

Combining Data Remapping and Voltage/Frequency Scaling of Second Level Memory for Energy Reduction in Embedded Systems

Sudarshan K. Srinivasan, Jun Cheol Park and
Vincent J. Mooney III

Georgia Institute of Technology
{darshan, jcpark, mooney}@ece.gatech.edu



Outline

- Introduction
- Motivation
- Related Work in Power Modeling
- Experimental Setup
- Data Remapping
- Voltage/Frequency Scaling of Off-chip Memory and Bus
- Experimental Results
- Conclusion

Introduction

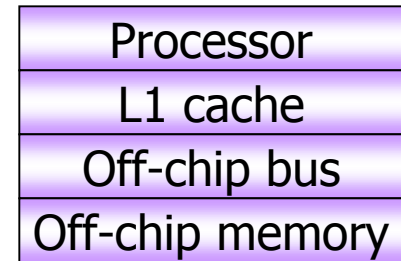
- Power/energy is a major issue in embedded systems
- Mobile devices require longer usage time



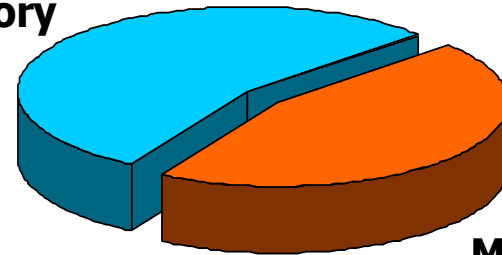
Introduction (Cont.)

- Memory consumes up to 45% of the total system power*
- Memory is a main target for power/energy reduction

Embedded System



Non-memory



Memory

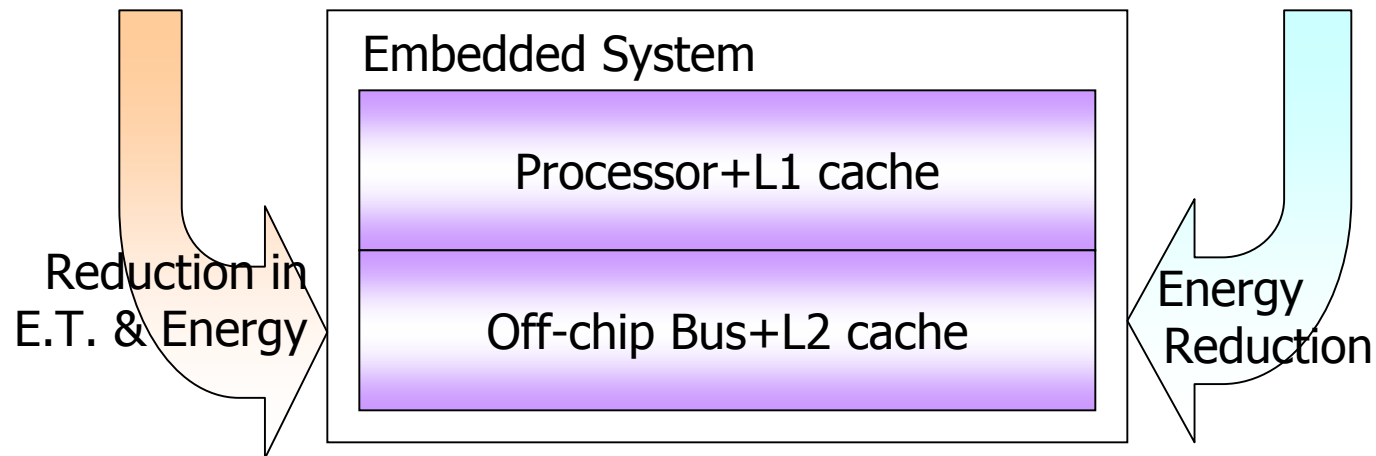
*P. Panda, N. Dutt, and A. Nicolau. *Memory Issues In Embedded Systems-On-Chip, Optimizations and Exploration*. Kluwer Academic Publishers, 1999.



