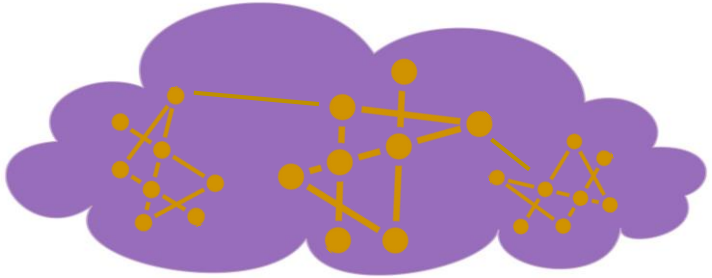
When cyber-physical systems meet HPC: Productivity and dependability through OpenMP

Dr. Sara Royuela
WHPC ISC22

Cyber-Physical System

CYBER



communication | computation | storage

sensor | actuator

PHYSICAL

Space

Automotive

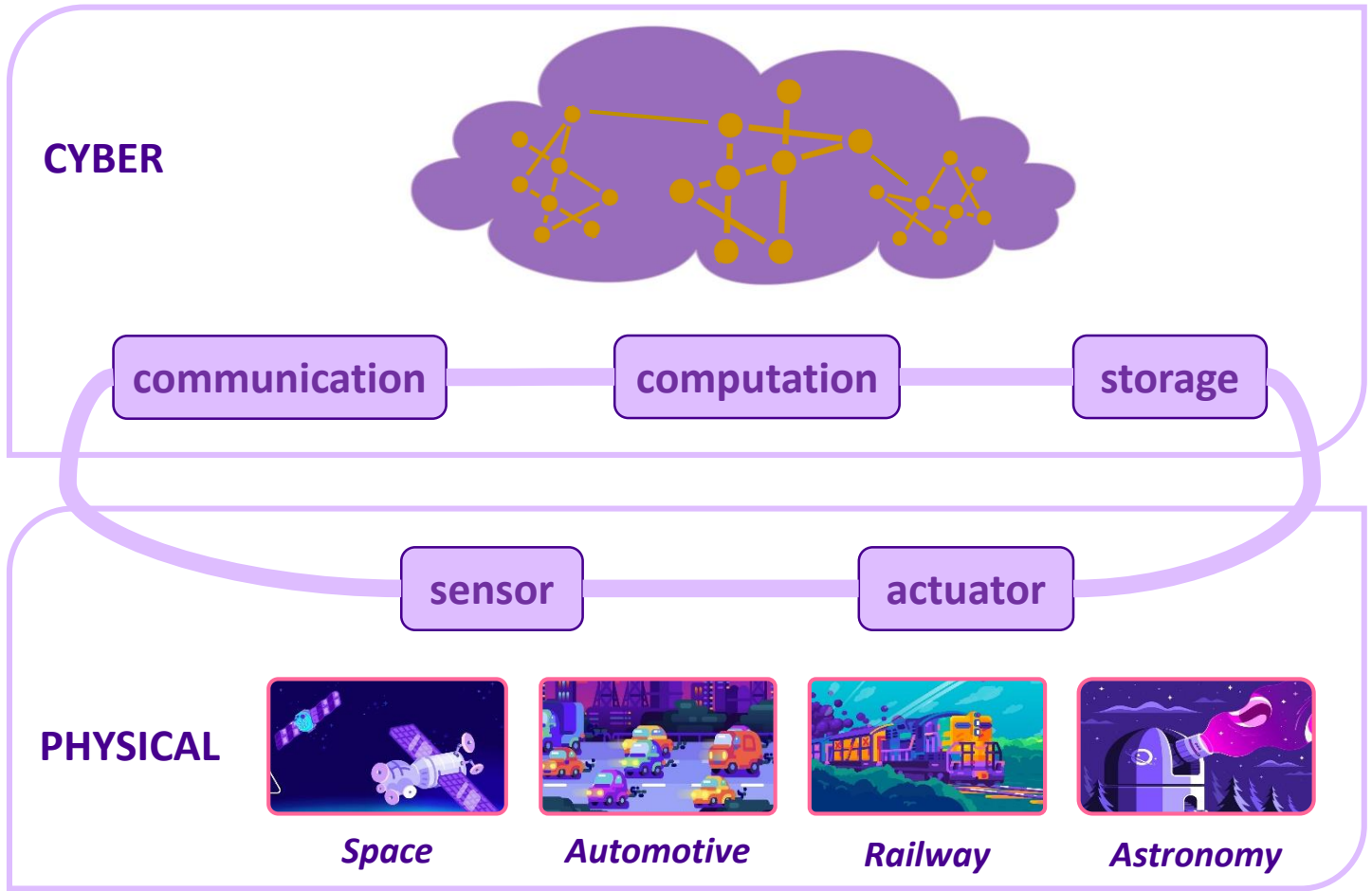
Railway

Astronomy

When cyber-physical systems meet HPC: Productivity and dependability through OpenMP

Dr. Sara Royuela
WHPC ISC22

Cyber-Physical System



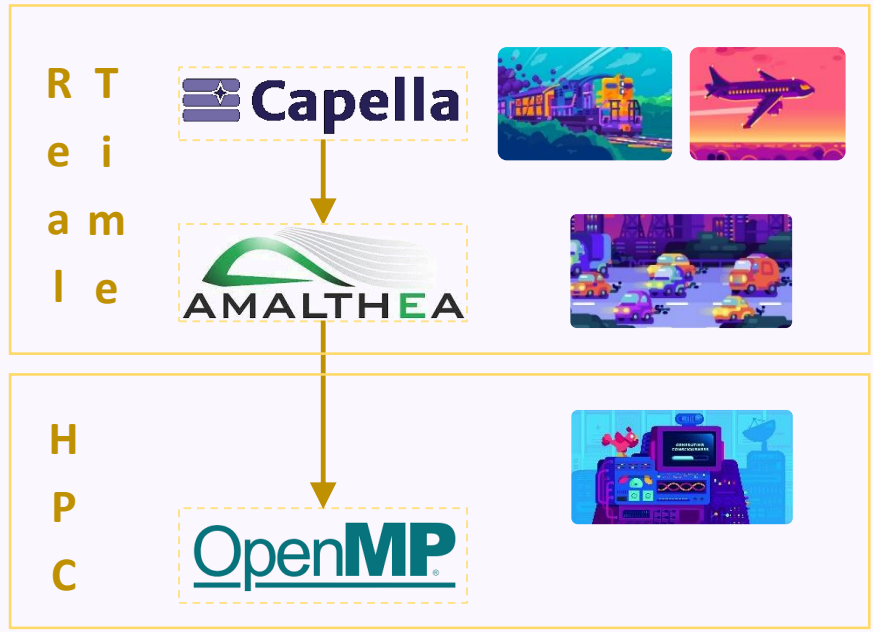
Images by Kurzgesagt

CONTEXT

- **Cyber-physical systems** (**CPS**) are convenient to exhibit the autonomy and intelligence of complex processes in safe mobility, giant scientific installations, smart manufacturing…
- **HPC** needs are added to **real-time** constraints and **non-functional** requirements, drastically increasing **programming efforts**.
- **OpenMP** is a de-facto standard for high-performance shared-memory systems with support for heterogeneity.
- The **tasking model** is a convenient mechanism to expose parallelism in a simple yet flexible manner.

