COIN-3D

Collaborative Innovation in 3D VLSI Reliability

Custom Tool Development Strategies for Chiplet Reliability

Task 2.3: Expert Webinars

M05 – M18

Date: Fri, 28th March 2025 12:00 - 13:15 CET

University of Glasgow

Funded by the European Union

Agenda

- **Introduction :** *Dr George Floros*
- **CoMeT: Interval Thermal Simulator for 3D ICs:** *Dr Anuj Pathania*
- **Physical Hierarchy Exploration of 3-D ICs:** *Prof Alberto Garcia-Ortiz*
- **PROTON: Physics-Based Electromigration Assessment on Modern VLSI Power Grids:** *Dr Olympia Axelou*
- **Q&A**

Funded by
the European Union

COIN-3D Webinar Series

- COIN-3D goal:
  - Research and knowledge exchange in the 3D VLSI domain.

- **Webinar 1 (M05)** – Machine Learning Applications in EDA for Chiplet Reliability ✅
  - Check our YouTube channel for the webinar
- **Webinar 2 (M07)** – Custom Tool Development Strategies for Chiplet Reliability
  - You are here
- **Webinar 3 (M15)** – Reliability Assessment in Emerging Technologies: Beyond 2.5/3D Chiplets (Upcoming)

Funded by
the European Union

CoMeT: Interval Thermal Simulator for 3D ICs

Funded by
the European Union

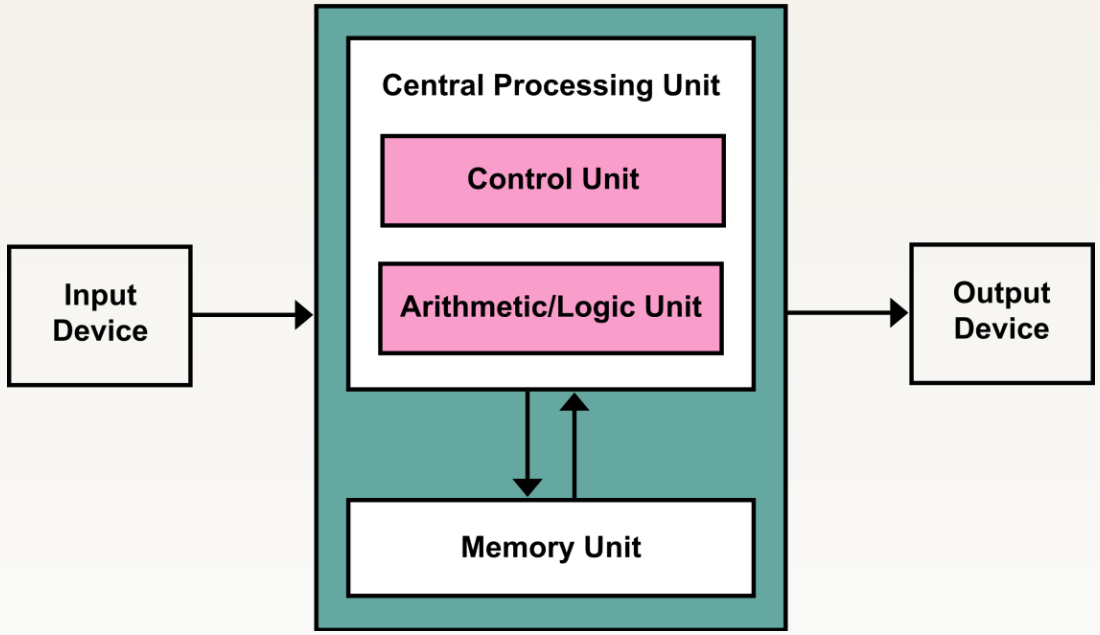
About

- Originally from New Delhi, India.
- Ph.D. from Karlsruhe Institute of Technology
- Started in University van Amsterdam (UvA) @ Sep 2020

- Area of Expertise
  - Low Power Design
  - Embedded Systems
  - Electronic Design Automation

Funded by
the European Union

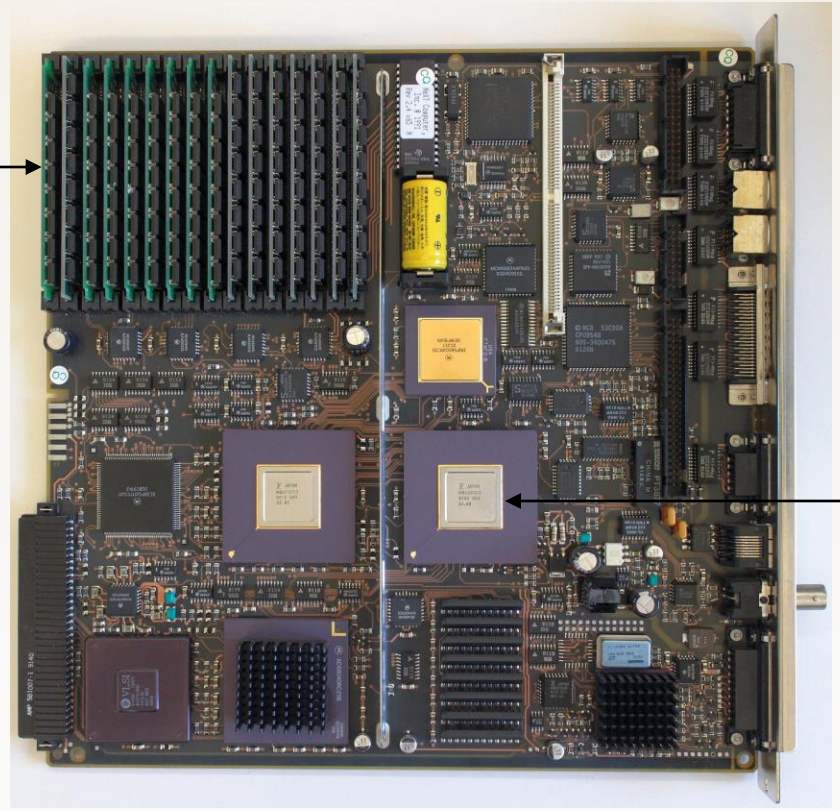
Von Neumaan Architecture



Funded by the European Union

Source: Wikipedia

2D Processors



Motherboard (Interconnect)

