GreenLight Technical Report

2010

Power and Thermal Management

Tajana Simunic Rosing: Power/Thermal Management

Research Activities

The goals of our work are to develop strategies to analyze the power, thermal and workload dynamics, and to design methods to reduce power consumption while mitigating the temperature induced problems in datacenter environments. Thus far we have implemented and tested a number of reactive and proactive thermal management techniques and evaluated the benefits of optimizing temperature while minimizing energy, versus only focusing on lowering the system power/energy consumption. We are in the process of extending Xen, a virtual machine monitor for IA-32 IA-64 and PowerPC 970 architectures, to enable power and thermal management within the virtualization technology.

Energy management in virtualized environments [9]

We have designed vGreen, a multi-tiered software system for energy efficient virtual machine management in a clustered virtualized environment. The system leverages the use of novel hierarchical metrics that work across the different abstractions in a virtualized environment, and capture power and performance characteristics of both the virtual and physical machines. These characteristics are then used to implement policies for scheduling and power management of virtual machines across the cluster. We show through real life implementation of the system on a state of the art test bed of server machines that vGreen can improve both average performance and system level energy savings by close to 40% across benchmarks with varying characteristics.

Accurate direct and indirect temperature sensing for thermal management [1]

Dynamic thermal management techniques require accurate runtime temperature information in order to operate effectively and efficiently. We designed two novel solutions for accurate sensing of system temperature. Our first technique is used at design time for sensor allocation and placement to minimize the number of sensors while maintaining the desired accuracy. The experimental results show that this technique can improve the efficiency and accuracy of sensor allocation and placement compared to previous work and can reduce the number of required thermal sensors by about 16% on average. Secondly we propose indirect temperature sensing to accurately estimate the temperature at arbitrary locations based on the noisy temperature readings from a limited number of sensors which are located further away from the locations of interest. Our runtime technique for temperature estimation reduces the standard deviation and maximum value of temperature estimation errors by an order of magnitude.

Predictive thermal management [2,8]

Conventional thermal management techniques are reactive in nature; that is, they take action after temperature reaches a predetermined threshold value. Such approaches do not always minimize and balance the temperature on the chip, and furthermore, control temperature at a noticeable performance cost. In this work, we investigate how to use predictors for forecasting future temperature and workload dynamics, and propose proactive thermal management techniques for multiprocessor system-on-chips (MPSoCs). The predictors we study include autoregressive moving average (ARMA) modeling and look- up table based predictors. We implement several reactive and predictive thermal management techniques, and provide extensive evaluations on an UltraSPARC T1 system as well as on an architecture-level simulator. By means of dynamic prediction, proactive methods can achieve significantly better thermal profiles in comparison to their reactive counterparts, and the performance cost associated with dynamic thermal management can be mostly avoided.

Energy and reliability aware workload scheduling [6,7]

Temperature-induced reliability issues are among the major challenges for multicore architectures. Thermal hot spots and thermal cycles combine to degrade reliability. As a part of this work, we designed new reliability-aware job scheduling and power management approaches for chip multiprocessors. Accurate evaluation of these policies requires a novel simulation framework that can capture architecture- level effects over tens of seconds or longer, while also capturing thermal interactions among cores from dynamic scheduling policies. Using this framework and a set of new thermal management policies, this work shows that techniques that offer similar performance, energy, and even peak temperature can differ significantly in their effects on the expected processor lifetime.

Furthermore, we also considered heterogeneous multiprocessor designs. Since these systems have various power and performance characteristics, they can customize their configuration to achieve higher performance per Watt. On the other hand, the inherent imbalance in power densities leads to non-uniform temperature distributions, which affect performance and reliability adversely. In this work, we propose a joint thermal and energy management technique specifically designed for heterogeneous systems. Our technique identifies the performance demands of the current workload. By utilizing job scheduling and voltage/frequency scaling dynamically, we meet the desired performance while minimizing the energy consumption and the thermal imbalance. In comparison to performance-aware policies such as load balancing, our technique simultaneously reduces the thermal hot spots, temperature gradients, and energy consumption significantly.

Cooling-aware workload scheduling [3,4,5]

We designed *GentleCool*, a proactive multi-tier approach for significantly lowering the fan cooling costs without compromising the performance. Our technique manages the fan speed through intelligently allocating the workload across different machines. The experimental results show our approach delivers average cooling energy savings of 72% and improves the mean time between failures (MTBF) of the fans by 2.3X compared to the state of the art.

In addition, we developed a detailed thermal model of the microchannels embedded between the tiers of 3D integrated systems. For energy-efficient cooling, we implemented a novel controller and job scheduler to adjust the liquid flow rate to meet the desired temperature and to minimize pump energy consumption. Our method guarantees operation below the target temperature while reducing the cooling energy by up to 30%, and the overall energy by up to 12% in comparison to using the highest coolant flow rate.

Energy efficient memory architecture [10]

To address the problem of memory energy efficiency, we designed a novel main memory architecture based on phase change random access memory (PRAM), a low power, fast access, non volatile memory. We propose a hybrid hardware/software solution. In order to maintain reliability for PRAM (because of write endurance problem), we introduce a cost efficient book keeping hardware technique that stores the frequency of writes to PRAM at a page level granularity. On the software side, we propose an efficient OS level page manager that utilizes the write frequency information provided by the hardware to perform uniform wear leveling across all the PRAM pages. Furthermore, the page manager intelligently allocates/migrates pages across DRAM/PRAM in order to minimize the impact of wear leveling on performance. Our experimental results indicate that our solution is able to achieve average energy savings of 30% at negligible overhead over conventional memory architectures.

