3.1 Publishable summary

The NANOC project aims at developing an innovative design platform for future network-on-chip (NoC) based multi-core systems. Strict component-oriented (and reuse) architectural design is targeted in the project, which enables faster design methods. The component-oriented method is out of reach of current design methods and tools due the introduction of the NoC concept in new and future designs and mostly because of new design constraints. Enhanced dynamism and flexibility in NoC composition is required to tackle the new requirements of future systems like virtualization, power management, thermal management, and application management. In addition, the continuous scaling of technology opens the door to new challenges like reliability and variability. Reliable systems will need to be designed out of unreliable components. NANOC provides design methods and prototype tools to address such challenges and embeds them into the final design platform taking the NoC as an additional component in the system, thus enabling the componentization approach. NANOC is built around a vertical design approach where system, circuit and processes are co-developed with silicon-aware decisions considered at every layer. A common exchange format (CEF) standard is co-developed in the project (and intends to span out of the project) to enable communication among all layers of the design platform.

The project builds around five technical WPs, each one with a particular and contributing goal to the final target of the project. The first two WPs deal with design methods to tackle the new challenges that systems are facing (manufacturing faults, process variation, run-time network management …). WP3 and WP4 deal with tools for the final platform at the front-end (logic level) and back-end (physical level), respectively. WP5 targets the final integration of both design methods and tools into the final design platform. During the first two years of the project, workpackages WP1 through WP4 have been almost completed, leaving the project ready for the implementation of WP5. These WPs address four main aspects:

- Development of design methods for supporting static irregularities in NoCs (WP1)
- Development of design methods for supporting dynamic irregularities in NoCs (WP2)
- Development of frontend NoC topology synthesis tools (WP3)
- Development of backend physical level design tools for NoC composability (WP4)

Design methods, developed in WP1, have been focused on tolerating irregularities in the NoC due to manufacturing defects. As technology scales down, the probability of manufacturing defects increases, affecting yield and thus manufacturing costs. There is an urgent need to tolerate such defects and, by that, increase the yield. In the two years, design methods have been designed to tolerate such defects in the NoC. In particular, full coverage of manufacturing defects has been achieved by the different LBDR and FDOR design methods developed in WP1. Both unicast (one to one traffic) and multicast/broadcast (one to many/all traffic) have been supported in the presence of any failure pattern in an initial regular mesh network. For broadcast and multicast communication, SBBM and FJE mechanisms have been developed, on top of LBDR and FDOR unicast counterparts. All these methods have been conceived taking into

account the area and power design constraints present in NoCs. An extension to FDOR (called iFDOR) to support dynamic irregularities (e.g., those caused by temporary shutdown of network components) has been designed. In addition, coherency protocols have been explored in WP1 in connection with unicast and broadcast routing mechanisms. New means of communication (gather operations) have been analyzed and rough implementation solutions have been provided. Invalidation-based, update-based, directory-based, and broadcast-based protocols have been evaluated and adapted to the underlying routing mechanisms developed within the project.

The effect of manufacturing process variability has a fundamental impact on NoC performance. It affects the maximum frequency that different parts of the system interconnect can achieve. An investigation regarding this effect has been performed for several existing and future process nodes. A new variability modelling framework has been elaborated, which is well suited for on-chip interconnection networks and for the system-level assessment of variability effects based on their characterization at a lower abstraction layer (physical links and logic gate delays). This provides valuable understanding of the impact of variability looking forward to advanced technology nodes, and helps to design architectural level techniques that reduce the uncertainty of post-silicon performance. Indeed, two strategies (phit reduction and space multiplexed channels) have been considered to address this issue. After a first feasibility analysis, the phit reduction mechanism has been implemented. The impact of process variation on routing in NoCs has also been covered. Indeed, a variability-aware process mapping algorithm has been designed to improve the overall chip multiprocessor (CMP) performance. This variability-aware process mapping algorithm has been enhanced so that frequency and voltage islands are considered.

