

Speculative Tag Access for Reduced Energy Dissipation in Set-Associative L1 Data Caches

Alen Bardizbanyan, David Whalley*, Magnus Sjölander*,
Per Larsson-Edefors

Chalmers University of Technology

*Florida State University

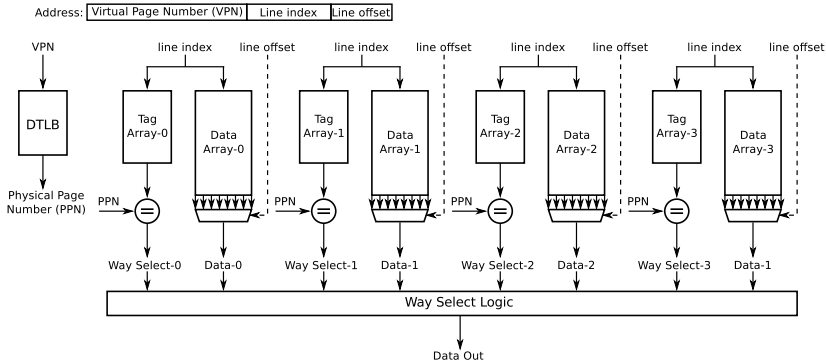


Energy Efficient Processor Design

- Need for energy efficient processors.
- Should also not negatively impact performance.
- Architecture features need to be reexamined with respect to energy efficiency.

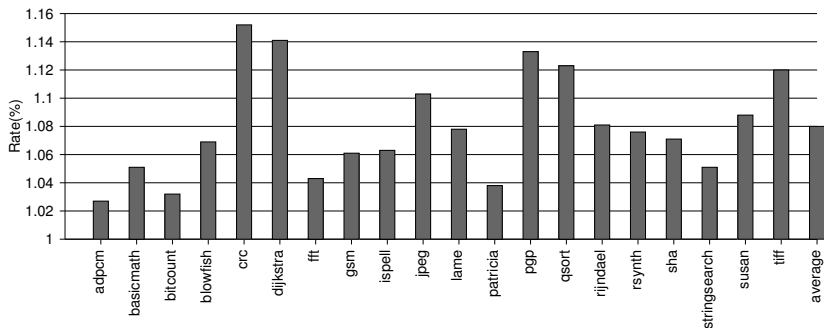
Set-Associative Data Caches

- Many set associative caches are virtually indexed and physically tagged.



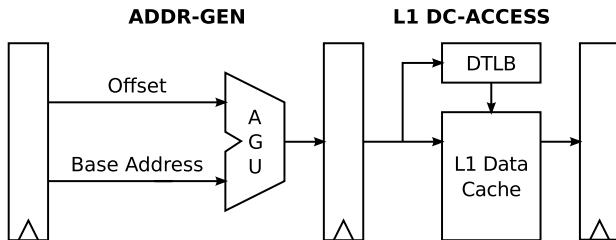
Loads from a Set-Associative L1 DC Are Energy Inefficient

- All L1 DC ways are accessed in parallel for loads to reduce stall cycles.
- The requested data can at most reside in one of the ways.
- We found an 8% execution time overhead on average when the L1 DC tag and data memories are sequentially accessed.



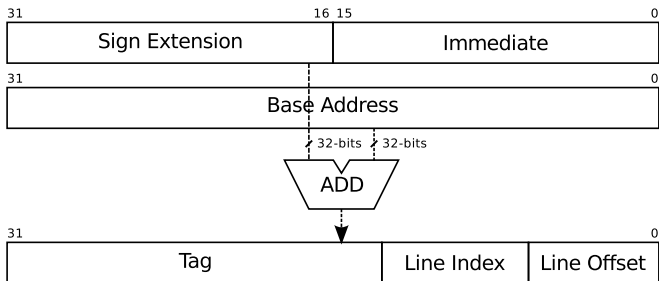
Address Generation and L1 DC Access

- The address generation unit calculates the effective address in a stage before the L1 DC is accessed.
- The address generation takes as input:
 - base address from a register value
 - offset from an immediate value in the instruction
- The figure below assumes a VIPT organization.



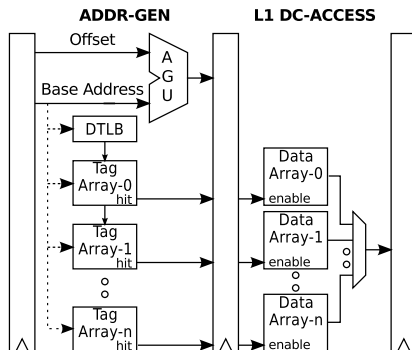
Data Memory Address Calculation

- If the offset is small, then it is possible that the tag and line index portions of the memory address will be the same as these fields in base address.



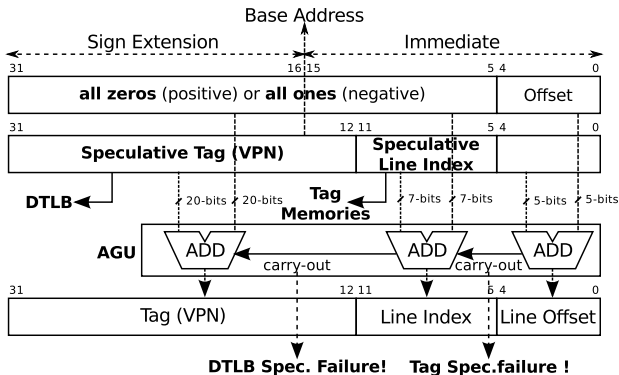
Speculative Tag Access

- Use the line index of the base address to speculatively access the L1 DC tags.
- Use the virtual page number (primarily the tag) of the base address to speculatively access the DTLB.
- One L1 DC data way is accessed after a successful speculation.



