Global Interconnect Trade-off For Technology Over Memory Modules To Application Level: Case Study

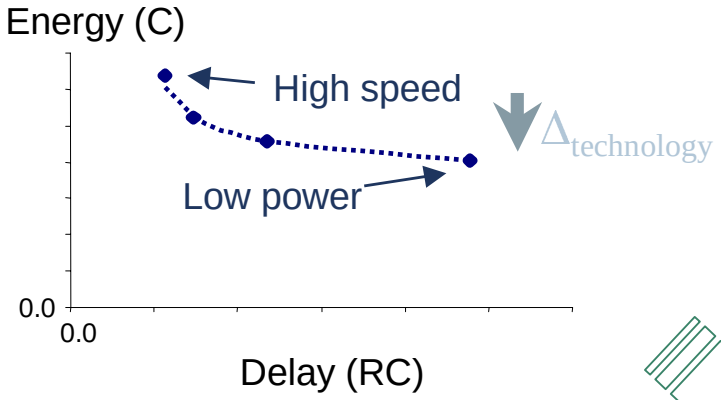
A. Papanikolaou, M. Miranda, F. Catthoor, H. Corporaal,
H. De Man, D. De Roest, M. Stucchi, K. Maex

SLIP 2003
5-6 April 2003
Monterey CA

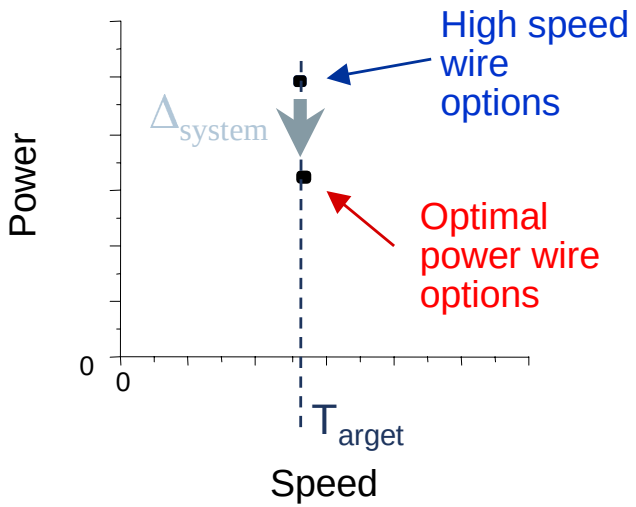
SEEDS FOR TOMORROW'S WORLD

Interconnect technology trade-offs can be exploited for crucial gains at system-level

Technology level trade-offs

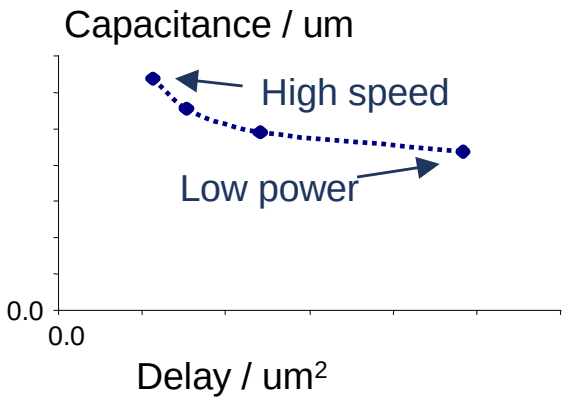


Application-level gains



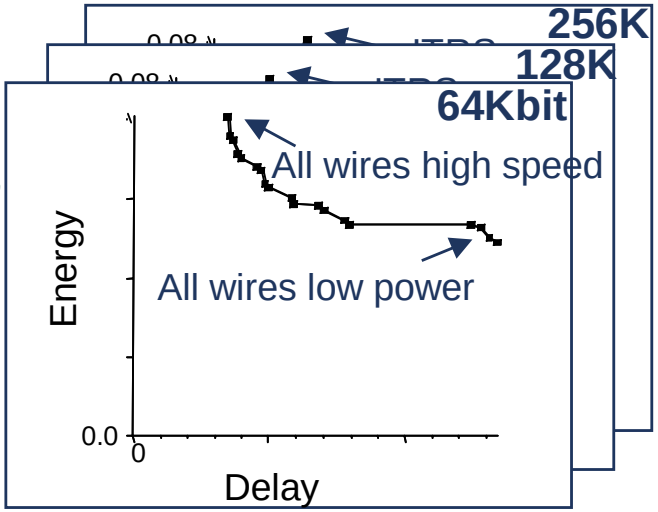
Methodology Overview

Parameterized wire models



3) Use application feedback for process parameter selection

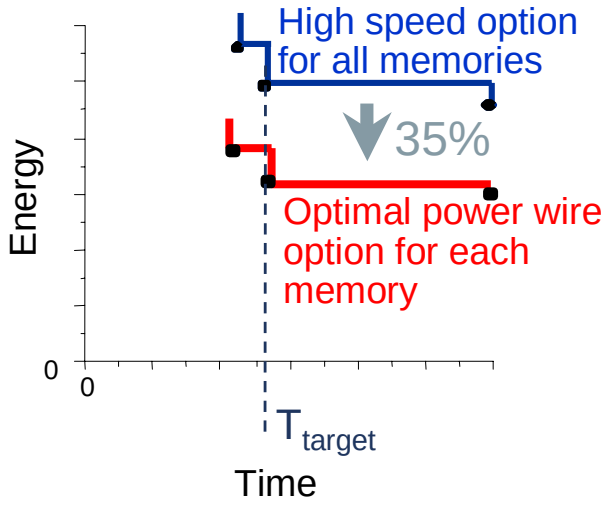
Parameterized IP models



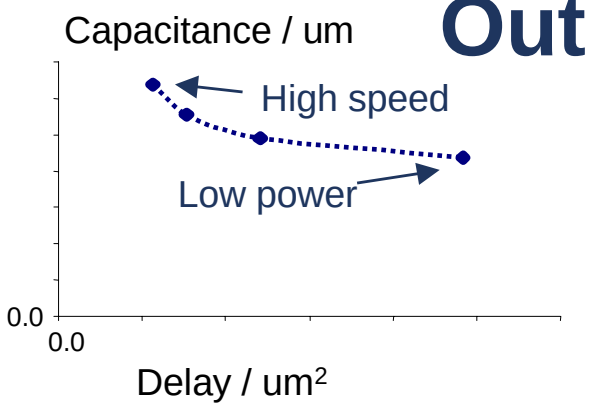
1) Add wire model to IP model

2) Use IP model + crude floorplan in application exploration (tools!)