Finally, static irregularities make it mandatory to engineer a testing infrastructure for NoCs. This is the challenge that WP1 tackled, by exploring the design space of testing strategies for NoCs. WP1 developed a built-in self-testing and self-diagnosis framework for NoCs relying on the principle of cooperation between switches for the sake of low test application time. In this context, the overhead vs. the coverage trade-off provided by different kinds of test patterns (deterministic vs. pseudo-random) was investigated. The approach was further extended to multi-synchronous NoCs, thus posing the foundation for a cost-effective, scalable and flexible testing framework for NoCs providing industry-standard coverage figures.

In WP2, further design methods, complementing those in WP1 have been explored. In this case, the focus was on dynamic methods capable of taking the proper course of action in the presence of low-level error detection, virtualization and power management strategies as well as for congestion management. The core of WP2 is a dynamic network reconfiguration framework which encompasses an hardware signaling infrastructure and a software algorithm for routing function reconfiguration. A dual-network was designed for control signaling between network switches and a global controller and vice versa. Low-overhead and fault-tolerance make such signaling infrastructure a suitable backbone to convey diagnosis and routing reconfiguration bits, intermittent fault notifications, debugging and congestion control information. As a result, the system was proved capable of full reconfiguration within a few hundred cycles. The developed mechanism builds on top of LBDR, thus being fully compatible with the routing engine of the project. This is an important achievement since will impact the final platform and will allow a truly dynamic system solution. NoC congestion has been extensively studied, both with respect to a shared CMP scenario, as

well as for bursty traffic in MPSoCs. For both scenarios solutions have been developed that allow high NoC utilisation regardless of adverse traffic patterns.

WP3 has seen large progress in the first two years of the project, and is now almost complete. An important initial contribution has been the release of CEF (Communication Exchange Format) 1.0, a file format specification that allows interconnect designers to describe their constraints and the outcomes of their work. The format is meant to be machine-consumable and to allow the interoperability of tools developed within NaNoC and outside. The main capabilities of the format allow for the description of:

- Frequency and voltage domains in which the chip is partitioned
- IP cores belonging to the design, and to be interconnected
- Communication requirements, including communicating pairs of cores and required performance
- Architecture of the chip interconnect
- Communication routing
- Physical properties of the design, e.g. basic floorplan, wire length
- Additionally, the file format is designed to be extensible with custom tags, so that each user of the format can customize it with additional specifications.

Starting from this platform, WP3 built on two main toolchains. The first one is a tool developed by Lantiq- for evaluating needs in NoC Design Space Exploration (DSE). Simulation is key for NaNoC, as it allows a preliminary evaluation of the different design methods under many different situations (traffic changes, topology changes). With the DSE tool, the huge design space is quickly and accurately cropped thus reducing tremendously design time and enabling the study impact of specification details over the overall design. The tool includes simulation facilities and floorplan awareness to accelerate turnaround times.

The other toolchain being leveraged and extended in WP3 is the iNoCs toolchain, which also performs design space exploration, but instead of primarily leveraging simulation of a given set of topologies, it focuses on the synthesis of optimal NoCs.

During the first year of NaNoC, the iNoCs tools were given the ability to support vertically stacked ("3D") chip designs, instantiating in this case NoCs with vertical communication links. Vertical connectivity is tuned to optimize multiple metrics, such as performance, power consumption, chip area and chip yield. The tools have been demonstrated to support four or more layers, as demanded by projections for 2015 designs. Routing design methods developed in the project (LBDR and FDOR in WP1) have also been evaluated for the use in 3D stacked chips as well.

With the second year, important steps were taken to improve the applicability of the iNoCs toolchain. First of all, they were made aware of multi-frequency designs, with the ability to account for use cases in which components may run in high-performance modes (with correspondingly high performance requirements on the NoC) as well as in low-power modes, in varying combinations. The NoCs are now generated to support such traffic loads and they automatically feature frequency conversion blocks where needed. In parallel, the iNoCs library of Network Interfaces that act as adaptors among the IP cores and the NoC was significantly extended, allowing for easy and

interoperable pluggability of blocks with AMBA AHB, AMBA AXI and OCP interfaces. These results are essential to approach the goal of transparent IP reuse.

