data from an input FIFO, the intermediate processing values are stored in a local memory and the results are written into an output FIFO. The output data-flow is read from the output

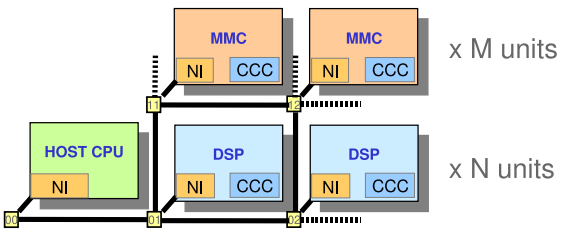

Fig. 1. Heterogeneous MAGALI architecture

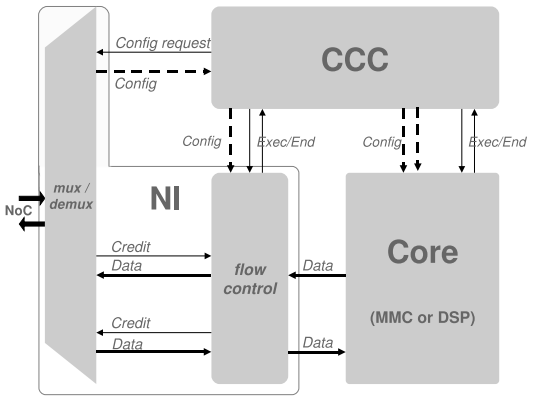

Fig. 2. Heterogeneous unit with DSP or MMC core

FIFO by a Read Process in the SME. The datapath has been optimized to perform intensive computing on a data-flow with a minimal power budget. Each DSP can perform 4 parallel 16-bits multiplications at 400MHz, i.e. 3.2 GMAC/s for the processing cluster.

This elementary unit of the processor array is highly programmable. The computing is DSP-based to provide the software flexibility. The SME block with micro-programmable Read Processes enables completely sofware-defined data manipulation.

In this architecture, presented in figure 3, reconfiguration and scheduling is performed by a Communication and Configuration Controller (CCC) as previously in the heterogeneous platform. So the computing is homogeneous, but the control is still shared between a host processor and the local CCC.

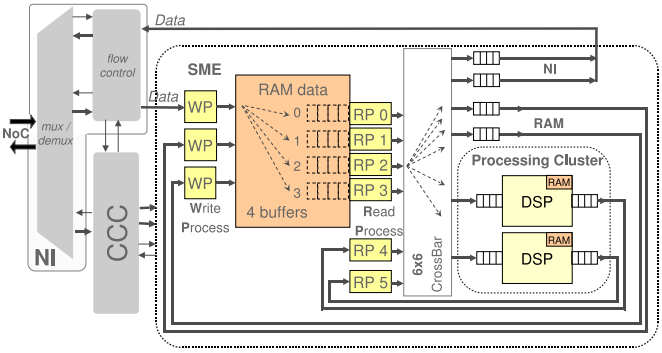

Fig. 3. Elementary unit (SMEP v0) of the homogeneous processor array with host processor

The implementation details on the SME and the DSP are not in the scope of this paper. Compared to the heterogeneous MAGALI, the SME block has the same data manipulation functionnality as the MMC unit. In this work, they can be considered as equivalent.

Based on the SMEP v0 unit, we define the homogeneous processor array GENEPY v0 with a host processor and N SMEP units interconnected by an asynchronous NoC. As we keep the same control mechanisms as in the heterogeneous approach, the host processor is still mandatory (see figure 4). The main advantages of this solution is a reduced NoC size (number of routers) for the same computing power. Actually, one SMEP unit is equivalent to one MMC plus two DSPs. But still, the global control through a host processor limits the scalability of the platform.

C. A fully Homogeneous Processor Array

A homogeneous platform is defined with a single unit's type instanciated several times. Each unit has to manage its processing resources, but also its configuration and its scheduling. We define a second architecture from the previous one, called SMEP v1. In this implementation, we keep the same processing and data management blocks as before and we change the control block. The architecture with a host processor and a local Configuration Communication Controller

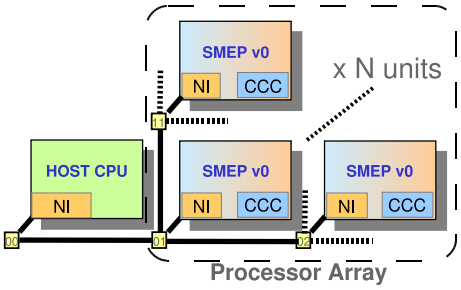

Fig. 4. GENEPY v0 platform

has evolved to a fully distributed control using a Control Processor as shown in figure 5 .