When cyber-physical systems meet HPC: Productivity and dependability through OpenMP

Dr. Sara Royuela
WHPC ISC22



Xilinx Zynq Ultrascale+ MPSoC

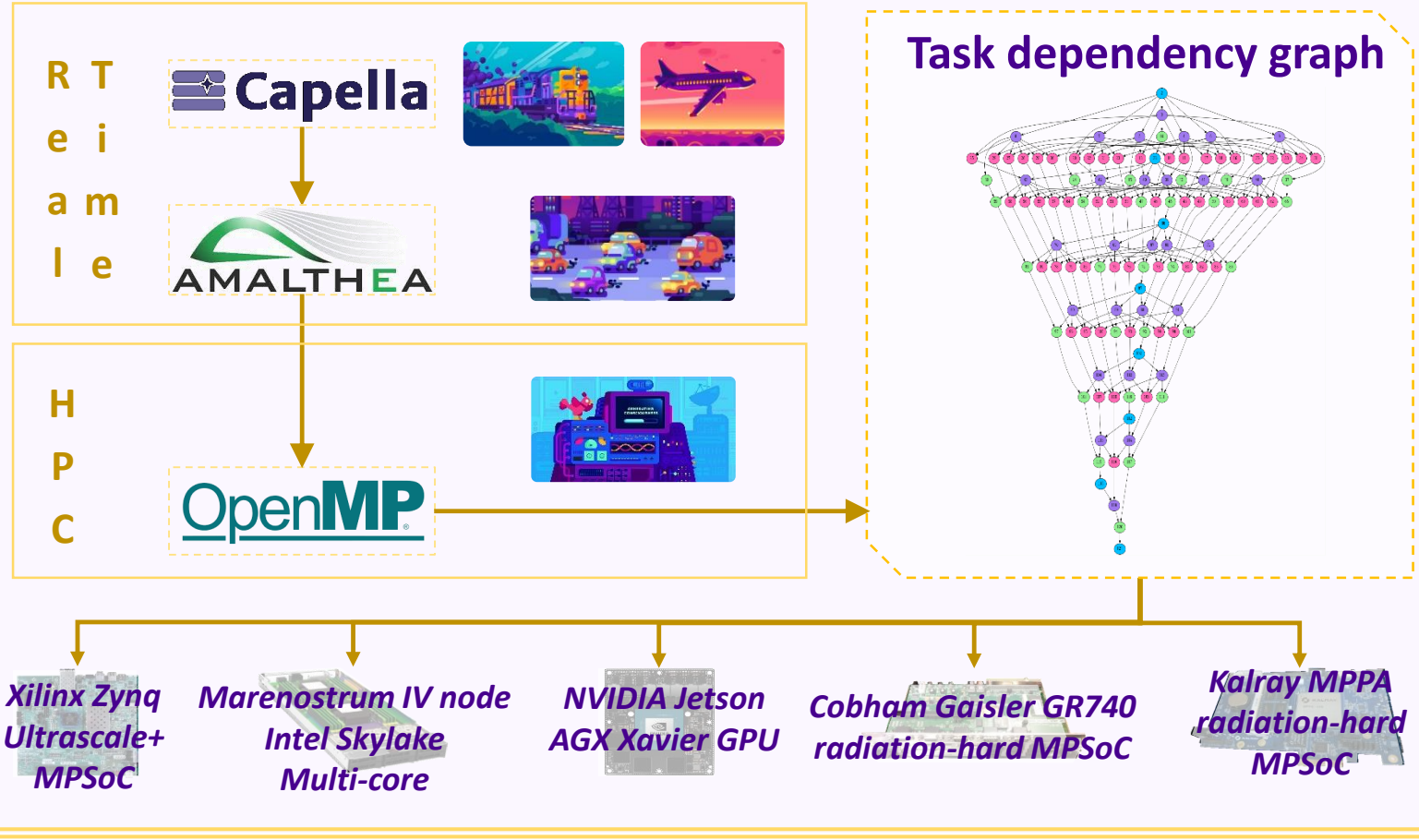
Marenostrum IV node Intel Skylake Multi-core

NVIDIA Jetson AGX Xavier GPU

Cobham Gaisler GR740 radiation-hard MPSoC

Kalray MPPA radiation-hard MPSoC

The TDG to match user needs with machine capabilities



When cyber-physical systems meet HPC: Productivity and dependability through OpenMP