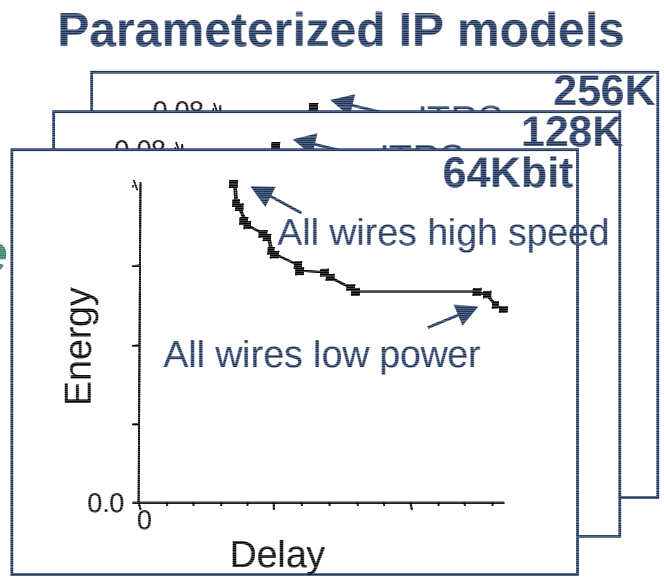
Application-level gains



Outline: Device level exploration

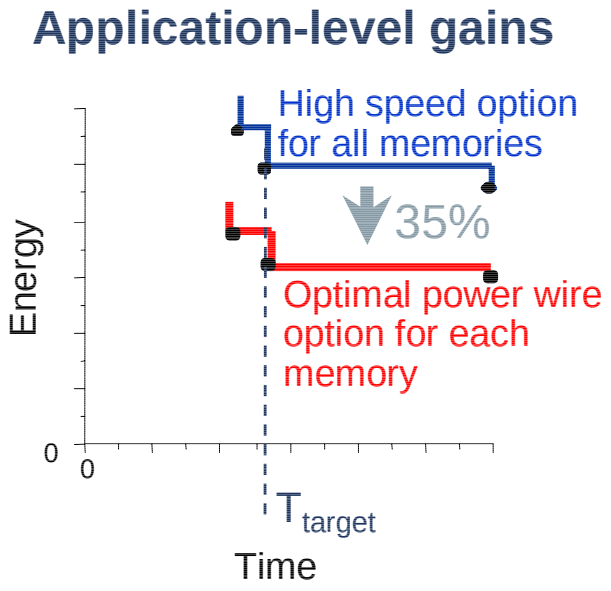


3) Use application feedback for process parameter selection

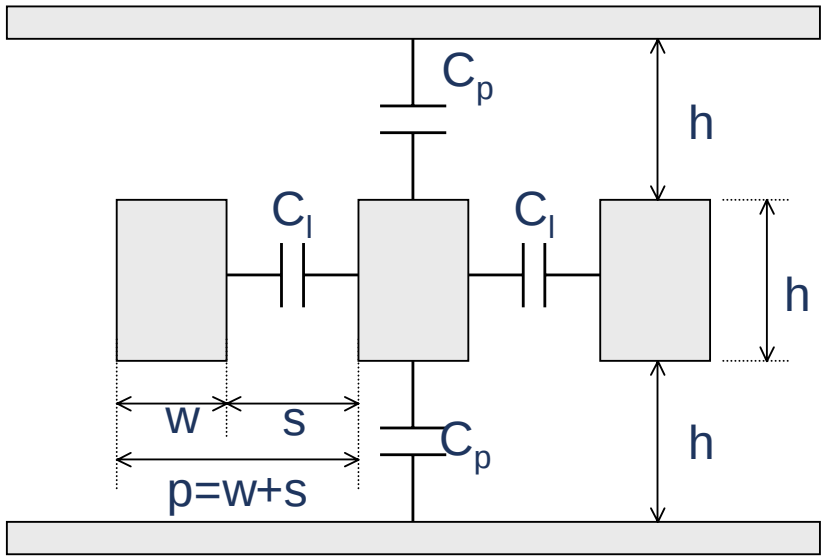


1) Add wire Model to IP model

2) Use IP model + crude floorplan in application exploration (tools!)



Modeling interconnect



ITRS roadmap projection:

- Geometrical parameters

  w,s,h,p

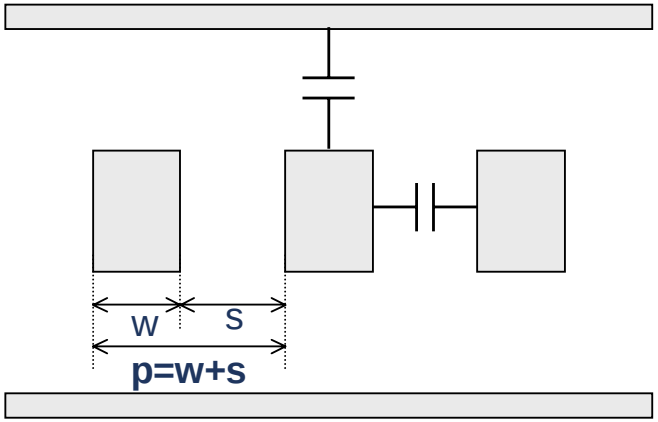
- Material parameters

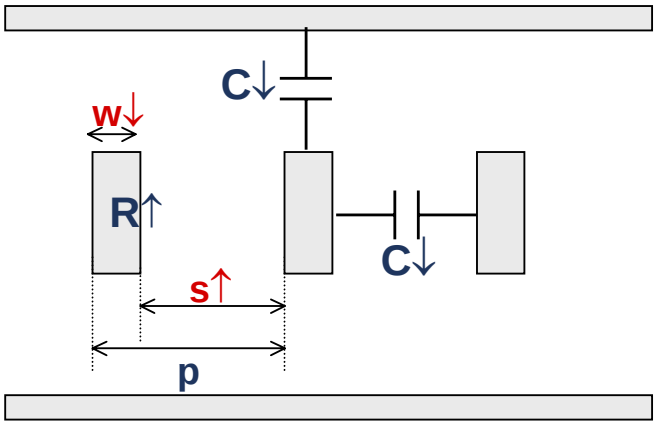
  k_{eff},ρ_{eff}

At least 2 options exist in theory to create energy vs. delay trade-offs at the technology level

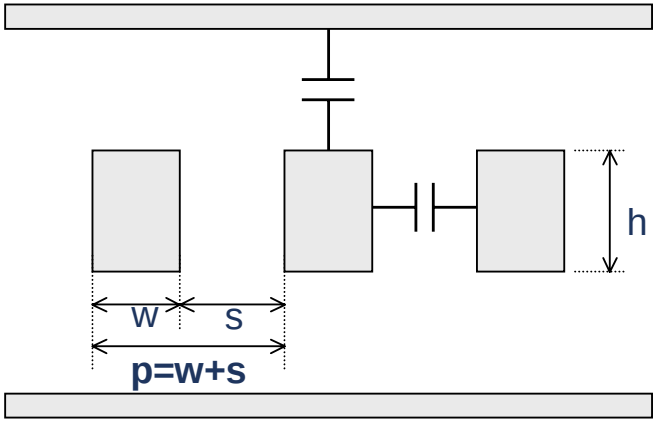
Possible options for wire exploration

Variable wire width (constant height)





Variable wire width and height (constant aspect ratio)