In a telecom platform, the scheduling is often quite complex due to dynamic modulation schemes. To support the LTE standard and even future standards, it is valuable to have more flexibility especially in a homogeneous approach to explore load balancing, fault-tolerant mechanisms, task migration, etc. The Control Processor is a 32-bit MIPS processor which manages dynamic reconfigurations, real-time scheduling, synchronizations. The CPU has several extensions to improve its efficiency:

- Input/Output extension to manage a control flow between units
- Timer extension to handle real-time constraints.
- Configuration handler to improve reconfiguration speed.

The Control Processor manages the NI, the SME and the processing cluster. Contrary to the CCC, the management with Control Processors is software-based. Therefore, the flexibility and the autonomy of each unit is increased, there is no need for a host processor.

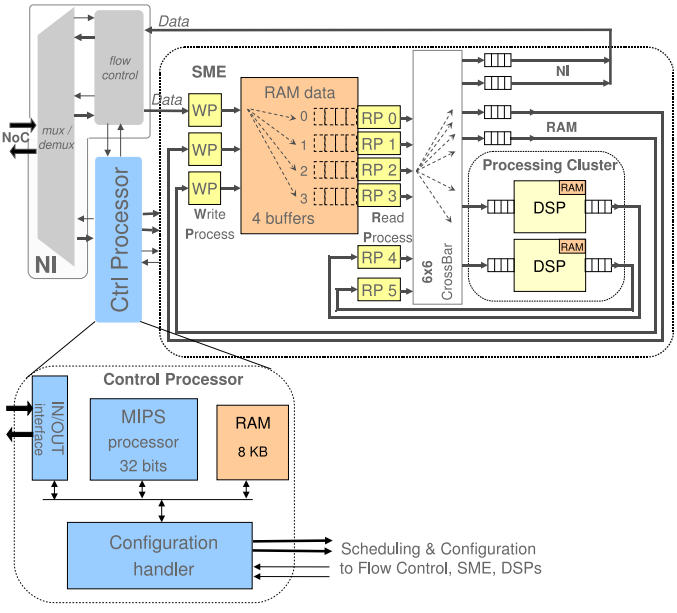

Fig. 5. Elementary unit (SMEP v1) of the fully homogeneous processor array

So based on this SMEP v1 unit, we have designed the homogeneous platform GENEPY v1 with only SMEP v1 units interconnected by a NoC (figure 6) .

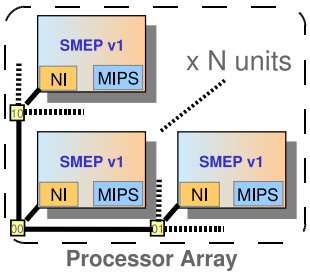

Fig. 6. Fully homogeneous processor array : GENEPY v1

At this stage, the platform is homogeneous and fully distributed, the SMEP unit can be replicated as needed to increase the computing capacity of GENEPY.

D. Design Results

The two SMEP units presented previously have been implemented at RTL level (VHDL). We compare these two designs with the equivalent in the heterogeneous platform as shown in figure 7 . In terms of functionnality and capacity, this three architectures are equivalent:

- 3 routers, 2 DSPs and 1 MMC
- 1 router and 1 SMEP v0
- 1 router and 1 SMEP v1

To perform a relevant comparison, we have kept the same characteristics (amount of memory, frequency, FIFO sizes, etc) in all implementations and prefered ignoring the silicon impact of the host processor in this study. The silicon area is extracted after logic synthesis with 65nm low-power CMOS technology at 400MHz. All designs include test mechanisms like scan chains and memory BIST.

As shown in table I, we save 17% silicon area between the reference model and the SMEP v0 unit. This result is mainly due to a reduced number of needed routers and the associated Network Interfaces. All communications are moved from the NoC to a local interconnect, defined by a 6x6 dynamic crossbar. The implementation of the control using a MIPS processor increases the silicon area by 4% comparing SMEP v0 between SMEP v1. This overhead is negligible compared to the added flexibility and is compensated by the absence of a host processor.