Motivation

Data Remapping
Software Technique

Voltage/Frequency Scaling
Hardware Technique





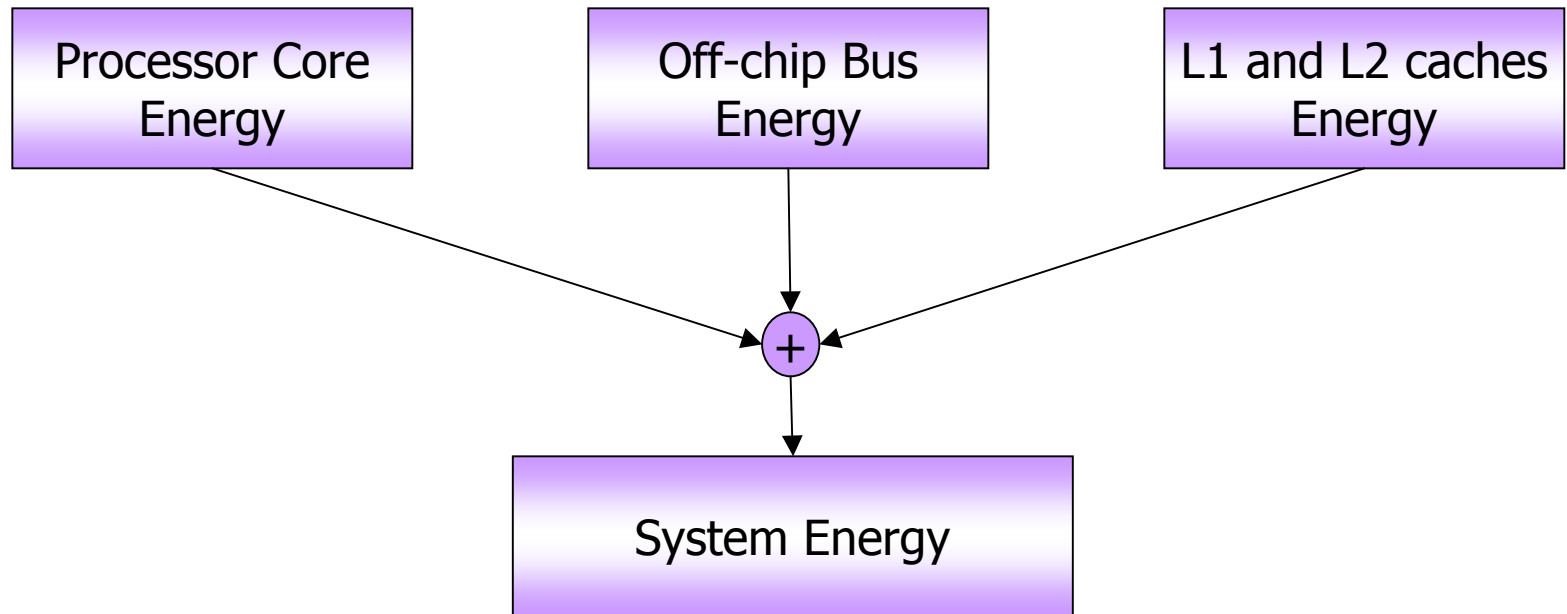
Related Work in Power Modeling

- **SimpleScalar/ARM PowerAnalyzer***
 - Cycle level power/performance simulator
- **SimplePower****
 - Architectural power estimation tool
 - Does not capture the energy of control unit of processor, clock generation

* <http://www.eecs.umich.edu/~jringenb/power/>

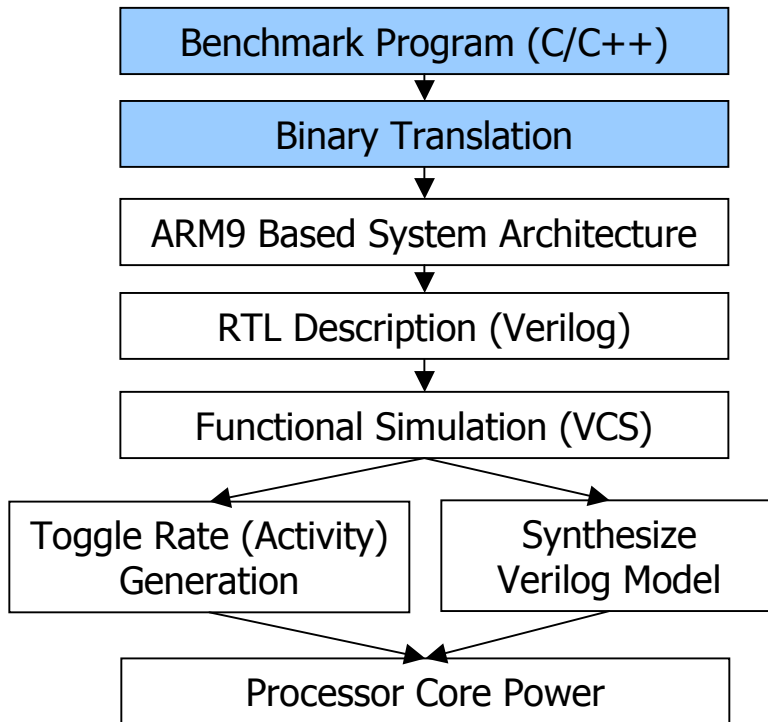
** <http://www.cse.psu.edu/~mdl/software.htm>