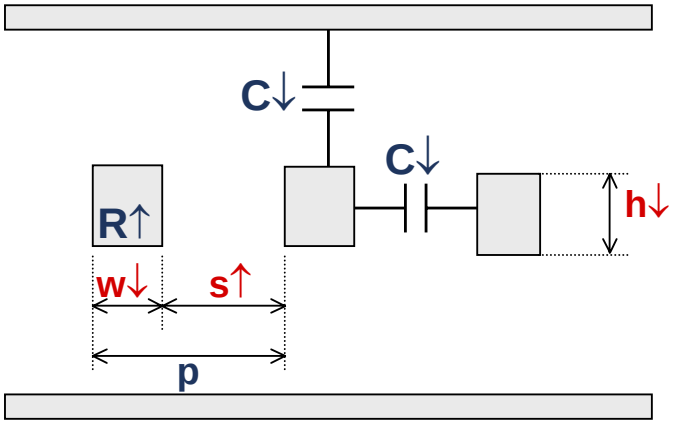
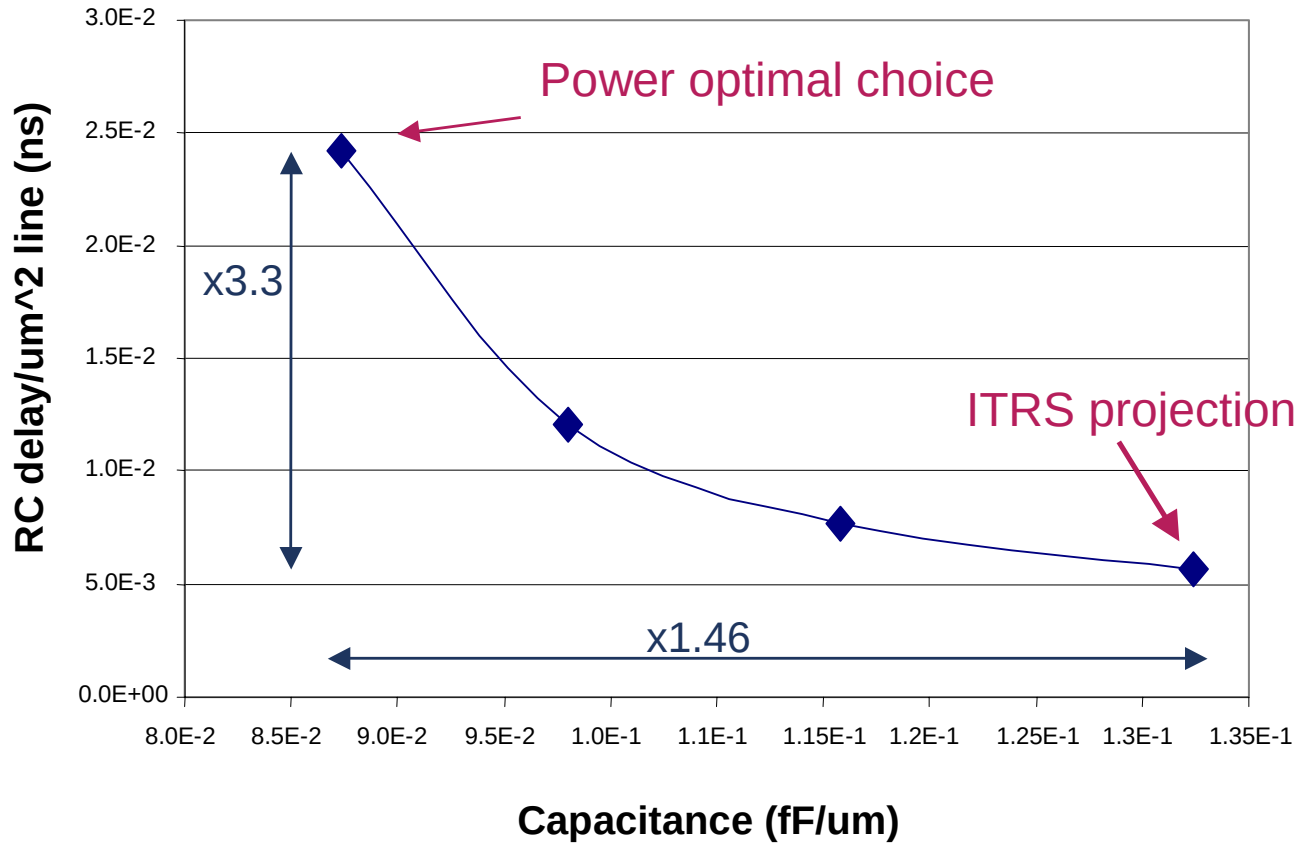
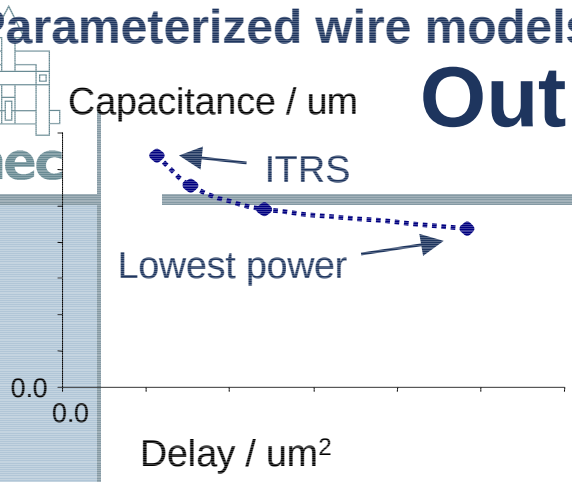


Illustration of technology level trade-offs for local interconnect



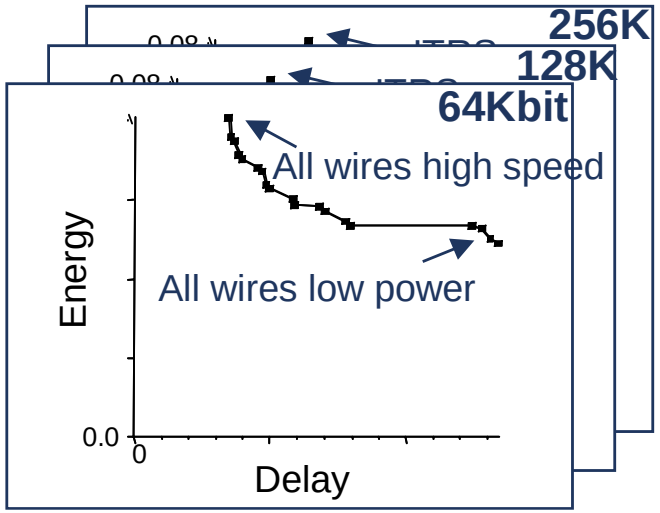
- Both wire exploration options exhibit similar trade-off ranges
- Litho can limit the number of choices but ranges exist

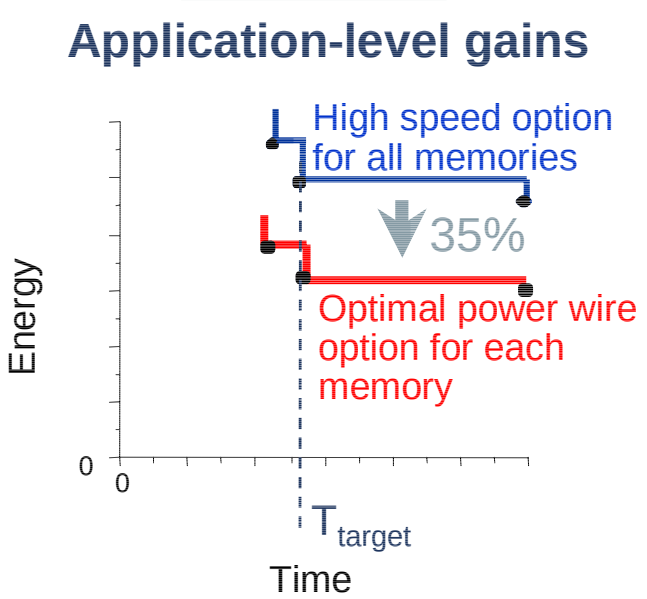
Outline: Module level exploration



3) Use application feedback for process parameter selection

1) Add wire Model to IP model





2) Use IP model + crude floorplan in application exploration (tools!)