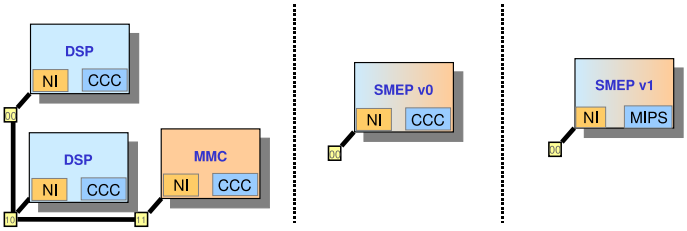

Fig. 7. Equivalent architectures

TABLE I

SYNTHESIS RESULTS - TECHNO ST65NM - 400MHZ

MAGALI	area (mm^2)	occurence nb.
Router (NoC)	0.159	3
DSP + CCC + NI	0.483	2
MMC + CCC + NI	1.323	1
TOTAL	**2.766**	

SMEP v0	area (mm^2)
Router (NoC)	0.159
2 DSPs	0.792
SME	1.226
NI + CCC	0.126
SMEP v0 + *NoC*	**2.303**
Area comparison	**- 17%**

SMEP v1	area (mm^2)
Router (NoC)	0.159
2 DSPs	0.816
SME	1.216
NI + MIPS	0.201
SMEP v1 + *NoC*	**2.392**
Area comparison	**- 14%**

III. POWER AND PERFORMANCE ANALYSIS FOR LTE MODEM

A. Reference LTE application

This study focuses on the downlink part of the LTE standard and more precisely on the demodulation side. Using the terminology defined in [12], data are transmitted in 10ms frames equally divided in 10 sub-frames also called TTIs (Time Transmission Intervals), *i.e.* the TTI aligns on sub-frame and equals 1 ms. The system is designed to transmit on 4 antennas and to receive on 2 antennas, wich requires a high performance processing, because of the implementation of diversity and spacial multiplexing schemes.

Our reference application is composed of 5 tasks (figure 8):

- 2 Channel Estimation Modules, one for each RX antenna based on Wiener filtering.
- 2 interpolation algorithms of the channel coefficients over the whole bandwidth.
- 1 MIMO MMSE decoder that implements a 4x2 double-Alamouti algorithm.

The modulation scheme depends on the user resource allocation. The application defines five operating modes from a low-quality (QPSK), low data-rate transmission to a high quality (64-QAM), high data-rate transmission. For this study, we

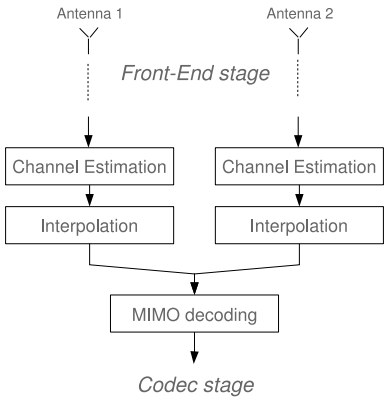

Fig. 8. Test-case LTE application

have considered the most demanding scenario; the computing demand requires :

- For heterogeneous MAGALI: 1 host processor, 2 MMC and 4 DSP units
- For GENEPY v0: 1 host processor and 2 SMEP v0 units
- For GENEPY v1: 2 SMEP v1 units

In each platform, we use the equivalent of 4 DSPs to execute the 5 tasks. Then to meet real-time constraints, the MIMO algorithm is duplicated on two DSPs. So, 6 functions are supported by 4 DSPs. Each function requires a set of reconfigurations to communicate data through the Network Interface and the SME. The control blocks (CCC or MIPS) can deal with dozens of reconfigurations and scheduling phases to process a TTI.

The mapping of the application on the three platforms is not discussed here but for comparison purpose, we chose an equivalent mapping that fulfills the hard real-time deadline.

B. Performance results

The LTE modem application is mapped onto all three different platforms from heterogeneous to fully homogeneous. The simulation environment integrates two data generators that emulate the incoming data-flow from the two antennas. To increase simulation speed, the NoC is modeled in TLM SystemC using post-layout parameters. All units of the three platforms are modelized at RTL level to provide cycle-accurate results. At the end, a System-C unit, called a recorder, records a trace of the output data-flow and compares it to a reference file to guarantee the righ execution.

The performance figures are summarized in table II. They match the processing time of a complete TTI including scheduling and reconfiguration phases. Each unit runs at 400 MHz clock frequency. Comparing MAGALI with GENEPY v0, the design of a homogeneous processing unit speeds up the execution time by 3%. In this case, our challenge was to keep at least the same performance to guarantee hard real-time constraints. This result proves that homogeneous processing unit in the LTE modem stage is as efficient as a heterogeneous implementation.

