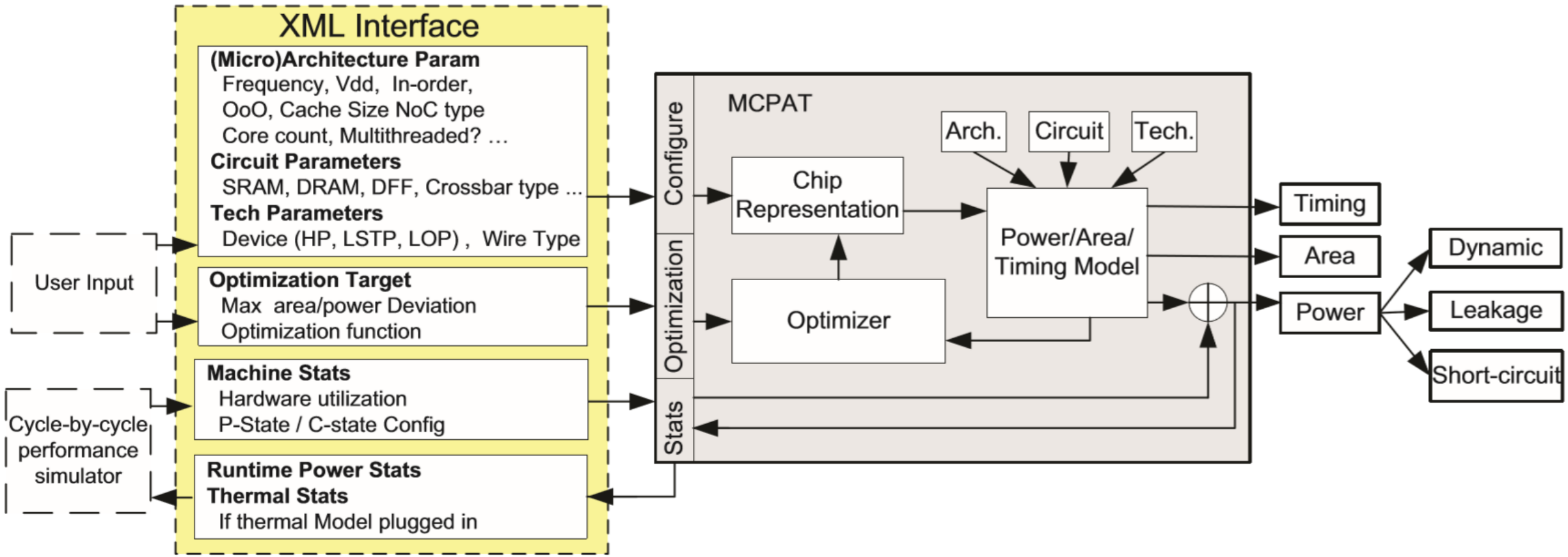
McPAT: An Integrated Power, Area, and Timing Modeling Framework for Multicore and Manycore Architectures

MICRO2009，
University of Notre Dame
University of California, San Diego


XML Interface
(Micro)Architecture Param
Frequency, Vdd, In-order,
OoO, Cache Size NoC type
Core count, Multithreaded? …
Circuit Parameters
SRAM, DRAM, DFF, Crossbar type ...
Tech Parameters
Device (HP, LSTP, LOP) , Wire Type
Optimization Target
Max area/power Deviation
Optimization function
Machine Stats
Hardware utilization
P-State / C-state Config
Runtime Power Stats
Thermal Stats
If thermal Model plugged in
User Input
Cycle-by-cycle performance simulator
Configure
Optimization
Stats
MCPAT
Arch.
Circuit
Tech.
Chip Representation
Optimizer
Power/Area/ Timing Model
Timing
Area
Power
Dynamic
Leakage
Short-circuit

Power Modeling:

dynamic

total load capacitance（decomposing it into basic circuit blocks, and using analytic models for each block with appropriately sized devices）, the supply voltage, the voltage swing during switching, the clock frequency, and the activity factor（access statistics from architectural simulation together with circuit properties）

short-circuit

Nose et al

leakage power

depends on the width of the transistors and the local state of the devices

Subthreshold leakage occurs because a small current passes between the source and drain of off-state transistors.
Gate leakage is the current leaking through the gate terminal, and varies greatly with the state of the device.

