The project aim is to provide theoretical and engineering results that will support the ICT community with new knowledge supporting design of energy-aware resource and job management systems (RJMS) capable of introducing guarantees for power consumption and application performance in data centers. Contributions in the area of energy-efficient distributed and parallel computing will support growth of the market of environment-friendly cloud services. The expected results may improve competitiveness of Polish ICT solutions as well as the involvement in the mainstream EU Exascale computing project. The project addresses the problem at the nexus of computer science, control engineering and communication, proving its interdisciplinarity.

The scientific objectives of the project include identification of data processing models, development of system observers and workload predictors, and synthesis of efficient control rules (green governors) compatible with the Linux kernel architecture. The main scientific challenge is to efficiently exploit the new knowledge regarding workload dynamics in the highly stochastic self-tuning environment of data center.

Motivation for the present research comes from current trends in development of modern computer architectures and tools for parallel programming. Prior research has applied control-theoretic approach to design power regulation and application performance control structures for data centers. However, the obtained solutions have been limited by the lack of software probes and hardware sensors supporting high-resolution performance and power consumption measurements. The approach based on machine learning has not been explored sufficiently neither. The ambition of this project is to make significant contributions to the field of energy-efficient computing by fully exploiting the advanced machine learning and approximate dynamic programming techniques and novel measurement sensors of modern computing hardware and software (Linux-OS). Precisely, kernel-level software and hardware register-based counters (providing measurements of instructions per cycle, cache-misses, memory-loads and running average power limits, etc.) will be observed in the course of specifically designed experiments. The collected data will be next used to develop maximally informative computing metrics and accurate dynamical system models. Due to high-frequency and fine-grained measurements, the identified models are expected to radically improve accuracy of system workload predictors, outperforming commonly used state-of-the-art solutions. In contrast to the research carried out so far, the project will follow the holistic approach combining tools of system identification, optimization, stochastic control and artificial intelligence.

Indeed, according to the analysis of current trends (gesi.org/SMARTer2020), the carbon dioxide emissions of the ICT industry are expected to exceed 2% of the global emissions, a level equivalent to the contribution of the aviation. Data centers, providing both cloud and high performance computing (HPC) services, consume increasing amounts of electrical energy. To be more specific, in the period from 2005 to 2010 the energy consumed by data centers worldwide rose by 56%, which was accounted to be between 1.1% and 1.5% of the total electricity use in 2010. The growth of energy consumption rises operating costs of data centers but also contributes to carbon dioxide (CO2) production.

Energy efficiency (FLOPS/W) of ICT systems continues to improve (www.green500.org). However, the rate of improvement does not match the growth rate of demand for computing capacity. Unless radically new energy-aware technologies are introduced, both in hardware and software domain, it will not be possible to meet DARPA's 20-MW exaflop goal (50 GFLOPS/W) by year 2020. Computational power improvements are, in fact, heavily constrained by the energy budget that is necessary for driving data centers. Limiting power consumption and related thermal emission has therefore become a key problem. Based on the projections of technology development, it has been argued that the continued scaling of the available systems will eventually lead to a data center consuming more than a gigawatt of electrical power (at Exaflop level), a level that violates economic rationale for providing cloud or HPC services. Optimization of energy consumption in data centers is necessary and must be addressed in response to the market and environment protection needs