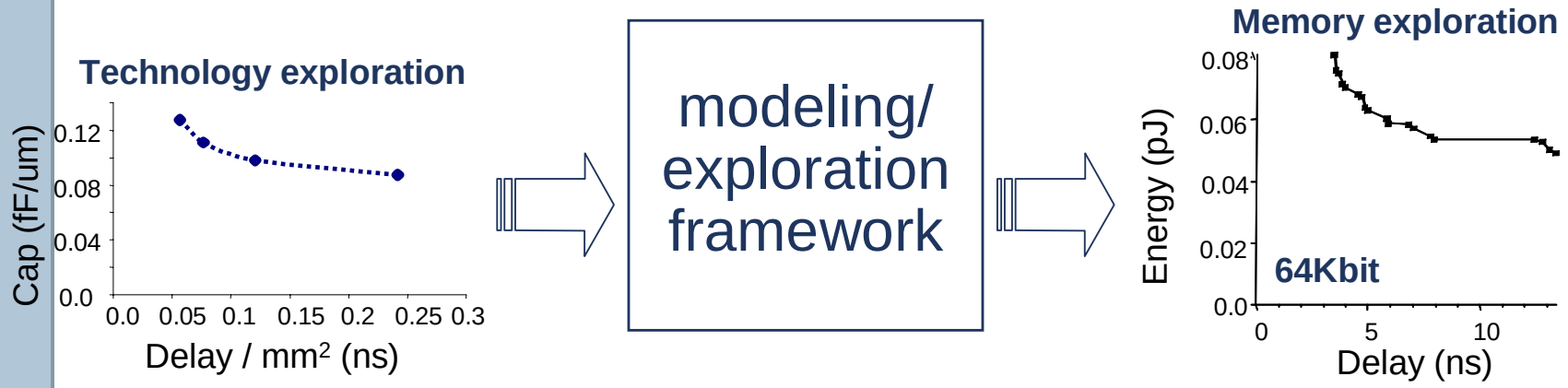
Modeling IP-modules

An IP block model is needed to show if & how technology trade-offs propagate to the module level

Use memories as a case study because they:

- are crucial IP blocks for data dominated systems
- are interconnect dominated
- have a controllable floor-plan and are very regular
  -> predictable topology & wire-lengths

Need for memory modeling/exploration framework



Goal is to evaluate impact of interconnect implementation trade-offs on memory module

Modeling/exploration framework :

- parameterized in terms of technology options (C, RC) & memory parameters (size, bitwidth)
- complex search-space (wire options, partitioning) for exploration -> Pareto optimal trade-offs needed
- analytical modeling of memory circuits for fast evaluation of options

Choice of CACTI as memory modeling/exploration framework

CACTI[1] is a parameterized **E**nergy/**T**ime/**A**rea model for cache memories, but:

- Very old device level parameters (1993, 0.8um)
- Outdated circuits for current technology

Extended:

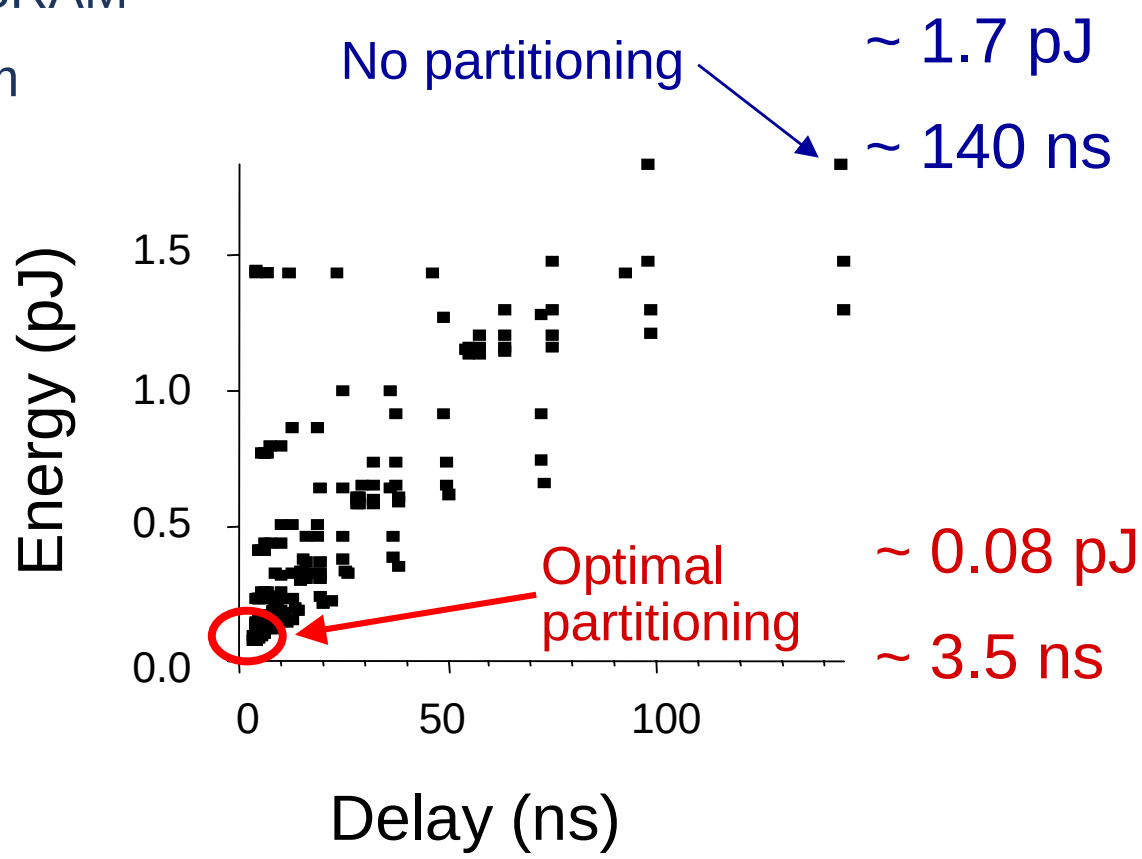
- Accurate wire models for current/future interconnect technologies
- Balanced contribution of silicon dominated blocks according to state-of-art design
- Exploration of interconnect implementation parameters for each possible long memory lines (bit/word lines, …)