Funded by the European Union

Source: Wikipedia

High-End Applications

3D Games

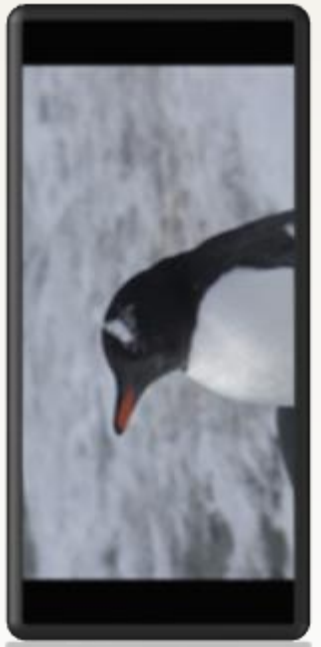
Image Recognition

Speech Recognition

Virtual Reality

Augmented Reality

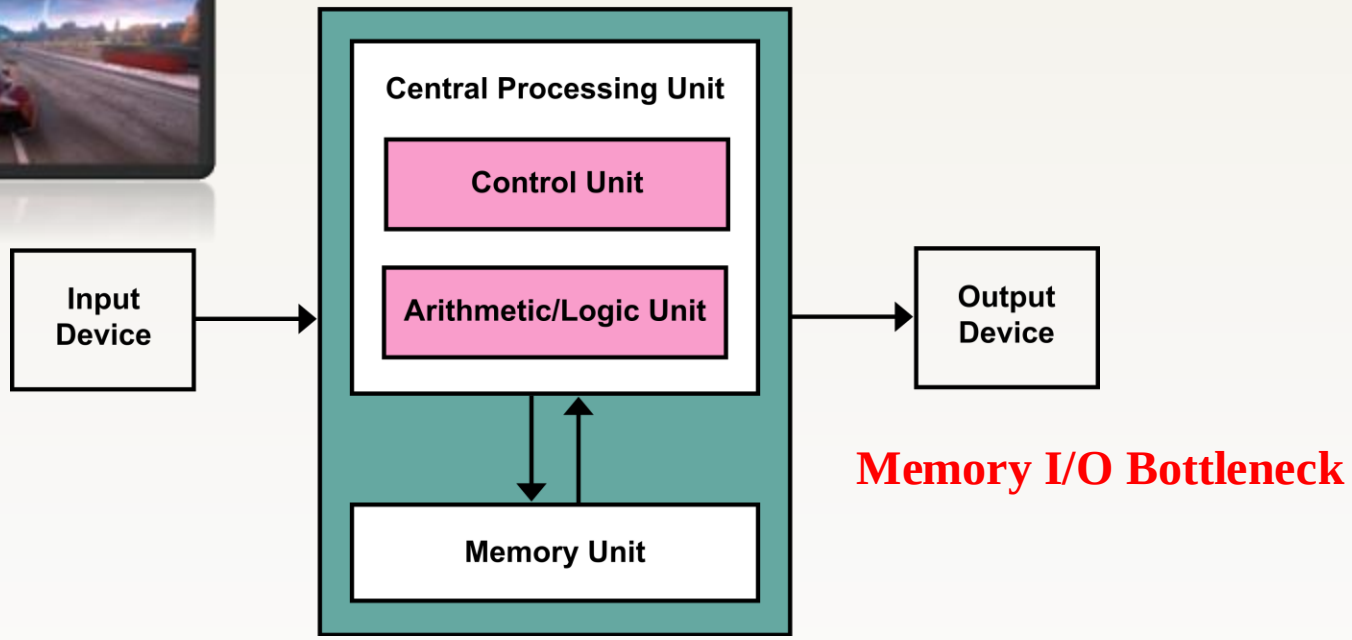
High-Def Video



Funded by the European Union

Source: Wikipedia

Memory Bottleneck

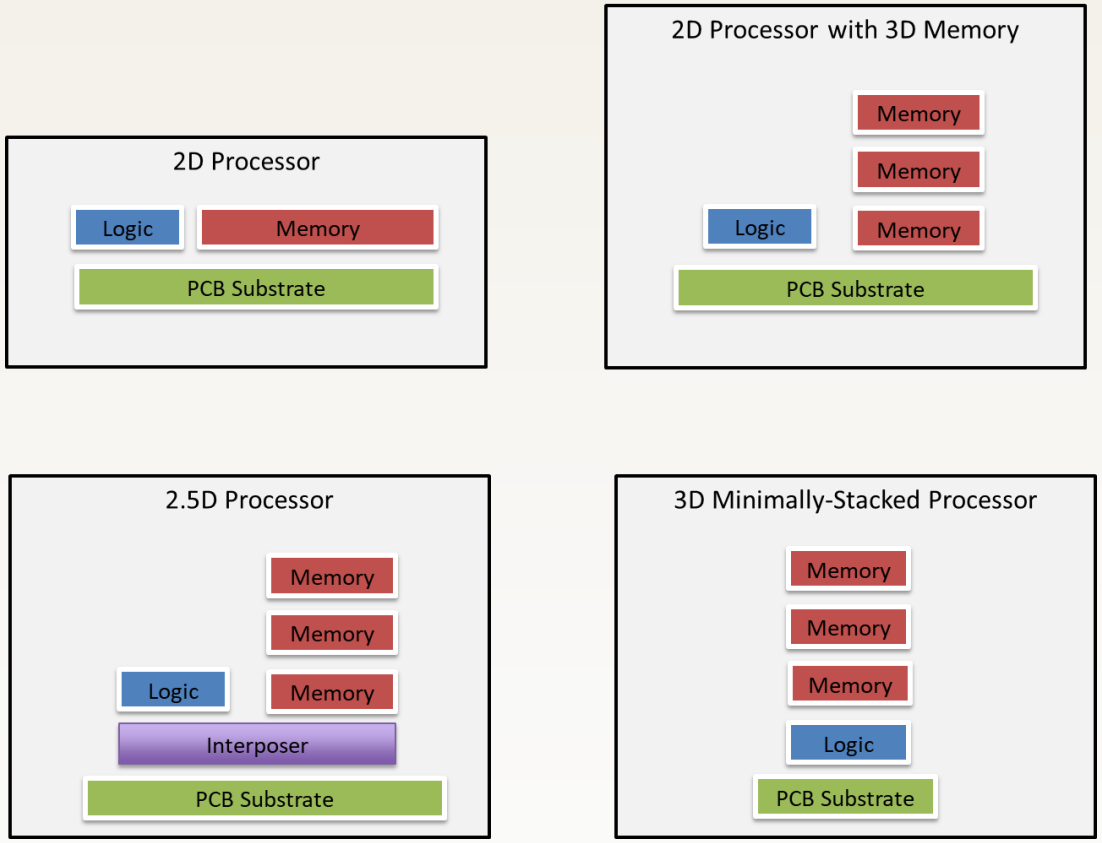


Memory I/O Bottleneck

Funded by the European Union

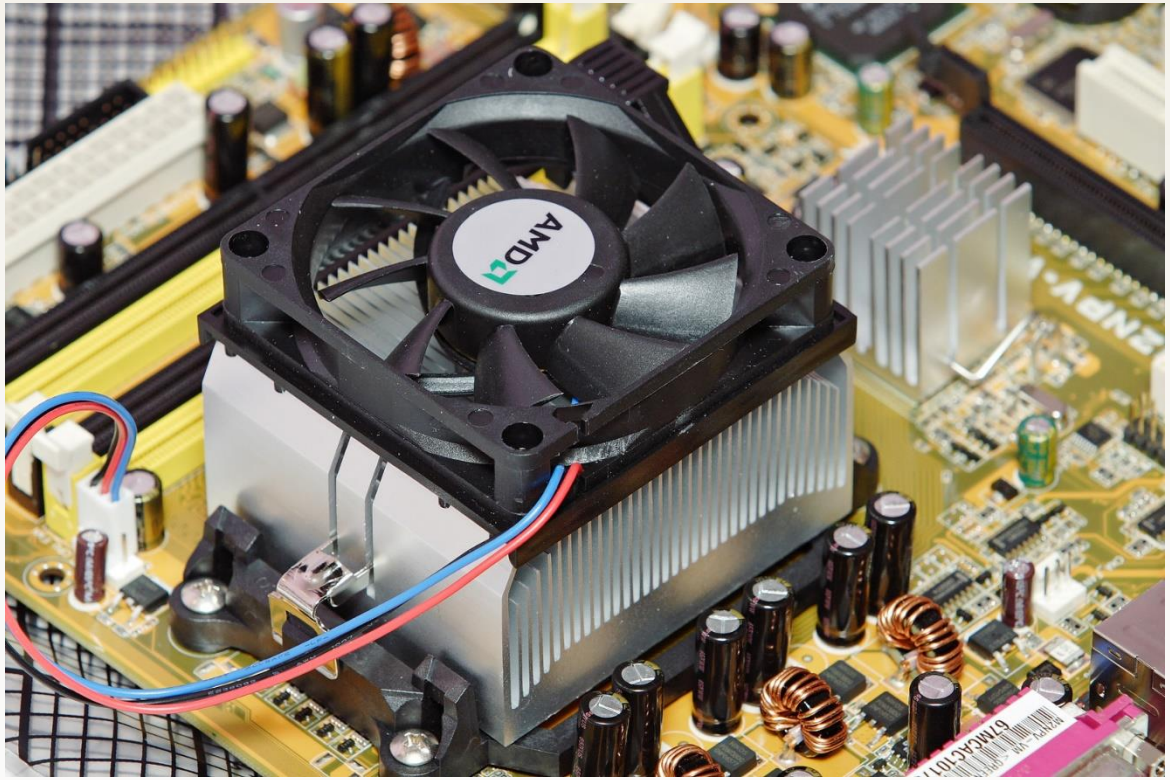
Source: Wikipedia

Memory-Wall Driven Processor Evolution



Funded by
the European Union

Thermal Problem

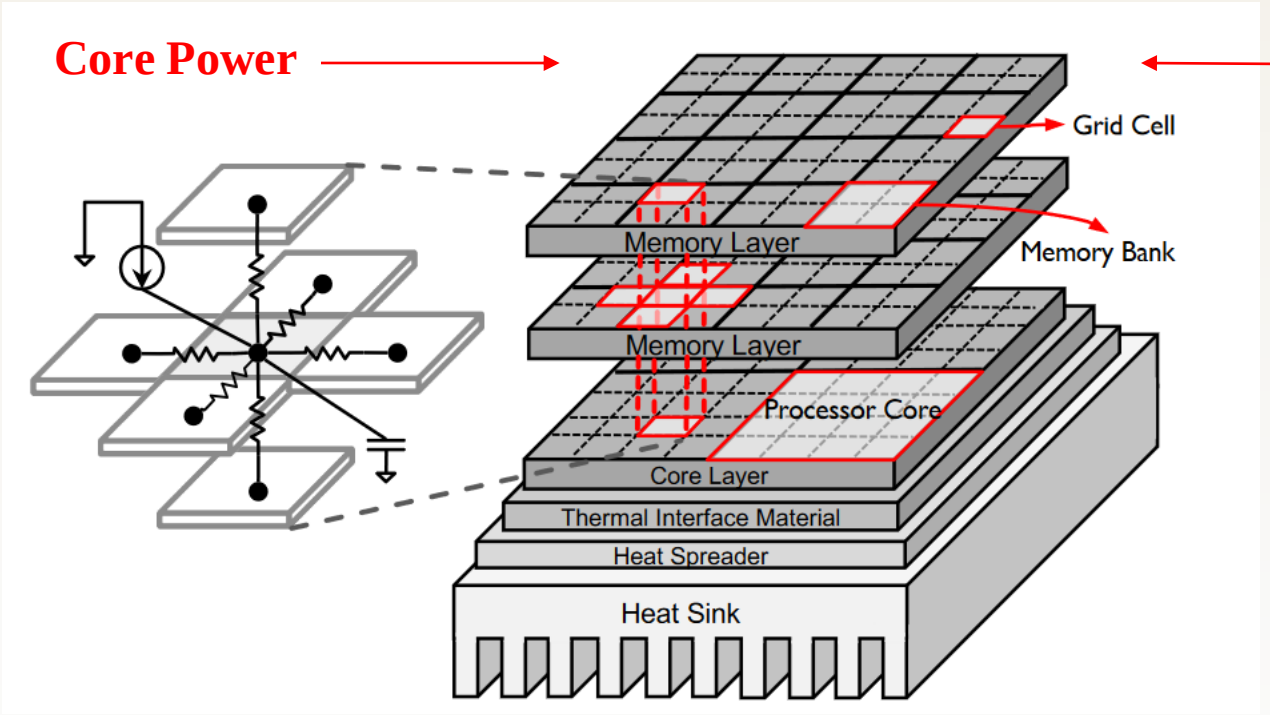


Heat is the **No.1** enemy of high-performance processors.

Funded by
the European Union

Source: Wikipedia

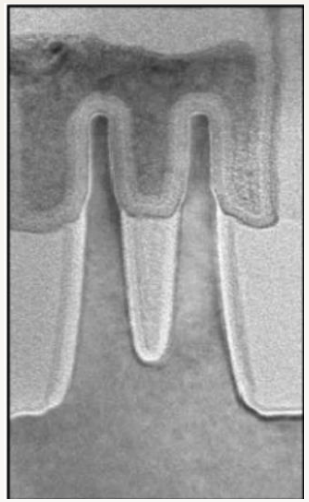
3D Stacked Processor



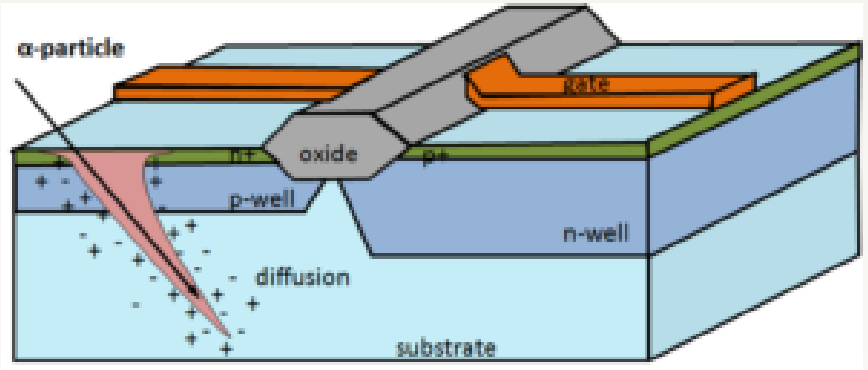
Funded by
the European Union

Source: Wikipedia

Reliability Challenges



Hard Reliability



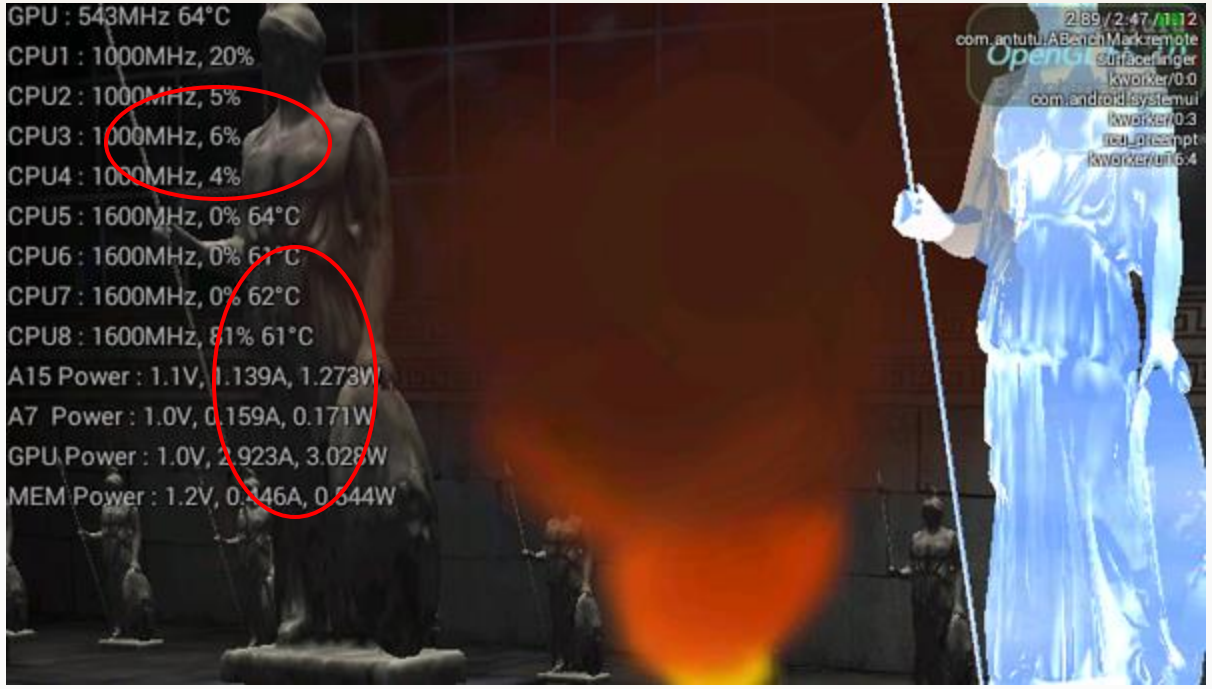
Soft Reliability

Funded by
the European Union

Source: Wikipedia

Need for Thermal Simulations

- Many real-world processors have thermal sensors.
  - However, resolution is not very high – spatially and temporally.
- Almost all thermal-related research done on Simulators.