Timing Modeling：

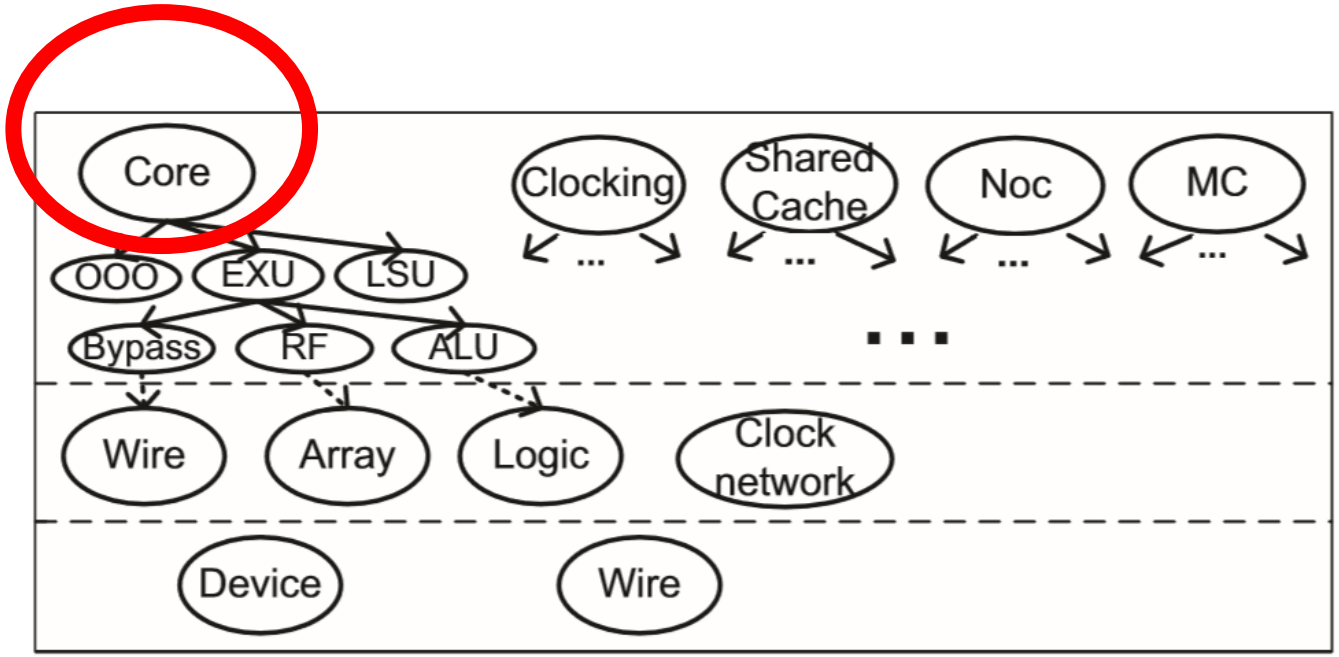
resistance and capacitance

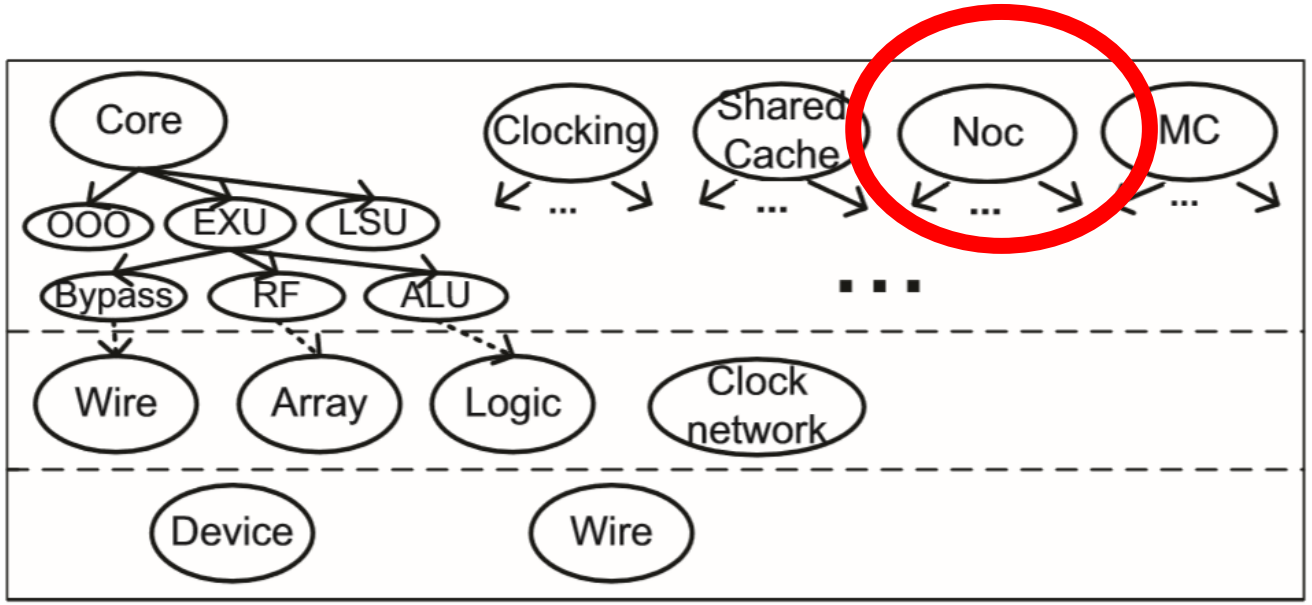
CACTI& Palacharla et al.