Additionally, work has been performed to account much more comprehensively for physical-level effects and in particular for wire lengths depending on the system floorplan. On one hand, the iNoCs toolchain was again leveraged and extended, making it capable of optimizing topologies for the floorplan (which is an optional input) and of segmenting wires as needed with pipeline stages. On the other hand, a comprehensive study on regular topologies, e.g. meshes, tori, etc., was performed, including logic synthesis, placement and routing. This study aimed at accurately understanding and characterizing the advantages and disadvantages of different topologies in view of physical properties among which wire length is, again, the principal. The two parts of this work are comprehensive and complementary since they target MPSoCs (mostly heterogeneous systems) and CMPs (mostly homogeneous systems) respectively.

In WP4, physical level issues of the NANOC design platform are addressed. Prototype tools and methods are developed to address issues relating to tying higher levels together with the lower levels of design abstraction, to avoid long iterations between frontend and backend and ensure convergence during physical implementation. The second year of the project has seen this WP close to completion.

A number of relevant contributions have been made:

- Analytical models of process variations and their application in NoC-wide performance variability maps have been developed.
- A prototype tool for improving SoC-level hold-time margins has been developed.
- A model and analysis engine for early-stage SoC dynamic power integrity analysis has been developed.
- A prototype tool for CEF-driven floorplanning in the context of NoC communication requirements has been developed.
- An exploration of the layout-level effects of NoC link implementation has been undertaken.

Together, these contributions provide a physical-level awareness in the NaNoC design platform that builds beyond the capabilities of existing backend tools, complementing current design flows to provide significant improvements in the backend convergence challenges related to NoC-based systems.

In the first year, a tool for improving hold-time robustness was developed and analytical process variability models were explored. These contributions have been described in detail in the first year progress report as well as in deliverables D4.1 and D4.3 respectively.

In the first year, the work on a tool for early-stage dynamic power grid analysis was also started. This work was successfully completed during the second year, with the objective of fast turn-around-time being achieved. The prototype tool developed was shown to be much faster than an existing commercially available tool and the inner loop of the underlying model and analysis engine executes in under 50 ms per iteration. The

purpose of reaching this performance goal has been to integrate the power grid model directly into the floorplanning tool also developed in WP4, and combine these in a common deliverable D4.2.

In the second year, a prototype floorplanning tool which takes a CEF input describing system-level blocks and abstract communication requirements was developed. The tool optimizes a floorplan with the objective to place blocks with high degree of communication closely together. The tool outputs an updated CEF, which can be used as input to the NoC synthesis tool of WP3. Interoperability was smooth due to the ground work done in defining the CEF standard. The floorplanner was integrated into an existing floorplan solution, to leverage support for standard design formats, visualisation, etc, and as such to enable interoperability with existing mainstream backend design implementation flows.

The power grid analysis models described above were integrated into the floorplanning engine, and as such allow the optimization of a SoC floorplan according to system-level communication requirements concurrently with power integrity.

Finally, the second year saw the completion of work started in the first year, relating to link inference techniques. The work aimed at constraining the non-determinism of post-place&route link statistics and at coming up with link-level area/performance trend, as the synthesis constraints and topology requirements are varied. Such trend was characterized for two major link inference techniques: repeater insertion and link pipelining. Furthermore, the role of non-routable obstructions was assessed on the derived results, thus accounting for the more or less limited routing channels available in actual chip layout for network link routing depending on the physical space budget. Moreover, the area/performance/power implications of utilizing such techniques were evaluated not just for the link in isolation, but for the network topology as a whole. This work resulted in deliverable D4.4.

As a summary of achievements for the two years of the project, we can list the following ones:
- Design and implementation of a static design method (LBDR) that achieves 100% coverage of manufacturing defects in system designs based on regular NoC topologies (deliverable 1.1).
- Support, with 100% coverage of manufacturing defects, for collective communication (bLBDR and SBBM design methods).
- Support for a complete reconfiguration mechanism able to reconfigure the network in hundreds of cycles without stopping traffic and without introducing deadlocks.
- Analysis of coherence protocols on top of LBDR approach, focusing on a variety of alternatives (invalidation-, update-, broadcast-).
- Development of a dual network for control signaling for use in many network functions (testing, fault-tolerance, debugging...).
- Development of a gather control network for efficient coherence protocol traffic handling.
- Development of BAHIA and BAHIA-2 mechanisms to remove the HoL blocking induced by bursty traffic.
- Development of an injection-limiting congestion management strategy.

