The Diopsis Multiprocessor Tile of ShApes

Pier Stanislao Paolucci

Technology Director
ATMEL Roma Advanced DSP

Permanent Staff Researcher (part time)
Istituto Nazionale di Fisica Nucleare
Roma – Italy

European Project Coordinator

Contact me at

pier.paolucci@atmelroma.it, pier.paolucci@roma1.infn.it

Abstract

- Nanoscale systems on chip will integrate billion-gate designs. The challenge is to find a scalable HW/SW design style for future CMOS technologies. A first problem is wiring, which threats Moore's law and prohibits monolithic architectures. The second problem is the management of the design complexity, which requires the reuse of smaller building blocks.
- Tiled architectures suggest a possible path: "small" processing tiles connected by "short wires".
- A typical SHAPES tile contains a mAgicV VLIW floating-point DSP (designed by Atmel Roma), a RISC, a DNP (Distributed Network Processor designed by INFN), distributed on chip memory, the POT (a set of Peripherals On Tile) plus an interface for DXM (Distributed External Memory).
- The SHAPES routing fabric connects on-chip and off-chip tiles, weaving a distributed packet switching network. 3D next-neighbours engineering methodologies is adopted for off-chip networking and maximum system density.
- The SW challenge is to provide a simple and efficient programming environment for tiled architectures.
- SHAPES will investigate a layered system software, which does not destroy algorithmic and distribution info provided by the programmer and is fully aware of the HW paradigm.
- For efficiency and QoS, the system SW manages intra-tile and inter-tile latencies, bandwidths, computing resources, using static and dynamic profiling. The SW accesses the on-chip and off-chip networks through a homogeneous interface.

Multi Processor Systems on Chip: Embedded System versus Personal Computer

- $ and # of embedded processors / persons increasing faster than conventional processors / persons
  - # of (phones, games, pdas, cars, home, medical, wearable) vs PC

- Collision/convergence on architectures is going to happen:
  - Because of changes on key driving markets
  - Because full systems can be integrated on a chip
  - Because of deep submicron technological facts:
    - WIRING,
    - COMPLEXITY,
    - POWER

Deep Sub-micron Architectures…

- ~160 MGate available on a 100 mm2 chip (45nm CMOS, 2008)
- Increasing GATES/CHIP vs Design Complexity Mngmt: embedded processors use a few million gates only, IP reuse possible;
- WIRING threatens Moore's law:
  - Wiring delay increases on new CMOS silicon generations
  - The full chip cannot be reached in a single clock cycle
  - Classic monolithic processor architectures do not scale
  - Locally Synchronous, Globally Asynchronous needed
  - Communication Centric SW and HW Architecture needed
- POWER DISSIPATION density approaching prohibitive values if high clock speed used; much better Oper/Watt at moderate clock (the human brain performs at 50 HZ!) (more details later…)

- … PROPOSED SOLUTION … TILED ARCHITECTURE…. HOW TO PROGRAM? … QUEST OF BEST TILE, ON-CHIP AND OFF-CHIP INTERCONNECT

The SW challenge of Tiled Architectures

- Long delays between distant tiles
- Hot Spots in communications
- Facilitate expression of parallelism
- Express real time constraints
- Avoid destroying information about available algorithm parallelism
- Compilation chain must fully aware of key architectural parameters: bandwidth, computational power, pipeline and latencies
- Exploit memory locality – efficient management of Distributed Memories
- Reduce RTOS overhead
- Networked RTOS
- Capture scalability in a library of characterized sw components
- Support for (semi)-automation of iterative design over HW, SW, Appl
- Monitor quality and real-time constraints
- Simulation speed of multi-tiled architectures

HW Background; Istituto Nazionale Fisica Nucleare
APE family of Massive Parallel Processors
custom Very Long Instruction Word Floating-Point Processors
and 3D first neighbour toroidal communication

INFN Istituto Nazionale di Fisica Nucleare

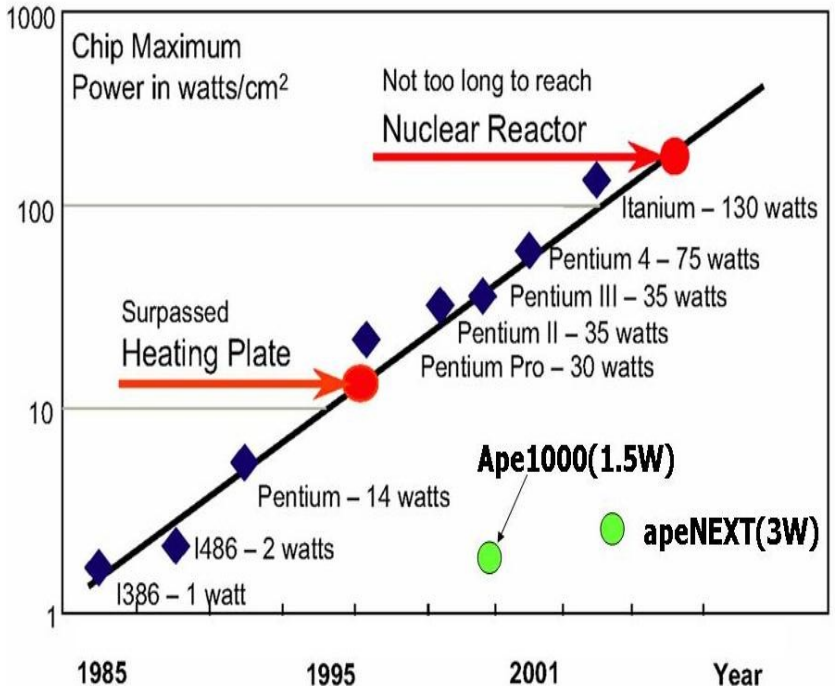
	APE (1984-1988)	APE100 (1988-1993)	APEmille (1994-1999)	apeNEXT (2000-2005)
Architecture	SIMD	SIMD	SIMD	SIMD++
# nodes	16	2048	2048	4096
Topology	flexible 1D	rigid 3D	flexible 3D	flexible 3D
Aggregated memory	256 MB	8 GB	64 GB	1 TB
# registers (w.size)	64 (x32)	128 (x32)	512 (x32)	512 (x64)
Clock frequency	8 MHz	25 MHz	66 MHz	200 MHz
Comp. Power/node	64 Mflops	50 Mflops	528 Mflops	1600 Mflops
Aggregated Comp. Power	1 GFlops	100 GFlops	1 TFlops	7 TFlops

TILED ACHITECTURES ARE LOW POWER

- **POWER Consumption**
  - **(Multi)Tiled SoCs and Systems are low power.**
    - **ATMEL D740 (2004 – 180 nm)**

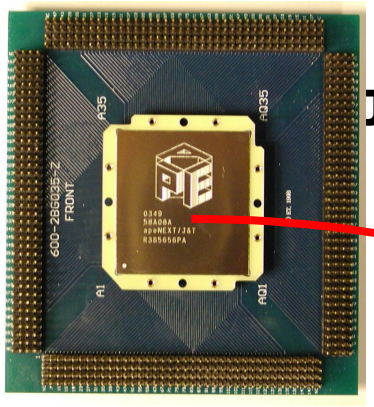
      ~500 mW/GFlops (40-bit)
    - **INFN apeNEXT**

      3W per 1.6GFlops (64 bit)
  - **good ratio of Flops/Watt**
  - **good ratio of computing power per volume**

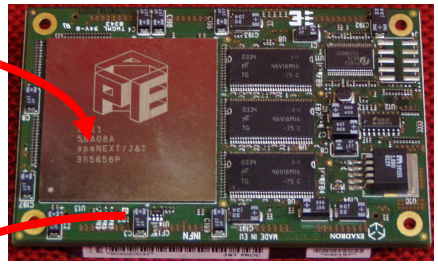


APENext (2005) 2048 processor system

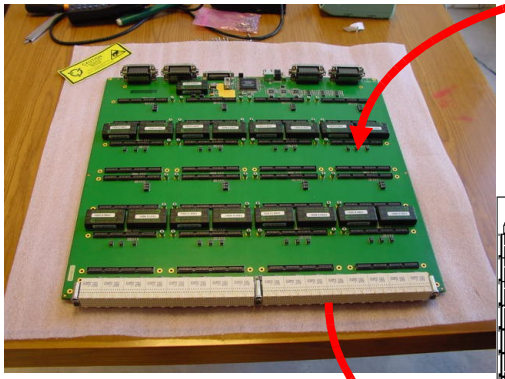
Assembling apeNEXT...