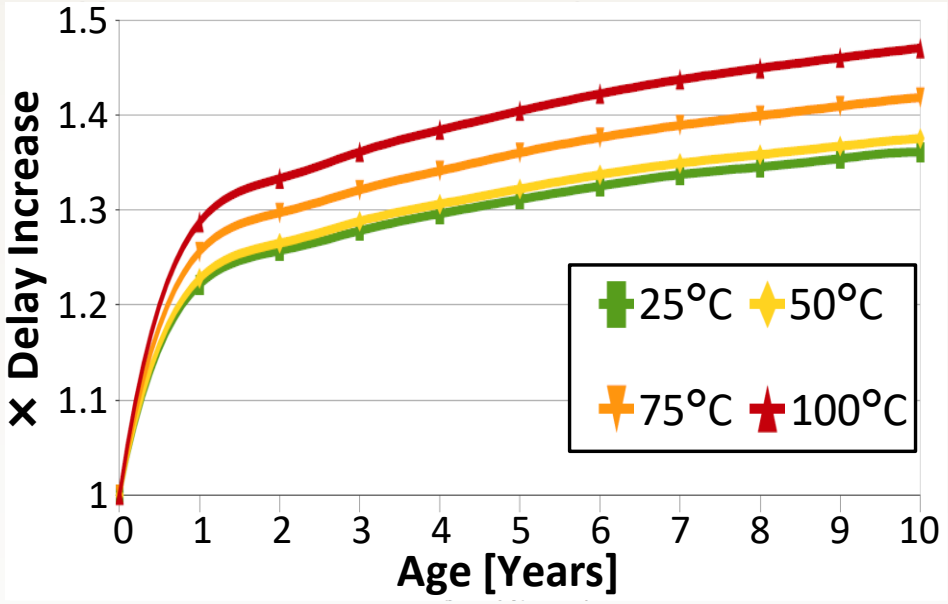


Funded by
the European Union

Source: Odroid

Need for Reliability Simulations

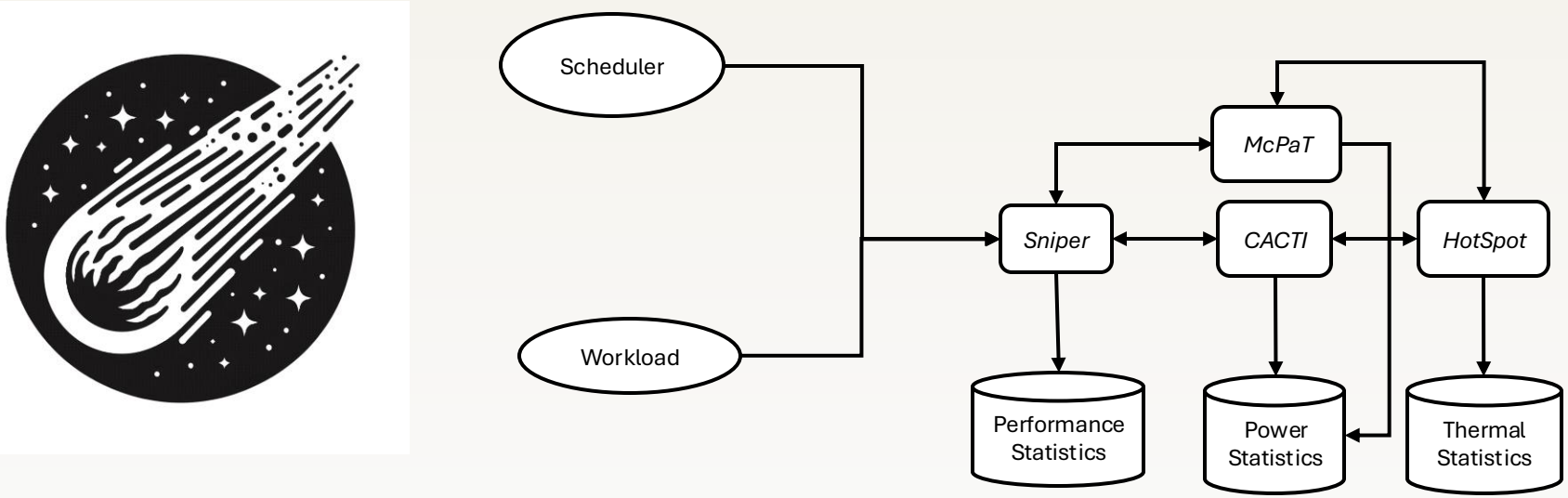
NBTI-Induced Aging for a CPU Pipeline



Funded by
the European Union

Source: Henkel et al. @ DAC'13

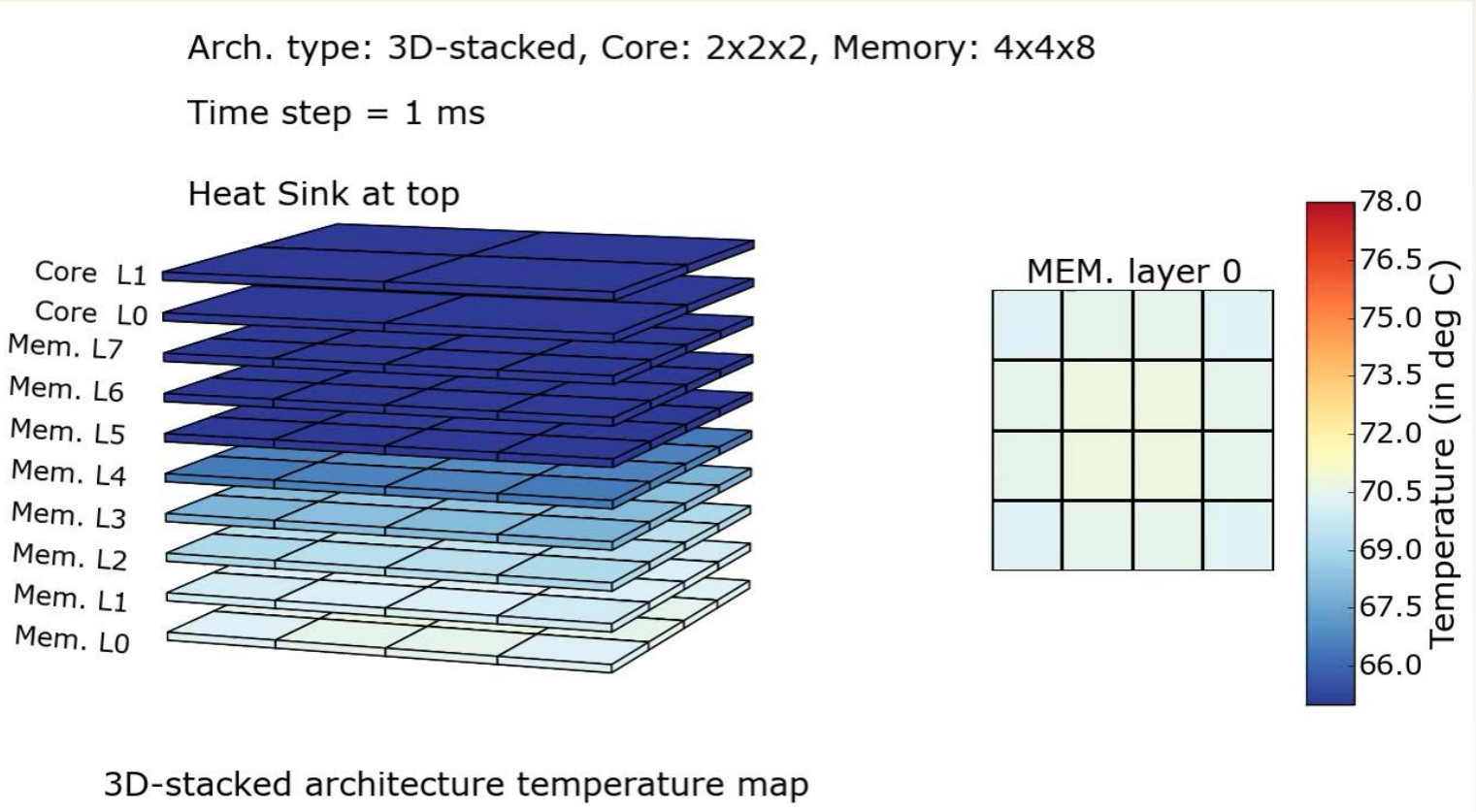
CoMeT: Interval Simulation Toolchain



An EDA toolchain for integrated core-memory interval thermal simulations of 2D, 2.5, and 3D multi-/many-core processors

Funded by the European Union

Simulation Example



3D-stacked architecture temperature map

Funded by
the European Union

Physical Hierarchy Exploration of 3-D ICs

Funded by
the European Union

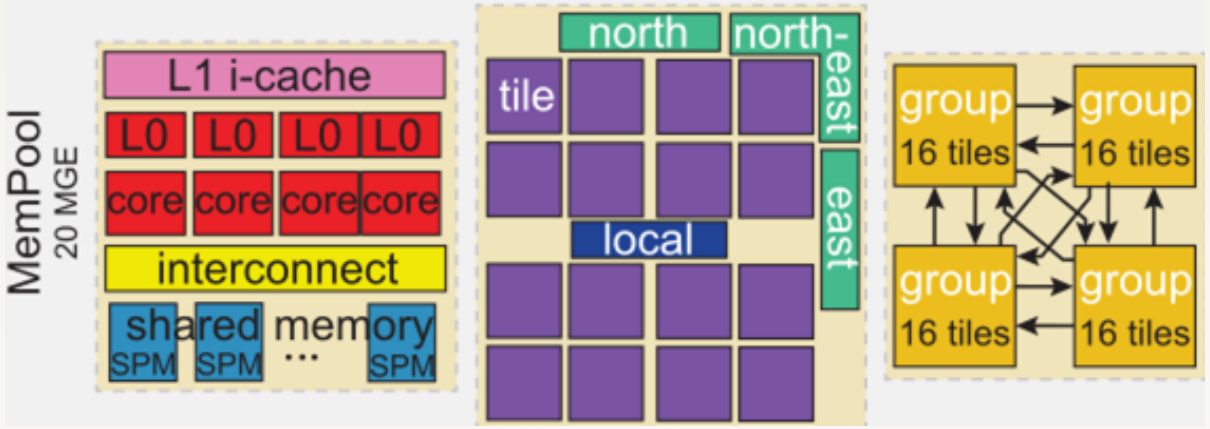
Acknowledge

- Nesara Eranna Bethur, Anthony Agnesina, Sung Kyu Lim
- Francky Catthoor, Dragomir Milojevic, Manu Komalan, Moritz Brunion
- Matheus Cavalcante, Samuel Riedel, Luca Benini
- Moritz Brunion, Alberto Garcia-Ortiz

Funded by the European Union

Hierarchical approach

- Subdivide complex, manycore, large-memory giga-scale designs into sub-blocks



- Independently synthesized and physically placed-and-routed as separate design units.
- Recombined into subsequent runs of higher-level blocks. Repeated as the hierarchy is traversed up to the top level.
- **Standard methodology in 2D. How to perform it in 3D?**

Funded by
the European Union

Benefits hierarchical approach

- Large designs can be implemented with acceptable **runtime** and memory usage
- **Reuse** of identical blocks (e.g. multicore)
- **Concurrent implementation** of design tasks across multiple development teams
- Elegant implementation of third-party intellectual property (**IP**) blocks

Easy design exploration

- **But**: budgeting for block-level timing constraints
- **But**: Setting appropriate hierarchical physical constraints

Funded by
the European Union

3D Design flows

- Sequential-2D flows
- Macro-3D flow
- Hier-3D flow

	Sequential-2D	Macro-3D [12]	Hier-3D
key idea	implement two 2D dies separately	holistic memory-on-logic	hierarchy scheme for 3D
3D stack	two separate dies	double metal stack	double metal stack
main strengths	reuse of 2D environment, macros and standard cells in both tiers	full utilization of metal resources	full utilization of metal resources, maximize silicon area utilization
main weakness	dies are designed separately: limited optimization	limited to memory-on-logic: best suited for equal die area	inherently hierarchical: best suited for large designs
restricted partitioning	no	yes	no
holistic routing	no	yes	yes
utilize unused space	no	no	yes

Funded by
the European Union

3D Design flows

- Sequential-2D flows
  - Separate implementation of stacks (defining vertical pins)
  - **Strength**: No restrictions for partitioning
  - Limitation: Critical partitioning step
  - Limitation: Complex modeling of constraints
  - Limitation: No reuse of BEOL resources
- Macro-3D flow
- Hier-3D flow

Funded by
the European Union

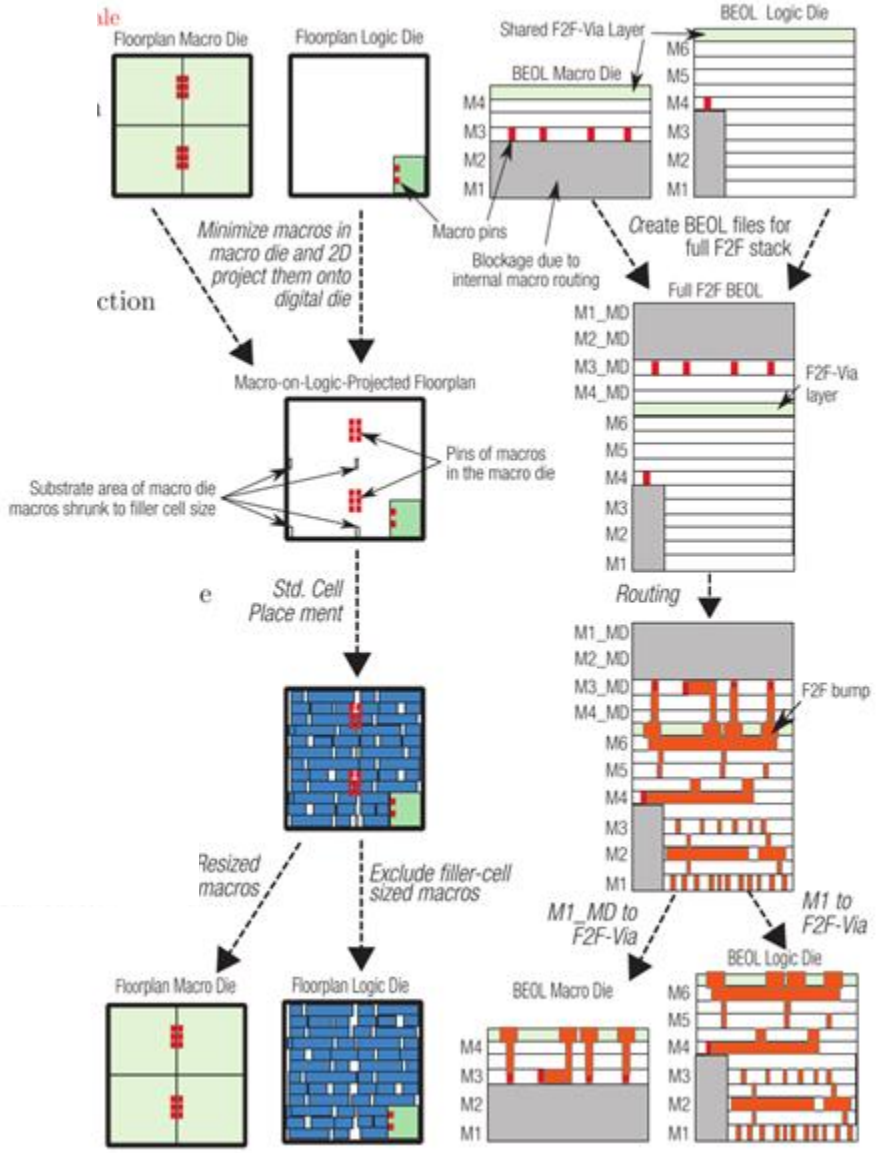
3D Design flows

- Sequential-2D flows
- Macro-3D flow
  - 2D tools are made aware of complete die stack
  - **Strength**: Complete view of timing and BEOL (metal reuse)
  - **Strength**: Full PR optimization (incremental synthesis)
  - Limitation: Focus on Memory-on-Logic
  - Limitation: Unused space in tier
- Hier-3D flow