Area Modeling：

CACTI

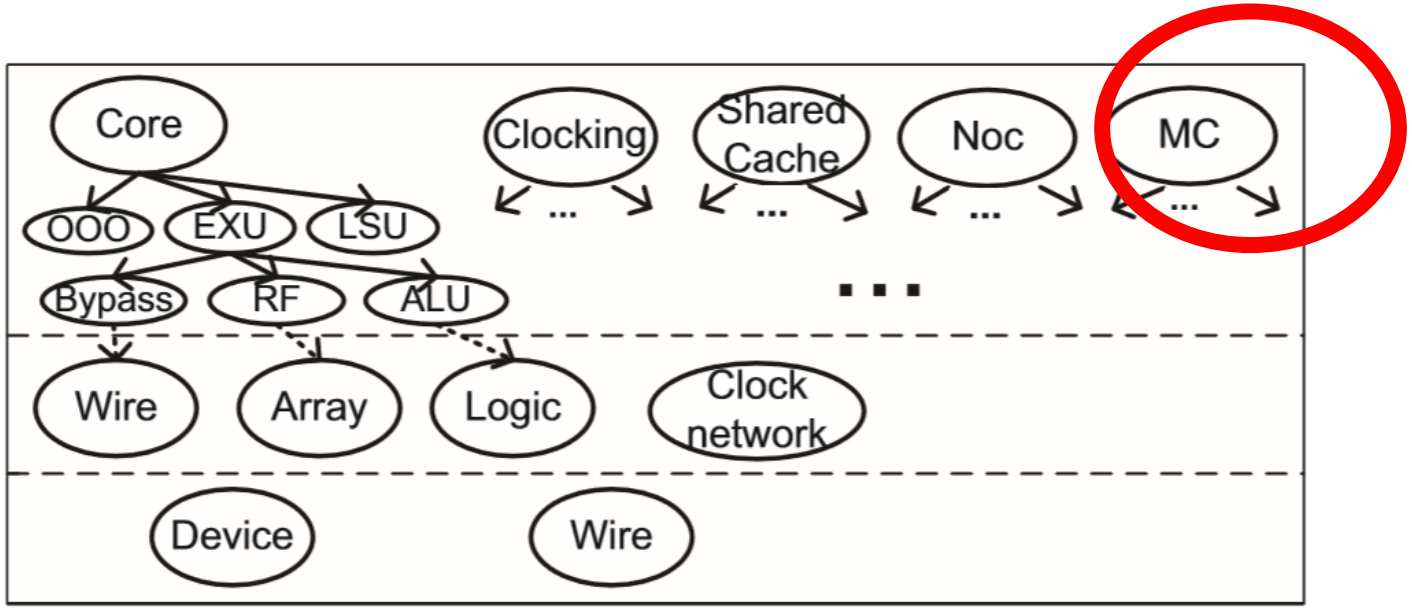
basic logic gates and regular structures, including memory arrays (e.g., RAM, CAM (content addressable memory), and DFFs (D flip-flop)), interconnects (e.g., router and link), and regular logic (e.g., decoder and dependency-checking unit).