Speculative Address Calculation and Failure Detection

- The carry-out signals from the line offset and line index are used to detect if there is a speculation failure due to an invalid line index to the L1 DC tag memory or due to an invalid virtual page number to the DTLB, respectively.



Speculative Access Benefits and Costs

- When the speculative tag access is successful:
 - On an L1 DC hit, the read energy of accessing $n-1$ data arrays of the n -way associative L1 DC is avoided.
 - On an L1 DC miss, the read energy of accessing all data arrays of the L1 DC is avoided and the next level of the memory hierarchy is accessed one cycle earlier.
- A speculation failure for the L1 DC index field requires the extra energy cost of unnecessarily accessing all the L1 DC tag arrays.
- A speculation failure for the L1 DC tag field requires the extra energy cost of unnecessarily accessing the DTLB.

Evaluation Framework

- 5-stage in-order processor
- The L1 IC and L1 DC are both 16kB, 4-way set-associative, and have a 32B line size.
- 16-entry fully associative DTLB
- RTL implementation synthesized using the Synopsis Design Compiler to obtain energy values for various events.
- Used the SimpleScalar simulator to count events and estimate total energy usage.

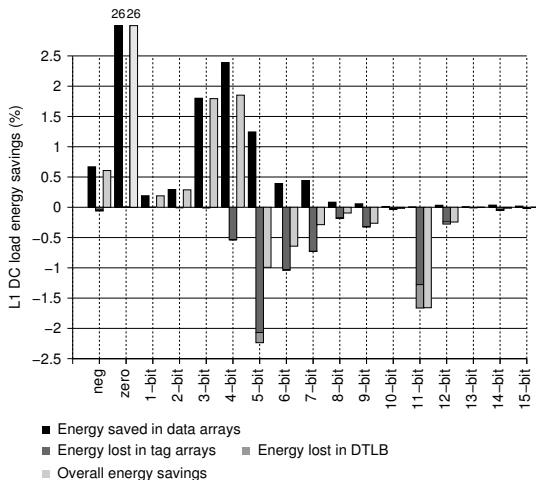
Benchmarks

- 20 benchmarks simulated from the MiBench benchmark suite.

Category	Applications
Automotive	Basicmath, Bitcount, Qsort, Susan
Consumer	JPEG, Lame, TIFF
Network	Dijkstra, Patricia
Office	Ispell, Rsynth, Stringsearch
Security	Blowfish, Rijndael, SHA, PGP
Telecomm	ADPCM, CRC32, FFT, GSM

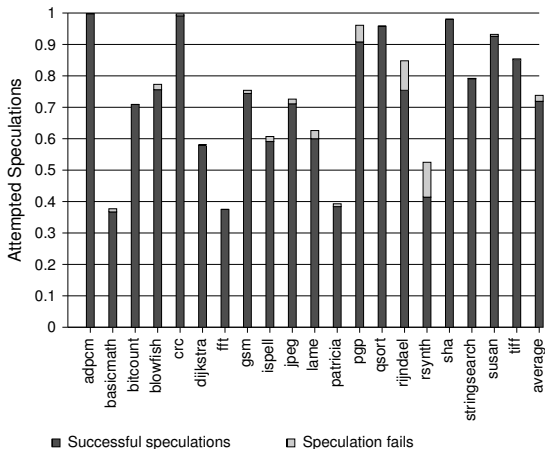
Impact of Offset Bit Width

- Speculation not attempted when negative offsets require more than five bits or positive offsets require more than four bits.



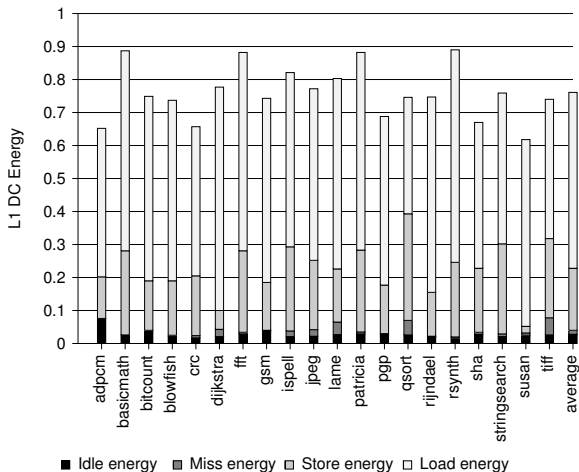
Load Speculation Success Rate

- 71.9% of all loads successfully access L1 DC tags early.
- 1.9% of loads cause speculation failures affecting tag access and 0.2% cause failures affecting DTLB access.



L1 DC and DTLB Energy Results

- Idle, store, and miss energy are unaffected.
- Energy due to loads decreased from 77.2% to 53.3%, resulting in a 24% L1 DC and DTLB energy savings.



Conclusions

- Speculative tag access reduces energy dissipated in a set-associative L1 DC with no execution time penalty.
- Benefits should increase as L1 DC line size increases.
 - Size of tag arrays compared to data arrays become smaller.
 - Fewer carry outs into the line index as line offset is larger.
- Benefits may increase as L1 DC associativity increases since the portion of L1 DC data being accessed on loads will decrease.