Funded by
the European Union

Macro-3D: Overview

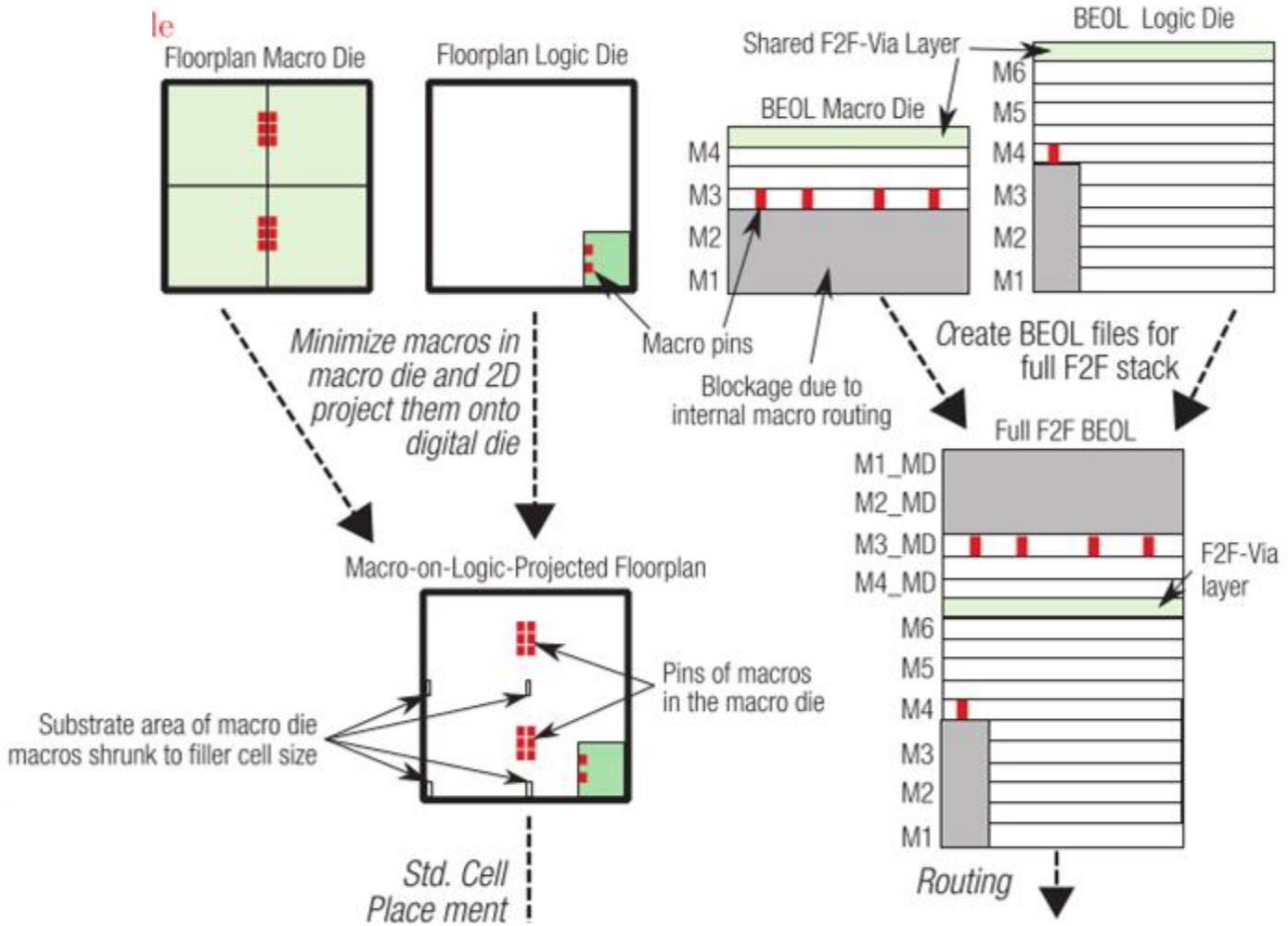
- Floorplan
- 2D Projection
- 2D P&R and sign-off
- Separation



- Create full BEOL
- Create IP with projection

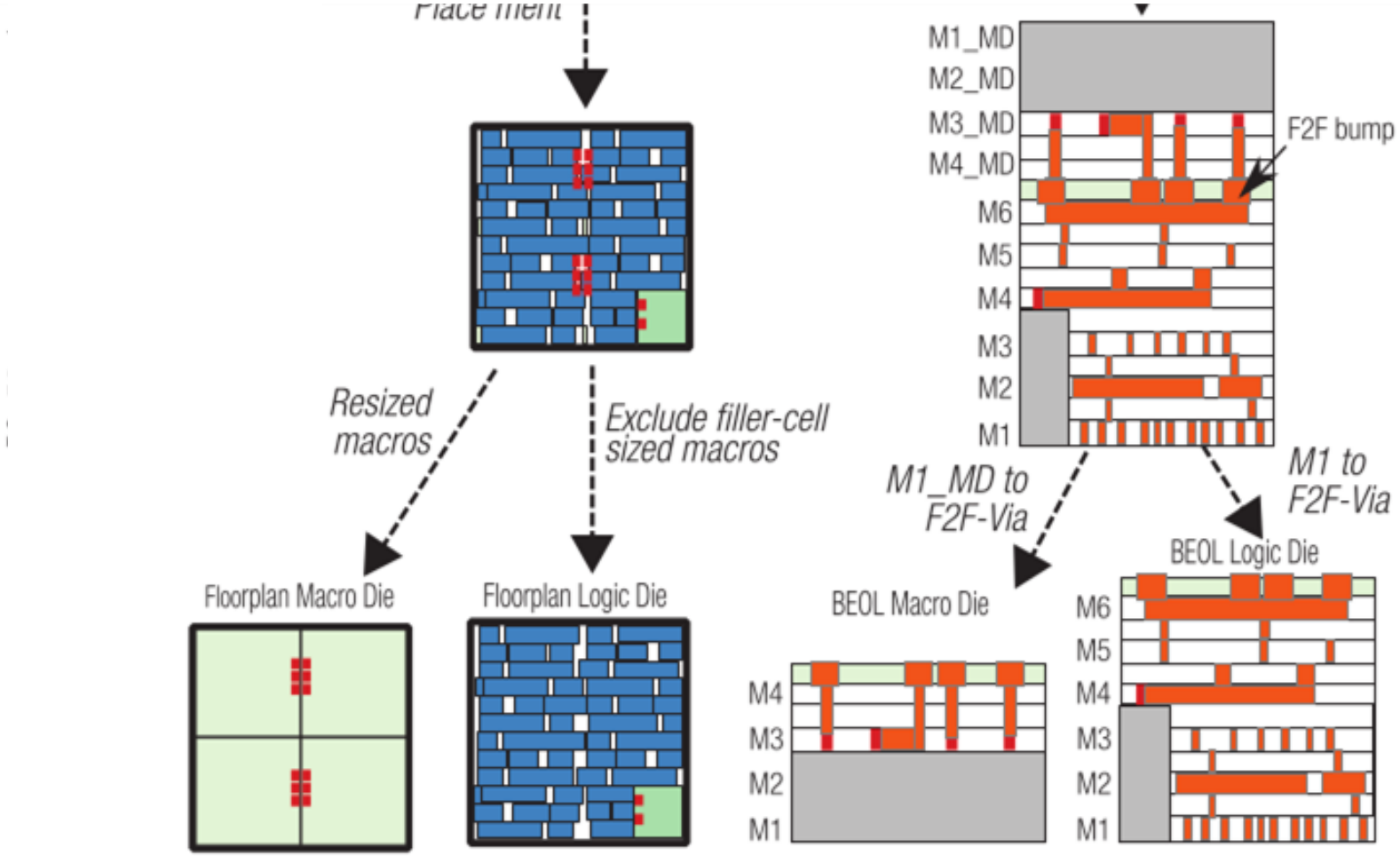
Funded by the European Union

Macro-3D: Preparation



Funded by
the European Union

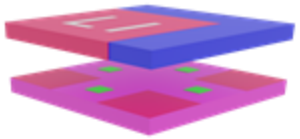
Macro-3D: Final step

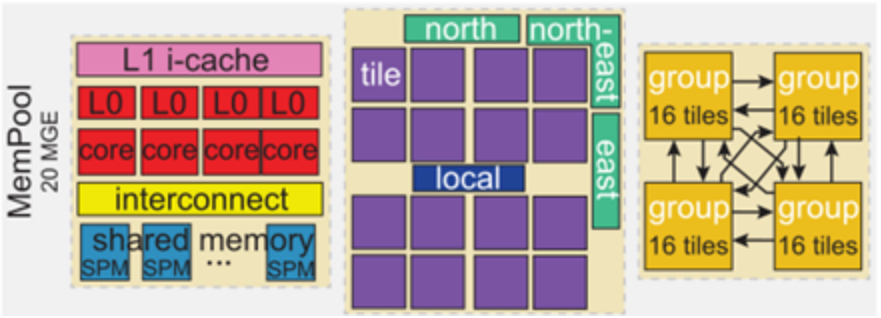


Funded by
the European Union

Hier-3D: Idea

- Combine advantages of Sequential-3D and Macro-3D
- 3-D bottom-up hierarchical implementations composed by cycles/iterations

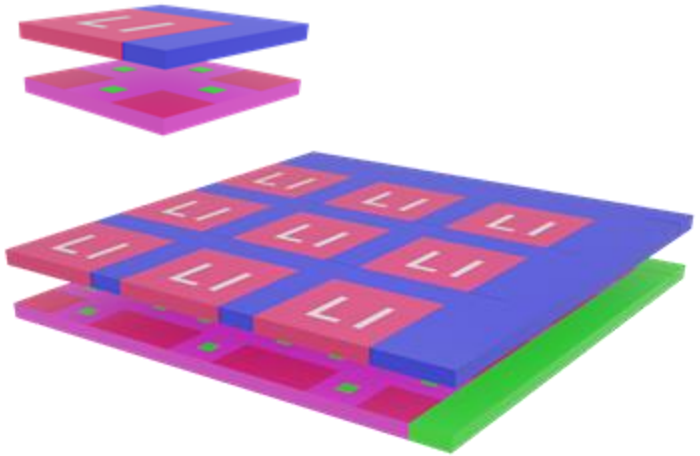


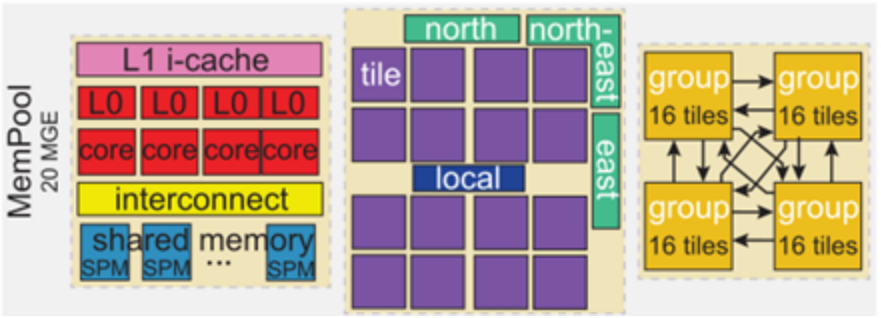


Funded by
the European Union

Hier-3D: Idea

- Combine advantages of Sequential-3D and Macro-3D
- 3-D bottom-up hierarchical implementations composed by cycles/iterations

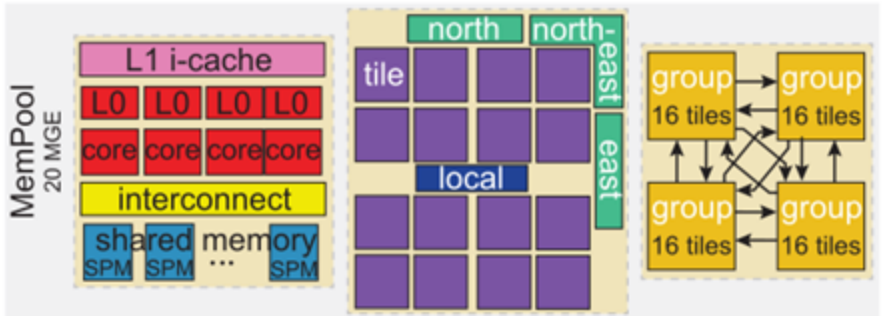


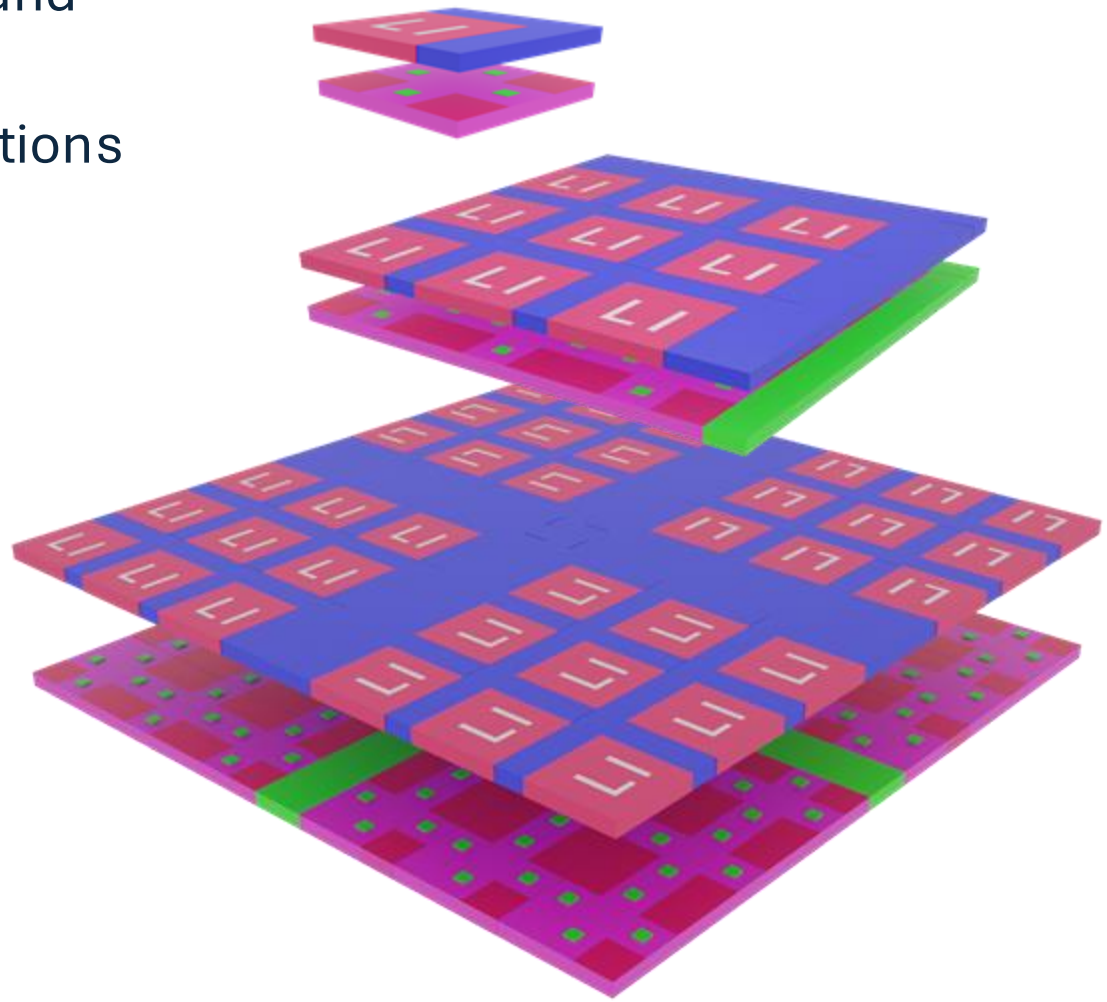


Funded by
the European Union

Hier-3D: Idea

- Combine advantages of Sequential-3D and Macro-3D
- 3-D bottom-up hierarchical implementations composed by cycles/iterations

