A Case Study on Matrix-Matrix Multiplication in the Alamode Lab

Brian Hoenes
Department of Electrical Engineering and Computer Sciences
Colorado School of Mines, Golden, CO 80401, USA
bhoenes@mines.edu

ABSTRACT

It has been demonstrated recently that ...

In this paper, we present ...

1. INTRODUCTION

While the peak performance of computer systems continue to grow exponentially, it is getting more difficult for scientific applications to achieve high sustained performance due to the increasing complexity of the underlying system architectures. This case study explores various methods to achieve a high ratio of sustained performance to peak performance for floating point matrix-matrix multiplication on the target architecture.

2. EXPERIMENTAL FRAMEWORK (QUESTION 1)

The target architecture is a single socket Intel®Core™i7-2600 "Sandy Bridge" processor clocked at 3.4GHz. The test systems are 24 Dell Optiplex 990 desktop workstations in the Colorado School of Mines Alamode Computer Science Linux Laboratory in Brown room 136. The systems are using an Ubuntu 10.04 operating system and all compilations are with the GNU gcc compiler version 4.4.3. All experiments are run serially on a single core of the four core processor. The i7-2600 is a four core processor launched by Intel in the first quarter of 2011. The processor die layout is shown in Figure 1 [1].

Each core of the i7-2600 has a 32KB data and instruction cache, a 256KB unified mid-level cache and 2 level Data Translation Lookaside Buffer (DTLB) system of 64 and 512 entries. There is a single, 32 entry large page DTLB. The cores in a socket share an inclusive last level cache that is 8MB and 16-way associative. [2]

The i7-2600 can perform four 64-bit floating point operations per cycle at peak performance. At the 3.4GHz clock this equates to 13.6 billion floating point operations per second (GFLOPS) per core. The processor has the ability, named Intel®Turbo Boost Technology 2.0, to increase its clock rate dynamically up to 3.8GHz if its operating below power, current, and temperature specification limits [3]. This results in a peak performance of 15.2 GFLOPS per core. The Turbo Boost ability is controlled in the system BIOS which is limited access in the Alamode Laboratory and could not be disabled in the test configuration. The occurrence of the clock frequency increase is not predictable and is dependent on many uncontrollable factors. Consequently, peak performance for experimental evaluation purposes will be reported as 13.6 GFLOPS but will actually range from 13.6 to 15.2 GFLOPS.

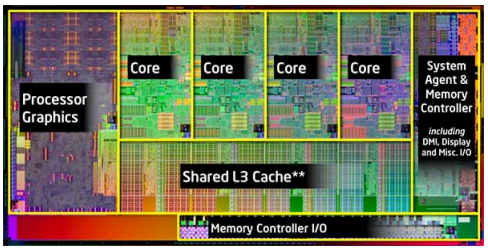

Figure 1: i7-2600 Die Layout

3. BASELINE MATRIX-MATRIX MULTIPLICATION (QUESTION 2)

Matrix-matrix multiplication ($C = C + A * B$) can be performed using the following simple triple loop:

Random 64-bit floating point square matrices of sizes 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, and 5000 were used in the experiments.

4. BASELINE BENCHMARK (QUESTION 3)

5. -O1 OPTIMIZATION BENCHMARK (QUESTION 4)

6. -O2 OPTIMIZATION BENCHMARK (QUESTION 5)

7. -O3 OPTIMIZATION BENCHMARK (QUESTION 6)

8. -O3 -FUNROLL-LOOPS OPTIMIZATION BENCHMARK (QUESTION 7)

9. SOFTWARE OPTIMIZATIONS QUESTION 8

9.1 Step00

9.2 Step01

9.3 Step02

9.4 Step03

9.5 Step04

9.6 Step05

10. CONCLUSION

...

11. REFERENCES

[1] D. Ramsey. (2011, Jan.) Intel core i7-2600k sandy bridge cpu review. web. [Online]. Available: http://benchmarkreviews.com/index.php?option=com_content&task=view&id=622&Itemid=63

[2] D. Levinthal, "Performance analysis guide for intel core i7 processor and intel xeon 5500 processors," Intel Corporation, Tech. Rep., 2009. [Online]. Available: http://software.intel.com/sites/products/collateral/hpc/vtune/performance_analysis_guide.pdf

[3] I. Corporation. (2012, March) Intel turbo boost technology. web. [Online]. Available: http://www.intel.com/content/www/us/en/architecture-and-technology/turbo-boost/turbo-boost-technology.html