[1] Wilton & Jouppi, 1996

Existing module level exploration: memory partitioning

64kbit SRAM

@ 45nm

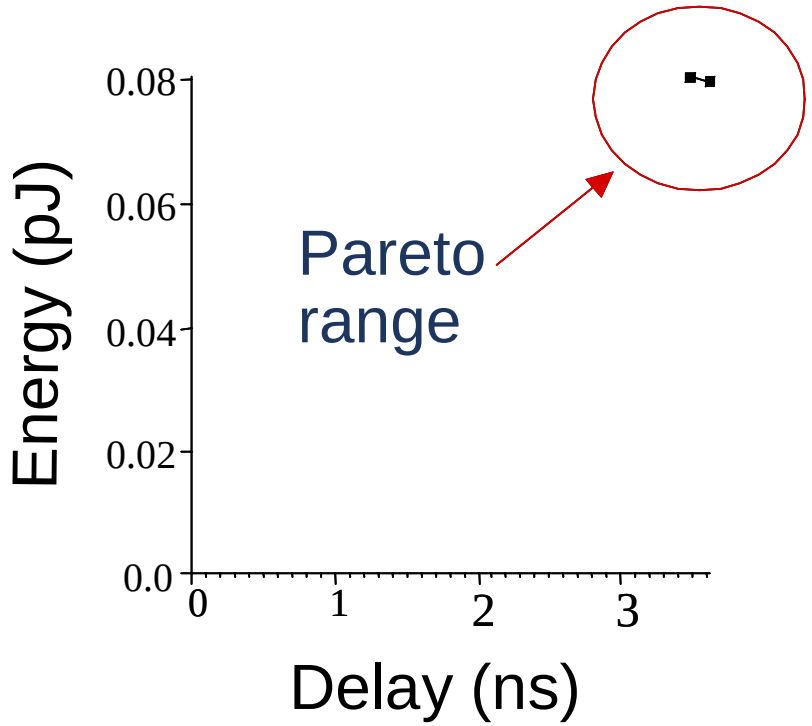


Partitioning a memory heavily improves energy and delay

Current design practice leads to very small trade-off range

64kbit SRAM

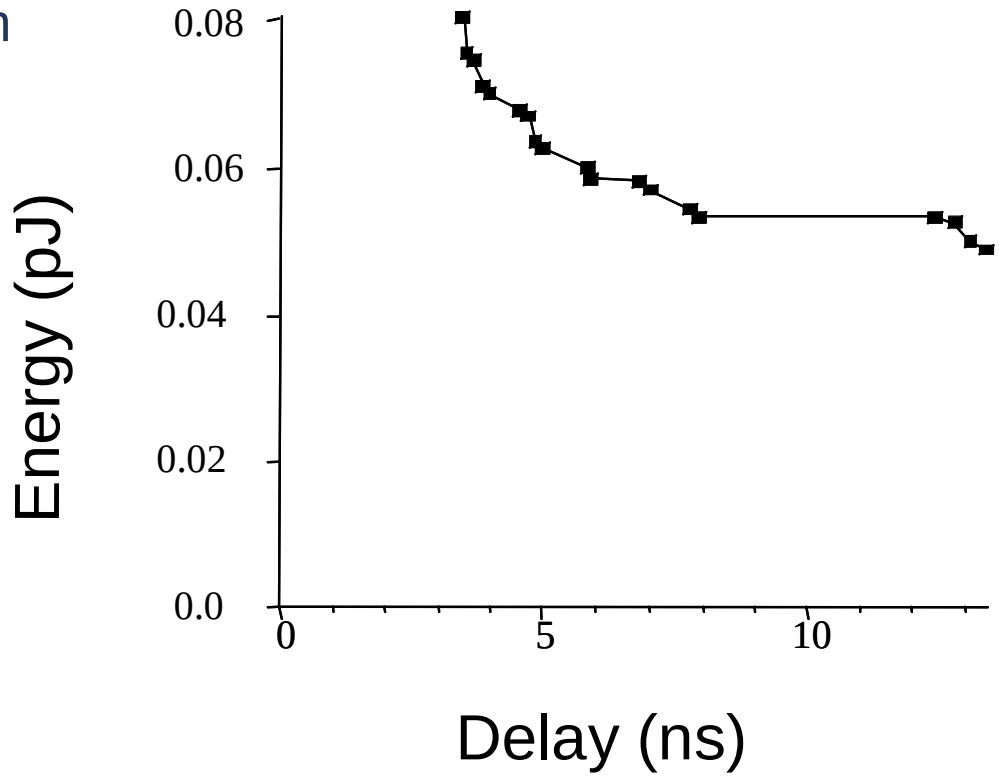
@ 45nm



Currently this is the main option considered by designers for energy-delay trade-offs

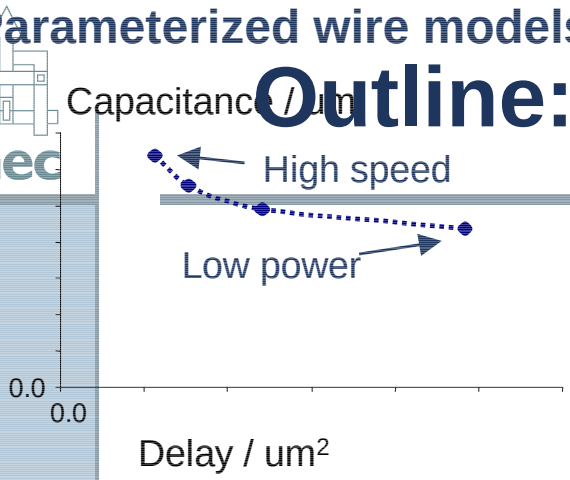
Combined technology&module level exploration gives good range

64kbit SRAM
@ 45nm



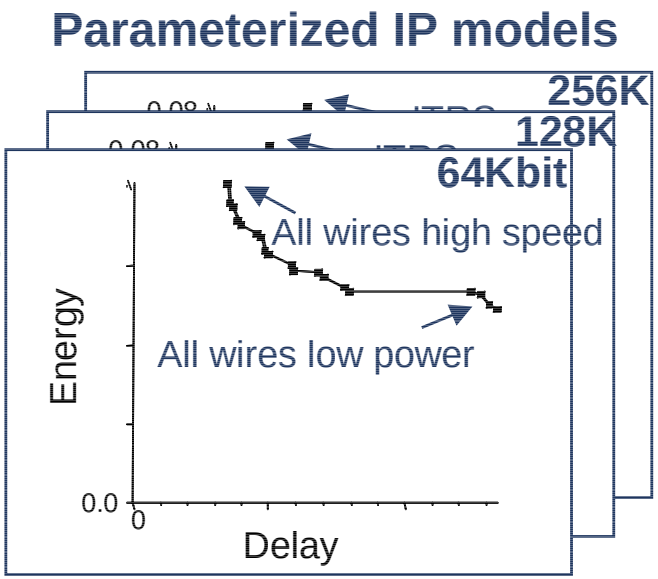
- wire_options * memory_line_types * partitioning_options
- Few technology options are sufficient for good range