Experimental Setup

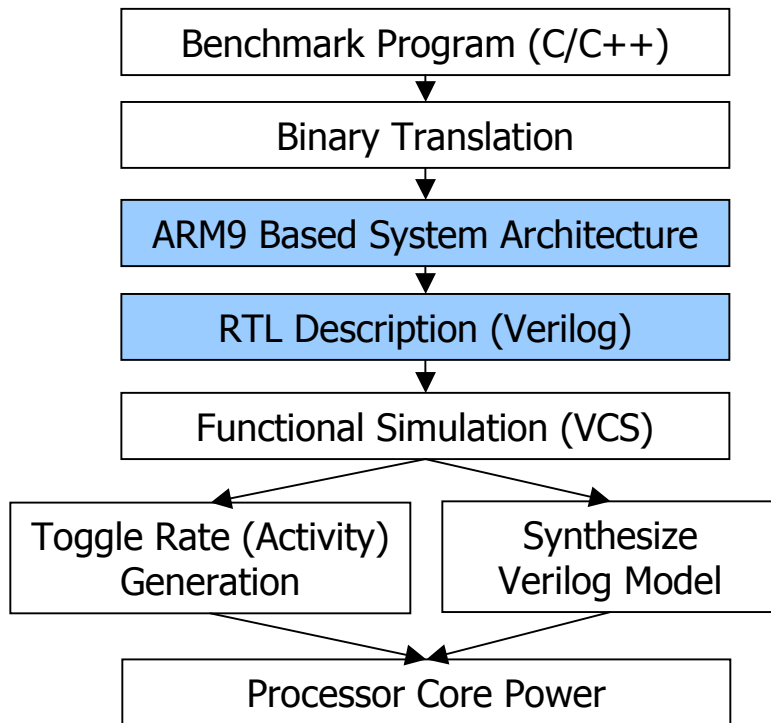


Experimental Setup (Cont.)

- Processor core power

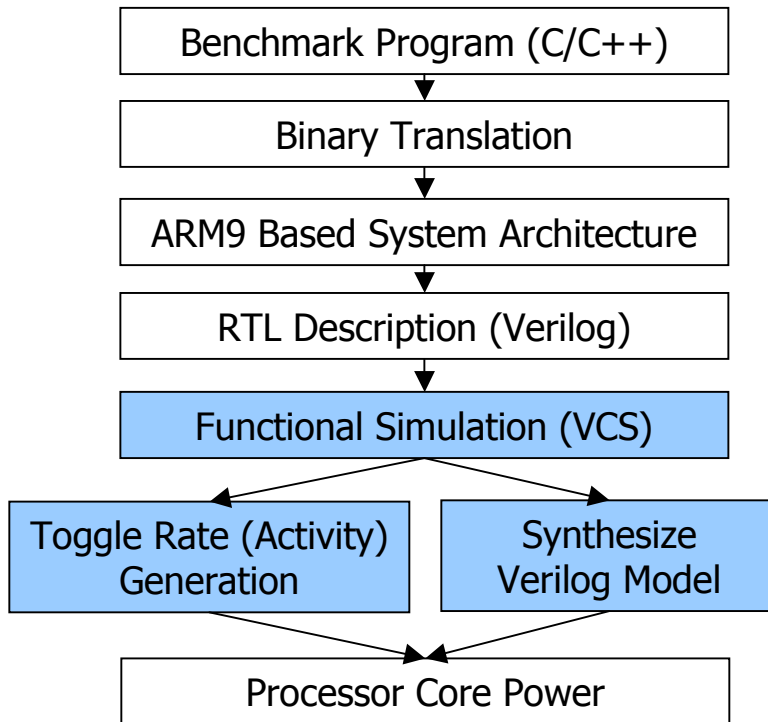


Experimental Setup (Cont.)



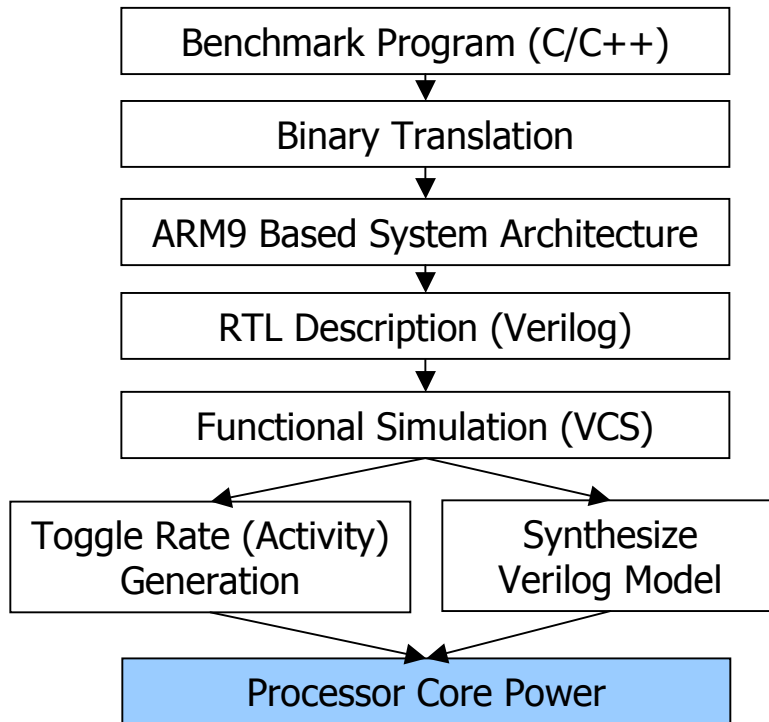
- Processor core power
 - MARS (Michigan ARM Simulator)
 - A cycle accurate verilog model of a RISC processor
 - Capable of running ARM instructions

Experimental Setup (Cont.)