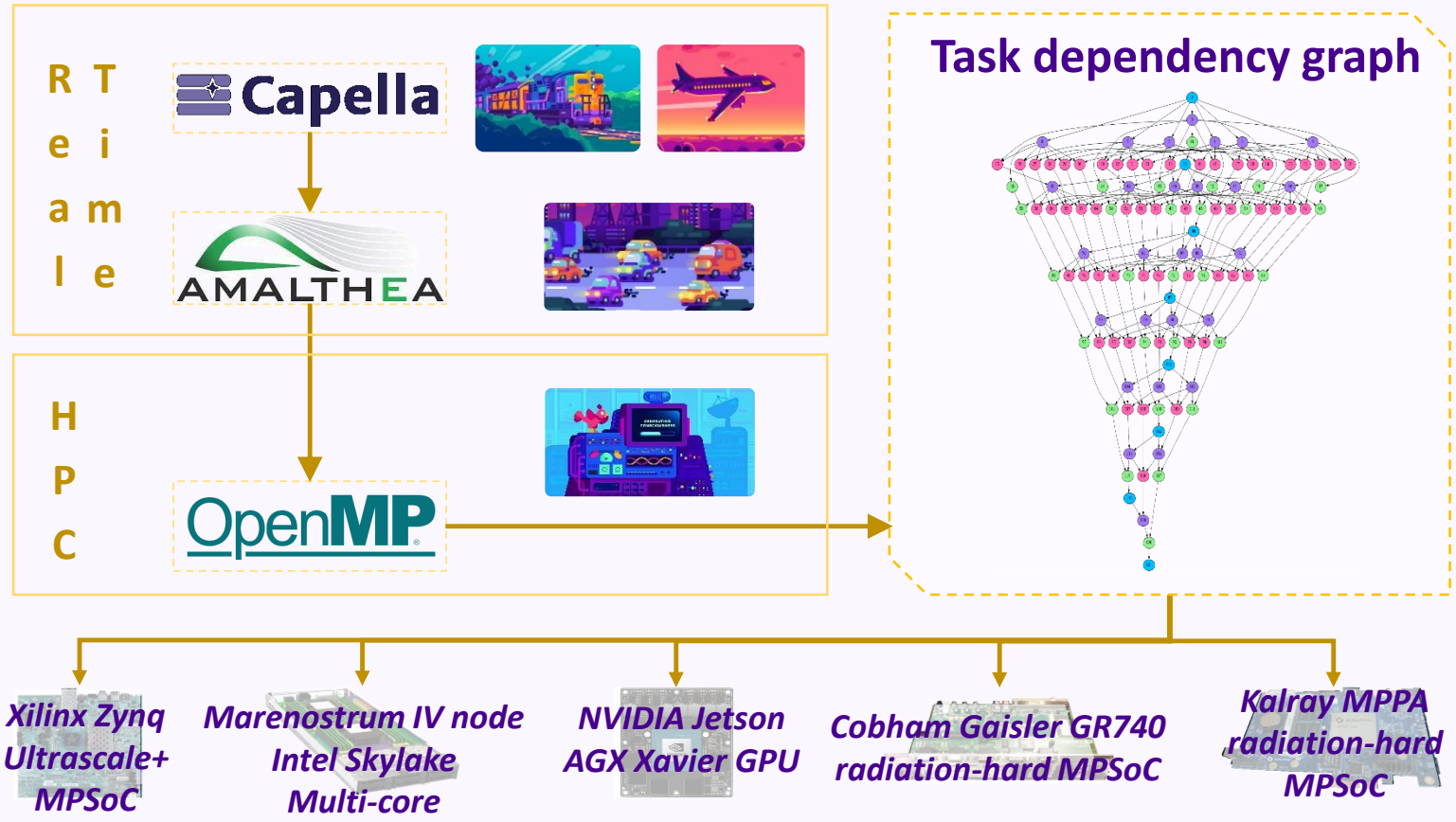
Dr. Sara Royuela
WHPC ISC22

OBJECTIVES

Exploit the **task dependency graph** to:

- **Introduce parallelism** in domains newly requiring HPC capabilities, e.g., automotive, space, etc.
- **Boost productivity** of the **tasking model** in terms of performance, programmability and (performance) portability.
- Provide **safety guarantees** / **QoS**.
- Enhance the **interoperability** of OpenMP with target-specific programming models, e.g., CUDA, to exploit heterogeneity.