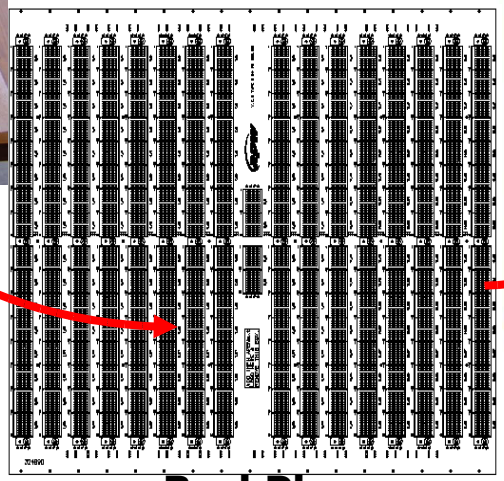
J&T Asic



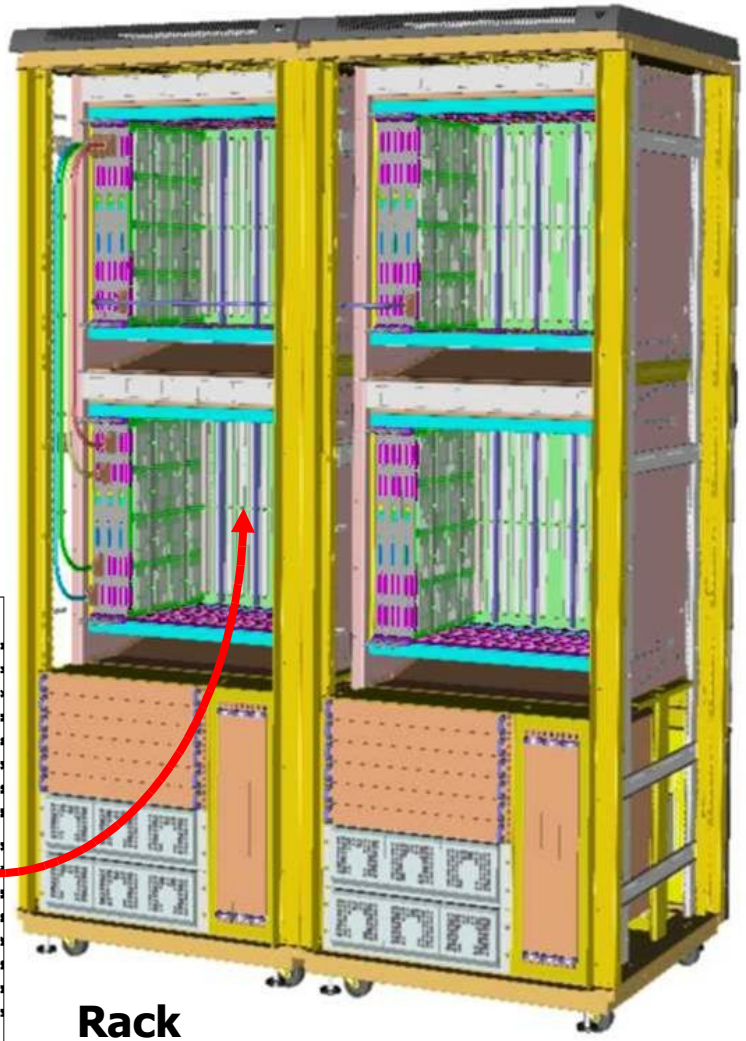
J&T module



PB



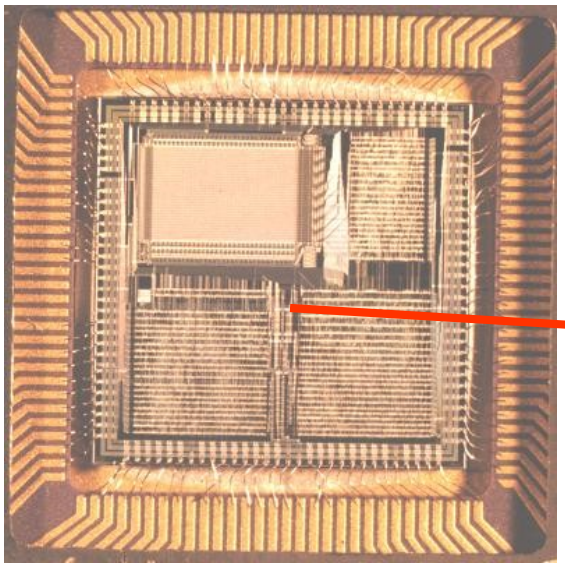
BackPlane



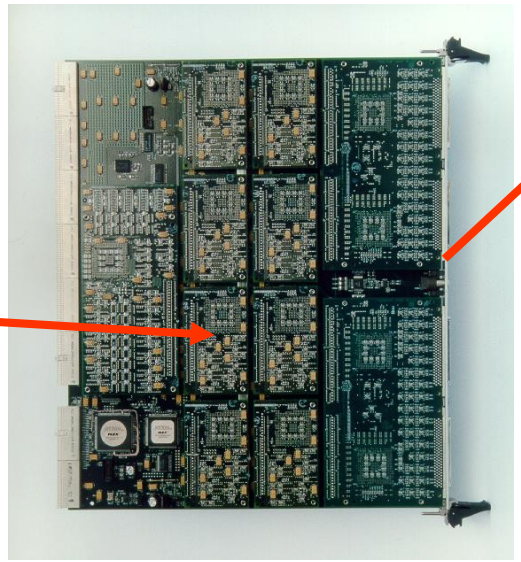
Rack

APEmille (1999) – 1 TFlops

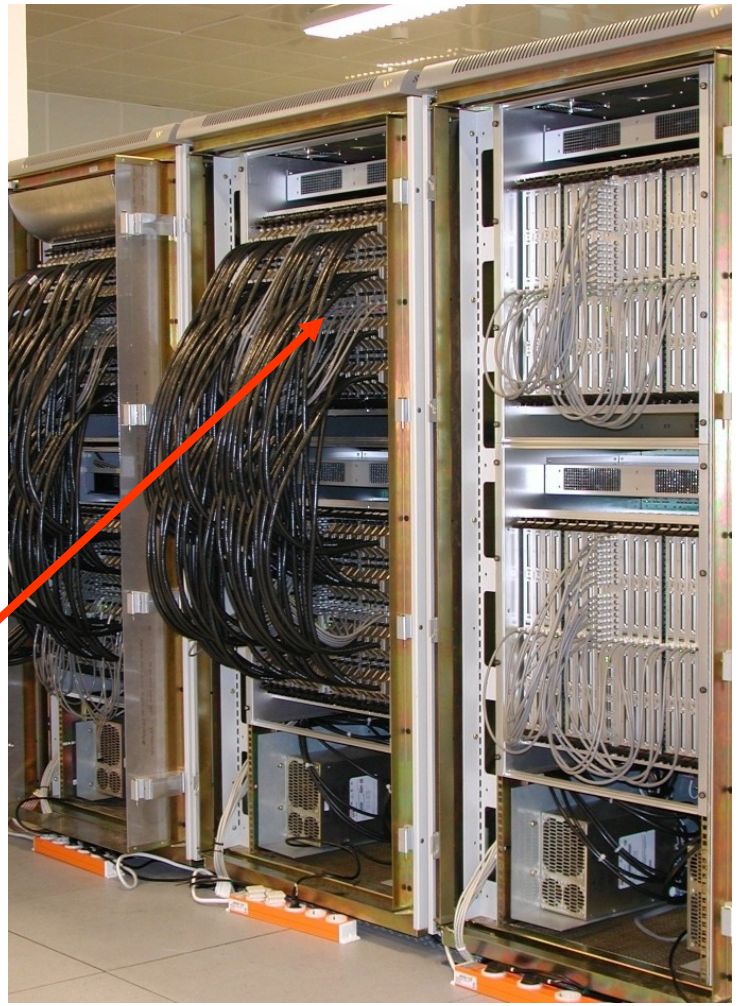
- 2048 VLSI processing nodes
- SIMD, synchronous communications
- Fully integrated "Host computer", 64 PCs cPCI based



