**COLLOQUIUM**

Department of Computer Science and Engineering

University of South Carolina

### **Energy Proportional Datacenters**

### **Daniel Wong**

### Department of Electrical Engineering

### University of Southern California

Date: **January 30, 2015**

Time: **0930-1100**

Place: **Swearingen 1A03 (Faculty Lounge)**

# Abstract

Datacenters provide the infrastructure backbone necessary to support big data analytics and cloud services, which are increasingly employed to tackle a diverse set of grand challenges. But datacenter power consumption is growing at an unsustainable pace. In order to keep up with the hyperscale growth in datacenter demand, it is imperative that datacenters become more energy efficient. Servers, the largest power consumer in datacenters, are optimized for high energy efficiency only at peak and idle load, but rarely operate in that region. Therefore, there is a need for energy proportional computing, where servers consume power in proportion to their utilization. How to achieve or surpass ideal energy proportionality is the focus of this talk.

Toward this goal, I will first present a historical trend analysis of energy proportionality, using novel metrics, in order to identify opportunities for proportionality improvements. Second, I will present KnightShift, a heterogeneous server architecture that tightly couples a low-power Knight node with a high-power primary server, which achieves near-ideal energy proportionality. Finally, I will present the implications of high energy proportional servers on cluster-level energy proportionality. We find that traditional cluster-level energy proportionality techniques may actually limit cluster-wide energy proportionality, and it may now be more beneficial to depend solely on server-level low power techniques such as KnightShift; a finding that is a major departure from conventional wisdom.

**Daniel Wong** is a PhD candidate in Electrical Engineering at the University of Southern California. His research focuses on energy efficient design of computer systems, from data center scale to micro-architecture scale, through cross-stack solutions spanning from circuits to runtime systems. His research has been recognized as one of *IEEE Micro*’s Top Picks in Computer Architecture for 2013. He held research internship positions at Samsung Semiconductor, Inc. and Lawrence Livermore National Lab. He earned his MS in Electrical Engineering in 2011, and BS in Computer Engineering and Computer Science in 2009, both from the University of Southern California. More information can be found at <http://www.danielwong.org>.