

Cristina Silvano<sup>1</sup>, Giovanni Agosta<sup>1</sup>, Andrea Bartolini<sup>2</sup>, Andrea Beccari<sup>3</sup>, Luca Benini<sup>4</sup>, João M. P. Cardoso<sup>5</sup>, João Bispo<sup>5</sup>, Pedro Pinto<sup>5</sup>, Carlo Cavazzoni<sup>6</sup>, Radim Cmar<sup>7</sup>, Jan Martinović<sup>8</sup>, Gianluca Palermo<sup>1</sup>, Martin Palkovič<sup>9</sup>, Erven Rohou<sup>9</sup>, Nico Sanna<sup>9</sup>, Kateřina Slaninová<sup>9</sup>

<sup>1</sup>DEIB – Politecnico di Milano, <sup>2</sup>IIS – ETH Zürich, <sup>3</sup>Dompé S.p.A., <sup>5</sup>FEUP – Universidade do Porto, <sup>6</sup>CINECA, <sup>7</sup>Sygit, <sup>8</sup>IT4Innovation National Supercomputing Center, <sup>9</sup>INRIA Rennes

### The Exascale Problem

- Exascale computing (10<sup>18</sup> FLOPs) at 20 Megawatts requires an energy efficiency “quantum leap”
- Use self-adaptivity and autotuning concepts to improve HPC efficiency
- ANTAREX proposes tools, languages and techniques to control all the decision layers



#### Accelerate

- Productivity
- Performance
- Innovation



#### Energy-Efficient Exascale

- Scalable Monitoring
- Power Management

### ANTAREX Approach



#### Parallel Programming Model

- OpenCL front-end for multi-target parallelism
- C/C++ transformation to add OpenMP directives and MPI API calls



#### Application Autotuning

- Enable application self-awareness by runtime tuning
- DSL-defined software knobs, control-loops and HW monitoring



#### Runtime Management

- HW-aware runtime for scalable management infrastructure
- Efficient communication of performance and energy metrics



#### DSL to Specify Adaptivity Concerns

- Separation of concerns: C/C++ for functionality; DSL for adaptivity
- Clang-based source-to-source compiler weaves adaptivity concerns

### Use Cases



#### Accelerated Personalized Drug Discovery

##### HPC Molecular Simulation

##### Huge exploration space

- Prediction of properties of protein-ligand complexes
- Verification of synthetic feasibility

##### Massively parallel but...

- Unpredictable imbalances in computation
- Dynamic load balancing is critical

**Platform:** CINECA Supercomputing infrastructure



#### Self-Adaptive Navigation System

##### Growing automotive traffic load

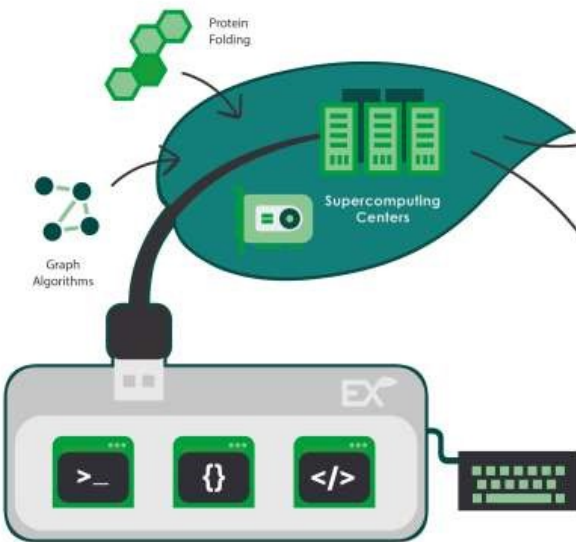
##### Smart City Challenge

- Serve drivers' routing requests with global best under a variable workload
- Efficiency dependent on balancing data collection and big data analysis

##### Exploit synergies

- Between server-side and client-side computation
- Many drivers – many routing requests to HPC system

**Platform:** IT4Innovations Supercomputing infrastructure



### Main Outcomes

- Achieve best energy-efficient, thermally-safe point
- System-wide adaptive approach for HPC systems
- Decouple functional and extra-functional concerns
- Cross-layer software stacks to exploit heterogeneity