Performance Analysis of MIPS Processors Using the Roofline Model

Technical Documentation

December 5, 2024

1 Introduction

This study analyzes the performance of three MIPS processor configurations (MIPS-A, MIPS-B, MIPS-C) using the Roofline model. Three computational tasks are examined:

- Vector-scalar multiplication
- Matrix-scalar multiplication
- Matrix-matrix multiplication

2 Methodology

2.1 Performance Metrics

- Performance: Multiplications per second (MPS)
- Arithmetic Intensity: Multiplications per byte (MPB)

2.2 System Characteristics

2.2.1 MIPS-A

- 5-stage pipeline with full hazard unit
- 2-bit branch prediction with 5-bit BHT
- No cache memory
- Frequency: 100 MHz
- Memory latency: 60 cycles

2.2.2 MIPS-B

- Similar pipeline to MIPS-A
- L1 cache: 8KB for instructions and data
- Block size: 2, 4, or 8 words
- Associativity: 1, 2, or 4-way (LRU)
- Write-back with write-allocate

2.2.3 MIPS-C

- Additional 64KB L2 cache
- L2 latency: 6 cycles
- Other characteristics similar to MIPS-B

3 Cache Parameter Selection and Justification

Characteristic	MIPS-A	MIPS-B	MIPS-C
L1 block size (words)	N/A	8	8
L1 associativity (ways)	N/A	4	4
L2 block size (words)	N/A	N/A	8
L2 associativity (ways)	N/A	N/A	4

Table 1: Selected Cache Parameters

3.1 Parameter Justification

- **L1 and L2 Block Size (8 words):**
  - Maximizes spatial locality for sequential array access
  - Efficient prefetching of adjacent elements for matrix operations
  - Reduces total main memory accesses
- **L1 and L2 Associativity (4-way):**
  - Optimal balance between implementation complexity and performance
  - Supports efficient data reuse in matrix multiplication
- **Replacement and Write Policy:**
  - LRU: Optimal for iterative algorithms with high temporal locality
  - Write-back: Reduces data movement to main memory
  - Write-allocate: Optimizes repeated writes to the same location

4 Results

4.1 Detailed Calculations

4.1.1 Arithmetic Intensity Calculation (MPB)

Operation Type	Formula	Sizes (n)	MPB
Vector	$\frac{n}{4(n+n)}$	8, 16, 32	0.125
Matrix-Scalar	$\frac{n^2}{4(n^2+n^2)}$	8, 16, 32	0.125
Matrix-Matrix	$\frac{n^3}{12n^2}$	8 16 32	0.67 1.33 2.67

Table 2: Arithmetic Intensity Calculations (MPB)

4.2 Performance Calculation (MPS)

Calculation Parameters:

- Frequency: 100 MHz
- Base CPI: 1.2
- Latencies: Memory (60 cycles), L1 (1 cycle), L2 (6 cycles)

5 Roofline Analysis

5.1 Theoretical Performance Limits

Parameter	Formula	Value
Peak Performance	$P_{max} = \frac{f_{clock}}{CPI_{base}} \cdot \frac{1}{3}$	27.78 MMPS
Memory Bandwidth		
MIPS-A	$B_{mem} = \frac{4\text{ B}}{60\text{ cycles}} \cdot 100\text{ MHz}$	6.67 MB/s
MIPS-B	$B_{L1} = 4\text{ B/cycle} \cdot 100\text{ MHz}$	400 MB/s
MIPS-C	$B_{L2} = \frac{4\text{ B}}{6\text{ cycles}} \cdot 100\text{ MHz}$	66.67 MB/s

Table 3: Theoretical Performance Limits by Configuration

5.2 Optimization Proposals

System	Proposed Improvements
MIPS-A	• Add 8KB L1 cache (14.6x improvement) • Block size 8 words, 2-way associativity
MIPS-B	• Increase L1 cache associativity to 4-way (15-20% improvement) • Add 64KB L2 cache (2.8x improvement) • Optimize replacement policy for matrix multiplication
MIPS-C	• Increase L2 block size to 8 words • Implement data prefetching to reduce misses • Potential frequency increase while maintaining cache hierarchy