Core
OOO
EXU
LSU
Bypass
RF
ALU
Clocking
Shared Cache
Noc
MC
...
Wire
Array
Logic
Clock network
Device
Wire



signal links and routers

use the same analytical approach used in modeling cores to model routers: breaking the routers into basic building blocks such as flit buffers, arbiters, and crossbars; then building analytical models for each building block

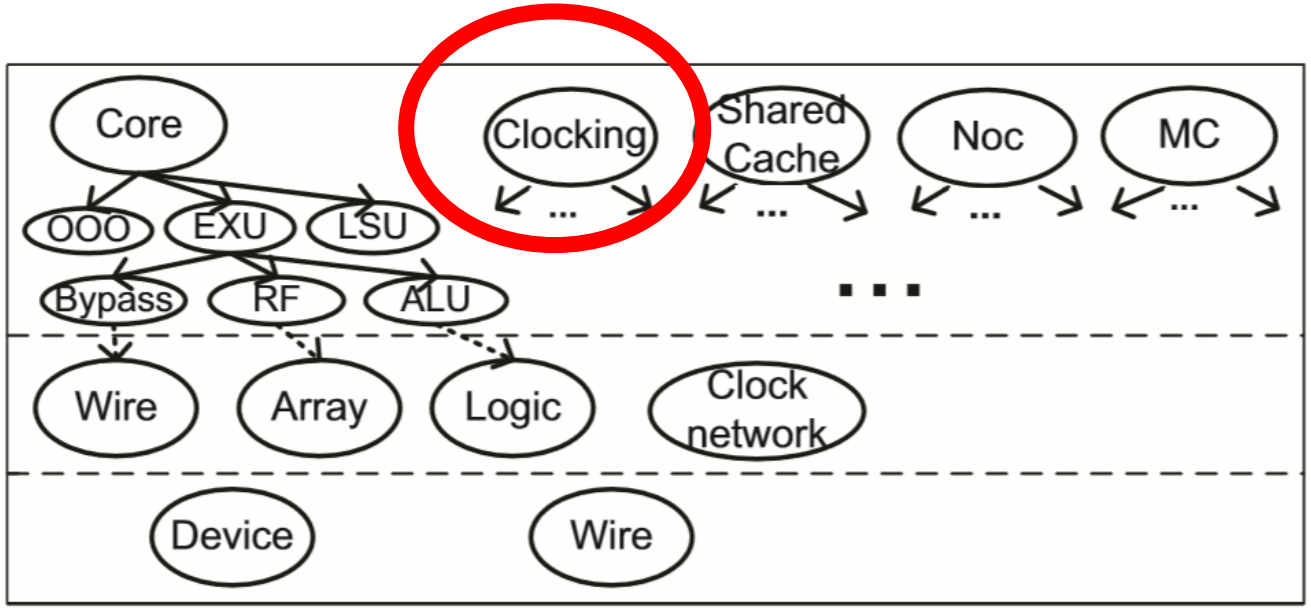


Denali

1) the front-end engine （CAM and RAM） responsible for rescheduling the memory requests

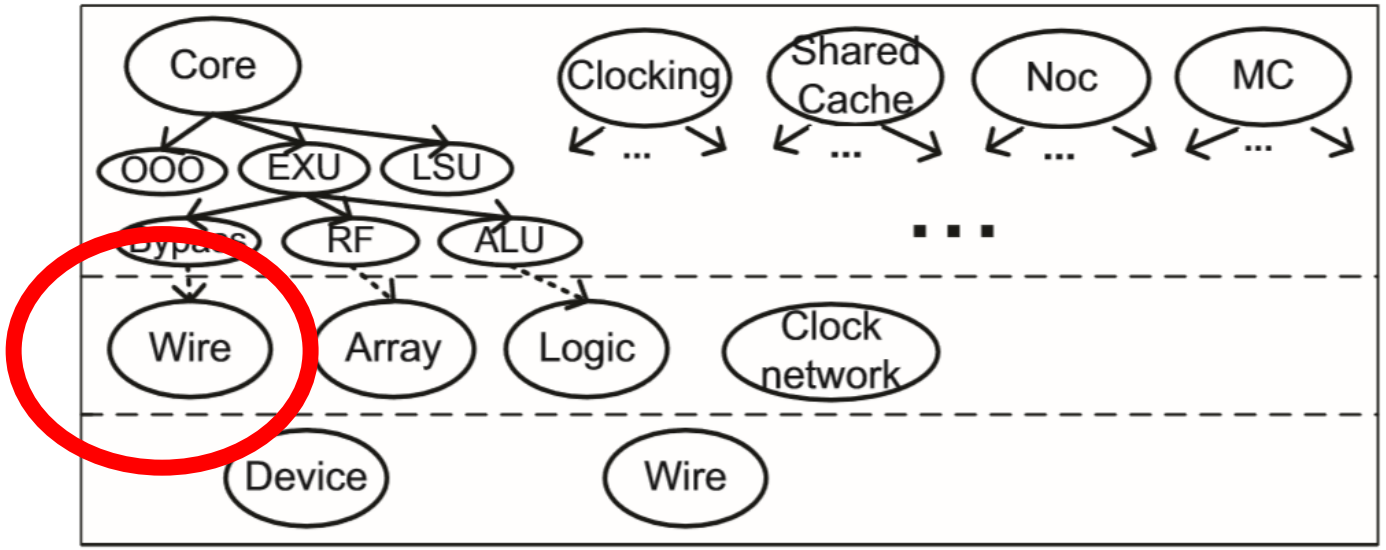
2) the transaction processing engine that has the logic and sequencer to generate the command, address, and data signal