From GENEPY v0 to GENEPY v1, the scheduling and reconfiguration managements have been transferred from a dedicated hardwired controller to distributed software programs on MIPS processors. By efficiently using the Control Processor, there is no time overhead due to the software management.

These results, from RTL simulation, show that a homogeneous platform with a highly flexible control is as efficient as the heterogeneous MAGALI platform.

TABLE II
TIME TO PROCESS A TTI

	Heterogeneous MAGALI	Homogeneous with host Proc. GENEPY v0	Fully Homogeneous GENEPY v1
execution time	551 μs	531 μs	530 μs
perf. speed up	**–**	**+ 3 %**	**+ 3%**

C. Power consumption results

To provide a full comparison, we have evaluated the power consumption of the three implementations. Each plaform has been placed and routed in 65 nm low-power CMOS technology. We have simulated a complete TTI processing with the placed and routed netlist. Table III presents the average power consumption of the three platforms at gate-level. We has chosen to ignore the consumption of the host processor for MAGALI and GENEPY v0 platforms. From MAGALI to GENEPY v0, the power savings are around 3%, which achieving a speedup of the same order. As two SMEP v0 units are sufficient for the whole application, the power consumption of the control blocks (NI + CCC) is reduced by 60% comparing to MAGALI with 6 units. So at platform level, GENEPY v0 is 10% more efficient than MAGALI.

On GENEPY v1, we have implemented a better control on the MIPS processor that saves 5.5% of the power consumption on processing and data manipulation. Concerning the power consumption on the control block, an automatic sleeping mode on the CPU has saved 47% energy compared to the CCC block. To perform scheduling and reconfiguration, the MIPS processor is only active 3% over the time of TTI period. So, the homogeneous GENEPY v1 is 18% more efficient than MAGALI.

Figure 9 shows a detailed view of the power consumption profile during a TTI processing, with separate contributions for processing (DSP), data reordering (SME) and control (MIPS) for the two units of GENEPY v1.

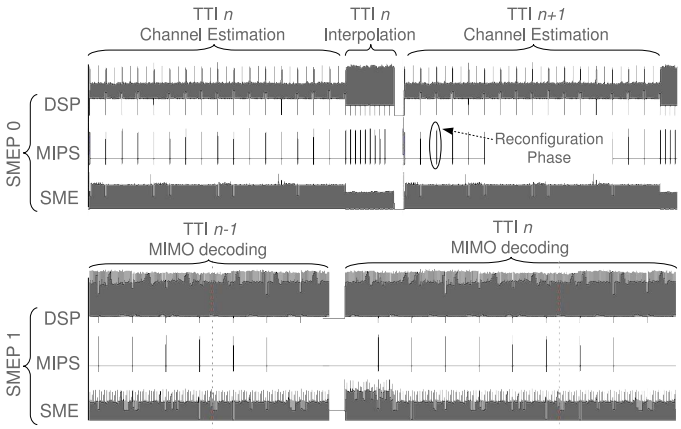

Fig. 9. Power Consumption Profile of GENEPY v1

TABLE III
DETAILED POWER CONSUMPTION TO PROCESS A TTI

	Heterogeneous MAGALI	Homogeneous with host Proc. GENEPY v0	Fully Homogeneous GENEPY v1
Processing	124.3 mW	121 mW	113,7 mW
Data Manipulation	74.8 mW	74.3 mW	70.8 mW
Control	34.5 mW	13.7 mW	7.2 mW
Total	**234 mW**	**209 mW**	**192 mW**
Power Consumption	**–**	**- 10%**	**- 18%**

D. Result Analysis

It is well-known that heterogeneous architectures are more efficient than homogeneous ones in general case. But in this study, we have shown that the homogeneous solution is 3% faster than the heterogeneous approach. We have design a homogeneous unit which substitutes at least 3 heterogeneous units. This architecture enables to drastically reduced the NoC sizes and the number of needed Network Interfaces. This choice leads to a reduced silicon area and also a limited number of NoC communications. NoCs offer a good flexibility but they also add a communication overhead.

A trade-off has been found between multiple small heterogeneous units and a more complex homogeneous one. Gathering functions in an optimized architecture reduce the power and performance overhead on communication and reconfiguration. The complexity of an homogeneous unit is limited by an exponential increase of the power consumption and a reduced maximum clock frequency .