Computing node

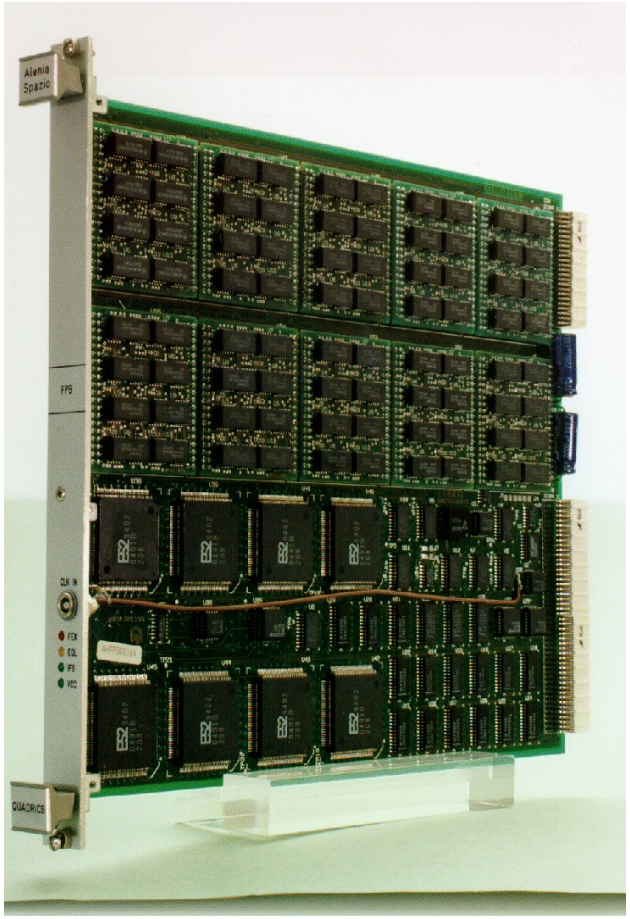


"Processing Board" (PB)
8 nodes, 4GFlops



"Torre"
32 PB, 128GFlops

APE100 (1993) - 100 GFlops

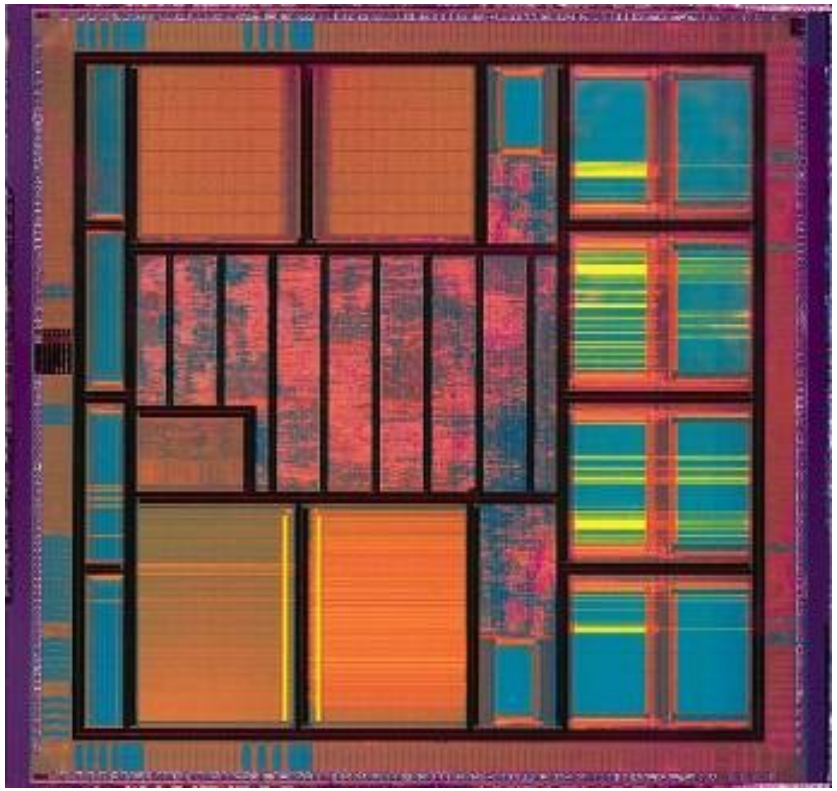


PB (8 nodes) ~ 400 MFlops

…toward MPSoC tile

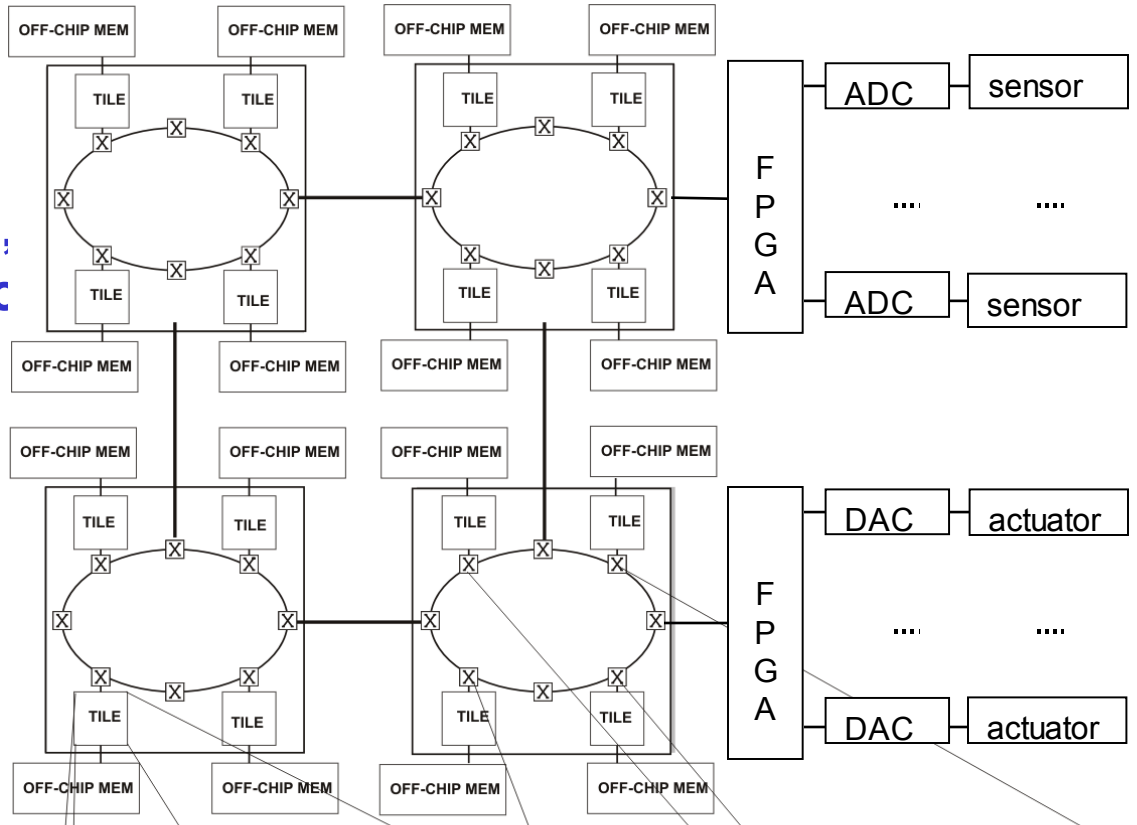
- 1997- 2001
  - Spin-off from INFN and Creation of IPITEC start-up (Intellectual Property Initiative for Tools and Embedded Cores) – (P.S. Paolucci, B. Altieri)
- 2002-2004
  - mAgic VLIW DSP synthesizable core
  - IPITEC becomes ATMEL Roma Advanced DSP Products ATMEL

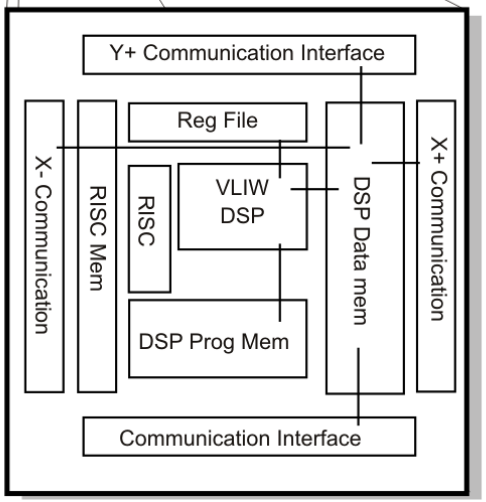
- Diopsis 740 tile: A gigaflops VLIW+RISC SoC Tile - HotChips 15 Conference – Stanford (2003)

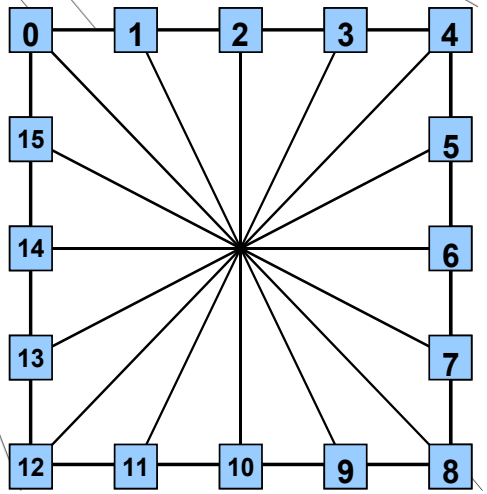


Tiled HW Architecture

- Communication Centric, not Processor Centric
- Homogeneous SW interface for on-chip and off-chip scalable connection and I/O
- Virtual tunnelling on packed switching
- Clustered toroidal 3D System Eng.
- HW support for Parallelism Aware System SW

