Cross-layer optimization for energy efficient datacenters

Prof. Tajana Simunic Rosing, UCSD
& MuSyC Datacenter Team

The Multiscale Systems Center

Focus Center RESEARCH PROGRAM

Summary of key team's results

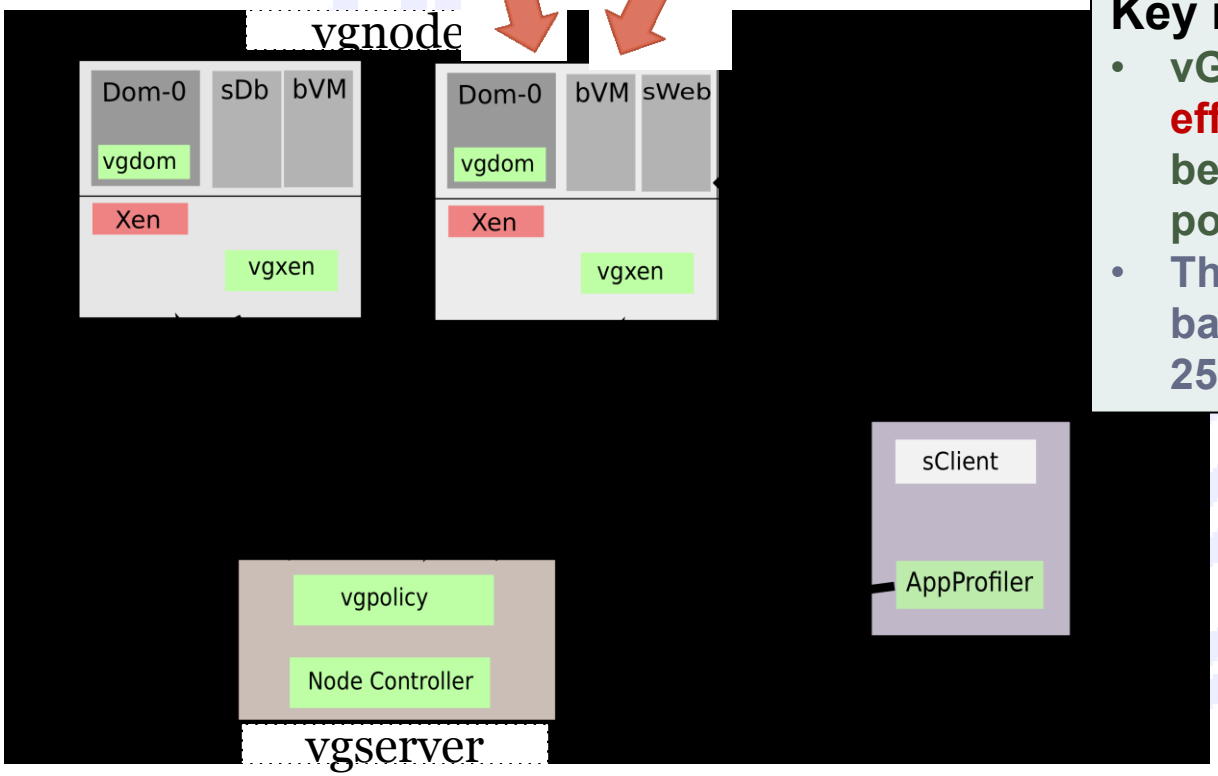
	Energy efficiency	Performance	Resilience
Applications	AS: 20% gain in energy efficiency with min perf. loss		RFL: up to 99% survival rate with double bit faults injected
Runtimes	RHK: 50% reduction in energy while meeting QoS requirements of web services and Hadoop workloads		TSR: 70% reduction in hot spots leads to reliability improvements via thermal prediction & OS/VM scheduling
OS/VM	VS: 35x energy and 17x performance improvement via heterogeneous job and HW scheduling		
Storage	TSR: 70% improvement in energy efficiency while meeting QoS and throughput requirements of jobs	TSR: 92% reduction in job perf. variability &17% power savings	RAMCloud: recovery from isolated crashes in 1-2 seconds
Network	RAMCloud: > 100x reduction in energy/op	RAMCloud: > 100x reduction in latency for small storage operations	JD: 30% energy reduction for resilient router
Powering	AV/GP: 6x energy reduction with Helios	AV/GP: 2.8x reduction in cost-performance with Helios	
Cooling	RKG: 68% energy savings/node	TSR: up to 40% energy savings with no performance hit with power gating	
Green energy	RKG: 40% HVAC Savings	TSR: 70% reduction in energy cost of cooling with 0.2% performance overhead & no change to system reliability	
SmartGrid	RHK: 62% penetration of renewables	TSR: 93% green energy usage efficiency and 22% performance improvement with green energy prediction	
	RKG: 20-40% energy savings by deferral across grid	TSR: 1.5x profit per server due to battery peak power shaving with no performance overhead	