Outline: Application level exploration

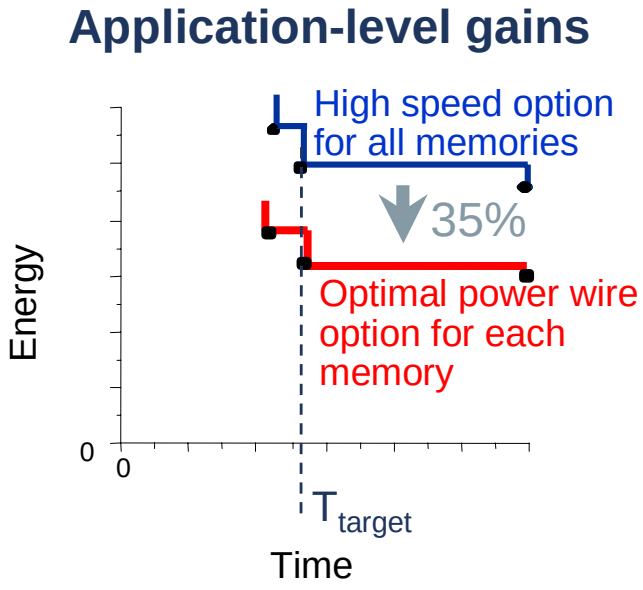


3) Use application feedback for process parameter selection

1) Add wire Model to IP model

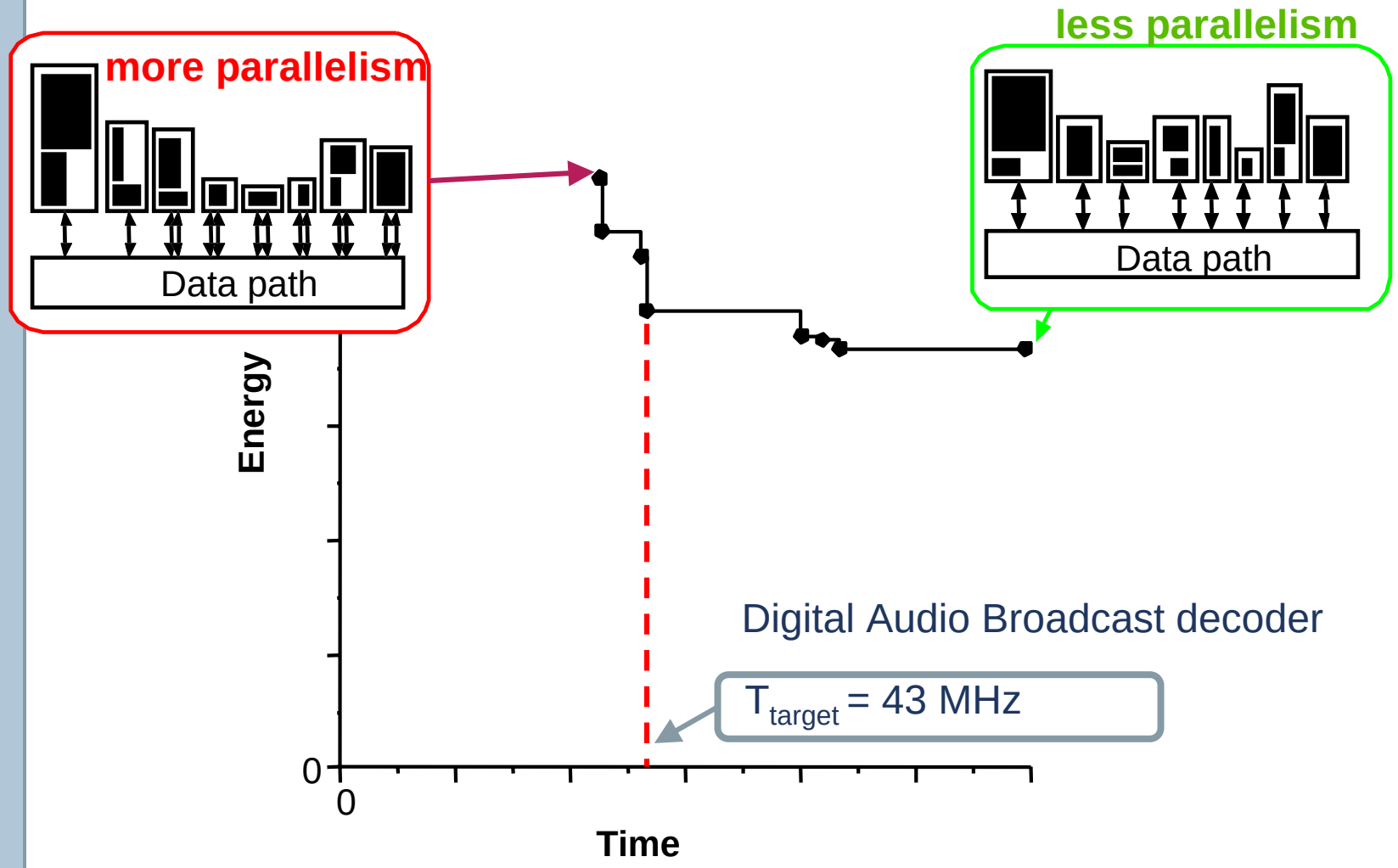


2) Use IP model + crude floorplan in application exploration (tools!)

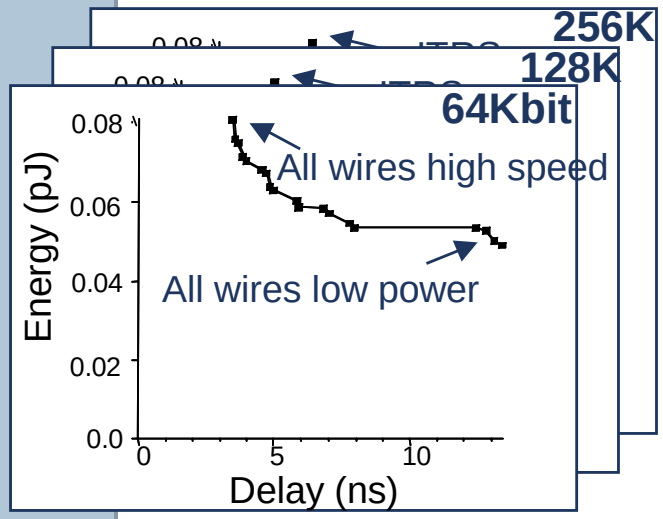


Complementary trade-offs at system-level for design exploration

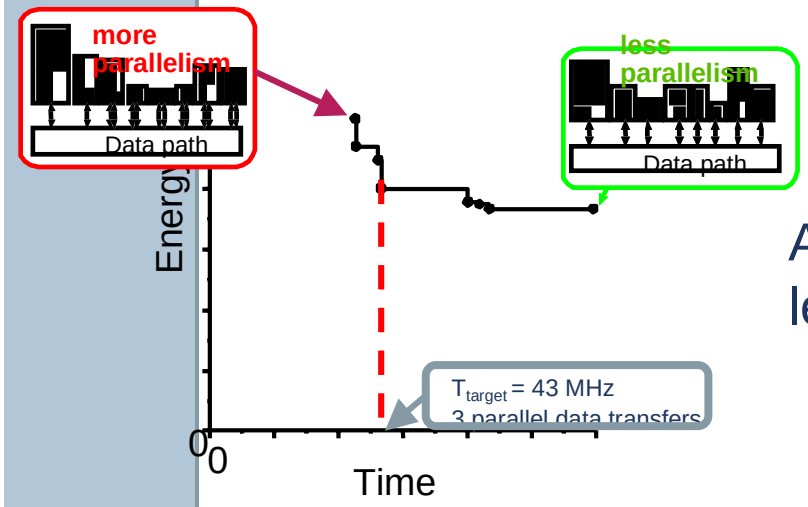
IMEC's Data Transfer and Storage Exploration Methodology