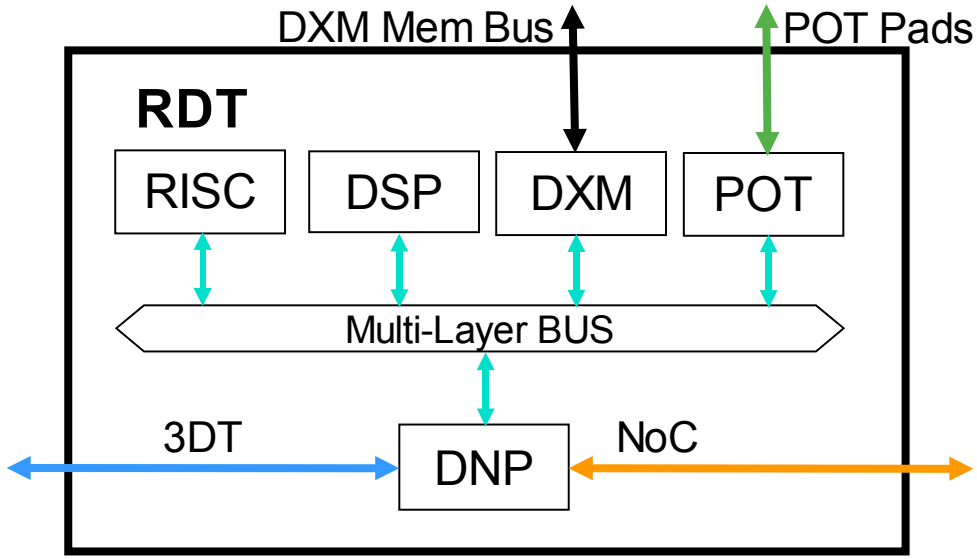




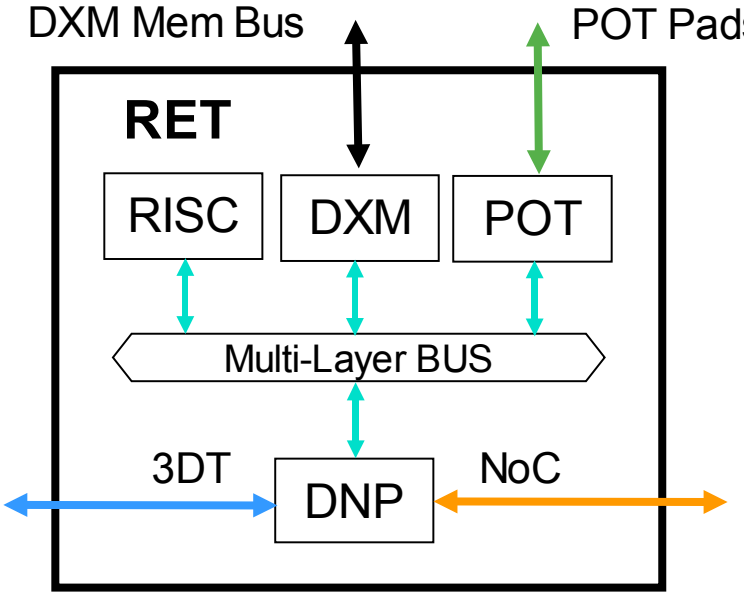


Shapes

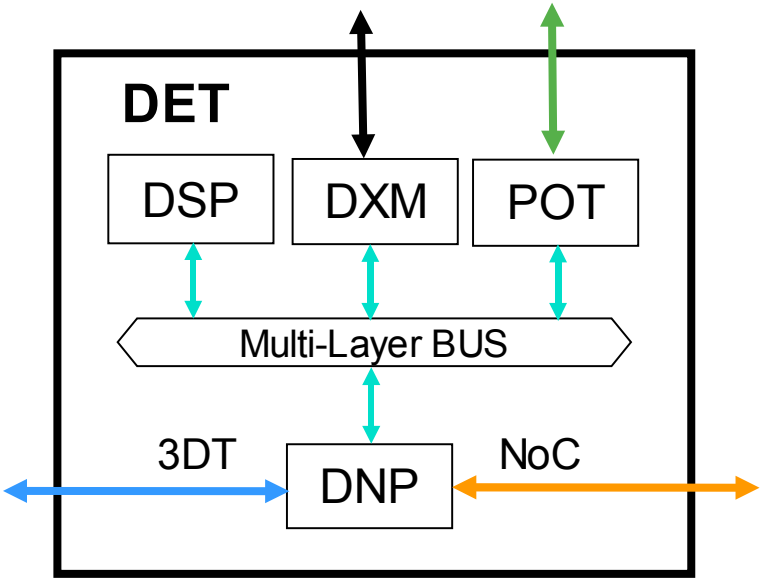
Different Types of Tiles



RDT: RISC + DSP Elementary Tile

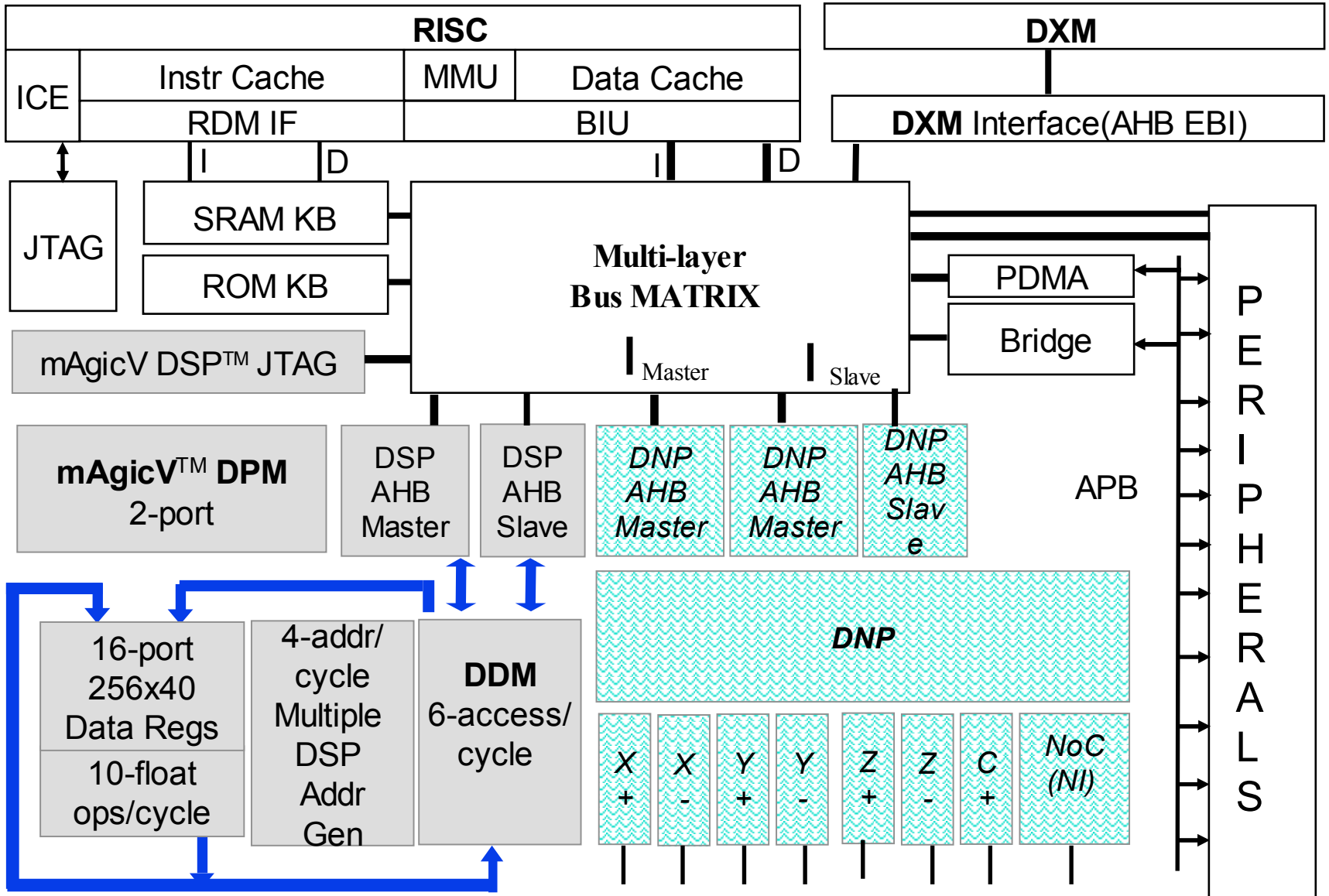


RET: RISC Elementary Tile



DET: DSP Elementary Tile

The tile: Diopsis +



SW Environment – Holistic Approach

- **Application specification: Kahn process networks –> network of actors**
  - Model application component…their interaction…available degree of parallelism
- **Model Compiler and Distributed Operation Layer**
  - Extracts source code and info about process interaction
  - Maps components on Processing and Networking Resources
    - Use of simul traces, analytic performance analysis and run-time monitoring
    - Multi-objective optimization (throughput, delay, predictability, efficiency,…)
  - Produces resources sharing strategies like arbitration and scheduling
- **Simulation Environment**
  - Uses component info plus…hardware characterization and component mapping
  - To perform simul at different levels of abstraction produce traces
- **Hardware dependent Software**
  - Generation of dedicated communication and synchronization primitives
- **Compiler**
  - Communication aware VLIW scheduling

SW Environment – Summary of Working Principles

Model Based Application Description

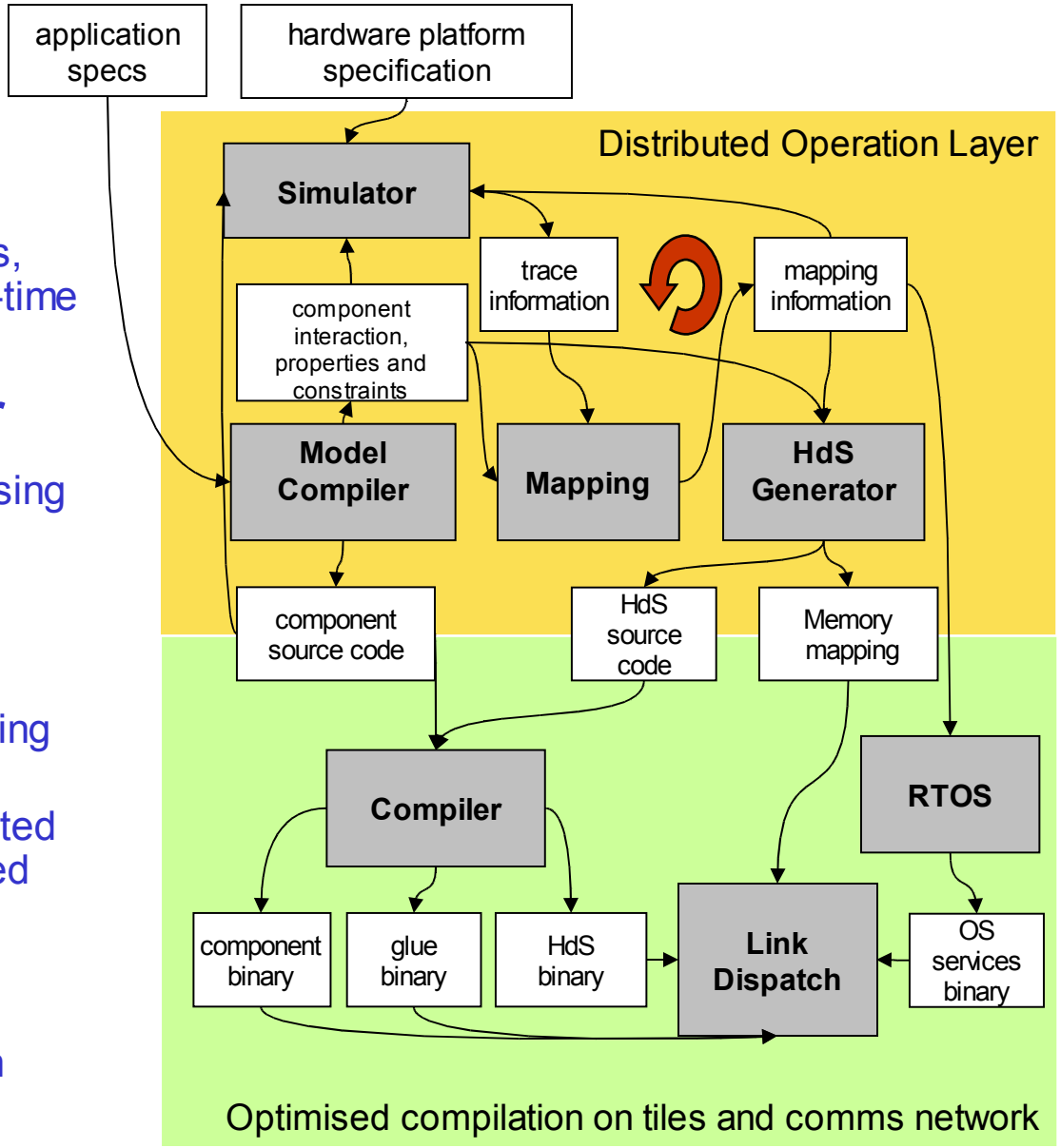
- Interacting Components incl. non-functional constraints, analytical predictions and run-time profiling