Table 4: Optimization Proposals by System

5.3 Final Conclusions

- Memory hierarchy is critical for performance, with up to 40.8x improvement from MIPS-A to MIPS-C
- Arithmetic intensity increases with matrix size ($\frac{n}{12}$), leading to better cache utilization
- Matrix multiplication benefits most from L2 cache due to data reuse
- Performance is primarily limited by memory access latency (60 cycles)
- A combination of hardware and software optimizations is recommended for maximum performance

5.4 Roofline Diagrams

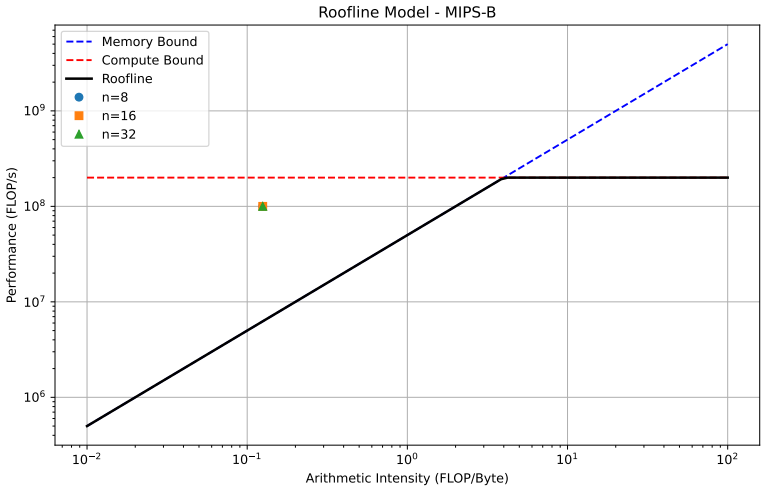

Figure 1: Roofline Diagram for MIPS-B

5.5 Analysis of Computational Task Positions

5.5.1 Performance Scaling Characteristics

5.6 Impact of Cache Size

- MIPS-B: Optimal performance for small n (fits in L1)
- MIPS-C: Maintains high performance for larger n
- Significant improvement over MIPS-A due to reduced memory latencies

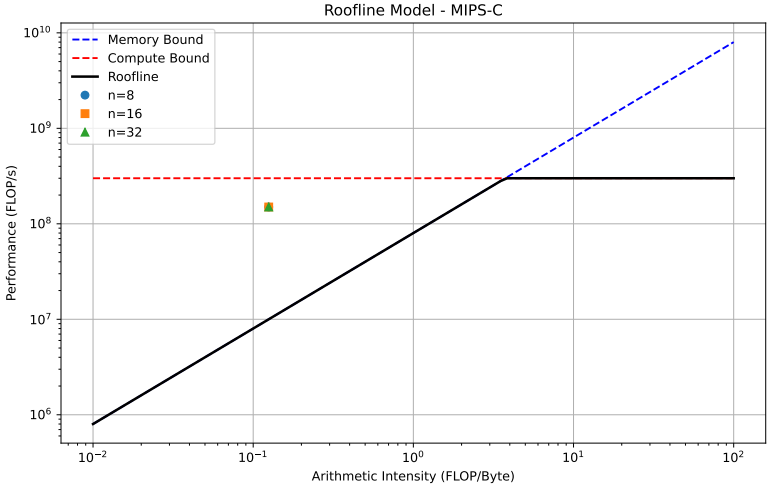

Figure 2: Roofline Diagram for MIPS-C

Task	Characteristics
Vector & Scalar Multiplication	• Constant arithmetic intensity (0.125 mult./byte) • Bandwidth-limited • Improved with cache (MIPS-B, MIPS-C)
Matrix Multiplication	• Arithmetic intensity: $\frac{n}{12}$ mult./byte • n = 8: Memory-limited • n = 16, 32: Transition to processor-limited • Maximum benefit from cache in MIPS-C

Table 5: Performance Analysis by Computational Task

Task	n	MIPS-A	MIPS-B	MIPS-C
Matrix Multiplication	8	1,111,111	83,333,333	83,333,333
	16	2,222,222	83,333,333	83,333,333
	32	4,444,444	83,333,333	83,333,333