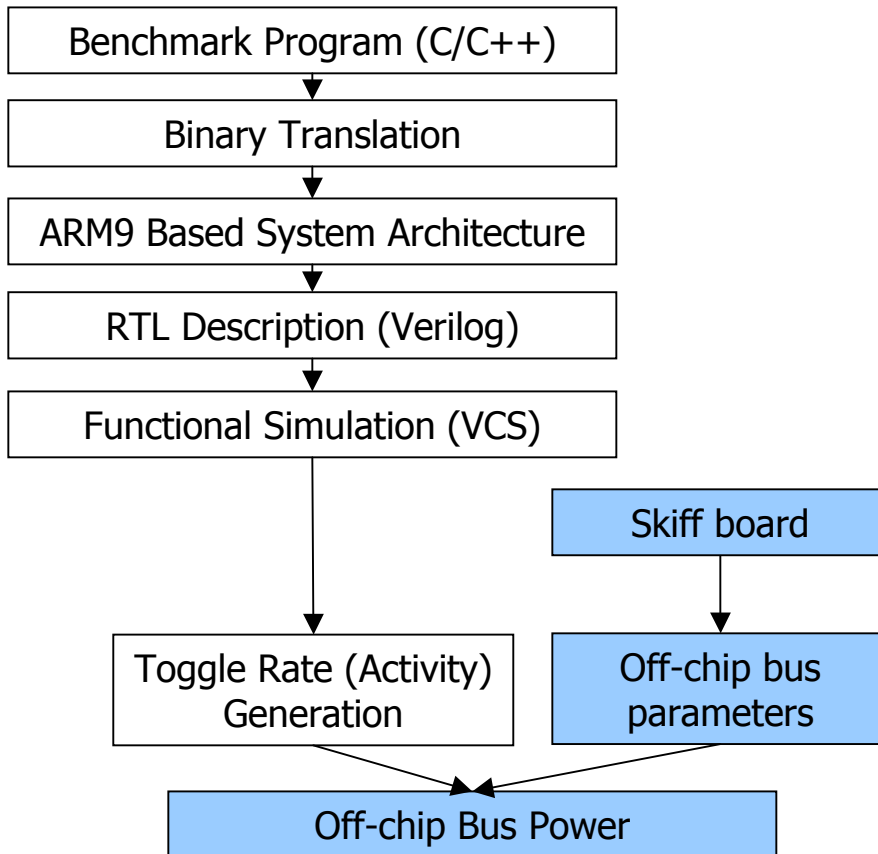
- Processor core power
 - Collect toggle rate of internal logic signals using Synopsys VCS simulation
 - Synthesize verilog model using TSMC .25 μ library

Experimental Setup (Cont.)



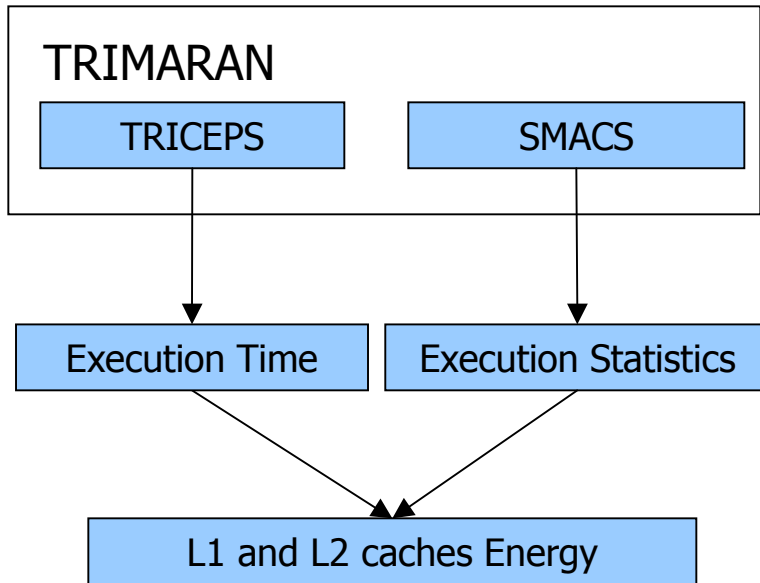
- Processor core power
 - Estimate power using Synopsys Power Compiler

Experimental Setup (Cont.)



- Off-chip bus power
 - Bus capacitance obtained from actual board
 - PCB board with SA1110 processor (Skiff board)

Experimental Setup (Cont.)