More than 100x energy savings while improving performance

vGreen: Maximize energy efficiency while managing SLAs at VM level

Batch Jobs:
•Throughput driven

Services:
•Latency sensitive



Key results:

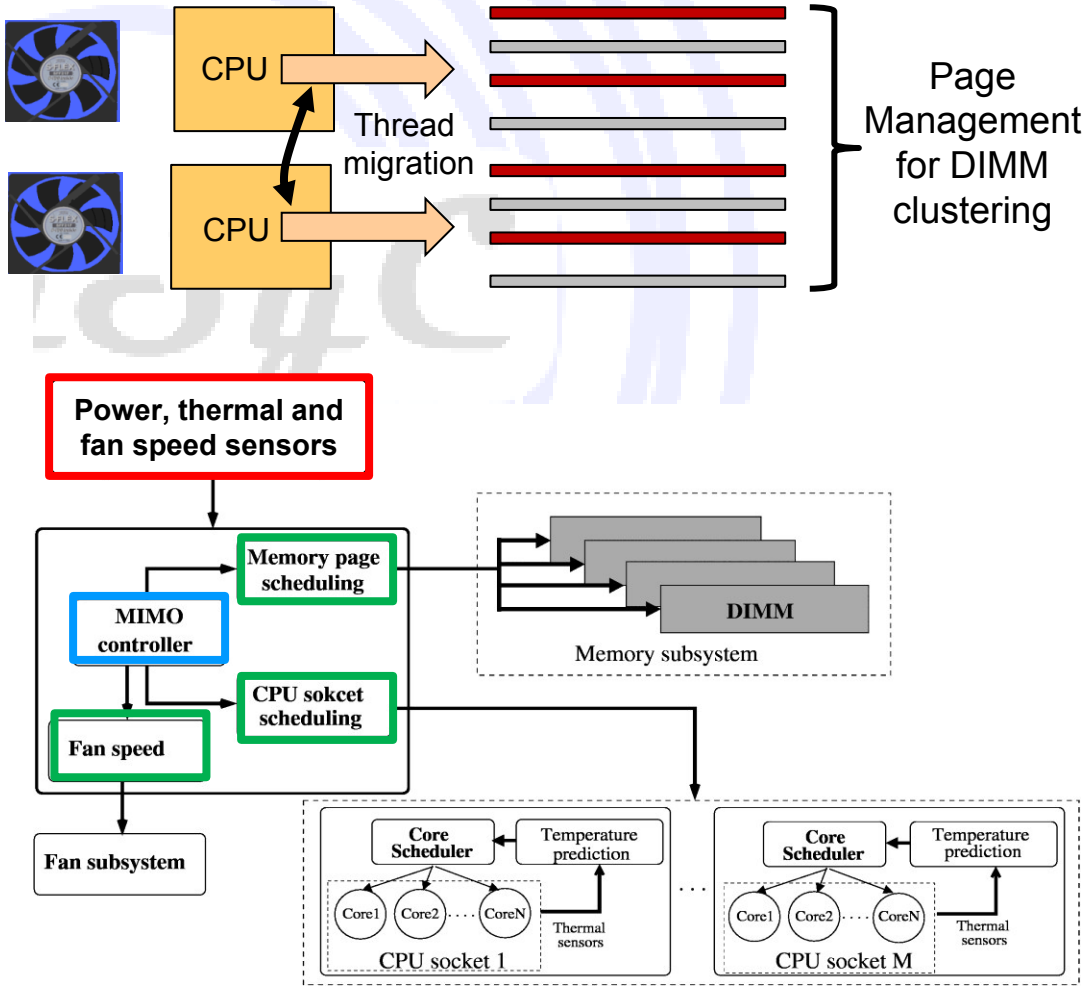
- vGreen is **70% more energy efficient** than the baseline & 35% better than ideal state of the art policies
- Throughput is within 7% of baseline with our controller vs. 25% for CPU capping

Jointly with: Gaurav Dhiman, Chris Sadler, Google

Integrating energy, temperature and cooling management

- **Key contribution**: fan control done *jointly* with job and memory scheduling
  - **fan power ~ fan speed^3**
  - today's fan controllers operate independently of workload scheduling
- Formal state-space control to ensure stability
- **Controller** decides the following on each tick:
  - CPU power distribution
  - DIMM power distribution
  - Desired temperature to control fan speed
- **Actuators**: CPU, mem & fan
- **Sensor inputs**: temperature, power, fan speed