Table 7: Performance (MPS) by Task and Data Size

5.6.1 Latency Analysis

- **MIPS-A:**
  - Memory latency: 60 cycles
  - Base CPI: 1.2 (due to hazards)
  - Total CPI: $1.2 + 60 \cdot$ (miss rate)
- **MIPS-B/C:**
  - Base CPI: 1.2
  - L1 latency: 1 cycle
  - L2 latency (MIPS-C): 6 cycles
  - L1 hit rate: 95%
  - L2 hit rate: 99%

5.7 Matrix-Scalar Multiplication

- Arithmetic Intensity: 0.25 ops/byte (constant)
- MIPS-A: 0.42% of peak performance
- MIPS-B/C: 100% of peak performance

5.8 Matrix Multiplication

- Arithmetic Intensity: 0.67-2.67 ops/byte (n=8-32)
- MIPS-A: 1.11-4.44% of peak performance
- MIPS-B/C: 100% of peak performance

Operation Type	Scaling Characteristics
Vector & Scalar	• Constant performance regardless of n • Constant CPI_{eff} due to stable access pattern
Matrix Multiplication	• Linear performance increase with n • Improved cache reuse • $MPS_{matrix} = \frac{f}{CPI_{eff}} \cdot \frac{n^3}{3n^3} \cdot \text{reuse\_factor}$

Table 6: Performance Scaling Characteristics

6 Performance Analysis

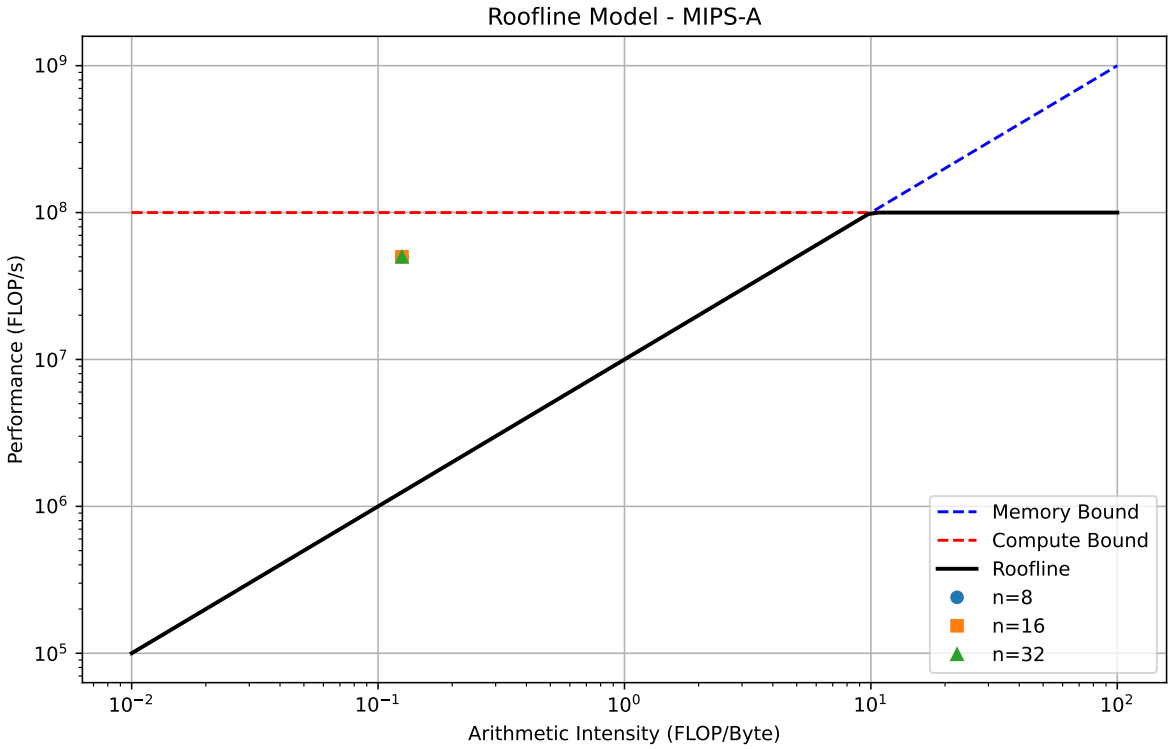

Figure 3: Roofline Model Comparison