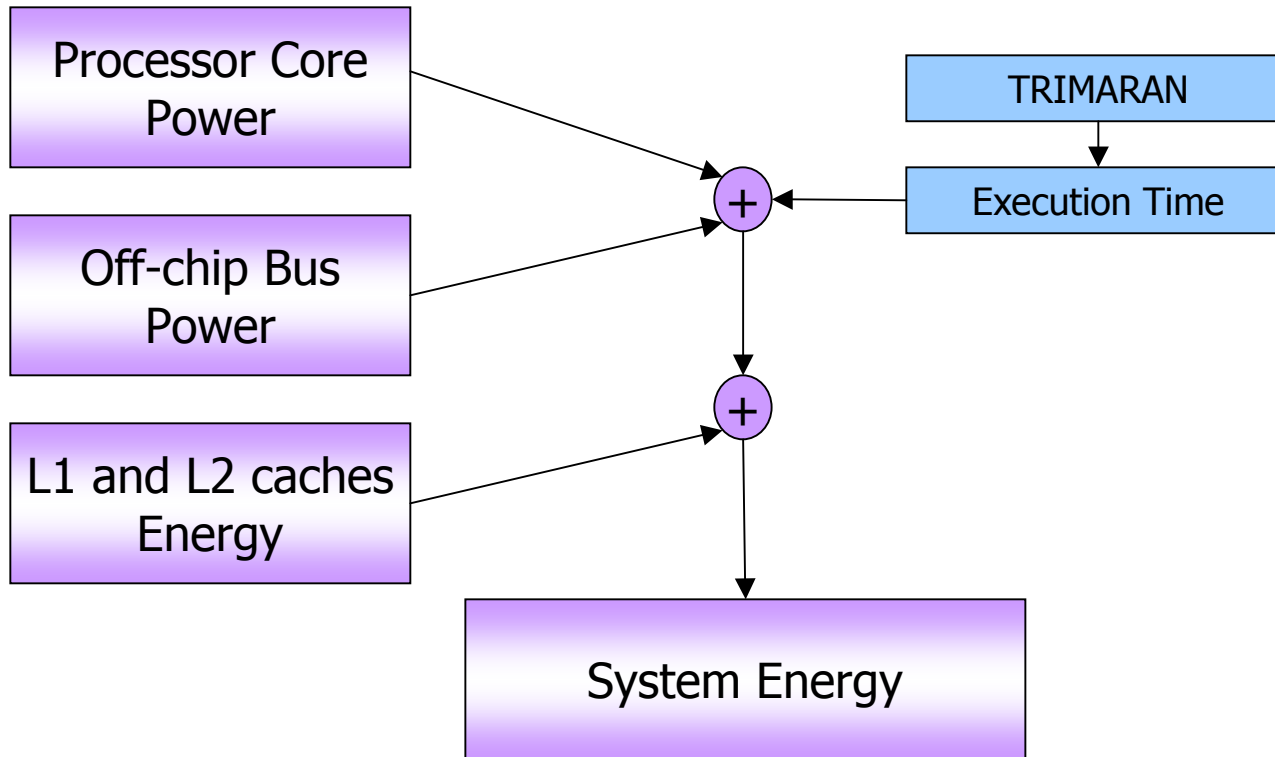


- L1 and L2 caches energy
 - TRIMARAN*
 - Integrated compilation and performance monitoring infrastructure
 - ARM-like processor simulator
 - TRICEPS
 - Generate ARM code
 - SMACS (Smart Memory and Cache Hierarchy Simulator)
 - cache activity statistics
 - Kamble and Ghose model**

*TRIMARAN <http://www.trimaran.org>

**M. Kamble and K. Ghose "Analytical energy dissipation models for low power caches,"
Proceedings of the International Symposium on Low Power Electronics and Design, pp.
143-148, Aug. 1997.

Experimental Setup (Cont.)