3) the physical interface (PHY) that serves as an actual channel of the memory controller for communicating off-chip to memory. （Rambus [23] and AMD ）



phase-locked loop (PLL

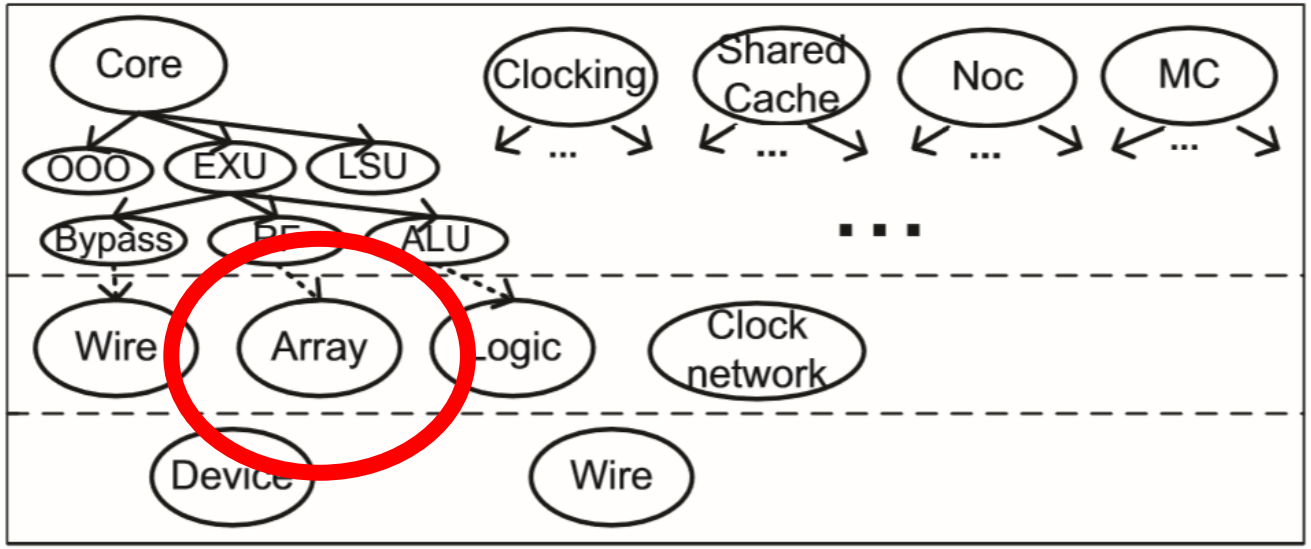
the clock distribution network



resistance and capacitance

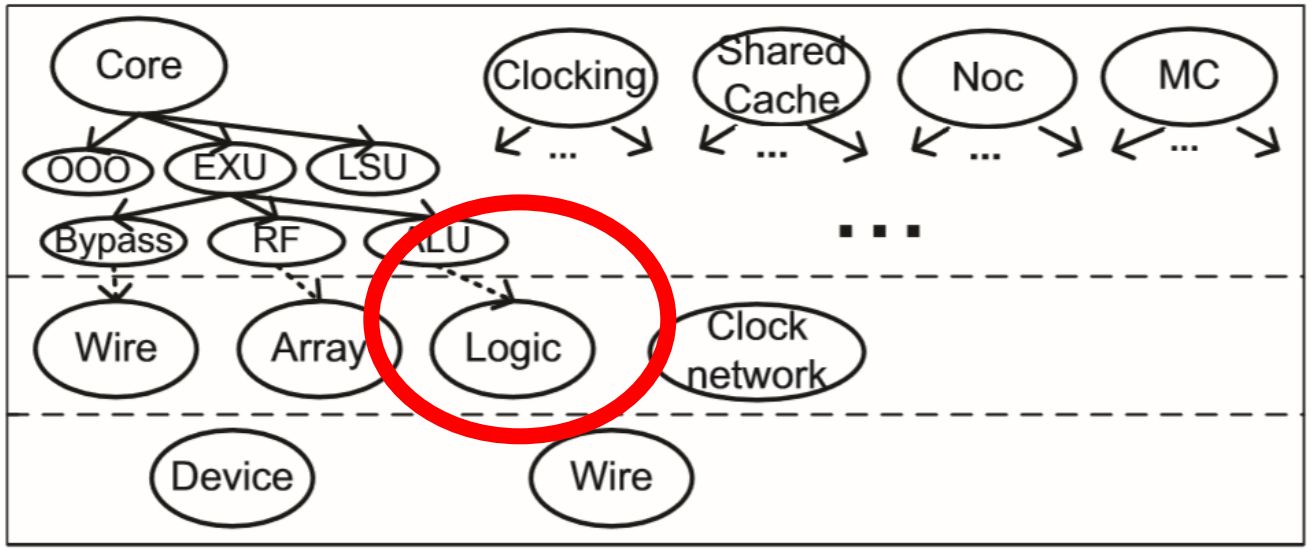