Module level exploration improves application-level trade-offs

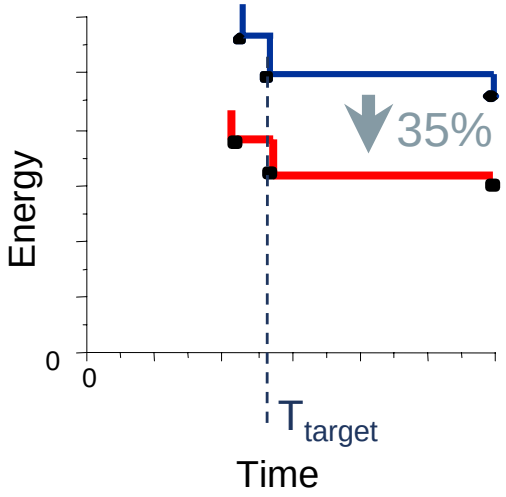


Module level trade-offs

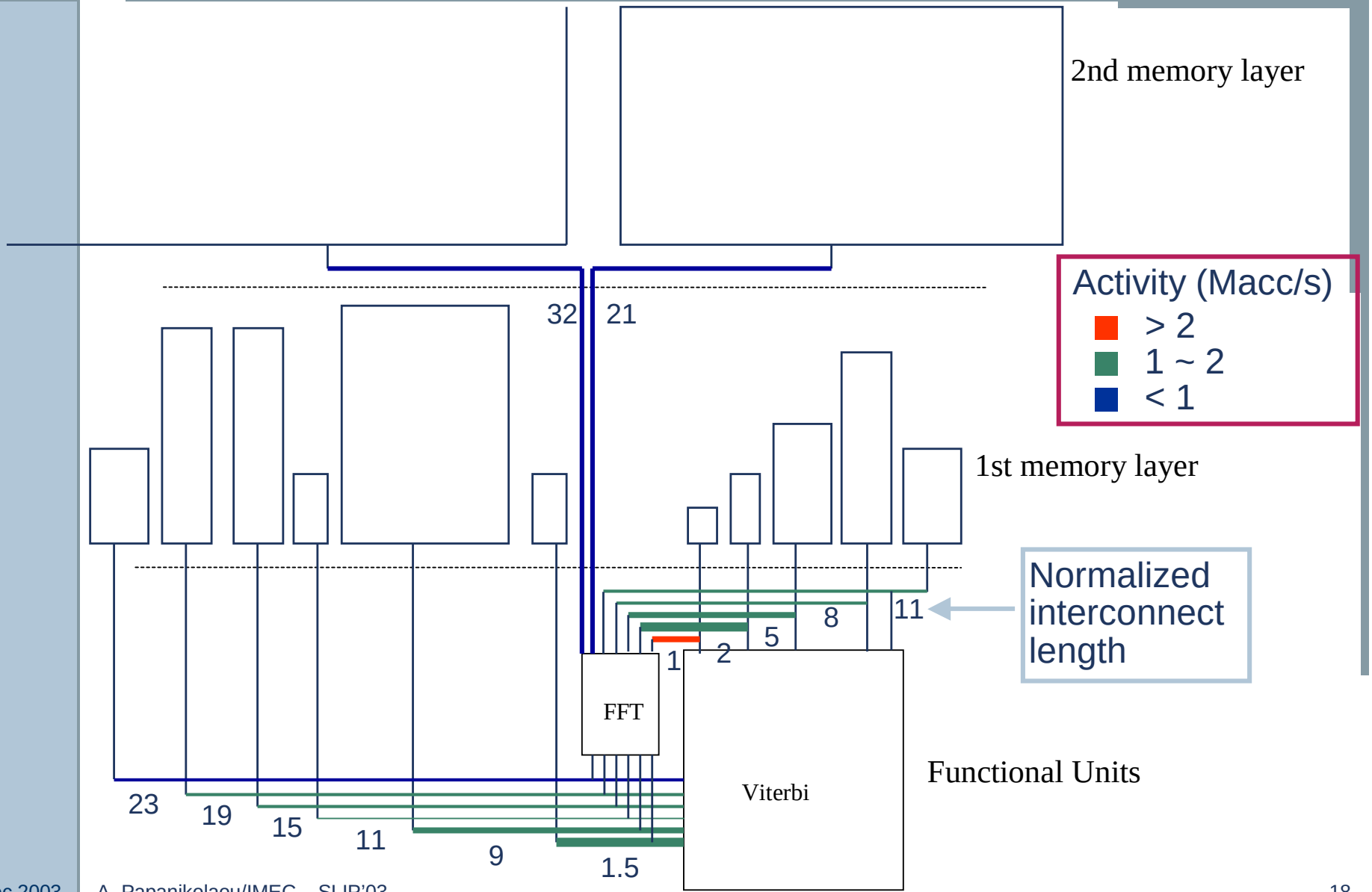


Application level trade-offs

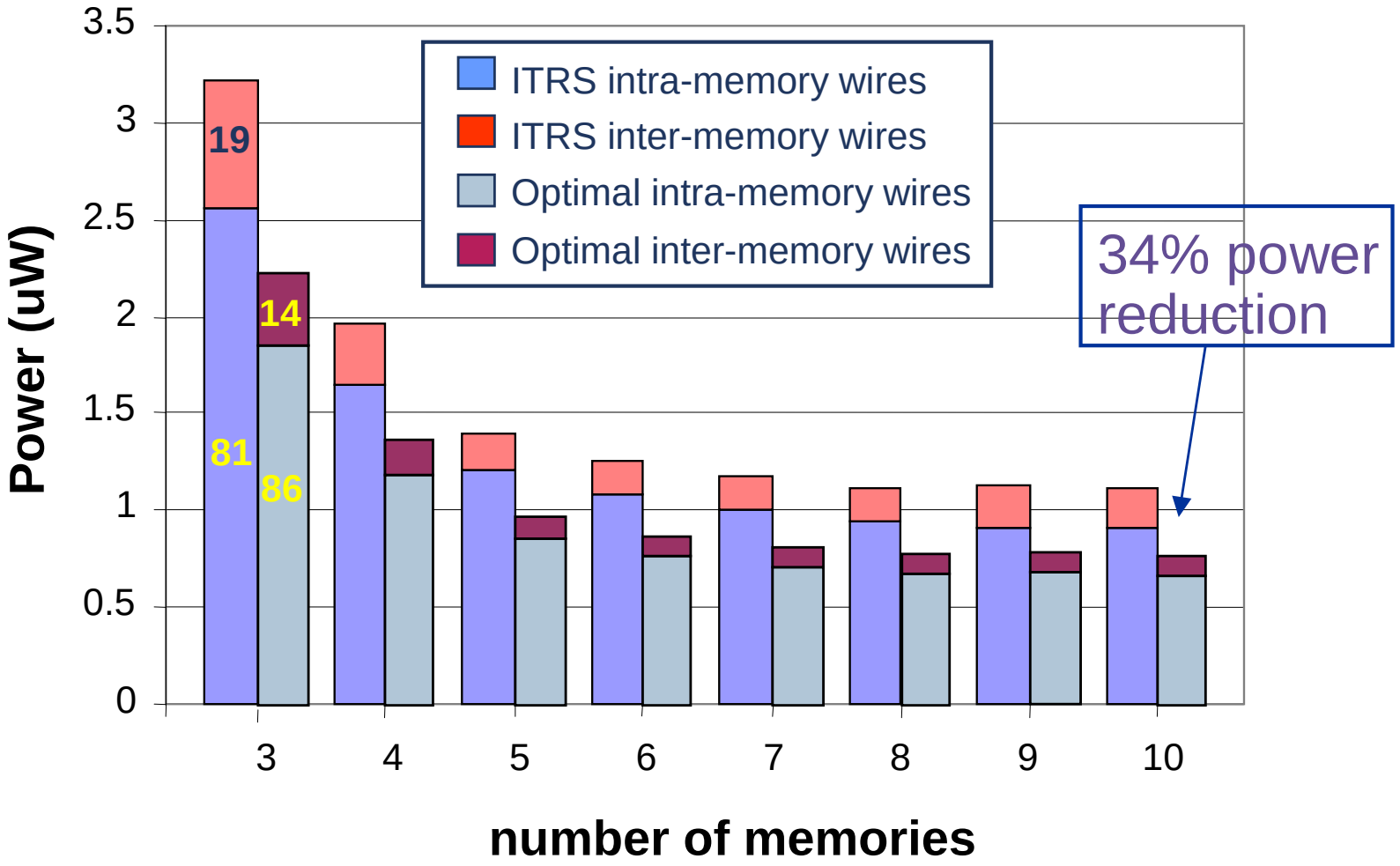
Application-level gains



Activity aware floorplan reduces power in inter-memory interconnect



Application-level power gains



Conclusions

- Technology level energy-delay trade-offs can propagate through the IP blocks to influence system level power consumption:
  - A 34% gain was achieved for the DAB
  - Feasible to steer technology related choices from the system level
- Inter-memory interconnect contribution is not important
- Wire parameter exploration could provide power or delay gains at system level without necessarily using advanced (and expensive) interconnect materials

SEEDS FOR TOMORROW'S WORLD

IMECNOLOGY

www.imec.be

Worldwide collaboration with more than 450 companies and institutes.

IMEC – Kapeldreef 75 – B-3001 Leuven – Belgium – Tel. +32 16 281211 – Fax +32 16 229400 – www.imec.be

Wire trade-off explanation

$$\frac{RC}{l^2} = 8k\rho\varepsilon_0\left[\frac{1}{p^2} + \frac{1}{p^2 - 4\Delta w^2}\right]$$