Distributed Operation Layer

- Maps components on Processing and Networking Resources
- Stepwise approach to semi-automated mapping:
  - By hand, assisted by simulation, run-time profiling and analytical models
  - By algorithms for automated multi-objective randomised search

Target Applications

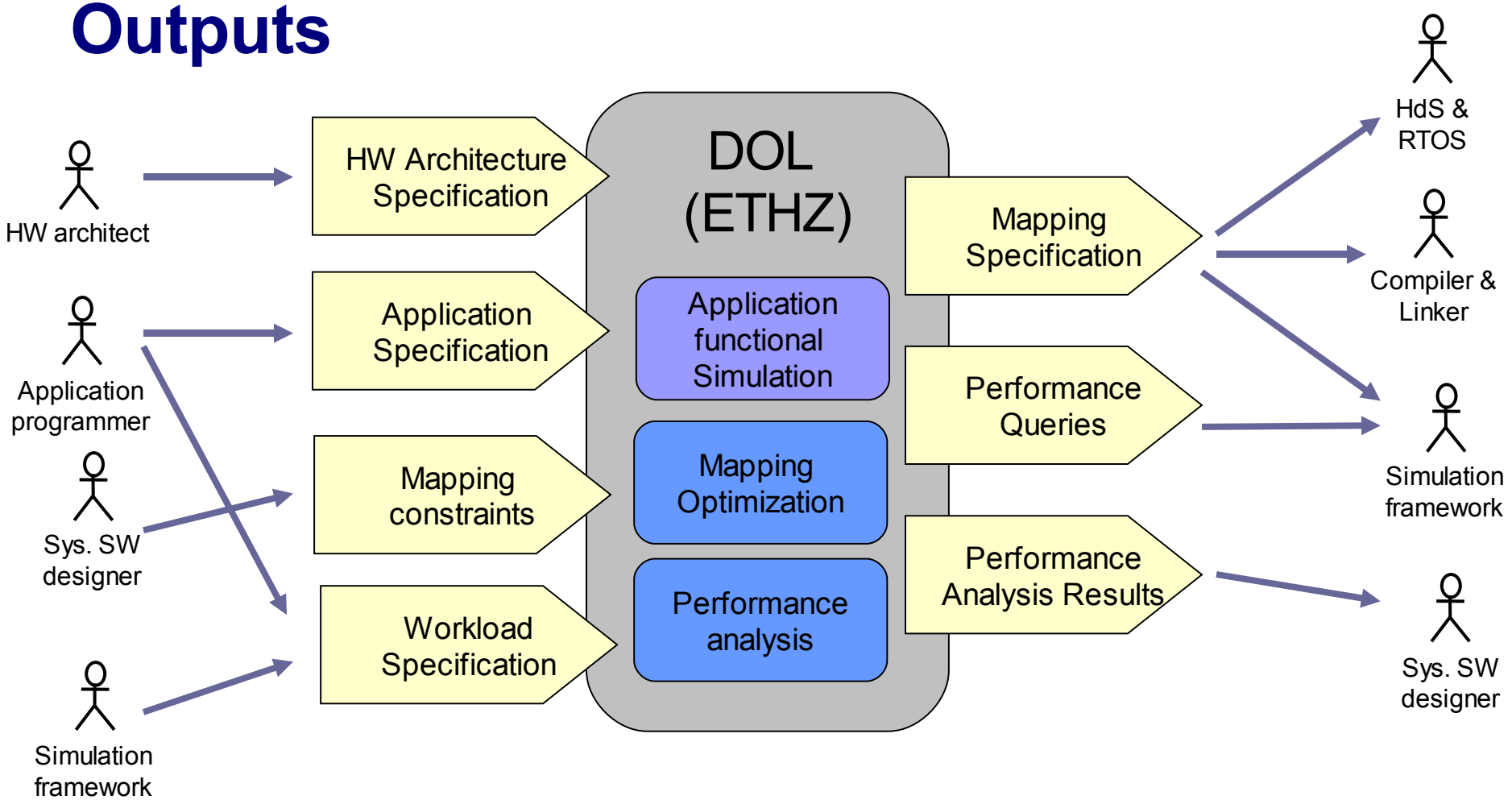
- Extensive inherent parallelism



Distributed Operation Layer

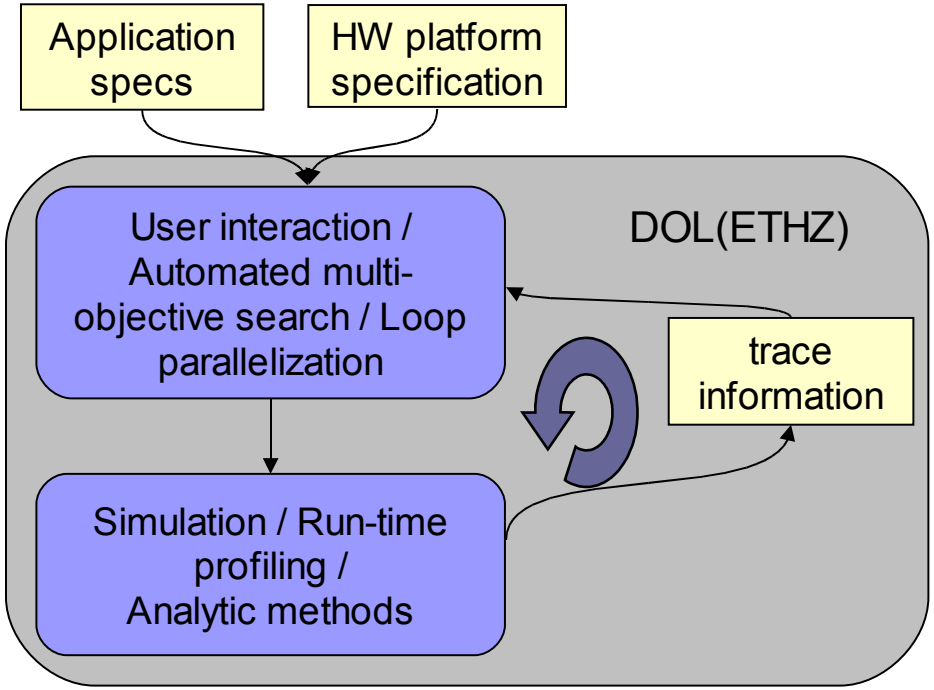
- The purpose of the DOL is to significantly reduce the effort associated with the mapping of applications (from a restricted domain) onto SHAPES platforms. It will:
  - help a programmer of a SHAPES platform to find an efficient mapping of application tasks and communication links between those tasks onto execution and communication resources of the platform.
  - support the programmer in designing distributed scheduling strategies for those resources.
  - support scalability, meaning that it has to minimize the effort necessary to re-map a given application onto the same or different SHAPES hardware architecture.

Distributed Operation Layer - Inputs and Outputs



Distributed Operation Layer – Mapping by Optimization

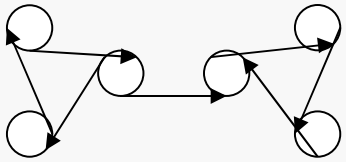
- Purpose
  - Spatial mapping
    - Comm links -> NoC & network
    - Components -> tiles & processing
  - Partial temporal mapping:
    - Arbitration & scheduling policies
  - HdS Generation
    - Expand communication API
- Tradeoffs conflicting quality criteria
  - Latency, throughput, energy
  - Bootstrap:
    - Simulation, run-time profiling, analytical prediction
    - Manual automation



Layered approach & Dependencies

WP1.11

Statistical Analysis



Tasks are represented as timing budgets:
- Very high simulation speed
- No functional modeling & verification

WP1.4

simulation speed

Virtual Processing Unit (VPU)

Generic abstract processor simulator:
- adaptable to arbitrary processor core
- high simulation speed
- functional validation
- user-dependent accuracy

Instruction Accurate (IA) Model

Instruction accurate Instruction Set Simulators (ISS):
- ARM9
- mAgic VLIW DSP
- DNP
- STM Spidergon Network-on-Chip

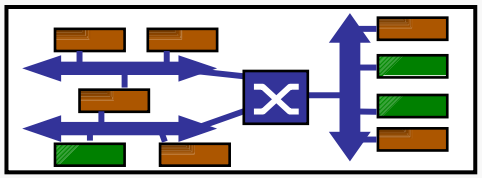
Cycle Accurate (CA) Model

Cycle accurate Instruction Set Simulators (ISS):
- ARM9 (commercially available)
- mAgic VLIW DSP (Target ISS)
- DNP
- STM Spidergon Network-on-Chip (STM model)