References:

S. Sharifi, T. Simunic Rosing, “Accurate direct and indirect on-chip temperature sensing for efficient dynamic thermal management,” to appear in IEEE TCAD, 2010.

A. Coskun, T. Simunic Rosing, “Utilizing Predictors for Efficient Thermal Management in Multiprocessor SoCs,” IEEE TCAD, 2009.

R. Ayub, S. Sharifi, T. Simunic Rosing, “GentleCool: cooling aware proactive workload

scheduling in multi-machine systems," DATE'10.
A. Coskun, D. Atienza, T. Simunic Rosing, "Energy-efficient variable-flow liquid cooling in 3D stacked architectures," DATE'10.

R. Ayoub, T. Simunic Rosing, "Cool and Save: Cooling Aware Dynamic Workload Scheduling in Multi-socket CPU Systems," ASPDAC'10.
S. Sharifi, A. Coskun, T. Simunic Rosing, "Hybrid Dynamic Energy and Thermal Management in Heterogeneous Multiprocessors," ASPDAC'10.
A. Coskun, R. Strong, D. Tullsen, T. Simunic Rosing, "Evaluating the Impact of Job Scheduling and Power Management on Processor Lifetime for Chip Multiprocessors, " SIGMETRICS'09.
R. Ayoub, T. Simunic Rosing, "Predict and Act: Dynamic Thermal Management for Multicore Processors," ISLPED'09.
G. Dhiman, R. Ayoub, G. Marchetti, T. Simunic Rosing, "vGreen: A System for Energy Efficient Computing in Virtualized Environments," ***Nominated for the best paper award*** at ISLPED'09.
[10] G. Dhiman, R. Ayoub, T. Simunic Rosing, "PDRM: A hybrid PRAM DRAM main memory
system", DAC'09.

Research Findings

The goals of our work are to develop strategies to analyze the power, thermal and workload dynamics, and to design methods to reduce power consumption while mitigating the temperature induced problems in datacenter environments. Thus far we have implemented and tested a number of reactive and proactive thermal management techniques evaluated the benefits of optimizing temperature while minimizing energy versus only focusing of lowering the system power/energy consumption. We show through real life implementation of the system on a state of the art test bed of server machines that vGreen, our energy management system based on virtualization technology, can improve both average performance and system level energy savings by close to 40% across benchmarks with varying characteristics [9]. To address the problem of memory energy efficiency, designed a hybrid hardware/software solution combining DRAM and PCM memory [10]. Our experimental results indicate that our solution is able to achieve average energy savings of 30% at negligible overhead over conventional memory architectures.

In the area of thermal management we have developed techniques for accurate thermal sensing. Our runtime technique for temperature estimation reduces the standard deviation and maximum value of temperature estimation errors by an order of magnitude [1]. Furthermore, we designed predictive methods for thermal management [2,8], which are capable of reducing the incidence of thermal hot spots by more than 60% relative to the state-of-the art techniques in modern multicore server systems, with negligible performance penalty. This has led to development of energy- and reliability-aware workload scheduling techniques [6,7] . By utilizing job scheduling and voltage/frequency scaling dynamically, we meet the desired performance while minimizing the energy consumption and the thermal imbalance. In comparison to performance-aware policies such as load balancing, our technique simultaneously reduces the thermal hot spots, temperature gradients, and energy consumption significantly. Lastly, we designed *GentleCool*, a proactive multi-tier approach for significantly lowering the fan cooling costs without compromising the performance [3]. The experimental results show our approach delivers average cooling energy savings of 72% and improves the mean time between failures (MTBF) of the fans by 2.3X compared to the state of the art. A similar idea has been applied to liquid cooling in 3D ICs. Here our method [4] guarantees operation below the target temperature while reducing the cooling energy by up to 30%, and the overall energy by up to 12% in comparison to using the highest coolant flow rate.

Published and referenced work:

[1] S. Sharifi, T. Simunic Rosing, "Accurate direct and indirect on-chip temperature sensing for efficient dynamic thermal management," to appear in IEEE TCAD, 2010.
[2] A. Coskun, T. Simunic Rosing, "Utilizing Predictors for Efficient Thermal Management in Multiprocessor SoCs," IEEE TCAD, 2009.
[3] R. Ayub, S. Sharifi, T. Simunic Rosing, "GentleCool: cooling aware proactive workload scheduling in multi-machine systems," DATE'10.
[4] A. Coskun, D. Atienza, T. Simunic Rosing, "Energy-efficient variable-flow liquid cooling in 3D stacked architectures," DATE'10.
[5] R. Ayoub, T. Simunic Rosing, "Cool and Save: Cooling Aware Dynamic Workload Scheduling in Multi-socket CPU Systems," ASPDAC'10.
[6] S. Sharifi, A. Coskun, T. Simunic Rosing, "Hybrid Dynamic Energy and Thermal Management in Heterogeneous Multiprocessors," ASPDAC'10.
[7] A. Coskun, R. Strong, D. Tullsen, T. Simunic Rosing, "Evaluating the Impact of Job Scheduling and Power Management on Processor Lifetime for Chip Multiprocessors, " SIGMETRICS'09. [8] R. Ayoub, T. Simunic Rosing, "Predict and Act: Dynamic Thermal Management for Multicore
Processors," ISLPED'09. [9] G. Dhiman, R. Ayoub, G. Marchetti, T. Simunic Rosing, "vGreen: A System for Energy Efficient
Computing in Virtualized Environments," ***Nominated for the best paper award*** at ISLPED'09. [10]
G. Dhiman, R. Ayoub, T. Simunic Rosing, "PDRM: A hybrid PRAM DRAM main memory system", DAC'09.

For more information please contact Tajana Rosing, trosing@ucsd.edu