$$\frac{C}{l} = 2k\varepsilon_0 \frac{2p^2 - 4\Delta w^2}{p^2 + 4p\Delta w}$$

if Δw increases => delay increases & capacitance decreases => Trade-off!

(p/2 – Δw) is the width of the interconnect wire,
p is the technology pitch

Comparison with existing implementation

DAB implementation

- @ 350nm consumes ~2.6 mW @ 43MHz
- @ 45nm is projected to consume ~0.7 uW @ 43 MHz

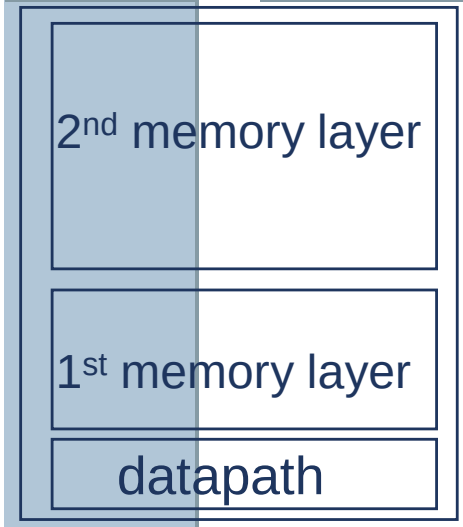
Projected reduction (x256) comes from:

- Vdd scaling (factor 16 = (2.0/0.5)^2) – one time drop
- scaling of physical dimensions (factor 8 =350/45) – 7 tech generations (huge investment)
- exploration in interconnect (factor 2) – smaller investment & extendable

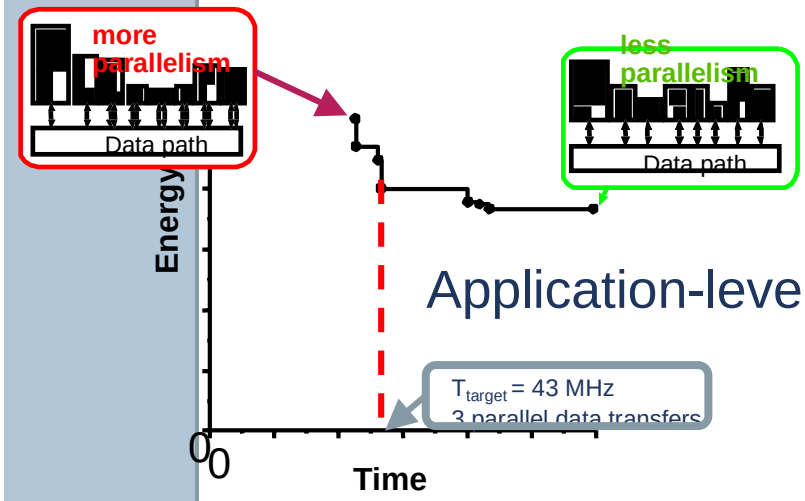
Still ~ factor 10 missing, may be due to:

- different memory types (DRAM v. SRAM for big 2 mems)
- different memory partitioning (factor 2 @ 350 nm)
- different memory component circuits and routing network

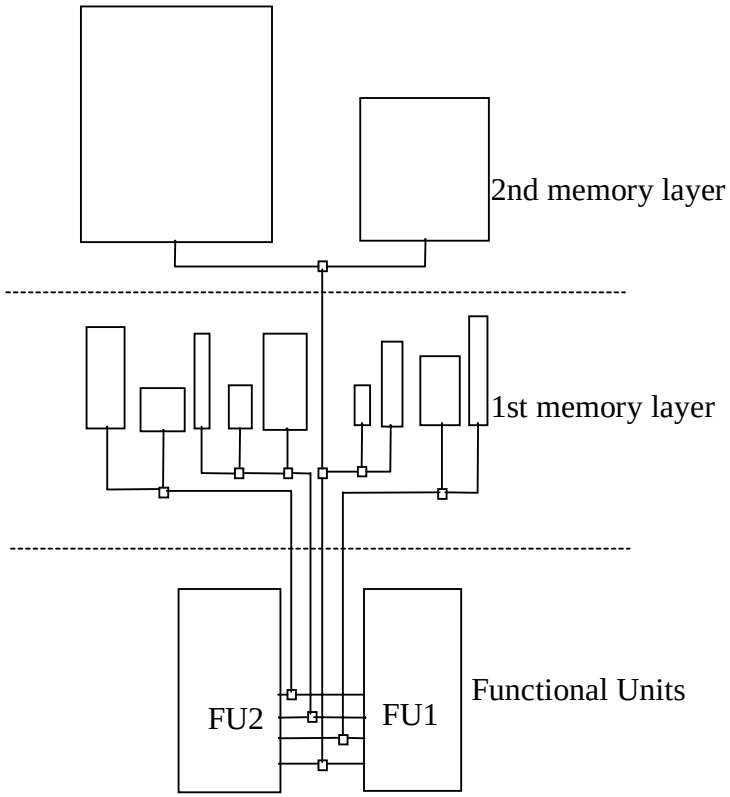
Application level exploration can steer physical design for power-aware systems



Floorplan template for inter-memory wire-length estimation



Application-level exploration



Inter-memory wire length depends on Signal to Memory Allocation&Assignment decisions (i.e., size,activity,…)