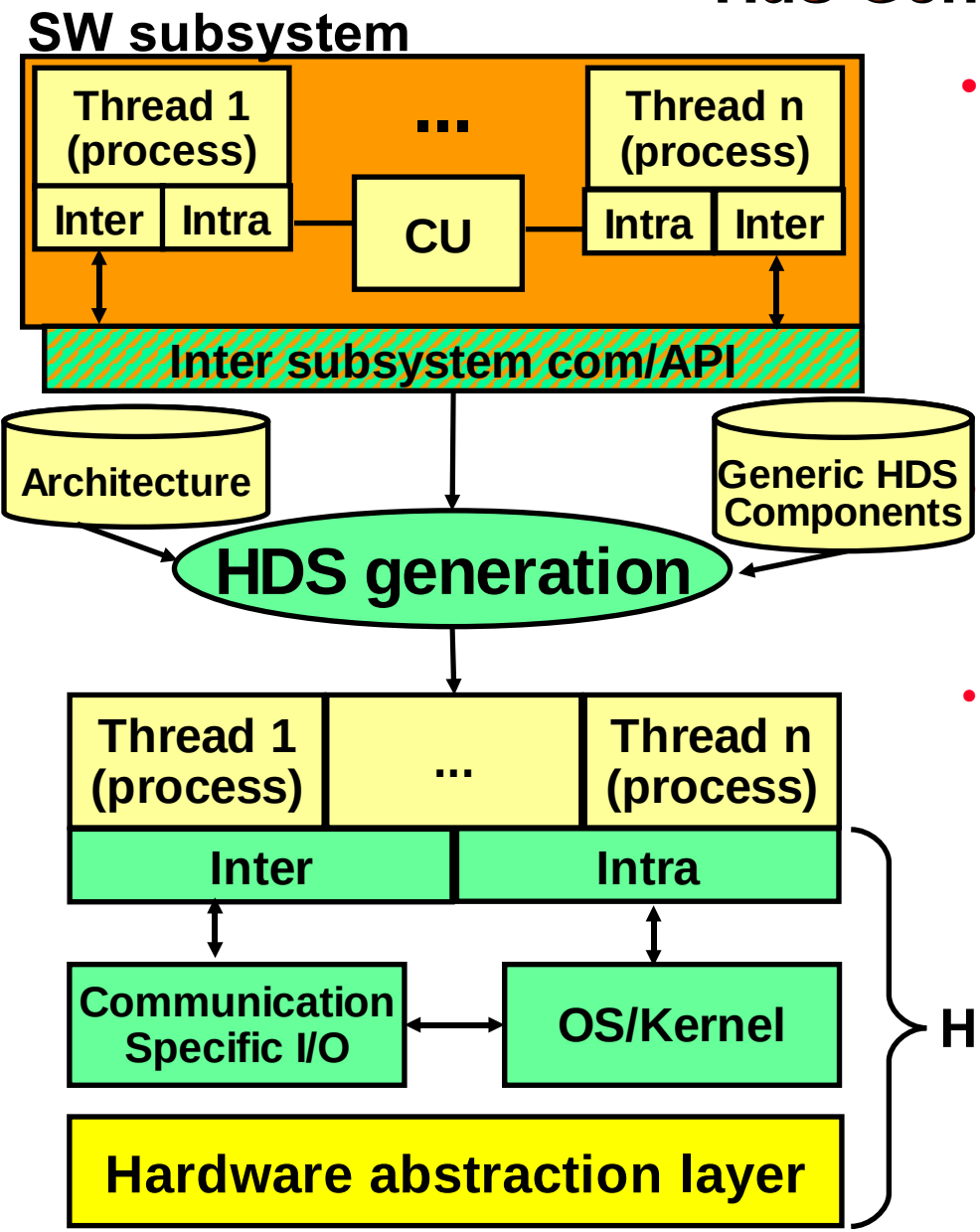
accuracy

WP1.1

SHAPES Hardware platform



HdS Generation



- SW Subsystem specification:
  - Threads
  - Explicit Communication Units
  - Communication API
    - Inter-subsystem
    - Intra-subsystem

HDS generation technique: customization of generic HDS components

- HDS components
  - Operating system
    - Flexible library
    - Application specific
  - Specific I/O
    - Custom API & communication layer
  - HAL: Basic HW access and addresing (e.g. SW to port OS)

Efficient Compilation (TARGET COMPILER TECH.)

Optimised scheduling

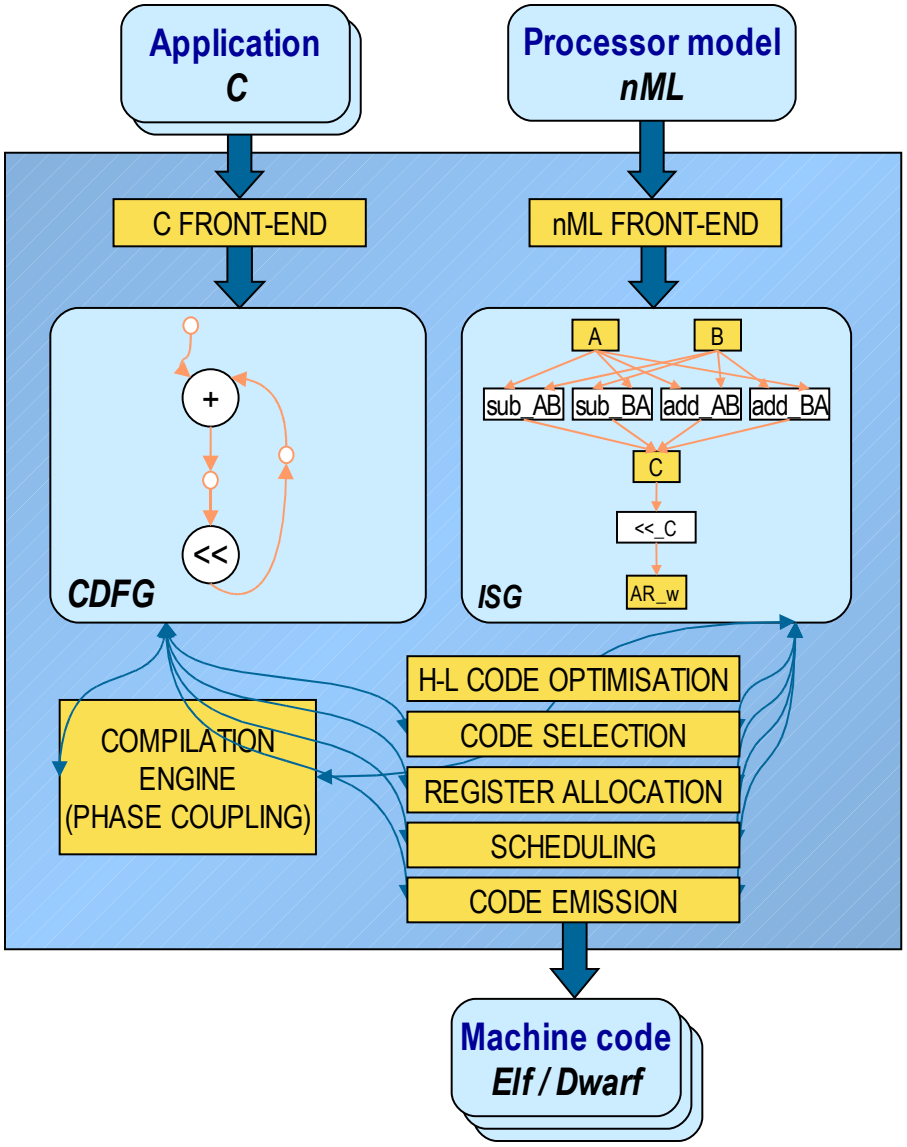
- Intra- and inter-tile communication, mixed with component code

RISC core: re-use existing compiler

VLIW core: advanced graph-based compilation technology

- Netlist-like processor model captures detailed HW resource utilisation and pipeline behaviour
- Graph-based optimisations exploit exact HW resources and timing: instruction and data-level parallelism
- Phase coupling

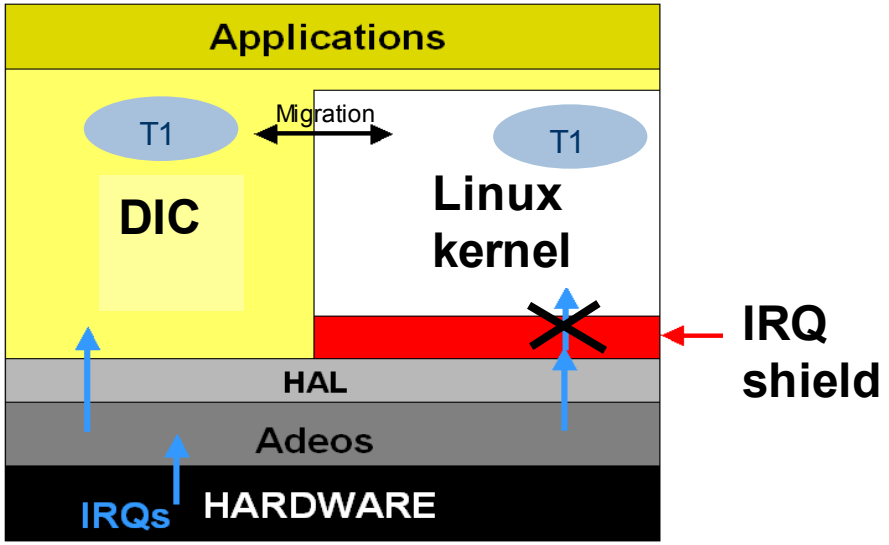
Retargetability enables architecture optimisation