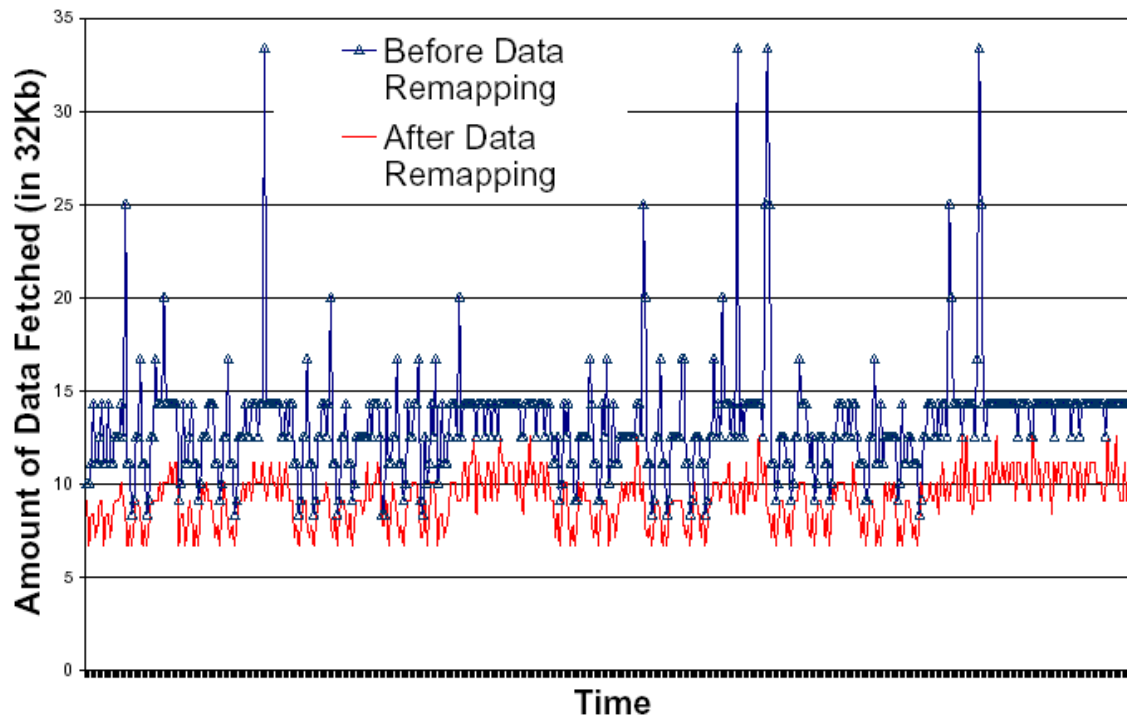
Data Remapping*

- A compile time technique for performance enhancement and energy reduction
- Remapping data into new set such that data items that are more likely to be used together are grouped together into the same cache block
- Enhancing spatial locality

*K. Palem, R. Rabbah, P. Korkmaz, V. Mooney and K. Puttaswamy,
"Design Space Optimization of Embedded Memory Systems via Data Remapping,"
Proceedings of the Languages, Compilers, and Tools for Embedded Systems (LCTES'02), pp. 28-37, June 2002.

Data Remapping (Cont.)

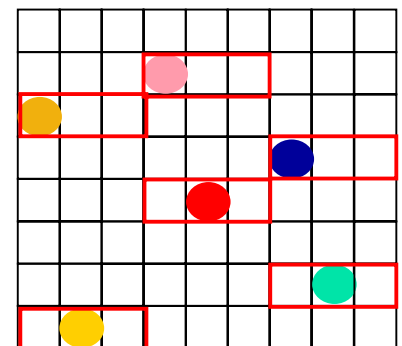
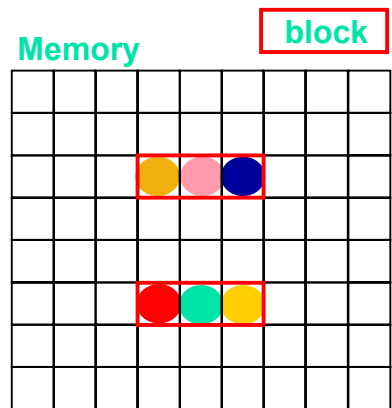
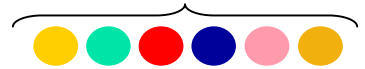
Amount of data fetched before and after remapping
(Traveling salesman problem in Olden Suite)



Data Remapping (Cont.)

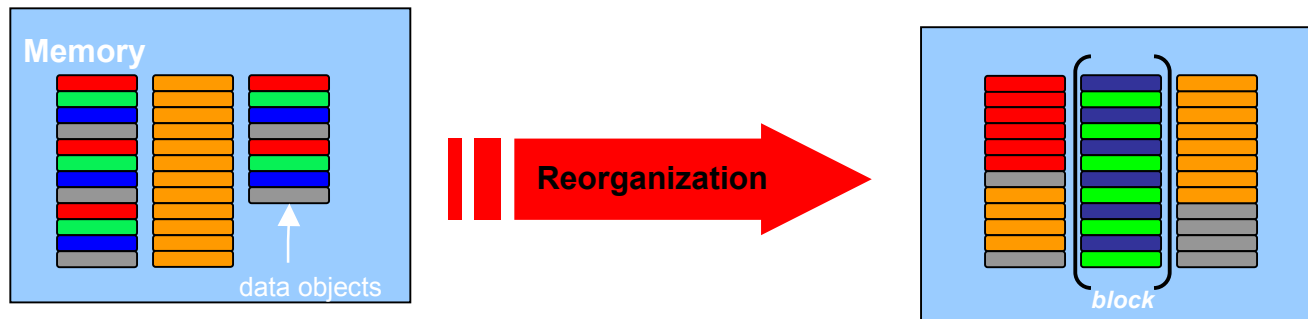
- An item in memory is accessed by initiating a load of the contents of a memory location or address
- Since a memory access is expensive, a set of adjacent memory locations are loaded at the same time and stored in a *cache*
 - The set of adjacent memory locations is known as a memory block
 - **Blocks do not overlap and have the same size**
- Each address can be mapped to a block in memory

