

Dynamic Branch Prediction in an Intel High-Performance Processor

Subjects: Computer Science, Hardware & Architecture

Contributor: Fahad Swilim Alqurashi, Muhammad Al-Hashimi

Power and energy efficiency are among the most crucial requirements in high-performance and other computing platforms. Extensive experimental methods and procedures were used to assess the power and energy efficiency of fundamental hardware building blocks inside a typical high-performance CPU, focusing on the dynamic branch predictor (DBP).

Keywords: dynamic branch predictor ; power efficiency ; energy

1. Introduction and Motivation

In the past, the primary interest of computer architects and software developers was in increasing performance. However, in the last couple of decades, power and energy efficiency have emerged as major requirements in computing. One of the reasons that led to power and energy efficiency interest in computing is the need to reduce the power consumption of HPC to achieve exascale supercomputers ^[1]. This goal motivated researchers to investigate some fundamental software building blocks, such as sorting, matrix multiplication, and shortest path algorithms, commonly used in HPC applications to find the factors that affect power consumption. Fundamental hardware components, such as arithmetic units, caches, and dynamic branch predictors (DBPs), should raise similar concerns.

There are several reasons to be interested in the energy behavior of DBP in particular. Firstly, it is an essential component in all modern CPUs used in HPC. Secondly, conditional jump instructions represent a significant percentage of most typical application instructions. Modern CPUs use DBP to guess the direction and the target address of jump instructions which means a heavy utilization of DBP during any application run. Consequently, any power savings related to this component may yield substantial benefits in the total energy consumption. Thirdly, according to some statistics, DBPs account for 10 to 40 percent of CPU dynamic power consumption ^[2]. Fourthly, the recent DBP security issues appeared, and the solutions proposed to mitigate them need investigation from a power and energy perspective ^{[3][4][5]}. Lastly, the lack of research papers studying DBP and its security issues from a power and energy perspective provided an extra justification for this investigation.

As many research papers pointed out, software style impacted program performance and power and energy consumption of computing devices in significant ways ^[6]. The impact demanded that developers find easy ways to measure power consumption. Intel's RAPL (Running Average Power Limit Energy Reporting), introduced in modern processors, is one of the most prominent software tools that can report the power in different CPU domains ^[7]. However, using RAPL for credibly measuring power consumption needs some precautions to avoid noise in measurement ^[8].

This report presents the researchers' experiences in empirically assessing the power consumption on the Intel Haswell platform and the precautions recommended to raise the credibility of measurements obtained from the RAPL tool. The main contribution of the research described here is the development of a methodology suitable for an empirical study of the power characteristics of DBP, a hardware building block. Previous work focused on the empirical study of software building blocks or studying DBP using simulation or mathematical models ^[9]. Power and energy consumption are directly related to the hardware. The DBP is a complex piece of hardware that can be difficult to simulate in detail or model adequately to get reasonable consumption assessments. Haswell incorporated separate sensors within the processor to measure power consumption at different levels, such as the package, individual cores, and DRAM providing real-time feedback on power usage and allowing for reliable higher-resolution readings of power consumption ^[8] without needing additional external devices like power meters or infrared imaging of relevant regions in the silicon. Therefore, RAPL on Haswell CPUs offered improved granularity in power measurement compared to previous generations. Empirical power estimation, using fine-grained instrumentation, such as in Haswell-class and later CPUs, should provide realistic insights into the power consumption of such complex units as DBP without getting into the complexity of mathematical modeling or

detailed simulation that requires a deep understanding of the internal workings and interactions with other components. RAPL provides an attractive alternative.

2. Dynamic Branch Prediction in an Intel High-Performance Processor

Researchers have recognized the impact of software on computing power and energy consumption since the 1990s. Mehtal et al. ^[10] proposed software techniques to reduce the energy consumption of a processor using compiler optimization, such as loop unrolling and recursion elimination. They also studied how various algorithms and data structures affect power and energy usage.

Capra and Francalanci ^[11] considered the design factors that influence the energy consumption of applications that perform the same functionality but have different designs. They experimentally assessed the energy efficiency of management information systems. In their study, they found that the application design has a significant impact on energy consumption.

Sushko et al. ^[12] studied the effect of loop optimization on the power consumption of portable computing devices. They applied their study to ARMv8 architectures. Their study showed the power efficiency gained by fitting data in the cache and using parallelization for loop optimization.