- An application-specific routing methodology capable of providing quality of service guarantees with minimal routing, which is deadlock free in terms of both packet and protocol deadlocks. Design of a dynamic design method (iFDOR) to adapt the NoC to sudden irregularities (deliverables 1.1 and 1.5).
- Tools to design NoCs for vertically stacked ("3D") chips having four or more layers (deliverable 3.1).
- Support in 3D chips for LBDR design method (deliverable 3.1).
- Initial multiple DVFS domain support in tools.
- Design of a tool-chain for Design Space Exploration to highly reduce exploration time (deliverable 3.3).
- Development of the first version of the Communication Exchange Format (CEF) standard to allow all the NANOC tools to exchange data in a common standard format (deliverable 3.4).
- Development of a system-level power grid modeling tool for SoCs enabling dynamic power grid performance at an early design stage and with fast turn-around-time.
- Development of an accurate variability model that takes into account the main sources of process variation (deliverable 4.3).
- Prediction of what the network link contribution will be to total area figures depending on the constraints of the physical synthesis (deliverable 3.3).
- A report on variability-aware network design mechanisms (deliverable 1.3).
- Extension of a library of Verilog Network Interfaces for multiple protocols (AMBA AHB, AXI, OCP) and a common packet format, allowing integration of IP cores with different interfaces in a platform. The area occupation can be as low as 0.01 mm^2 and operation at or above 500 MHz is achieved (deliverable 3.2).
- Extension of the INOCS NoC synthesis tools to generate NoCs suitable for multi-frequency, multi-voltage systems, including the instantiation of appropriate converters at the frequency island interfaces (deliverable 3.2).
- Extension of the INOCS NoC synthesis tools to generate NoCs with floorplan awareness, minimizing total wire length and pipelining wires where needed, accounting for simplified routing models (deliverable 3.3).
- Developments of a CEF-driven floorplanning engine that floorplans according to system-level communication requirements, as well as system-level power integrity (deliverable 4.2).
- Report on exploration of link inference techniques for physical NoC link design (deliverable 4.4).

All the work performed in these two years of the project is in line with the initial schedule and no major deviations and roadblocks have been found. Therefore, the partners expect the best achievements obtained will be fully integrated in the final design platform in the third year of the project.

In WP6 an effort has been made to forecast the future use of the project's results. There are mainly three exploitation categories, which are feasible for such a use. The first is the possible use within academia. In other words, the results can be used in lectures and courses offered, for example, first by the two project's partners University of Valencia and University of Ferrara. Furthermore, there are two possible uses within the semiconductor industry. The first one is by the CAD tools providers, which will make use of the algorithms and methodologies resulting from the project. The second use is

directly by the semiconductor manufacturers, where they will apply designs and structures, which are made available through the project.

European universities and companies are still competitive in research and development of complex systems-on-a-chip. Especially the use of ESL tools is accepted more openly in European companies compared to the US. The design tools of the NaNoC project may help to keep this edge throughout the next years. This improved productivity is vital to be able to achieve first-time-right SoC designs with more and more processors and other components communicating over more complex interconnect solutions facing more variations caused by shrinking process geometries. Such tools help to do such designs cost-effectively in European countries despite of their higher cost of labour. This has a double effect on economy and society: first, the directly associated jobs can be kept in Europe, and second, a lot of related jobs will remain in Europe as well in industries that have a growing part of electronics contributing to their success (e.g. automotive, medical equipment, industrial control).

With respect to the dissemination effort of the consortium during the two years, 28 publications have been published mostly in highly-rated and peer-reviewed conferences (e.g. DATE conference, NOCS conference) and further 6 publications have been already accepted in the third year. Two books have been edited with six chapters involving the NaNoC research, thus potentially reaching a larger audience. Several press publications and press interviews have been performed by UPV and UNIFE. Poster presentations of the project have been held also in important venues like the HiPEAC network of excellence and contributions to a summer school have been provided. A promotional video was also prepared and posted on Youtube, in the NaNoC website and in CORDIS. The dissemination effort can be considered as high.

The project consortium maintains a website (http://www.nanoc-project.eu) with the most recent achievements and the list of public deliverables. All the NANOC-related info can be found at the website.