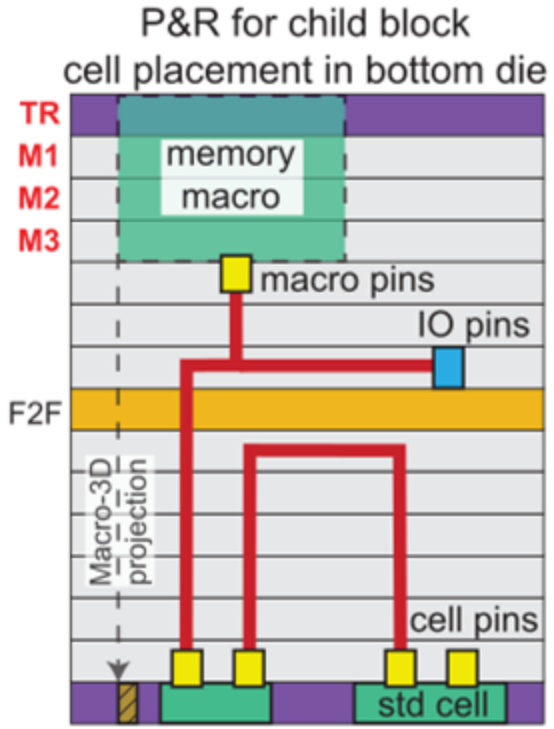




Funded by
the European Union

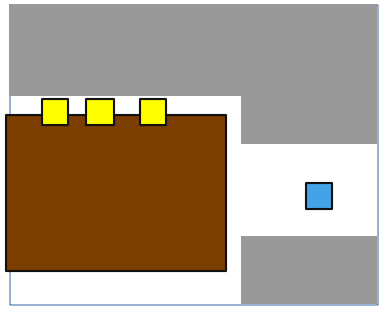
Hier-3D: Idea

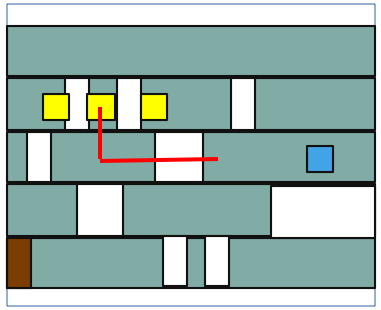
- Combine advantages of Sequential-3D and Macro-3D
- 3D bottom-up hierarchical implementations composed by cycles/iterations
- Each iteration like Macro-3D
  - Full routing in one tier
  - Stack inversion for next iteration

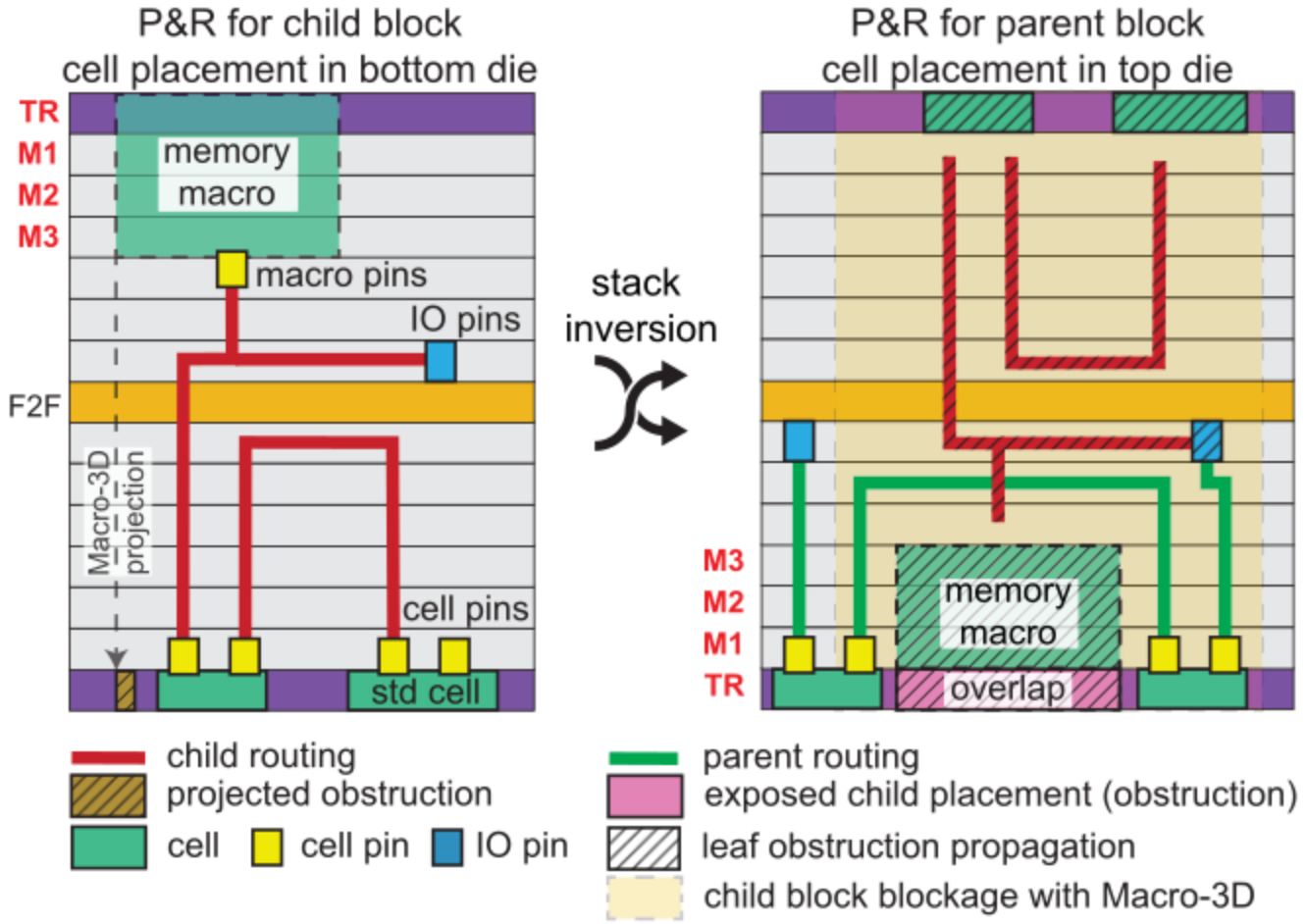


Funded by
the European Union

Hier-3D: Physical View







Funded by
the European Union

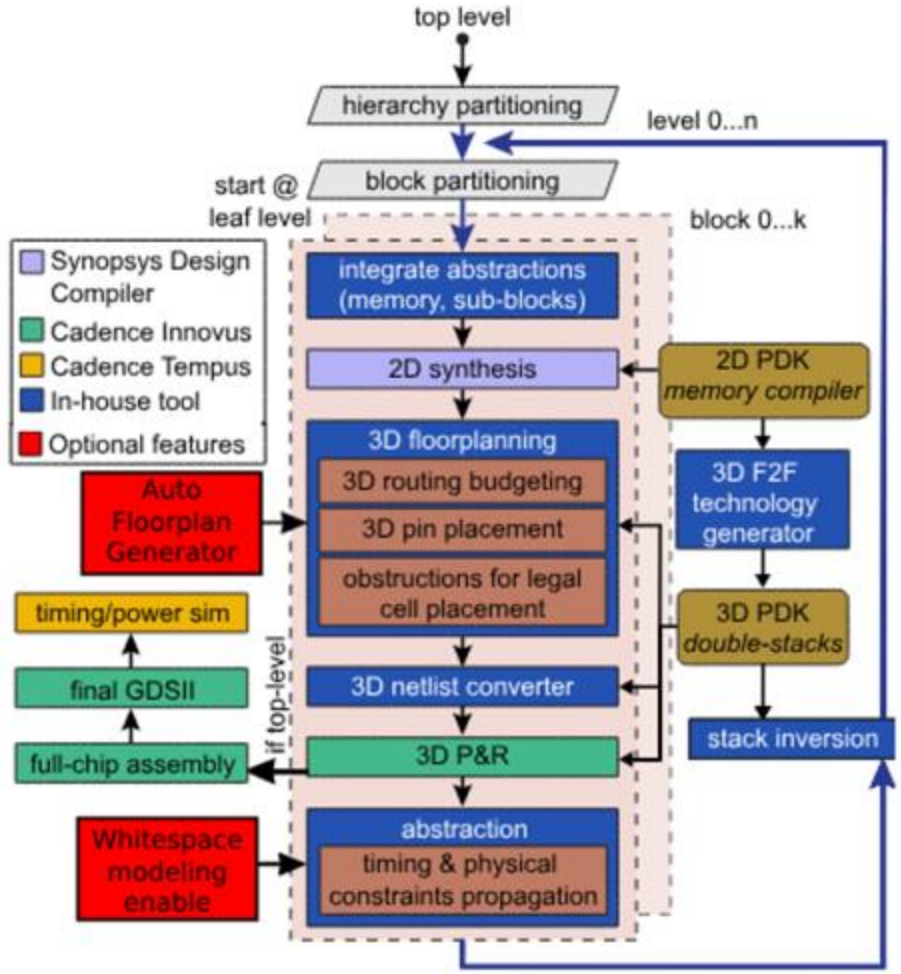
Hier-3D: Summary of methodology

- 3D **bottom-up hierarchical** implementations
- **Reusing the unassigned area** of preceding hierarchy levels
- Hierarchical **multitier** standard cell placement, greatly expanding the design space of 3D ICs
- Requires:
  - RTL clustering, 3D logic-on-logic floorplanning, and holistic 3D physical implementations
- Optional:
  - Automatic floorplaning, whitespace modeling to propagate the physical shapes of the standard cells

Funded by
the European Union

Hier-3D Methodology. Key Steps

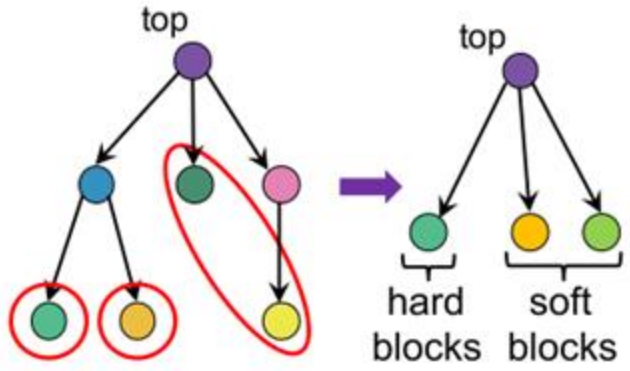
- Define number of cycles/iterations according to the hierarchy depth
- Per iteration, synthesized block-level designs are prepared with 3-D floorplanning, 3-D P&R, and abstracted through physical and timing constraints propagation.
- Stack and abstractions are inverted to enable standard cell placement o the opposite die at the upper hierarchy level
- Optional automati floorplan generator and whitespace modeling



Funded by
the European Union

Semi-automated floorplaning: Clustering

- Block generation can be done manually (hierarchy) of automatically
- Sub-block creation using clustering based on size of soft-blocks



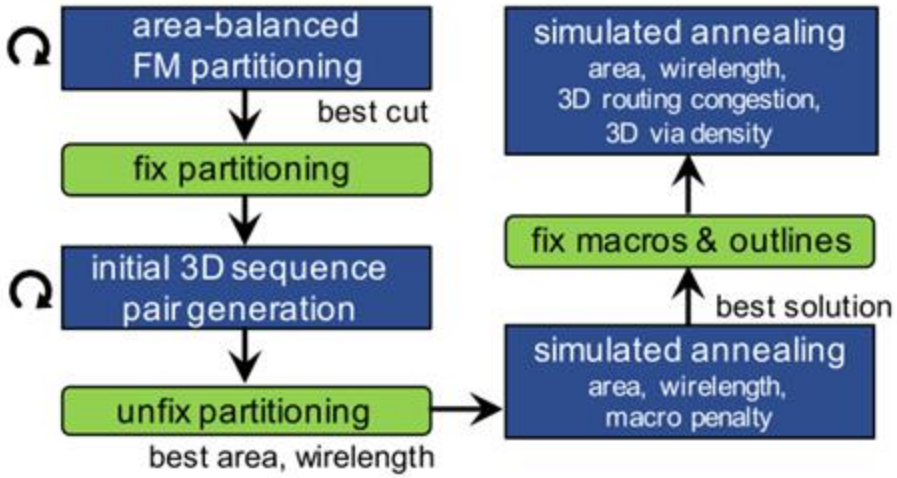
input: netlist N, area threshold t
output: clusters from **cluster**($N \backslash macros(N)$, t) $\cup$ $macros(N)$ $\cup$ remaining set of unclustered cells

```
cluster(O, t):
C = ∅
if area(O) ≥ t then
    Q = sub-modules(O); sort Q by decreasing area
    if area(Q) ≥ t then
        for q in Q do
            R = sum of area(i) for i after q in Q
            C = C ∪ cluster(q, t)
            if R < t then
                break
            end if
        end for
    else
        C = O
    end if
else
    C = O
end if
return C
```