Average energy savings of 70% relative to state-of-the-art



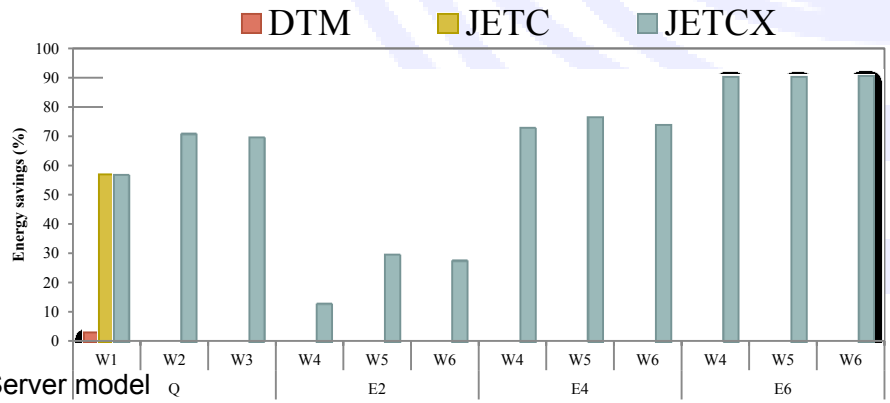
Jointly with: Raid Ayoub, Intel

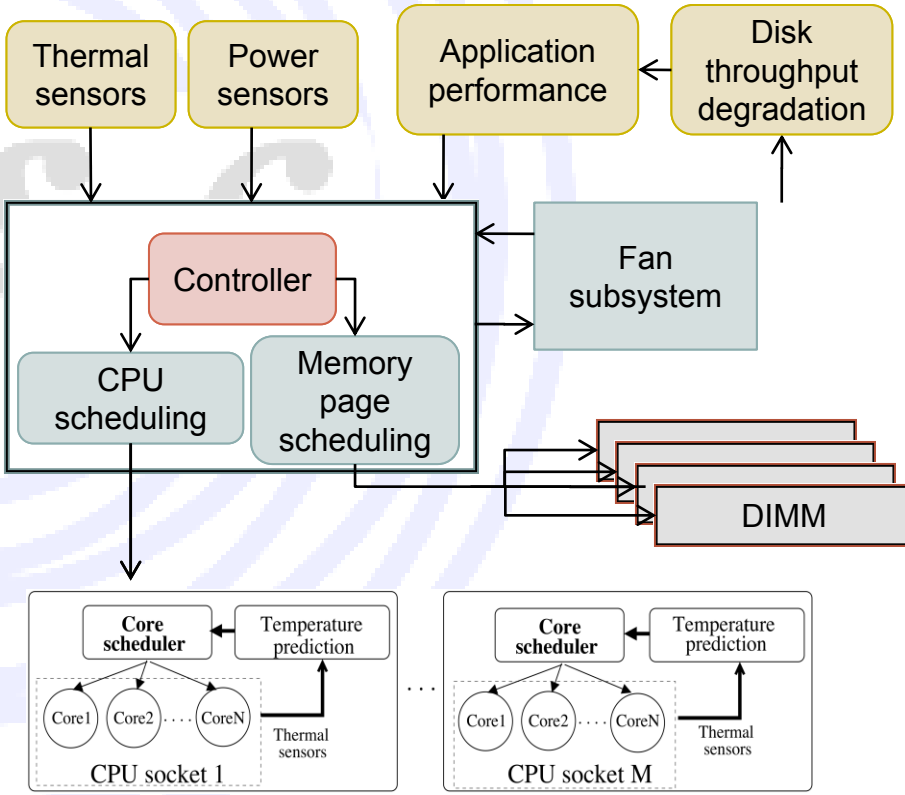
Energy & performance side effects of cooling

- Coupling between air-cooling system and hard disks
  - Mechanical vibrations from the fans lower the hard disk throughput
  - Overall performance degrades, especially for disk intensive workloads
- TPC-H queries running on Intel Xeon:

Query	1	13	19
Ave. disk BW	59%	67%	99%
Performance hit with disk BW degradation:			
80% BW available	24%	69%	20%
60% BW available	81%	105%	57%
40% BW available	108%	170%	100%

- Up to 90% energy savings by better balancing performance, temperature, cooling and vibrations





Jointly with: Kenny Gross, Oracle

Big problem to solve:

Complexity…..