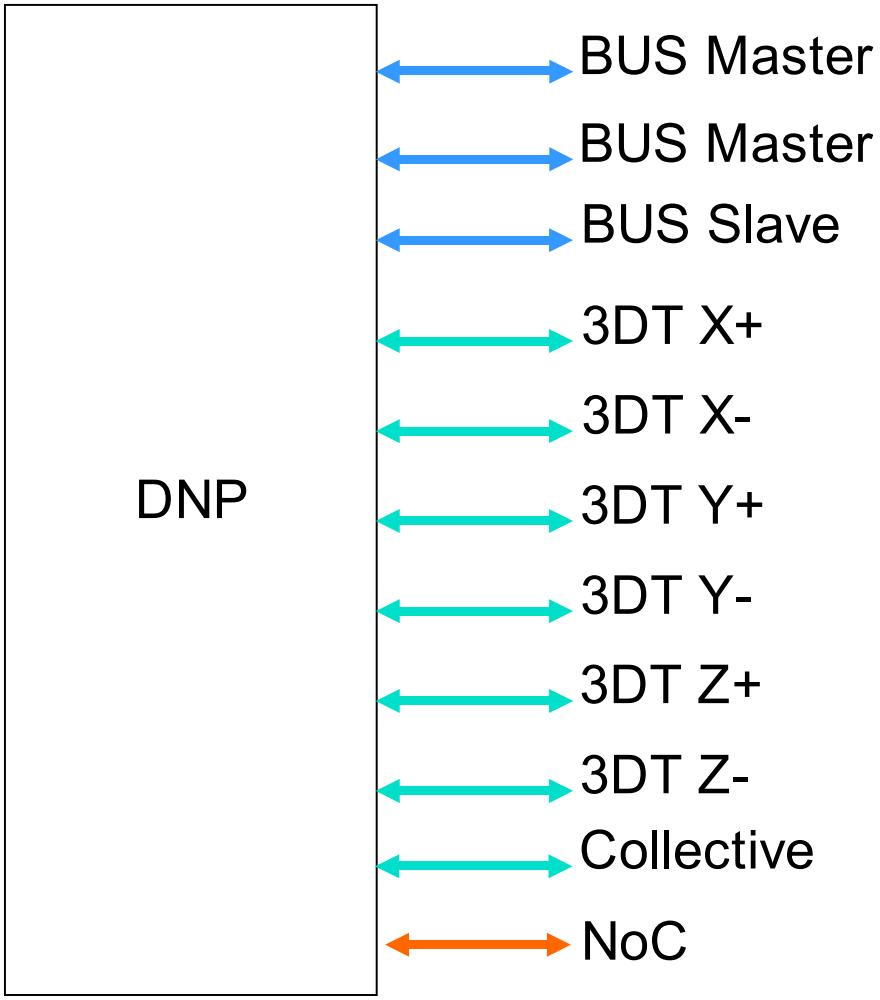
RTOS on a Tiled Multi-processor Architecture (THALES)

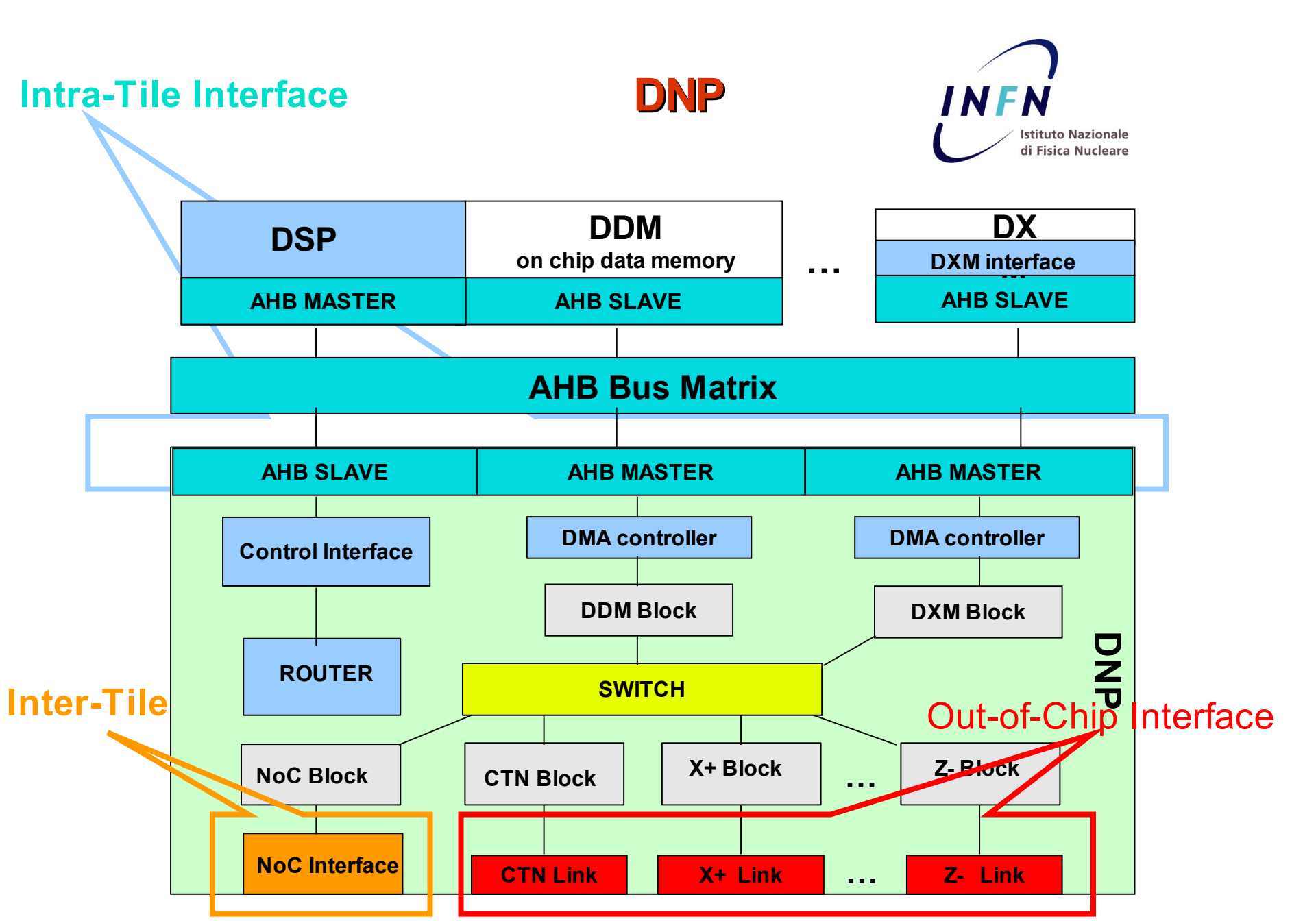
Main Activities in Shapes:

- Port of a pre-emptible Linux kernel on the multi-tile heterogeneous architecture.
- Design and implementation of POSIX compliant real time extensions
  - Adeos (interrupt pipeline layer)
  - Real time domain: DIC (Deterministic Intensive Computing)
- Definition of compiler requirements related to intra-tile communication optimizations



Distributed Network Processor
DNP: Interface

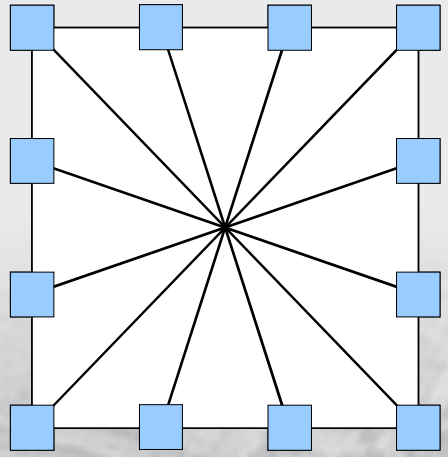


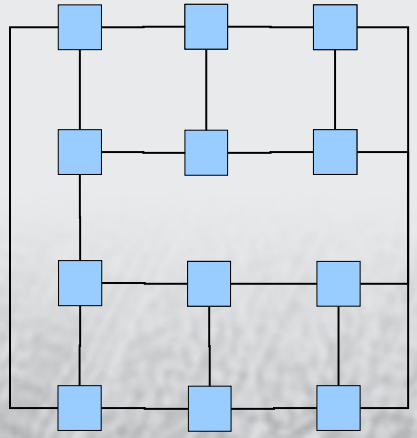

Intra-Tile Interface
DNP
INFN
Istituto Nazionale
di Fisica Nucleare
DSP
AHB MASTER
DDM
on chip data memory
AHB SLAVE
...
DX
DXM interface
AHB SLAVE
AHB Bus Matrix
AHB SLAVE
AHB MASTER
AHB MASTER
Control Interface
DMA controller
DMA controller
DDM Block
DXM Block
ROUTER
SWITCH
DNP
Inter-Tile
Out-of-Chip Interface
NoC Block
CTN Block
X+ Block
...
Z- Block
NoC Interface
CTN Link
X+ Link
...
Z- Link