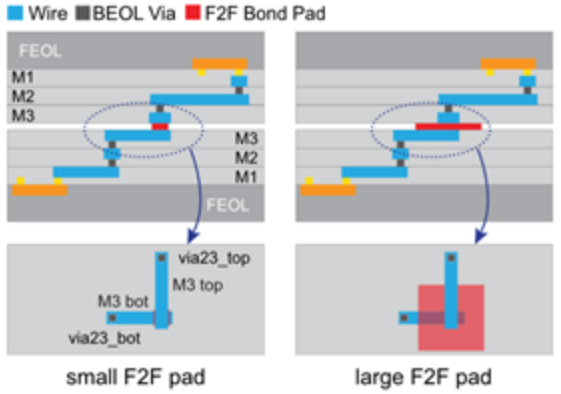
Funded by
the European Union

Semi-automated floorplaning

- Initial FM partitioning, initial 3D placement
- Simulated annealing: macros
- Simulated annealing: full



- Assign location to bumps minimizing cost (as Sequential-3D)
- Automatic placement of bumps may require KoZ and dedicated optimization



Funded by the European Union

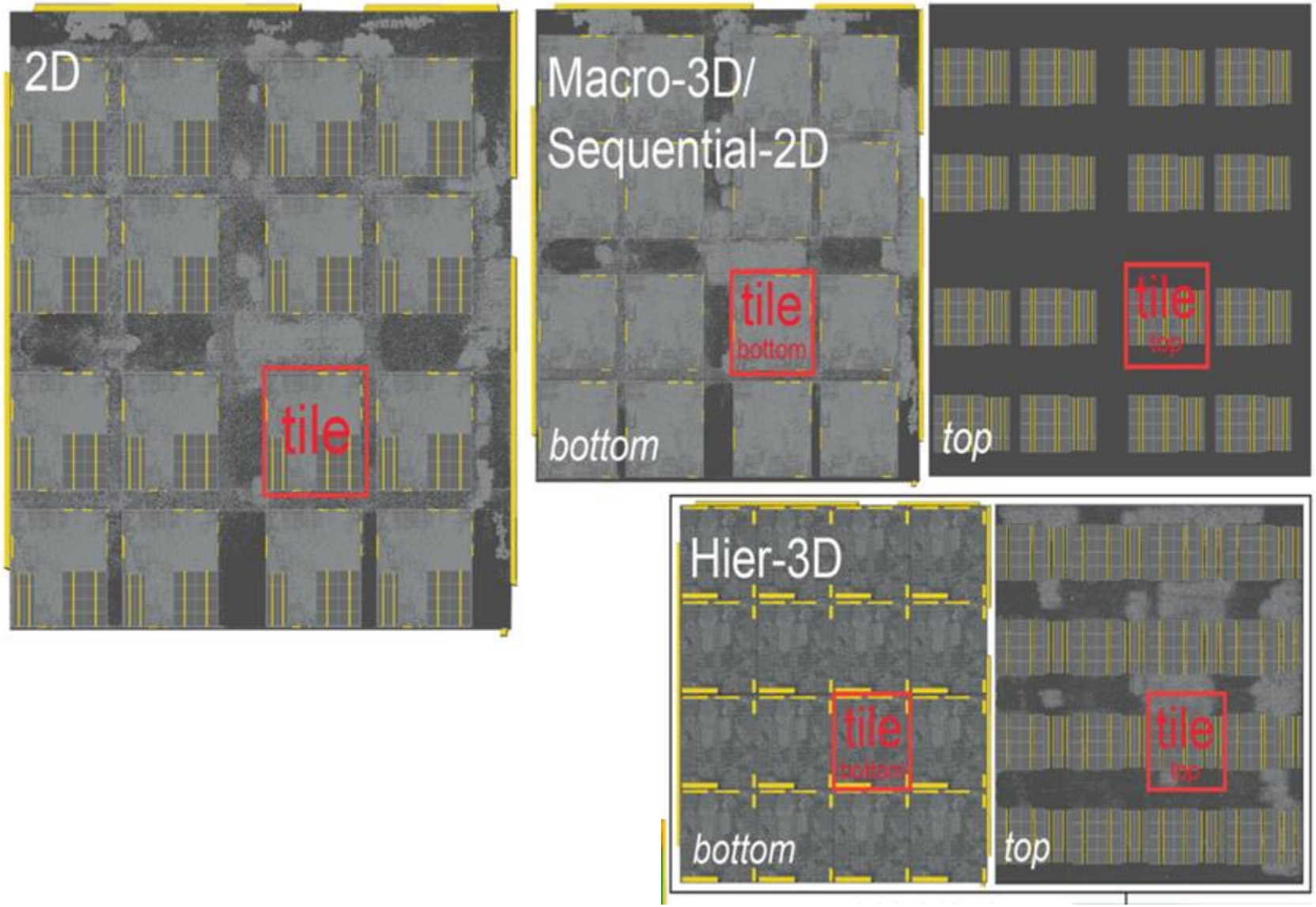
Results MemPool

- Large PPA improvements
- Shared BEOL
- Reduction of congention in cluster
- Repeaters to reduce detours

MemPool Cluster	2D	Seq-2D	Macro-3D	Hier-3D
metals used	6	6 (bot) 6 (top)	6 (bot) 6 (top)	6 (bot) 6 (top)
silicon area	1	1.49	1.14	0.75
total WL	1	0.87	0.80	0.71
density (%) bot/top	56.2	50.8/22.6	62.4/29.6	84.4/53.4
buffer count	1	0.91	0.67	0.61
# F2F bumps	-	130K	813K	482K
effective freq	1	1.00	1.05	1.42
total power	1	0.98	0.89	0.80
power × delay	1	0.98	0.85	0.57
die cost	1	1.57	1.19	0.79
power perf cost	1	0.65	0.99	2.22
runtime	1	1.09	0.91	0.68

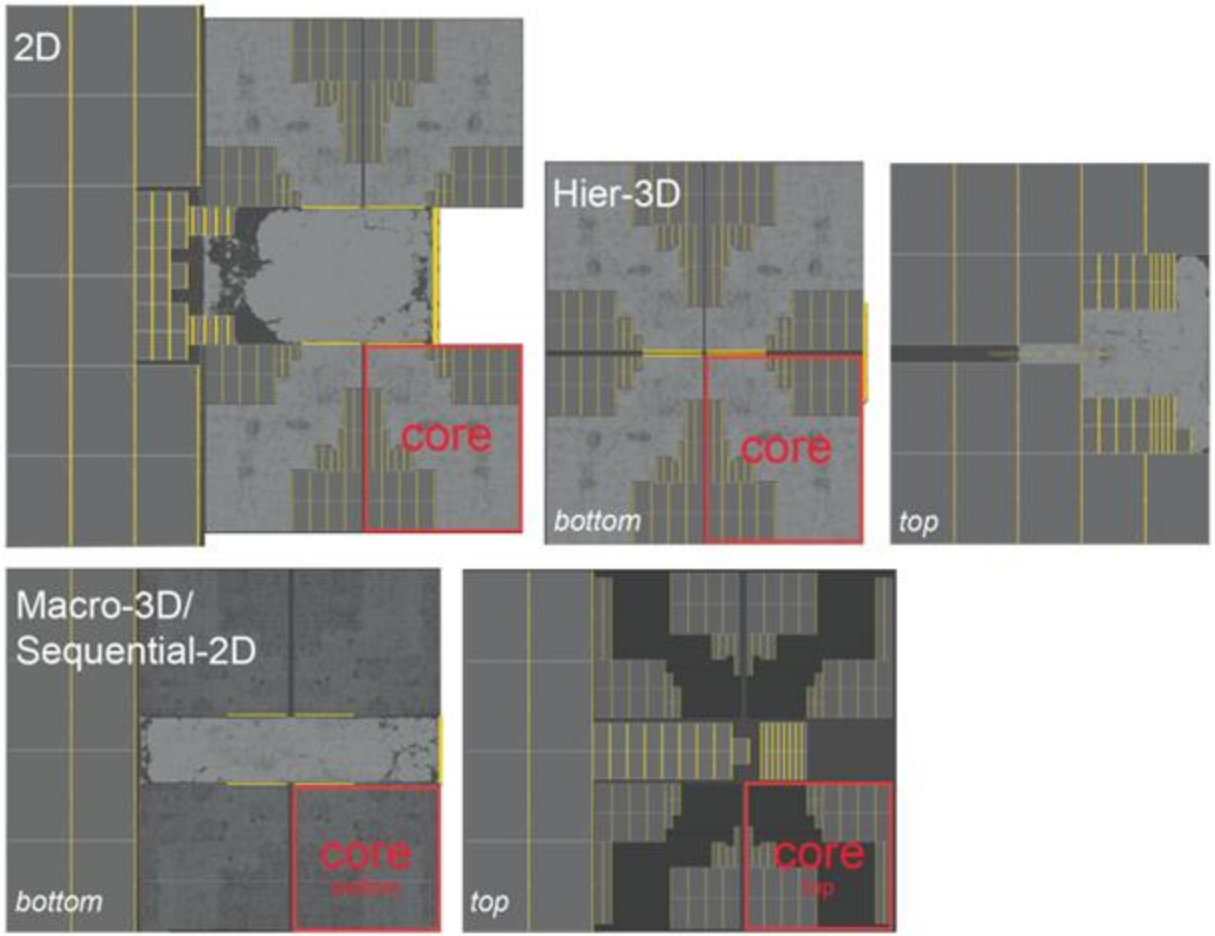
Funded by
the European Union

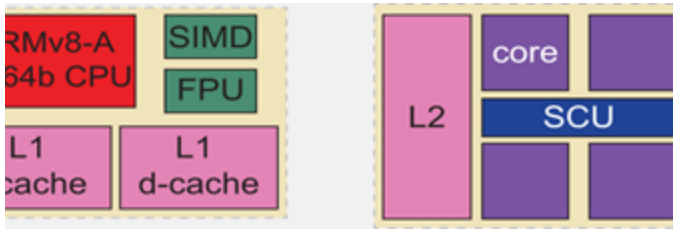
Results MemPool



Funded by
the European

Results: Cortex-A53 4-cores

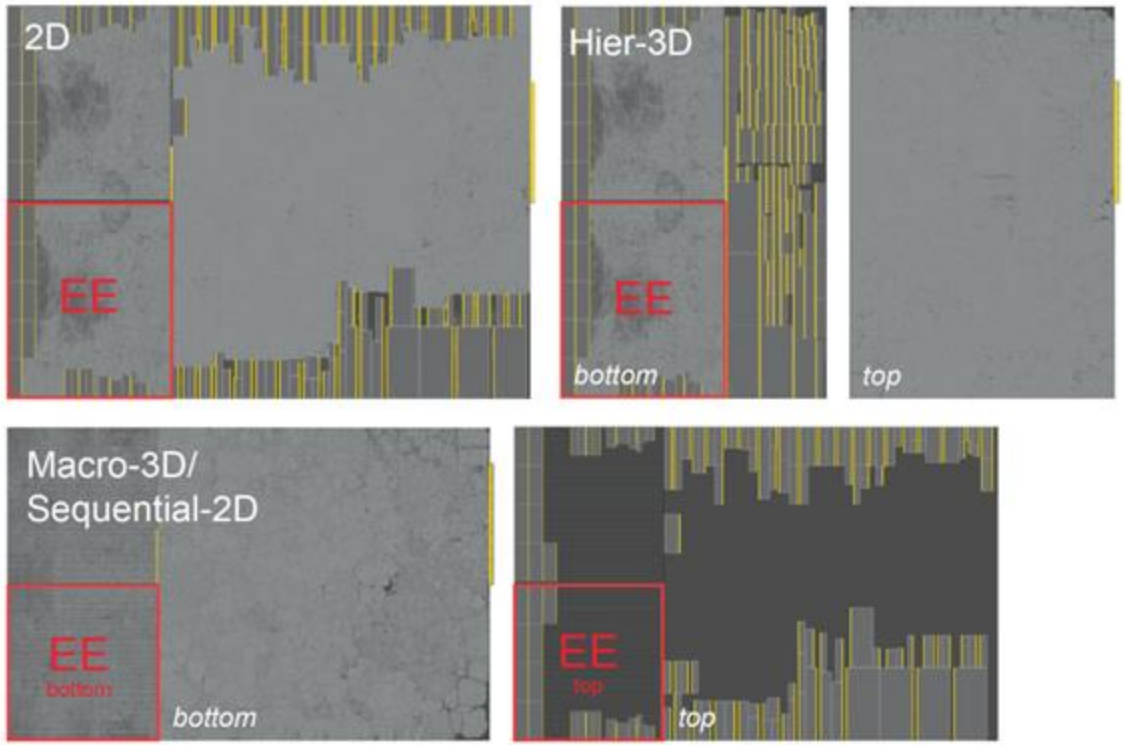




Cortex-A53	2D	Seq-2D	Macro-3D	Hier-3D
metals used	8	7 (bot) 6 (top)	7 (bot) 6 (top)	6 (bot) 7 (top)
silicon area	1	1.21	1.21	0.95
total WL	1	1.02	0.97	0.94
density (%) bot/top	77.5	73.7/62.6	72.1/62.5	81.0/90.7
buffer count	1	1.15	1.05	0.98
# F2F bumps	-	22K	81K	74K
effective freq	1	0.93	0.95	1.33
total power	1	1.14	0.97	0.97
power × delay	1	1.22	1.02	0.73
die cost	1	1.13	1.13	0.91
power perf cost	1	0.78	0.87	1.51
runtime	1	1.12	0.89	0.82