Data Objects



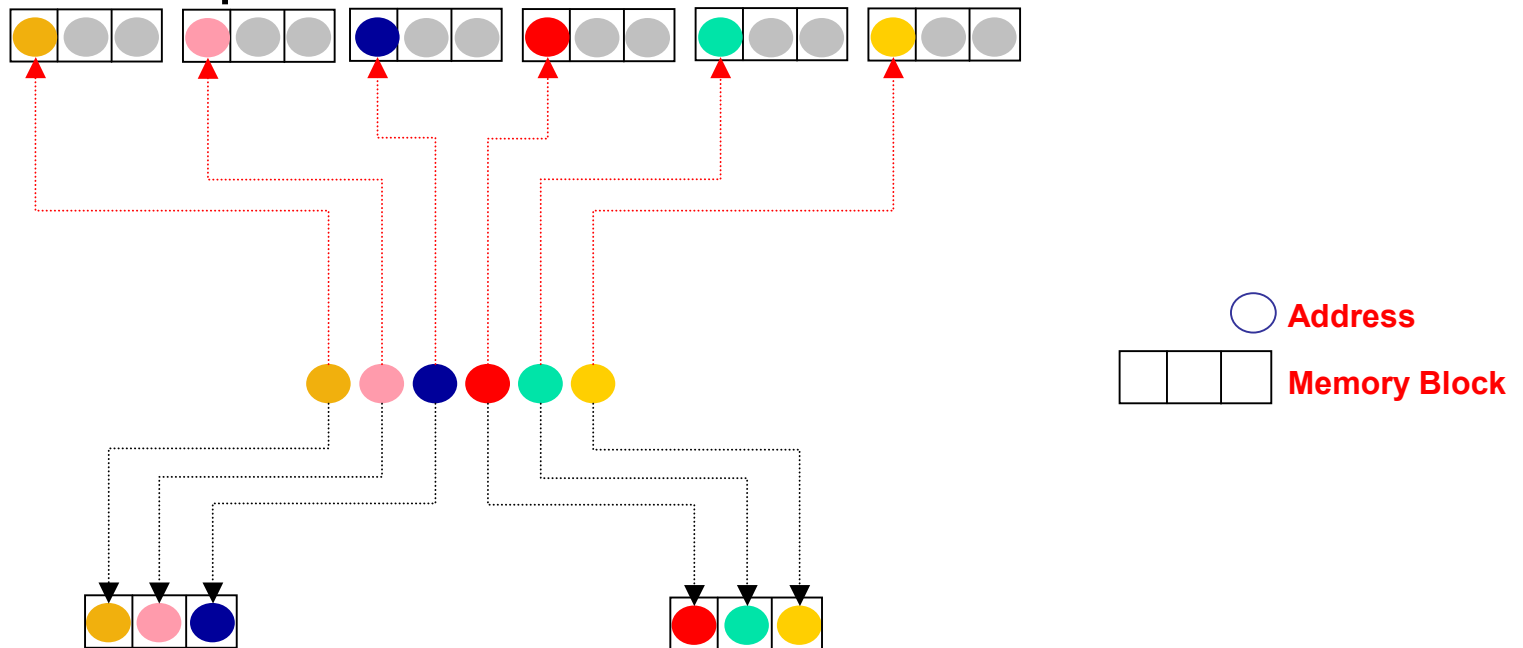
Data Remapping (Cont.)

- Data reorganization is the relocation of data objects in memory



Data Remapping (Cont.)

- Analyze application memory access pattern then remap data



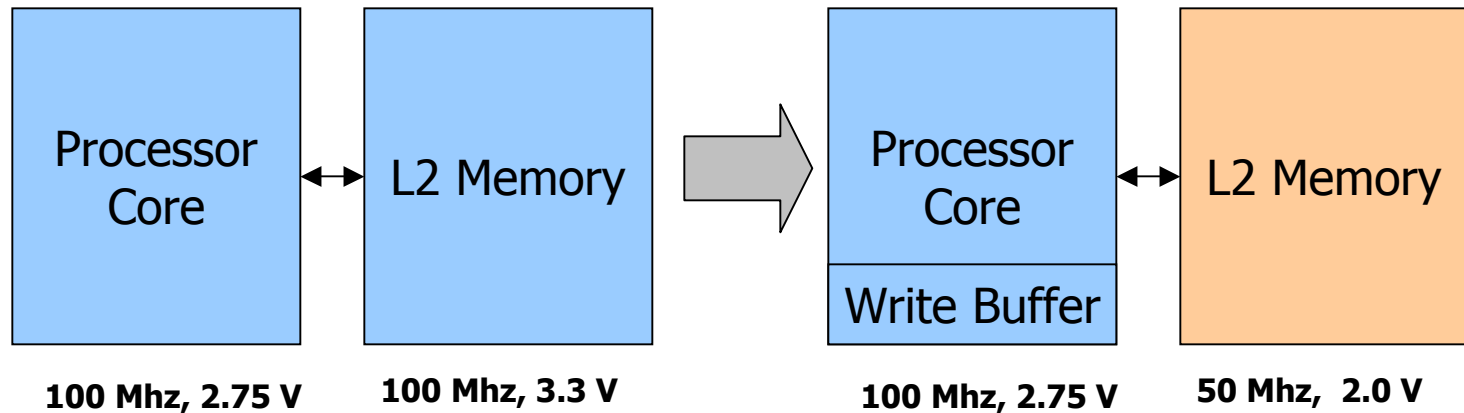


Voltage/frequency scaling of off-chip memory and bus*

- Scaling down supply voltage of off-chip bus and memory (L2 cache)
 - P is proportional to V^2
- Significant energy saving in L2 cache
- Doubling the memory access latency
- L2 cache miss rate affects system performance significantly

*K. Puttaswamy, K. Choi, J. C. Park, V. J. Mooney III, A. Chatterjee and P. Ellervee, System Level Power-Performance Trade-Offs in Embedded Systems Using Voltage and Frequency Scaling of Off-Chip Buses and Memory," *Proceedings of International Symposium on System Synthesis*, to appear, October, 2002, Kyoto, Japan.

