



# Modeling Energy across Memory Hierarchy

Aditya Deshpande, Jeffrey Draper {adeshpan, draper}@isi.edu

MARINA Group - Information Sciences Institute :: Ming Hsieh Department of Electrical Engineering



## Motivation

- Energy consumption is a big concern for designers of VLSI systems at both extremes of computing-
  - Embedded – Due to limited battery life
  - Exa-scale – No one willing to pay millions in electricity bills
- The first steps in reducing energy of the system is to identify how and where energy is used in each sub-system
  - Prior research has focused intensely on processor energy, but little has been done to address other system components
- Caches /memory-subsystems are a potential candidate for energy consumption identification because –
  - They occupy large percentages of chip area
  - Active majority of time
  - Increasing trend of putting more & more caches in system exacerbates the problem
  - Have a different activity profile compared to processor core
  - Transistor size shrinking, increases the leakage power
  - No known energy reduction techniques for memory systems

We propose to measure the energy expended in various levels of the memory hierarchy so as to accurately estimate memory sub-system energy for an application execution and then propose ways to reduce its contribution to total system energy

## Experiment Setup

### Steps in Modeling energy in caches –

- Compile the program
- The executable is then instrumented with CrayPat- a performance analysis tool
- Instrumented Code is executed which captures the performance data counter or PAPI counters
- The report generated contains information for the performance counters which are then used in energy estimation

### CACTI

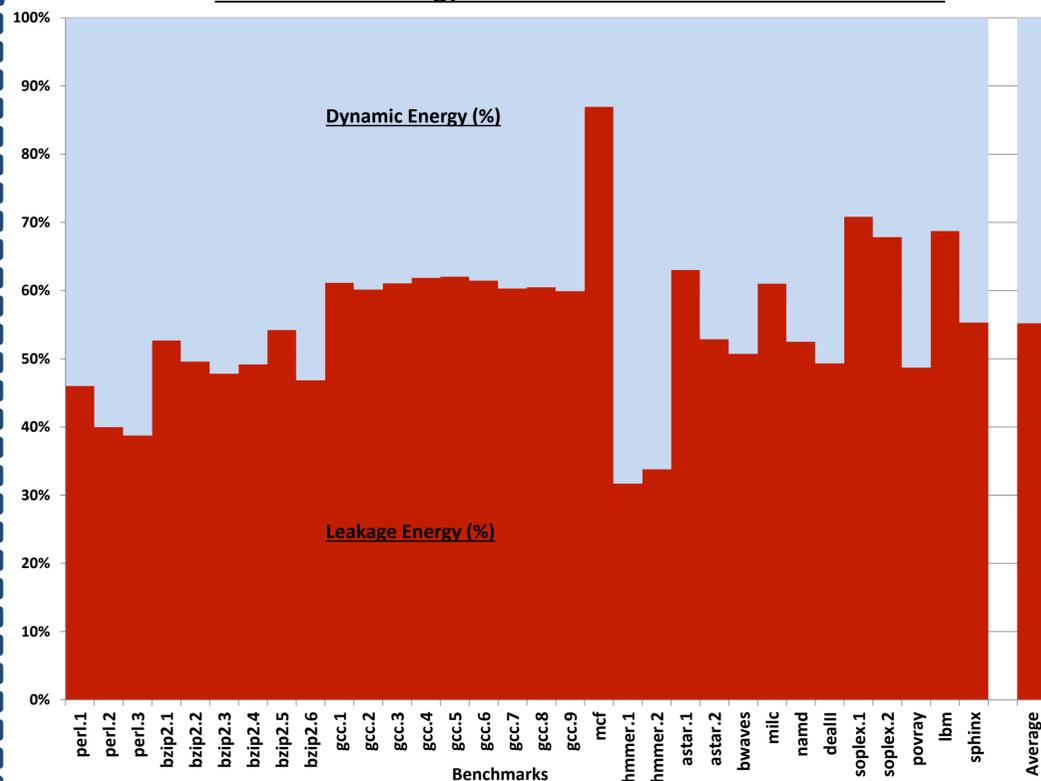
- CACTI is an integrated cache and memory access time, cycle time, area, leakage, and dynamic power model.
- Developed by HP Labs and is widely used in computer architecture field for modeling caches and memories

### Experiment Environments

- Hopper, a Cray XE6 machine located at National Energy Research Scientific Computing Center (NERSC), Berkeley, CA
- Hopper uses AMD processor cores with following specs
  - L1 Data/ Instruction (C = 64KB, B = 64B, A = 2)
  - L2 Unified (C = 512KB, B = 64B, A = 16)
  - L3 Shared (C=10240KB, B=64B, A = 96)
- We model the dynamic energy per access and leakage energy for the caches in 45nm technology with 2 Read-Write ports at 350K temperature

## Results

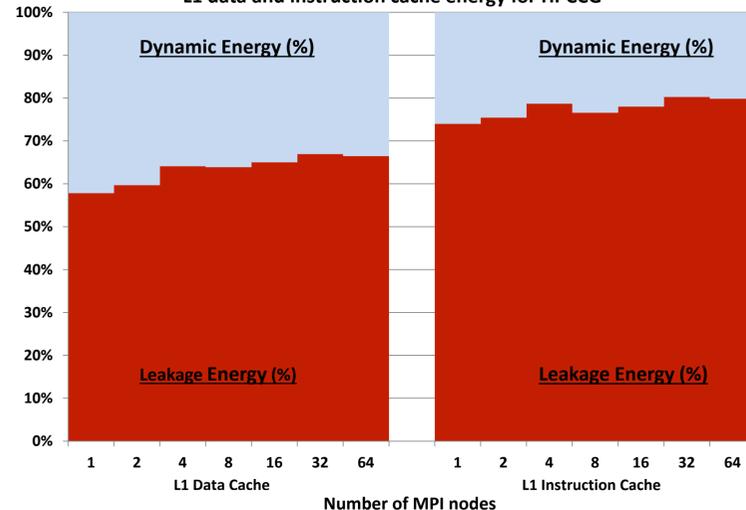
L1 data cache energy measured for SPECCPU2006 Benchmarks



### HPCCG

- One of the Mini-apps for Exascale study
- The HPCCG program is a linear system solver using the Conjugate Gradient (CG) method on an arbitrary number of processors.
- Program is from the diffusion problem on 3D chimney domain by a 27 point implicit finite difference scheme with unstructured data formats and communication patterns.
- Strong scaling with problem size constant at  $(64 \times 64 \times 1024 = 4194304)$  with 150 iterations set for results

L1 data and instruction cache energy for HPCCG



## Observation

- In L1 data cache leakage accounts for 55% of total L1 cache energy for SPECCPU2006 benchmarks
- For HPCCG, moving data for MPI runs resulted in more caches being active for longer time leading to higher leakage energy
- Memory sub-systems are becoming bottlenecks in reducing system energy –
  - Various DVFS techniques reduce energy for cores
  - Data movement in memory-systems now require more energy than instruction execution
  - No Dynamic/Leakage reduction techniques across memory hierarchy exists for reducing memory system overall energy

## Conclusion/Future Works

- Need to study more applications/benchmarks, levels of memory hierarchy and how different cache architectures affect energy profile for an application execution
- Leakage in caches need to be studied more closely, especially since the problem is expected to get worse as technology scales
- Leakage studies must be incorporated more closely in the architecture space design decision and not as afterthought where dynamic and leakage reduction mechanisms may contradict each other