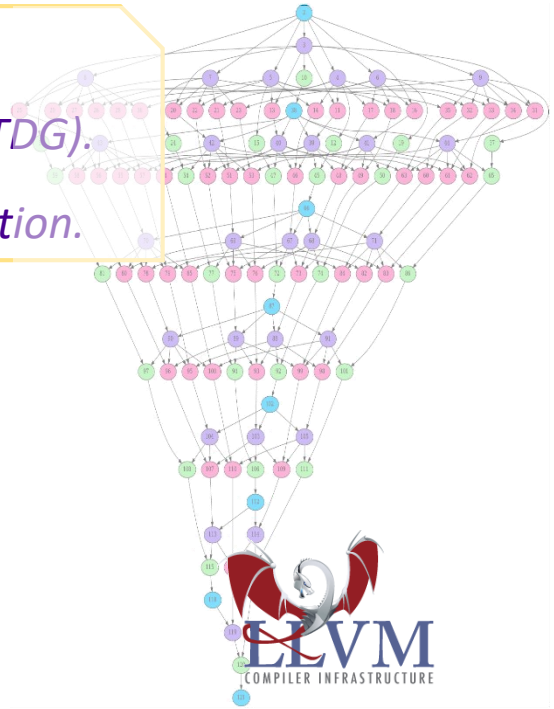
The TDG to match user needs with machine capabilities



Optimizations built on top of the TDG

Memory:

- ✓ *Memory bounding (heuristic based on TDG).*
- ✓ *Task data preallocation + lazy task creation.*



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

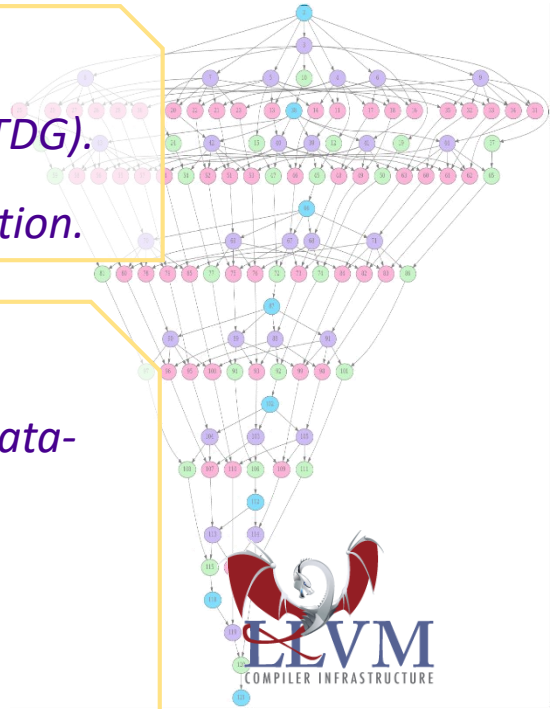
Optimizations built on top of the TDG

Memory:

- ✓ *Memory bounding (heuristic based on TDG).*
- ✓ *Task data preallocation + lazy task creation.*

Safety:

- ✓ *Correctness analysis (race conditions, data-sharing/dependencies consistency,...).*
- ✓ *Timing/schedulability analysis.*
- ✓ *Replication towards fault-tolerance.*