Spidergon topology

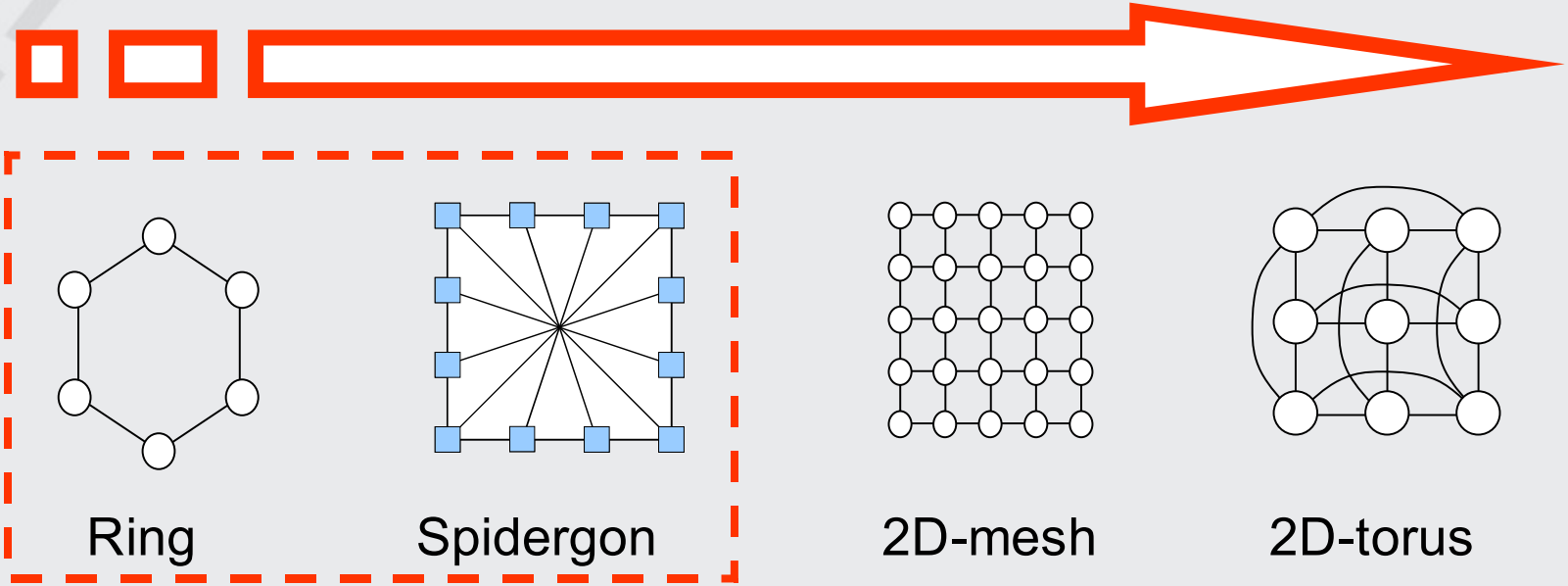
- It's a family of *regular/symmetric* topologies

- We look for a *complexity/performance* trade-off
  - Low degree (router cost)
  - Low number of links (wire cost)
  - Symmetry (homogeneous building blocks; simple routing)
  - Low diameter (performance)
  - Good scalability (small network size granularity)





Topologies overview

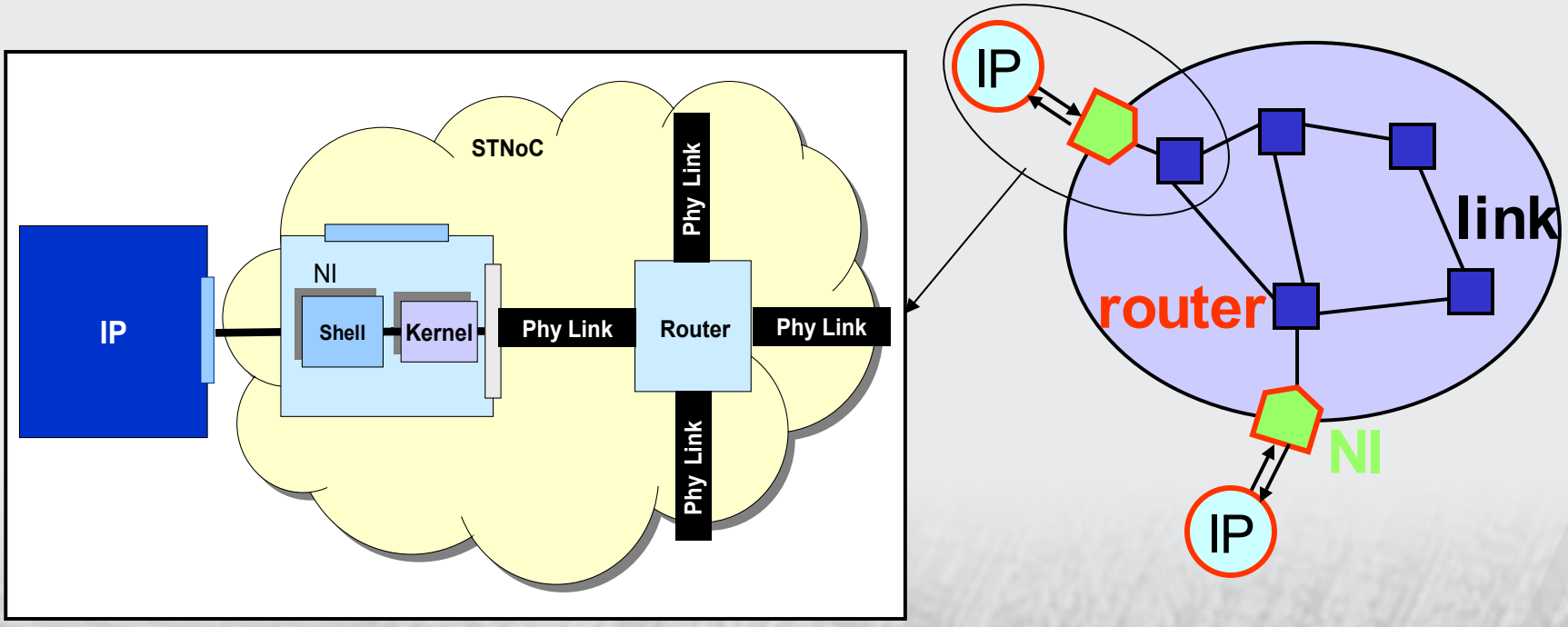


supported by STNoC

STNoC key components

Network on Chip is a set of **on-chip routers** (up to layer 3),

Network Interfaces (NI) (layer 4) and physical Link



Benchmarking through parallel applications

- **Audio Wave Field Synthesis: the equivalent of a 3D sound hologram of a multitude of moving objects for theater, home and car**
- **Extraction and treatment of voice signals from noisy environment through benchmarking on microphone arrays (hand-free, vocal command, ambient intelligence)**
- **Ultrasound scanners: echo graphic beam-forming in SW and graphical rendering**
- **Physical Modelling: Lattice Quantum ChromoDynamics and BioComputing**

Summary

- **Tiled Approach for management of wiring on deep submicron technologies and billion gate design complexity**
  - **RISC + floating-point VLIW DSP + DNP Elementary tile**
  - **Communication Centric HW Architecture**
  - **Low end single module hosting 4-32 tiles for mass market applications**
  - **Classic digital signal proc. systems e.g. radar and medical equipments (2 K tiles)**
  - **High-end systems requiring massive numerical computation (32 K tiles)**
- **Target Applications with extensive inherent parallelism**
- **Model Based Parallel Programming Environment with Mapping Exploration and Communication Aware HdS Layer and Communication Aware Compilation System**
- **www.shapes-p.org**

HW GLOSSARY

INSIDE THE TILE

- RISC max one per tile
- DSP max one per tile
- DNP: Distributed Network Processor (always one per tile)
- DDM: Distributed Data Mem (inside the DSP)
- DPM: Distributed Progr Mem (inside the DSP)
- DXM: Distributed eXternal Mem Interface (max one per tile, outside the RISC and DSP)
- POT: Peripherals On Tile
- RDM: Risc (tightly coupled) Data Memory
- RPM: Risc (tightly coupled) Program Memory
- RCM: Risc Cache Memory
- DCM: DSP Cache Memory (future improvement)

FUNDAMENTAL TYPE OF TILE

- RDT includes:
  - RISC: (includes RDM and RPM) +
  - DSP(includes DDM and DPM)
  - DNP + DXM + POT

POSSIBLE TILE VARIANTS

(subset of RDT)

- RET := RDT minus DSP
- DET:= RDT minus RISC
- DDT:= DET minus DXM

AT THE CHIP LEVEL

- MTC: Multiple Tile Chip (composed of multiple Tiles)
- NOC: Network On Chip (connecting Tiles)
- 3DT: 3 Dim Toroidal Connection (outside the chip)

Acknowledgments

- Lothar Thiele, Kai Huang – ETH Zurich
- Rainer Leupers, Torsten Kempf – RWTH-Aachen - ISS
- Ahmed Amine Jerraya – TIMA Lab. Grenoble
- Gert Goossens – Target Compiler Technologies
- Marcello Coppola – STMicrolectronics
- Piero Vicini, Davide Rossetti, Mersia Perra, Alessandro Lonardo – INFN Roma
- Luigi Raffo, Gianni Mereu – Università di Cagliari
- Philippe Kajfasz - THALES

- European Proj. – FET – FP6 – IST - 4 2.3.4(viii) Adv. Comp. Arch.

Research Lines

Scalable Software Hardware Architecture Platform for Embedded Systems

System SW

- **ETH Zurich - *Distributed Operation Layer*; manage application parallelism**
- **RWTH Aachen Univ. - *Simulation of Heterogeneous Multi Proc. Systems***
- **TIMA Lab and THALES - *Hardware dependent Software Layer and OS***
- **TARGET Compiler Tech. - *Retargetable VLIW Compilers***

System HW

- **ATMEL Roma - *Tile*:**
  - **Evolution of Diopsis™: mAgicV VLIW DSP™ + RISC + INFN DNP™**
- **STMicrolectronics + Univ. of Cagliari and Pisa –**
  - **Evolution of Spidergon™ *Packet Switching Network on Chip***
- **INFN Roma – DNP™ *Distributed Network Processor + 3D Toroidal Eng.*:**
  - **Evolution of APE Massive Parallel Processors**

Parallel application benchmarking

- **Fraunhofer IDMT,IGD - Audio Wave Field Synthesis and Graphic Algorithm**
- **PIE, MedCom - Ultrasound scanner**
- **INFN - Physical Modelling**