Voltage/frequency scaling of off-chip memory and bus (Cont.)





Experimental Results

- Two Olden benchmarks (Health and Perimeter) are used
- The supply voltage for L2 cache and buses are scaled down to 2V, 50Mhz
- The benchmarks are remapped and simulated with 50Mhz L2 cache
- Half size L1 and L2 cache system is simulated
 - Data remapping can achieve same execution time with half cache resources



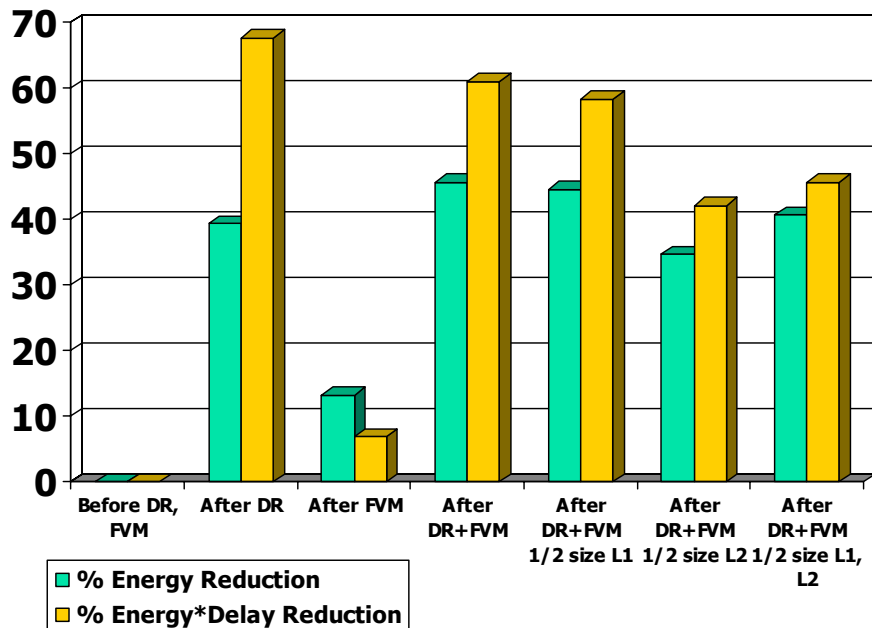
Experimental Results (Cont.)

Energy delay with frequency/voltage scaling of memory (FVM) and data remapping (DR) for health benchmark (L1 32KB 16B/line, L2 1MB 32B/line)

| | Before DR, FVM | After DR | After FVM | After DR+FVM | After DR+FVM 1/2 size L1 | After DR+FVM 1/2 size L2 | After DR+FVM 1/2 size L1,L2 |
|---------------------------|----------------|-----------|-----------|--------------|-----------------------------|-----------------------------|--------------------------------|
| Execution Cycles | 803645821 | 479612138 | 892552982 | 578046486 | 603275469 | 711151104 | 736311686 |
| Delay (Execution Time)(s) | 8.036 | 4.796 | 8.926 | 5.78 | 6.033 | 7.112 | 7.363 |
| Energy(J) | 17.076 | 10.360 | 14.316 | 9.274 | 9.468 | 11.158 | 10.134 |
| Energy*Delay | 137.231 | 49.687 | 127.778 | 53.608 | 57.118 | 79.35 | 74.618 |
| % Energy Reduction | 0 | 39.33 | 16.16 | 45.69 | 44.55 | 34.66 | 40.65 |
| % Energy*Delay Reduction | 0 | 63.79 | 6.89 | 60.94 | 58.38 | 42.18 | 45.63 |

Experimental Results (Cont.)

Energy delay with frequency/voltage scaling of memory (FVM) and data remapping (DR) for health benchmark (L1 32KB 16B/line, L2 1MB 32B/line)

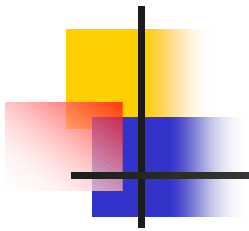


- Maximum of 46% of energy reduction
- Energy consumption of the cache reduced by half after halving L1 and L2 cache without performance loss



Conclusion

- Combine of two techniques (HW & SW) to maximize energy reduction
- Achieve 46% of energy reduction without performance loss
- Achieve 1/2 energy consumption with half size cache, same performance



Thank you.