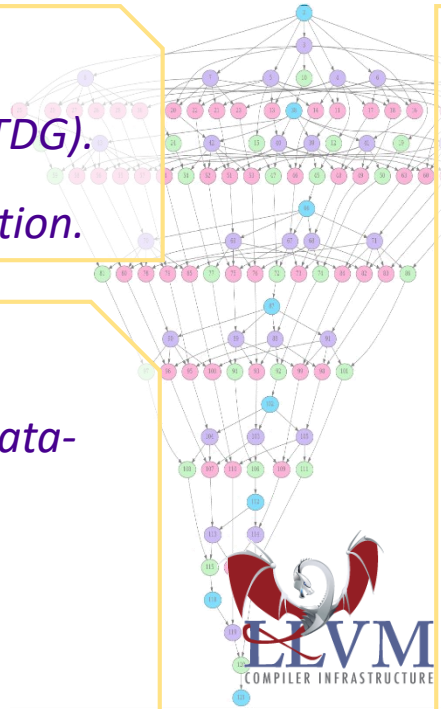
Optimizations built on top of the TDG

Memory:

- ✓ *Memory bounding (heuristic based on TDG).*
- ✓ *Task data preallocation + lazy task creation.*

Safety:

- ✓ *Correctness analysis (race conditions, data-sharing/dependencies consistency,...).*
- ✓ *Timing/schedulability analysis.*
- ✓ *Replication towards fault-tolerance.*



Performance:

- ✓ *Define-once-run-repeatedly* execution model.
- ✓ *Scheduling* optimizations (e.g., data affinity, critical path, fixed/static scheduling).
  - → Data affinity, critical path,...
- ✓ *Iteroperability / Heterogenenity (FPGA/GPU):*
  - → OpenMP TDG to CUDA graphs (GPU).
  - → OpenMP TDG to FRED/DART (FPGA).

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

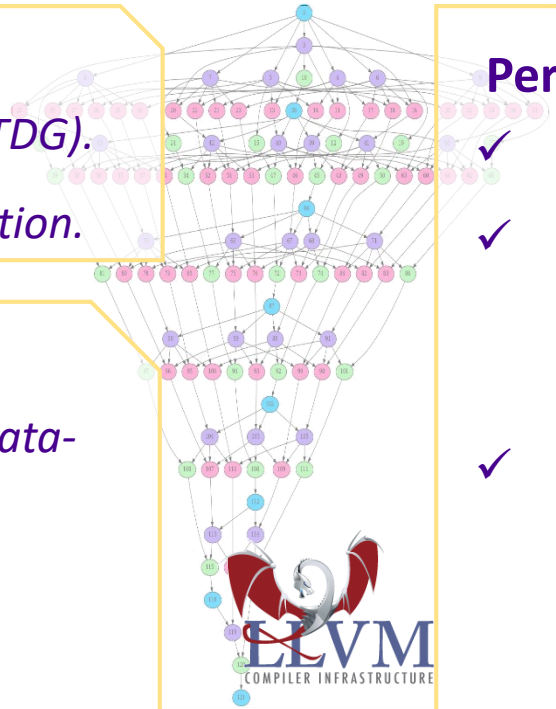
Optimizations built on top of the TDG

Memory:

- ✓ *Memory bounding (heuristic based on TDG).*
- ✓ *Task data preallocation + lazy task creation.*

Safety:

- ✓ *Correctness analysis (race conditions, data-sharing/dependencies consistency,...).*
- ✓ *Timing/schedulability analysis.*
- ✓ *Replication towards fault-tolerance.*



Performance:

- ✓ *Define-once-run-repeatedly* execution model.
- ✓ *Scheduling* optimizations (e.g., data affinity, critical path, fixed/static scheduling).
  - → Data affinity, critical path,...
- ✓ *Iteroperability / Heterogenenity (FPGA/GPU):*
  - → OpenMP TDG to CUDA graphs (GPU).
  - → OpenMP TDG to FRED/DART (FPGA).

Rising STARS — RISE International Network for Solutions Technologies and Applications of Real-time Systems

AMPERE — A Model-driven development framework for highly Parallel and EneRgy-Efficient computation supporting multi-criteria optimisation

HP4S — High Performance Parallel Payload Processing for Space

BSC — Barcelona Supercomputing Center — Centro Nacional de Supercomputación

This work has received funding from the European Union-s Horizon 2020 research and innovation programme under grant agreement No **871669** and the Marie Sklodowska-Curie grant agreement No **873120**.