Funded by
the European Union

Results: Mali-G52



Mali-G52	2D	Seq-2D	Macro-3D	Hier-3D
metals used	8	7 (bot) 6 (top)	7 (bot) 6 (top)	6 (bot) 7 (top)
silicon area	1	1.45	1.45	0.99
total WL	1	0.88	0.86	0.98
density (%) bot/top	77.7	74.5/31.3	74.3/31.3	78.1/70.2
buffer count	1	0.93	0.93	0.97
# F2F bumps	-	56K	325K	149K
effective freq	1	0.98	0.95	1.15
total power	1	1.14	0.96	0.98
power × delay	1	1.16	1.01	0.85
die cost	1	1.35	1.35	0.95
power perf cost	1	0.64	0.73	1.24
runtime	1	0.99	1.08	0.81

Funded by
the European Union

Conclusions: Hier-3D

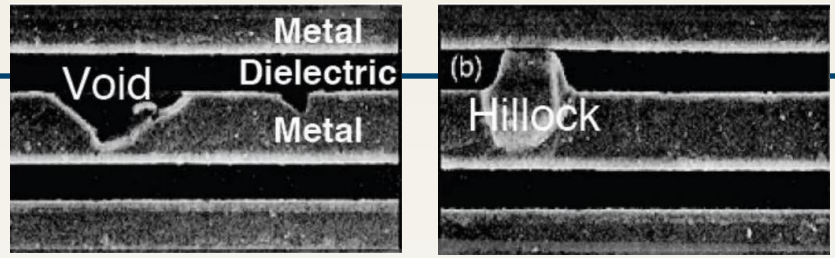
- Full-chip RTL to GDSII physical design solution for the exploration of 3D design using commercial tools
- 3-D **bottom-up hierarchical** implementations
- **Reusing the unassigned area** of preceding hierarchy levels
- Hierarchical **multitier** standard cell placement, greatly expanding the design space of 3-D ICs
- Contributions:
  - RTL clustering, 3-D logic-on-logic floorplanning, and holistic 3-D physical implementations
  - Automatic floorplaning, whitespace modeling

Funded by
the European Union

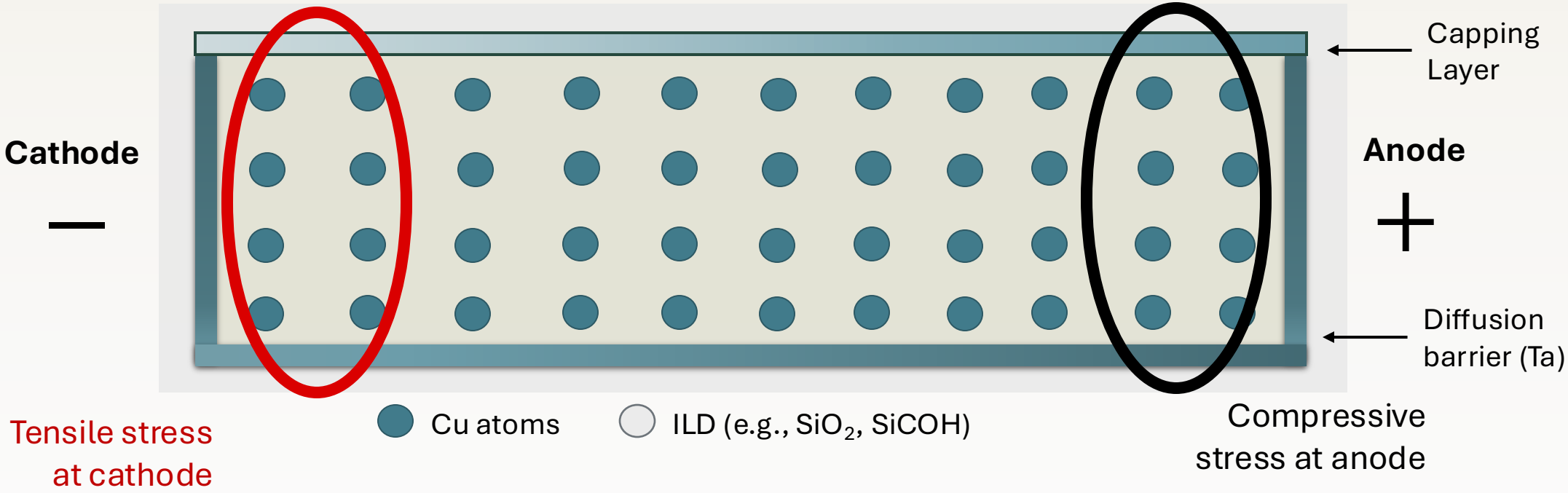
PROTON: Physics-Based Electromigration Assessment on Modern VLSI Power Grids

Funded by
the European Union

Electromigration (EM) Stress



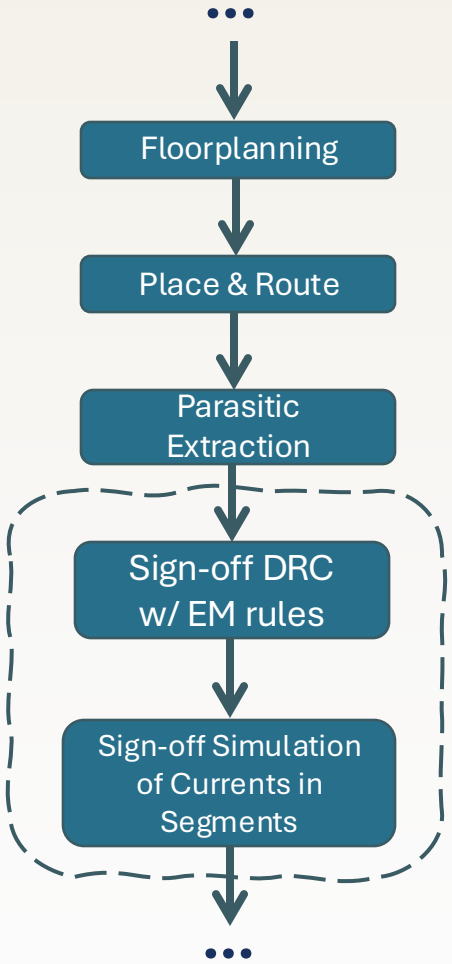
[1]



[1] J. Lienig and M. Thiele, Fundamentals of electromigration-aware integrated circuit design. Springer International Publishing, 2018.

Long-term Reliability of Power Grids

- Phenomenon of EM
  - Caused by the continuous downscaling of the ICs and the simultaneous increase of currents
- Major concern for nanometer-scaled Integrated Circuits (IC)
  - Process variation in modern technology nodes
- Has been included in the IC design flow
  - Traditional methods are incorrect for modern interconnects (e.g., Black's equation and Blech's criterion)



Addressing the gap: A free physics-based EM

SYNOPSYS®

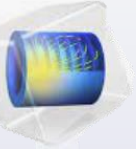
IC Compiller II

(RedHawk Fusion)

Ansys

COMSOL

COMSOL MULTIPHYSICS®



cadence®

Voltus-XFi

Custom Power Integrity Solution

Need for a dedicated EM tool

Empirical method

Bundle of analyses

Heuristic method

Black's equation

Need for free EDA software

Inaccurate for modern ICs

Closed-source

Funded by the European Union

PROTON: Physics-Based Electromigration Assessment on Modern VLSI Power Grids

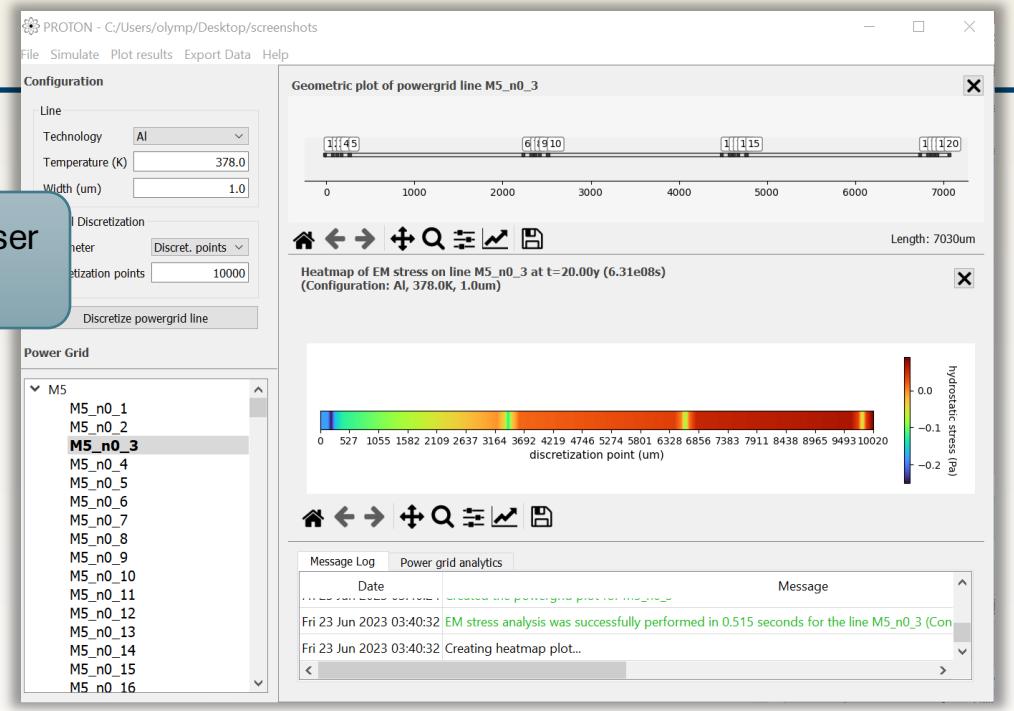
Democratizing EM Analysis

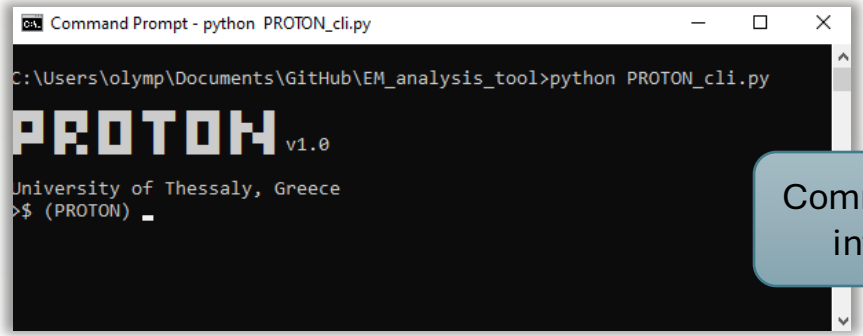
- PROTON's goal aligns with open-source strategies from EU
  - (e.g., EuroEXA, CONVOLVE) and large-scale EDA projects in US (e.g., OpenROAD)
- Enabling smaller design teams and individuals to benefit from cutting-edge EM analysis
  - Fostering innovation and collaboration within the EDA community

PROTON – Overview

- Very fast EM analysis
- Open-source
- Cross-platform
- Intuitive UI
- High level of integration

Graphical-user interface





Command-line interface

Funded by the European Union

PROTON: Physics-Based Electromigration Assessment on Modern VLSI Power Grids

The core of the tool

✓ **Robust methods**

✓ **Low complexity**

Korhonen's equation w/ line boundary conditions + **FDM Discretization** into n discretization points

$$\dot{\boldsymbol{\sigma}}(t) = \mathbf{A}\boldsymbol{\sigma}(t) + \mathbf{B}\boldsymbol{j}$$

Line Analysis

$$\boldsymbol{L}(t) = diag\left(t, \frac{e^{\lambda_2 t}-1}{\lambda_2}, \dots, \frac{e^{\lambda_n t}-1}{\lambda_n}\right)$$

$$\boldsymbol{\sigma}(t) = \boldsymbol{V}\boldsymbol{L}(t)\boldsymbol{V}^T\boldsymbol{B}\boldsymbol{j}$$ [1]

$$\boldsymbol{V}^T\boldsymbol{r} \Leftrightarrow DCT(\boldsymbol{r})$$
$$\boldsymbol{V}\boldsymbol{r} \Leftrightarrow IDCT(\boldsymbol{r})$$

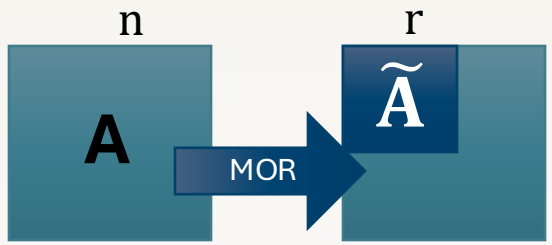
$O(nlogn)$

Completely **parallelizable** for different lines in the under-test power grid

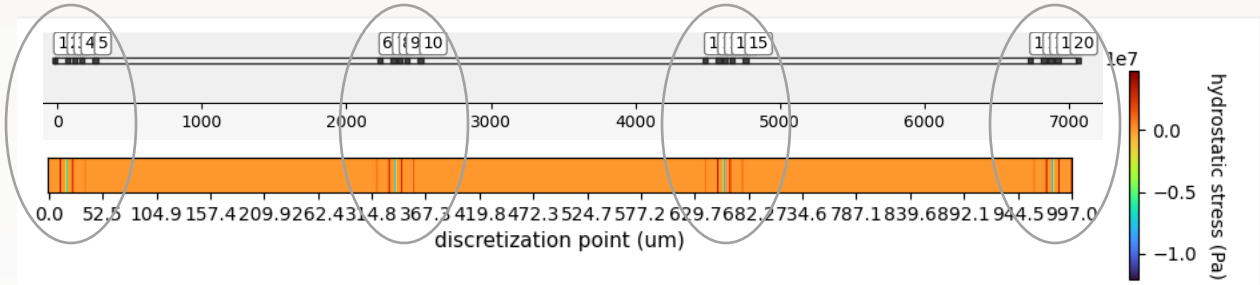
Vias Transient Analysis

State-space model

$$\dot{\boldsymbol{\sigma}}(t) = \mathbf{A}\boldsymbol{\sigma}(t) + \mathbf{B}\boldsymbol{j}$$ [2]
$$\boldsymbol{y}(t) = \boldsymbol{L}\boldsymbol{\sigma}(t)$$



Stress is more pronounced at vias and endpoints



[1] O. Axelou et al., A Novel Semi-Analytical Approach for Fast Electromigration Stress Analysis in Multi-Segment Interconnects. ICCAD, 2022.

