EFICAS

Finding the balance between AI performance and power consumption

The AI Call 2020 project EFICAS (Energy Efficient Heterogeneous AI-Framework for Smart Mobile and Embedded Systems) will create a software solution which can run on heterogeneous computation environments. This will uncover the right balance between performance and energy efficiency for AI algorithms in various domains.

Addressing the challenge

AI can enable greater performance in domains such as mobility and automation but is not power-efficient. As the increased functional complexity of mobile and autonomous applications increases power demands on embedded platforms, the computational load is raised; AI-driven concepts therefore necessitate lower energy consumption. Although various parties are competing to achieve this, the results are dedicated solutions with special market participation. Instead of focusing on improved AI performance, there is an enormous need for greater resource efficiency for available AI frameworks and algorithms.

Proposed solutions

EFICAS will create an AI-driven software framework that merges heterogeneous technologies via a cognitive, power-sensitive middleware which utilises performance and consumption markers at runtime. By dynamically assigning tasks to the best-suited architecture (based on timing and geographical constraints), this will create an optimal trade-off between performance and energy consumption. This framework will also include a library of functions and services to motivate the adoption of a variety of AI/ML methodologies and algorithms.
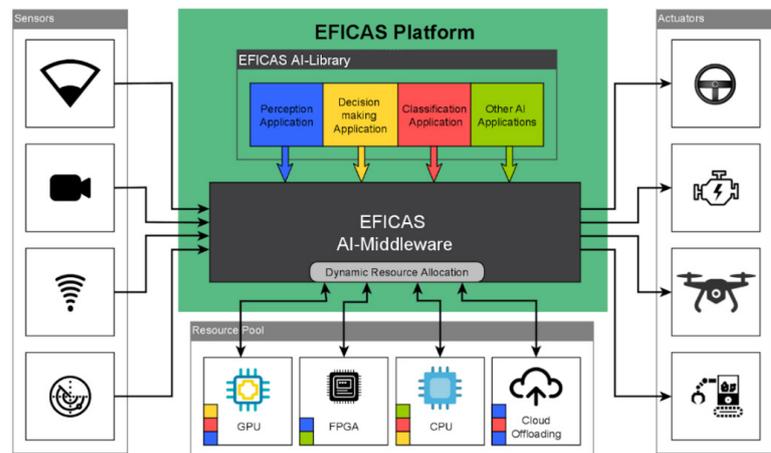
EFICAS capitalises on the fact that Central Processing Units (CPUs) are flexible but work sequentially, Graphics Processing Units (GPUs) are less flexible but work in parallel, and dedicated accelerators such as Field Programmable Gate Arrays (FPGAs) have little flexibility but combine high parallelism and execution speed. The use of heterogeneous architectures can thus offer advantages over monolithic architectures based solely on FPGAs or

⌃ EFICAS Framework Structure

GPUs. The resulting reductions to energy requirements will be demonstrated in use-cases on autonomous agricultural vehicles, autonomous mobile robots and automation systems.

Projected results and impact

EFICAS's energy efficiency will be measured in floating point operations per second (FLOPS/W) against a baseline of 10-20 GFLOPS/W for a GPU-based solution. In a low computational load scenario combining FPGA and GPU, 25 GFLOPS/W can be achieved with 30-100% energy savings. For high computational load, 40 GFLOPS/W can be reached with 20-60% energy savings. In addition, the implementation time for ML algorithms with the EFICAS framework will be approximately the same as with monolithic platforms.

