<table>
<thead>
<tr>
<th>ANTAREX</th>
<th>AutoTuning and Adaptivity appRoach for Energy efficient eXascale HPC systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Call:</td>
<td>H2020-FET-HPC-1-2014</td>
</tr>
<tr>
<td>Type of action:</td>
<td>H2020: Research &amp; Innovation Actions (RIA)</td>
</tr>
<tr>
<td>Topics:</td>
<td>HPC Core Technologies, Programming Environments and Algorithms for Extreme Parallelism and Extreme Data Applications</td>
</tr>
<tr>
<td>Subtopic</td>
<td>b) Programming methodologies, environments, languages and tools</td>
</tr>
<tr>
<td>Project Coordinator</td>
<td>Cristina Silvano, Politecnico di Milano</td>
</tr>
<tr>
<td>EC Contribution</td>
<td>3, 115, 251 euro</td>
</tr>
<tr>
<td>Project start:</td>
<td>September 1st, 2015 (duration 3 years)</td>
</tr>
</tbody>
</table>
Kick-off Meeting held at CINECA (Italy)  
9-10 September 2015
To reach the DARPA’s target of 20MW of Exascale supercomputers projected to 2020, current supercomputers must achieve an energy efficiency “quantum leap”, pushing towards a goal of 50 GFlops/W.

Heterogeneous systems currently dominate the top of the Green500 list and this dominance is expected to be a trend for the next coming years to reach the target of 20MW Exascale supercomputers.

Energy-efficient heterogeneous supercomputers need to be coupled with a radically new software stack capable of exploiting the benefits offered by heterogeneity at different levels (supercomputer, job, node) to meet the scalability and energy efficiency required by the Exascale era.
Target Scenario

- To reach the DARPA’s target of 20MW of Exascale supercomputers projected to 2020, current supercomputers must achieve an energy efficiency "quantum leap", pushing towards a goal of 50 GFlops/W.

- Heterogeneous systems currently dominate the top of the Green500 list and this dominance is expected to be a trend for the next coming years to reach the target of 20MW Exascale supercomputers.

- Energy-efficient heterogeneous supercomputers need to be coupled with a radically new software stack capable of exploiting the benefits offered by heterogeneity at different levels (supercomputer, job, node) to meet the scalability and energy efficiency required by the Exascale era.

The main goal of the ANTAREX project is to provide a breakthrough approach to express by a Domain Specific Language the application self-adaptivity and to runtime manage and autotune applications for green and heterogeneous High Performance Computing (HPC) systems up to the Exascale level.
ANTAREX Main Objectives

1. **Dynamic self-monitoring** and **self-adaptivity or «autotuning»** HPC applications with respect to changing workloads, operating conditions and computing resources.

2. **Programming models and languages to express self-adaptivity and extra-functional properties.** Introducing a **separation of concerns** between extra-functional (self-adaptivity, parallelisation, energy/thermal management) strategies and application functionality by the design of a new aspect-oriented **Domain Specific Language**.

3. Exploiting heterogeneous computing resources in Green HPC platforms by **runtime resource and power management**
The ANTAREX Approach

- Introducing a new DSL (as and extension of the LARA language) for expressing **adaptivity and autotuning strategies**.
- Enabling performance/energy control capabilities by introducing **software knobs** (application parameters, code transformations and code variants).
- Designing optimal control-loops to manage **performance/energy metrics** (job dispatching, resource management and dynamic voltage/frequency scaling) at different time scale (compile time, deployment time, runtime).
- Monitoring the evolution of the supercomputer as well as the **application status and requirements** to provide this information to the ANTAREX energy/performance-aware software stack.
Compiling and Optimizing with LARA DSL: Motivations

- Enable **separation of concerns**: functional and non-functional descriptions are decoupled
- Enable sophisticated **strategies** for code instrumentation and synthesis/compiler optimizations
- Enable users and DSE engines to fully explore **compiler optimizations**
- Enable users to apply the most suitable **compiler sequence** according to code and target architecture
- Enable more advanced **control** than using pragmas/directives/switches
LARA DSL: design benefits

Reusable Strategies

app 1 + aspect = design 1

app 2 + aspect = design 2

app N + aspect = design N

Custom Targetability

app + aspect = design

Design Exploration

original app + aspect = design N

design parameters

design 1

design 2
Application Autotuning

- One or more application parameters, code transformations and code variants (*application knobs*) can be tuned at runtime.
- Adaptivity to adjust the application behavior to the changing operating conditions, usage contexts and resource availability.
Application Autotuning

- One or more application parameters, code transformations and code variants (*application knobs*) can be tuned at runtime
- Adaptivity to adjust the application behavior to the changing operating conditions, usage contexts and resource availability
- Approximate computing: output just needs to be "good enough" trading off accuracy and throughput
• Define **mechanisms** and **strategies** to be adopted at application-level to on-line adapt the application behaviour and the platform configuration.

• To use the **ANTAREX DSL** to instrument the application for monitoring and to support the autotuning.

• To search for the best combination of the **software knobs** (i.e. application parameters, code transformations and code variants) impacting the performance and energy efficiency of the application w.r.t. the execution environment and dynamic workload.
• Autotuning framework will implement a **collect-analyse-decide-act loop** to make the application behaviour self-aware.
ANTAREX Toolflow
Progress beyond state-of-the-art

- Programming Models, Language & Compilers
- Application Scenarios
- Self-Adaptivity & Application Autotuning
- Runtime Resource & Power Management
- Target HPC Systems
Two HPC application scenarios

1. A biopharmaceutical HPC application for accelerating drug discovery provided by Dompé and to be deployed on the 1.2 PetaFlops heterogeneous NeXtScale Tier-1 Intel-based IBM system at CINECA.

2. A self-adaptive navigation system provided by Sygic to be used in smart cities and deployed on the server-side on an heterogeneous Intel-based PetaFlops class system given by IT4Innovations Supercomputing Center.

- Performance metrics extracted from the two use cases will be modelled to extrapolate these results towards *Exascale systems* expected by the end of 2020.
Use Case 1: Computer Accelerated Drug Discovery System

- **Personalized Medicine** will enable to "treat the right patient with the right drug at the right dose at the right time." <FDA>

- **Need of HPC in Drug Discovery**
  - HPC Molecular Simulations
LiGen Project: A High Performance Workflow for Chemistry Driven de Novo Design
Use Case 2: Self-adaptive Navigation System

- **Sygic Company** develops world’s most popular offline navigation application and provides [professional navigation software](#) for business solutions
Use Case 2: Self-adaptive Navigation System

- Many drivers – many routing requests to HPC system
- Smart City Challenge – serve all city drivers’ requests with global best