short wires using a one-section π-RC model

For long wires, we use a buffered wire model



RAM-, CAM-, and DFF-based arrays

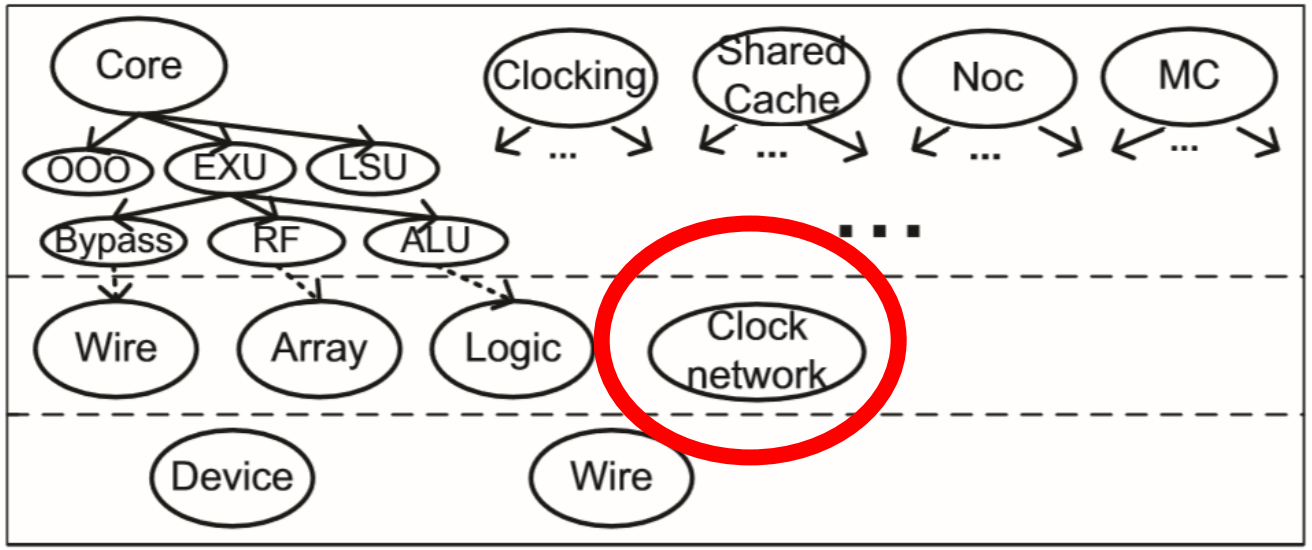
CACTI



For highly regular blocks with predictable structures, such as memories or networks, McPAT uses the algorithmic approach of CACTI