To implement a fully homogeneous platform, we have substituted a dedicated controller by a MIPS processor in each unit of the processor array. For a complex unit, this MIPS processor adds a small silicon overhead around 4% compared to a dedicated controller. This Control Processor is more flexible and can manage resources more efficiently to save energy. For data flow applications with high computing demand, the scheduling is critical to guarantee real-time constraints but can be executed in parallel of the computing task. In our study, MIPS processors are only active 3% of the time to reconfigure and schedule the application, the power consumption of the control part is only 2.5 mW per unit in the processor array. In that case, a MIPS processor with very short active periods is more energy efficient than a dedicated programmable controller, active all the time to cope with a host processor.

IV. Conclusion

We have presented the GENEPY platform, a low-power homogeneous MPSoC for 4G Mobile Terminals. The major component is the SMEP unit, able to provide data manipulation at 77 GBits/s and computing at 3.2 GMAC/s at a 400MHz operating frequency. Due to separate data handler and data processing blocks, this architecture is efficient and configurable. The control over the platform is fully distributed on MIPS processors: this solution is highly flexible and scalable. Compared to the reference heterogeneous MAGALI platform, the homogeneous GENEPY platform is about 14% smaller in terms of silicon area. For a LTE application, the comparison has shown that we achieve a performance speed-up around 3% with a power saving of 18%. We have highlighted that homogeneous MPSoC approaches make sense for future Mobile Terminals.

Our future research efforts comprise the enhancement of the SMEP unit towards a better power management as well as an extended support for FFT and frame synchronization algorithms. As the MIPS processor is used only at 3% of its capacity, we will explore its use to support distributed power management algorithms, task migration, fault tolerance, load balancing, etc.

References

[1] J. Berkmann, C. Carbonelli, F. Dietrich, C. Drewes and W. Xu, "On 3g LTE Terminal Implementation - Standard, Algorithms, Complexities and Challenges," *Wireless Communications and Mobile Computing Conference*, IWCMC '08, pp. 970-975, Aug. 2008

[2] C.H. van Berkel, "Multi-Core for Mobile Phones," *Design, Automation & Test in Europe Conference (DATE'09)*, pp. 1260-1265, Apr. 2009

[3] W. Wolf, A.A. Jerraya and G. Martin, "Multiprocessor System-on-Chip (MPSoC) Technology," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 10, pp. 1701-1713 , Oct. 2008

[4] D. Lattard *et al.*, "A Reconfigurable Baseband Platform Based on an Asynchronous Network-on-Chip," *Journal of Solid-State Circuits (JSSC)*, vol. 43, pp. 223-235, Jan. 2008

[5] F. Clermidy, R. Lemaire, X. Popon, D. Kténas and Y. Thonnart, "An Open and Reconfigurable Platform for 4G Telecommunication: Concepts and Application," *Euromicro Conference on Digital System Design (DSD)*, Aug. 2009

[6] D. Pulley, "Multi-core DSP for Base stations: Large and Small," *Design Automation Conference (ASPDAC'08)*, pp. 389-391, March 2008

[7] U. Ramacher, "Software-defined radio prospects for multistandard mobile phones," *Computer*, vol. 40, no. 10, pp. 62-69, 2007

[8] J. Glossner *et al.*, "The sandbridge sb3011 platform," *EURASIP J. Embedded Syst.*, vol. 2007, no. 1, pp. 16-16, 2007

[9] J. Martin, C. Bernard, F. Clermidy and Y. Durand, "A Microprogrammable Memory Controller for High-Performance Dataflow Applications," ESSCIRC '09, Sept. 2009, in press

[10] Y. Thonnart, E. Beigné and P. Vivet, "Design and Implementation of a GALS Adapter for ANoC based Architectures," *IEEE symp. on Asynchronous Circuits and Systems (ASYNC '09)*, pp. 13-22, May 2009

[11] F. Clermidy, R. Lemaire, Y. Thonnart and P. Vivet, "A Communication and Configuration Controller for NoC based Reconfigurable Data Flow Architecture," *ACM/IEEE Symp. on Networks-on-Chip (NoCS '09)*, pp. 153-162, May 2009

[12] 3GPP TSG-RAN, 3GPP TS36.211, *Physical Channels and Modulation (Release 8)*, v8.1.0 (2007-11)