[2] O. Axelou et al., Fast Electromigration Stress Analysis Using Low-Rank Balanced Truncation for General Interconnect and Power Grid Structures. Integration Journal, 2023.

Funded by the European Union

PROTON: Physics-Based Electromigration Assessment on Modern VLSI Power Grids

PROTON – C++ Kernels

sparse modules

Line Analysis

https://fftw.org/

Fast complete power grid analysis

Vias Transient Analysis

state-of-the-art direct solver

State-of-the-art libraries

Low-memory footprint

Fast sparse solvers

Suitable for large-scale problems

✓ **Robust**

Accurate, fast and robust

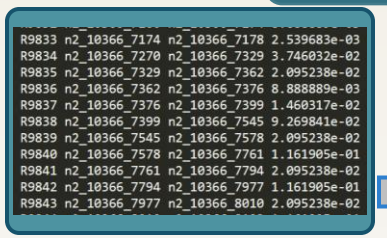
✓ **Applicable** for all power-grid sizes

PROTON Flowchart

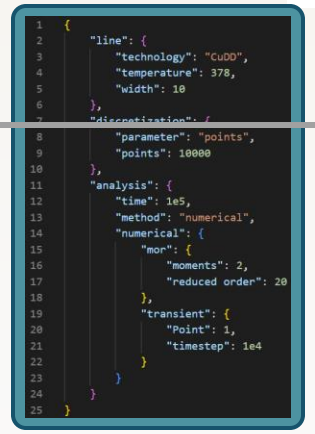
✓ Easily Integratable

✓ Highly automated

SPICE file

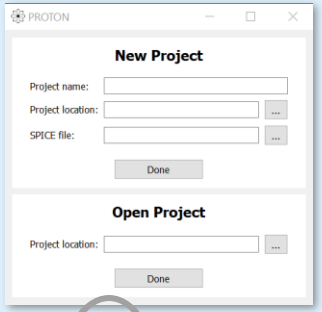


JSON configuration file



On *Open*

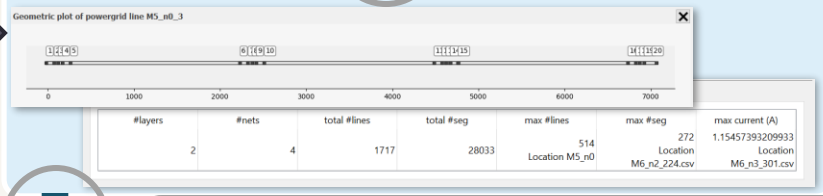
IMPORT POWERGRID & CONFIGURATIONS



SPICE FILE PARSING & POWERGRID ANALYSIS

SPICE Parsing & Line extraction

Line visualization (e.g., vias, segments)



LINE SPATIAL DISCRET

Finite Difference Method

Closed-source

cadence Spectre

SYNOPSYS PrimeSim HSPICE

Open-source

Xyce

Segment currents

Choice from configuration

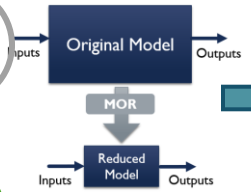
PROTON ENGINE (C++ Kernel)

LINE ANALYSIS

Eigendecomposition of system matrix & DCT-II Transform

EM stress calculation at all spatial points independently with near linear complexity

VIAS TRANSIENT ANALYSIS

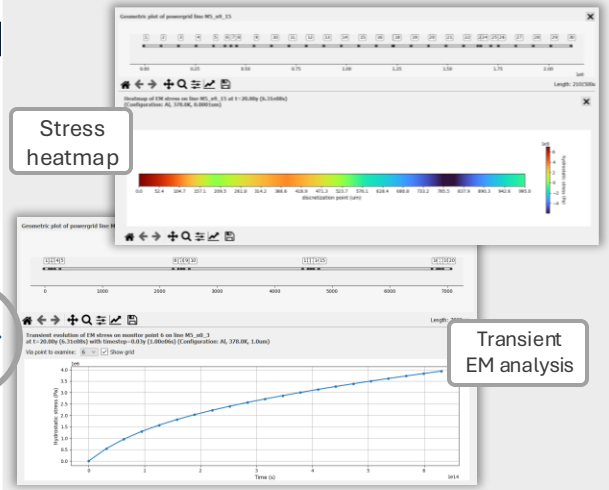


Transient EM stress analysis on via and blocking points with an implicit Backward Euler formula

PROTON

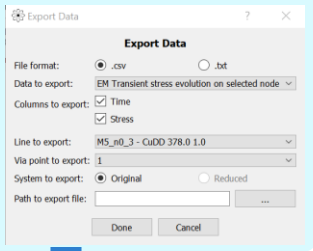
STRESS PLOTTING & RESULTS ANALYTICS

Stress heatmap



Transient EM analysis

STRESS RESULTS EXPORT



Input/Output

User-driven act

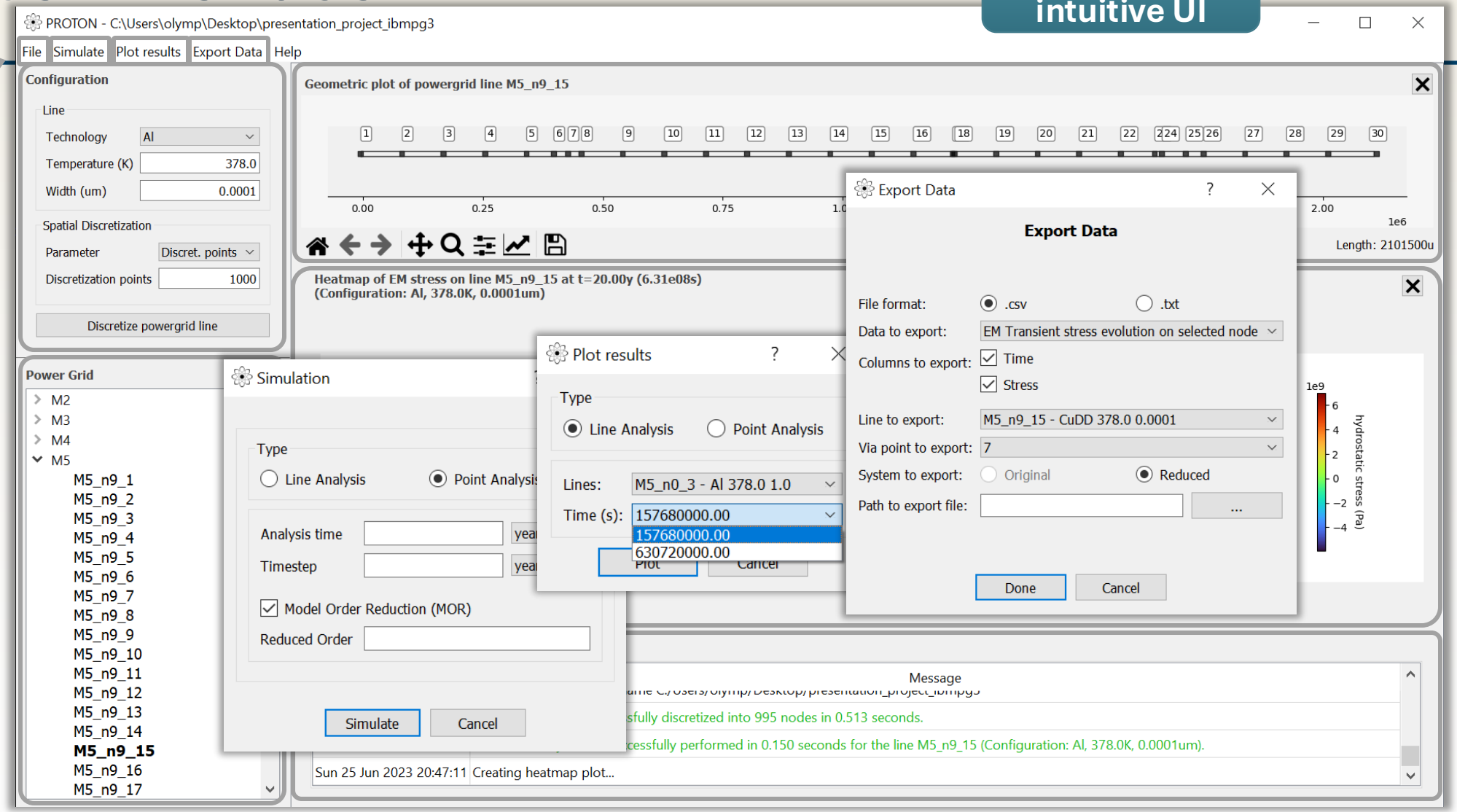
Fully-automated act

Interaction w/ external tools (Fully-automated)

Funded by the European Union

User Interface

✓ Simple and intuitive UI



Analysis parameters

Selection of line to analyze

Visualization of selected line

Analysis results

Message log & power grid analytics

Command-Line Version

✓ Fully automated

```
$ powergrid_EM_analysis.tcl
...
#  An automated script that parses a SPICE powergrid file and performs
# Electromigration (EM) analysis on the complete powergrid at specific
# time (e.g. chip lifetime) and finds the EM-susceptible lines.
#  Then, some statistics on the parsed powergrid are reported as well
# as the line with the maximum stress.
#
# Author: Olympia Axelou
# Affiliation: University of Thessaly, Greece
# Date: 19/6/2023

# Create new project
set_powergrid C:\Users\olymp\Documents\GitHub\EM_analysis_tool\benchmarks\ibmpg1.
set_project_path C:\Users\olymp\Desktop
set_project_name automated_ibmpg1
parse_powergrid
# open_project C:\Users\olymp\Desktop\automated_ibmpg1

# Set technology, temperature, line width
set_technology CuDD
set_temperature 378
set_line_width 1

# Complete powergrid EM stress check at 20y (6.38e8s) for a random sample of 100
# and check for lines that exceed the critical stress of 10KPa
analyze 6.38e8 --sample 100 --critical 10000

# Export stats on the powergrid and EM stress results
report_powergrid_stats --file C:\Users\olymp\Desktop\powergrid_stats.txt
report_line_stress --maxstress --file C:\Users\olymp\Desktop\critical_line_stress.csv

# Quit
quit
```

source TCL scripts

```
Command Prompt - python  PROTON_CLI.py

PROTON v1.0

University of Thessaly, Greece
>$ (PROTON) set_powergrid example.spice
>$ (PROTON) set_project_path C:/Users/olymp/Desktop
>$ (PROTON) set_project_name example_cli_project
>$ (PROTON) parse_powergrid
Found layer M3, net 1
Found layer M4, net 2
Found layer M5, net 5
Found layer M6, net 6
Found layer M3, net 0
Found layer M4, net 3
Found layer M5, net 4
Found layer M6, net 7
Spice file was successfully parsed in 1.921 seconds and the project has been created at C:\Users\o
lymp\Desktop\example_cli_project.
>$ (PROTON)
```

Simple command-line interface

We invite you to take a look!

- Installation guide
- Integration of custom kernels guide

https://github.com/oaxelou/PROTON

Code available on GitHub

Olympia Axelou

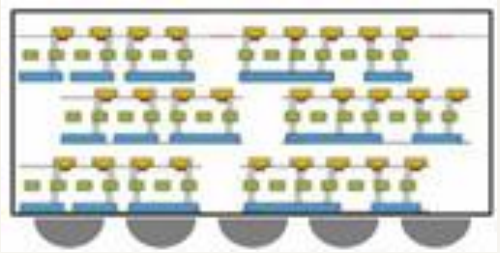
oaxelou@e-ce.uth.gr

University of Thessaly, Greece

UNIVERSITY OF THESSALY

Extending PROTON to 2.5/3D Chiplets

- Challenges with Chiplet-Based Designs:
  - Increased interconnect density and complexity
  - Heterogeneous integration: non-uniform stress profiles



- Adapting PROTON for 2.5/3D:
  - Extend power grid line analysis to vertical interconnects (TSVs, micro-bumps)
  - Seamless transition from traditional 2D analysis

- Collaboration with Industry
  - Validation on chiplet benchmarks
  - Adaptation in 2.5/3D Design Flows

Q&A

Funded by
the European Union