Overall, the EFICAS middleware's performance/energy consumption trade-off gain is expected to be up to 15%. This translates, for instance, into 12.5 hours more operating time for autonomous mobile robots or an 11% greater range for auton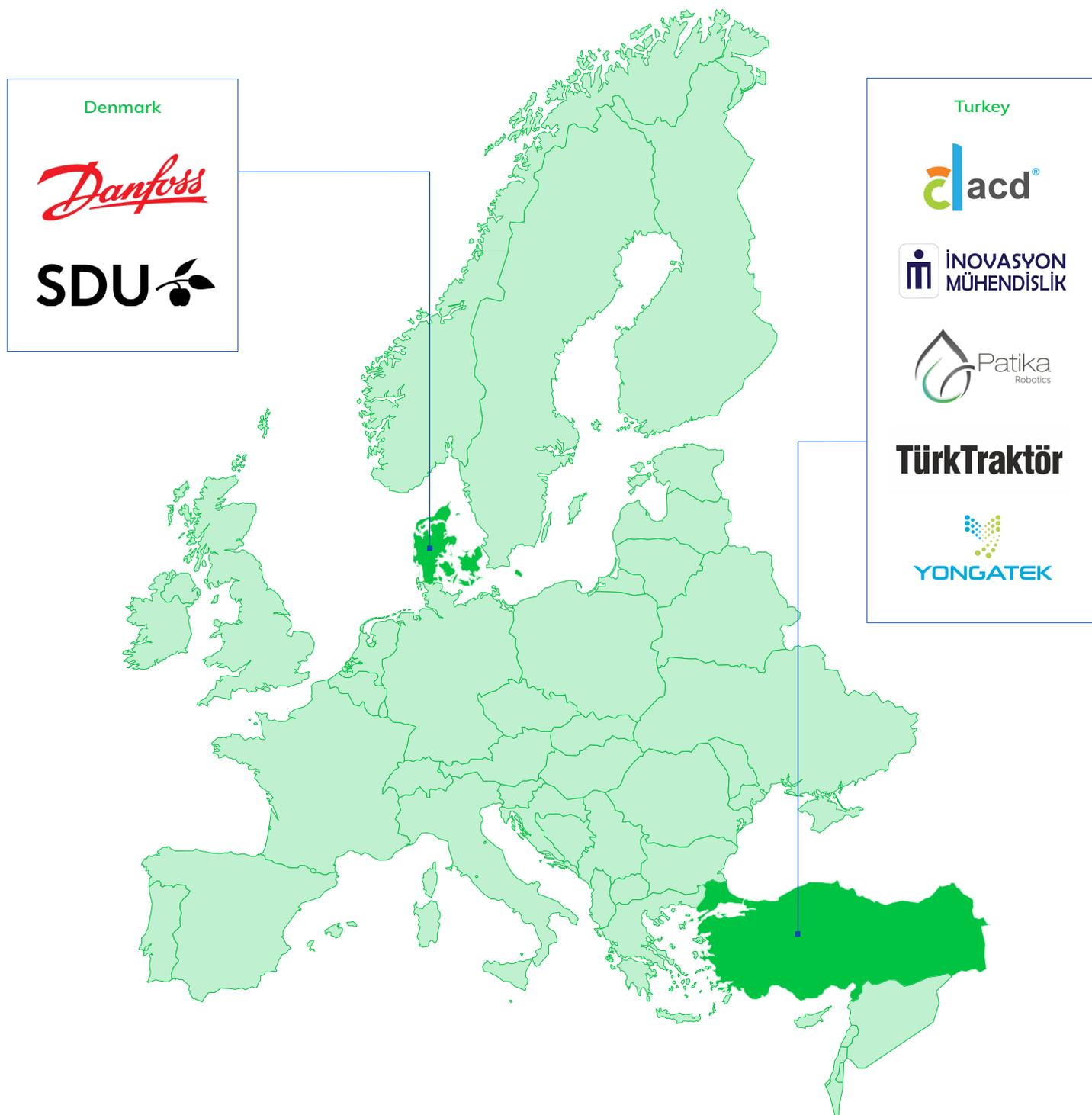omous agricultural vehicles. As the GPU market will reach USD 157.1 billion in 2022 (a compound annual growth of 35.6% since 2016) and the FPGA market will hit USD 9.5 billion in 2023 (more than double the 2015 value), EFICAS offers a golden opportunity to establish a position in these rapidly growing domains as well as the emerging market that combines them.

Project partners

Denmark

Danfoss

SDU

Turkey

acd

İNOVASYON MÜHENDİSLİK

Patika Robotics

TürkTraktör

YONGATEK

ITEA is the Eureka R&D&I Cluster on software innovation, enabling a large international community of large industry, SMEs, start-ups, academia and customer organisations, to collaborate in funded projects that turn innovative ideas into new businesses, jobs, economic growth and benefits for society. ITEA is part of the Eureka Clusters Programme (ECP).

https://itea4.org

ITEA4

Σ eureka