Guest Editorial: Special Issue on Self-Aware Systems on Chip

Axel Jantsch, Nikil Dutt

September 25, 2017

The core insight that motivates research on self-aware Systems on Chip (SoCs) is that good decision making requires a sound understanding of the situation. As there are many decisions to be taken for the resource and infrastructure management of complex, heterogeneous SoCs, the collection of data about the chip and its environment, and their interpretation is well justified. If the chip itself has gained a comprehensive and sufficiently complete understanding of its state and of its environment, it is called self-aware.

The current trends make this issue ever more urgent. While the number of cores grows slowly in IoT devices, modestly in mobile devices and fast in data centers, the heterogeneity is on the rise in all devices due to the relentless pursuit of efficiency and, in particular, the tight constraints on power consumption. Already today, with an energy of 2.52 fJ per switching activity only 10% of the transistors are allowed to switch concurrently in a chip with 1 billion transistors clocked at 1 GHz, if the power consumption should be limited to 100W. Current trends will accentuate this constraint as the device density grows faster than the switching energy decreases. According to the International Roadmap for Devices and Systems (2016 edition) the accumulated increase of device density between 2017 and 2021 is 4.6x while the switching energy is expected to decrease by 63% which leaves us at 70% more power consumption per unit area than today.

This is only one motivation for improved resource management on chip. Other technology factors are the increasing probability of transient and permanent faults, wear out and aging effects, and increasing PVT (Process, Temperature, Temperature) variations. If that were not enough, we also see an increase of diversity and dynamism in applications. Application workloads vary a great deal in terms of performance requirements but also in terms of service levels. Some require high throughput or low latency while others require adherence to strict deadlines or user experience. Even the same application may change its characteristics over time.

Putting all these considerations together it becomes obvious that on chip resource management in terms of voltage and frequency scaling, power gating, supply and body biasing, task mapping, scheduling, traffic shaping, memory allocation and cache management, fault detection and management is bound to gain in importance.

The basis of good resource management decisions is a complete and correct assessment of the current situation. Many sensors can be and are embedded in the chip to monitor the power consumption and temperature, delays, functional integrity, buffer occupancy, link and network

load, processor utilization, deadline misses, and many more attributes. Measurements are taken at dozens or hundreds of locations, sometime with high frequency, and on all relevant levels from the device to the architecture, middleware and application.

The vision of a self-aware SoC is to combine all these data, abstract and assess them properly in the light of given objectives in order to provide a solid basis for decision making.

This is the first of two special issues on self-aware SoCs which show that many of the methods necessary to achieve this vision are already there or are under active study. But they also demonstrate that some topics urgently require more attention. In particular the integration of point techniques into a comprehensive self-assessment is still incomplete.

The first paper is a survey of the state of the art, which reviews work that use self-awareness in systems on chip, but also relates to several threads of relevant work in autonomic computing, organic computing, adaptive systems and control theory.

This is followed by the article "Health Management for Self-Aware SoCs based on IEEE 1687 Infrastructure" by K. Shibin et al., which is motivated by faults and fault tolerance and provides a rather complete approach to health monitoring and management of an SoC.

The third paper by J. J. Davis et al., "KOCL: Power Self-awareness for Arbitrary FPGA-SoC-accelerated OpenCL Applications", introduces a tool for FPGA designs developed with OpenCL to monitor and manage power consumption from the system level.

The fourth paper "A Self-aware Architecture for PVT Compensation and Power Nap in Near-threshold Processors" by D. Rossi et al. describes software-controllable body biasing to compensate PVT variations and for implementation of low-power modes in near-threshold processors.

The last paper of this special issue, "Self-Adaptive Timing Repair" by H. Giesen et al., addresses timing variations in FPGAs by measuring delays during operation, identifying slow links and speeding them up by trying alternative interconnect and gate configurations.

While the article by K. Shibin et al. describes a system solution to monitor and manage the SoC health, the last three papers present point techniques that each offer a solution to a specific problem of power or performance monitoring and management. Each of them helps to make the system as a whole more aware of itself in order to manage itself.

The second special issue on Self-Aware Systems on chip, to appear in Design & Test of Computers, will continue this pattern by providing a survey on self-test and self-diagnosis in SoCs, and point techniques for the monitoring and management of power, reliability, temperature and quality of service.

These articles provide a snapshot of the state of the art and hopefully can also give an impression of the direction and dynamics of this important research theme. Enjoy reading!