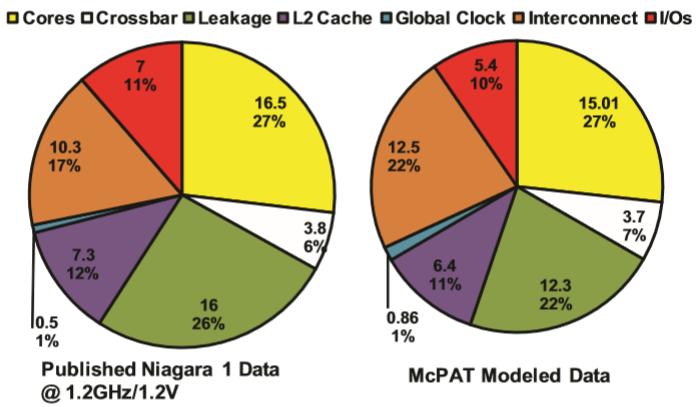
less regular but can still be parameterized modeled after existing processors from Intel, AMD, and Sun

highly customized blocks such as functional units, empirical models based on published data for existing designs scaled to different technologies

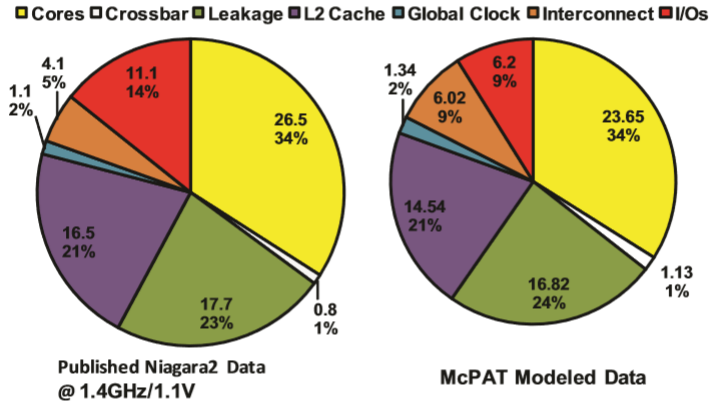


global, domain, and local

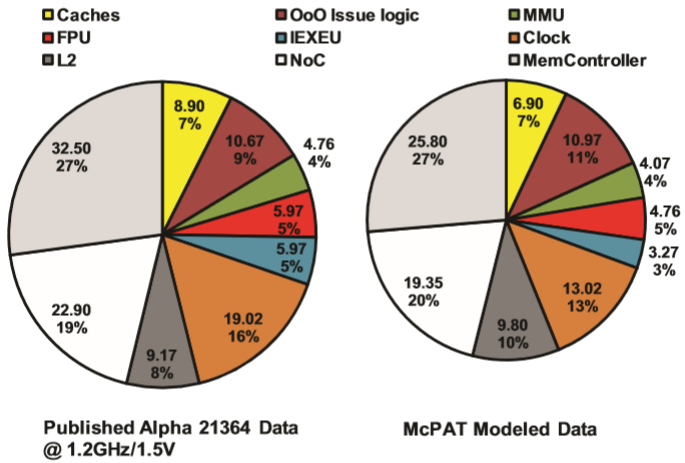
H-tree topology for global-level and domain-level networks and a grid topology for the local networks



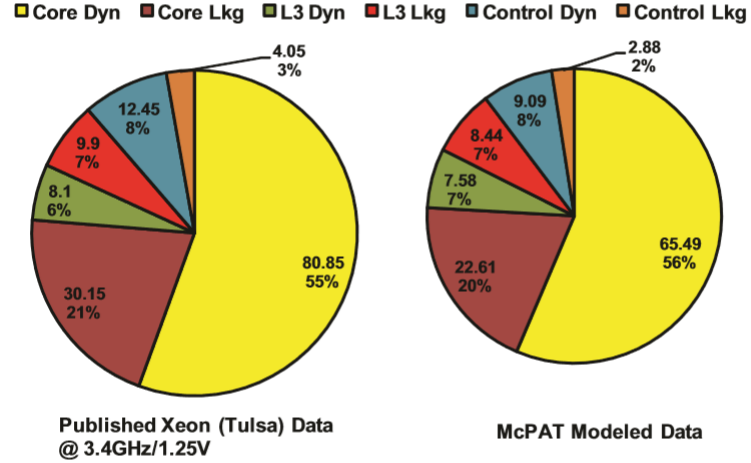
(a) Validation against Niagara processor



(b) Validation against Niagara2 processor



(c) Validation against Alpha 21364 processor



(d) Validation against Xeon Tulsa processor