Al-Hashimi et al. ^[13] studied the effects of three iteration statements on the system power consumption: For loop, While loop, and Do-While. For each case, they measured the average time, power, and temperature; the value of the maximum temperature; and the number and percentage of times reached. They found that the For loop was the most power-efficient and that the While loop had the worst power efficiency.

Abulnaja et al. ^[14] analyzed bitonic mergesort compared to an advanced quicksort on the NVIDIA K40 GPU for power and energy efficiency. They introduced the factors that affected power consumption and studied those that led to higher energy and power consumption, such as data movement and access rate. They concluded that bitonic mergesort is inherently more suitable for the parallel architecture of the GPU. This study triggered the investigation of more software building blocks, such as spanning tree algorithms and binary search algorithms.

Aljabri et al. ^[15] conducted a comprehensive empirical investigation into the power efficiency of mergesort compared to a high-performance quicksort on the Intel Xeon CPU E5-2680 (Haswell), which is more commonly used in HPC and has more accurate sensor readings than the previous-generation Intel Xeon E5-2640 CPU (Sandy Bridge) utilized in an earlier work ^[16]. The research was motivated by the fact that divisions by powers of two, the most frequent operation in mergesort, may be performed by a power-efficient barrel shifter. Mergesort's procedure applies a divide-and-conquer strategy in which the original list (or array) is divided, recursively, into two equal lists. The study concluded that mergesort had an advantage over quicksort in terms of power efficiency, with comparable time efficiency between the two algorithms. This study encouraged more investigation into other algorithms that perform similar tasks but have different time efficiencies from a power perspective.

NZ Oo et al. ^[17] studied Strassen's matrix multiplication algorithm from a performance vs. power perspective. In their study, they found a way to enhance performance and reduce energy consumption by using loop unrolling on the recursive level of the algorithm to minimize cache misses and increase the data locality. They claimed that their method increased performance by 93 percent and reduced energy consumption by 95 percent.

Jammal et al. ^[18] studied the power efficiency of three matrix multiplication algorithms, i.e., definition-based, Strassen's divide-and-conquer, and improved divide-and-conquer, on the Intel Xeon CPU E5-2680. The main finding of this work is that the fastest divide-and-conquer algorithm is power-efficient only for small matrix sizes. For larger sizes, the definition-based algorithm turned out to be more power-efficient. They also studied the effect of every cache level miss on power consumption.

References

1. Nain, S.; Chaudhary, P. Branch prediction techniques used in pipeline processors: A review. *Int. J. Pure Appl. Math.* 2018, 119, 2843–2851.
2. Hicks, M.A. Energy Efficient Branch Prediction. Ph.D. Thesis, University of Hertfordshire, Hertfordshire, UK, 2007.
3. Kiriansky, V.; Waldspurger, C. Speculative buffer overflows: Attacks and defenses. *arXiv* 2018, arXiv:1807.03757.

4. Dong, X.; Shen, Z.; Criswell, J.; Cox, A.; Dwarkadas, S. Spectres, Virtual Ghosts, and Hardware Support. In Proceedings of the 7th International Workshop on Hardware and Architectural Support for Security and Privacy (HASP'18), Los Angeles, CA, USA, 2 June 2018.
5. Kocher, P.; Horn, J.; Fogh, A.; Genkin, D.; Gruss, D.; Haas, W.; Hamburg, M.; Lipp, M.; Mangard, S.; Prescher, T.; et al. Spectre Attacks: Exploiting Speculative Execution. In Proceedings of the 2019 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 19–23 May 2019; IEEE: San Francisco, CA, USA, 2019; pp. 1–19.
6. da Silva, W.G.; Brolsara, L.; Corrêa, U.B.; Carro, L. Evaluation of the impact of code refactoring on embedded software efficiency. In Proceedings of the 1st Workshop on Embedded Systems, Gramado, Brazil, 25 May 2010; pp. 145–150.
7. Khan, K.N.; Ou, Z.; Hirki, M.; Nurminen, J.K.; Niemi, T. How much power does your server consume? estimating wall socket power using RAPL measurements. *Comput. Sci.—Res. Dev.* 2016, 31, 207–214.
8. Khan, K.N.; Hirki, M.; Niemi, T.; Nurminen, J.K.; Ou, Z. RAPL in action: Experiences in using RAPL for power measurements. *Acm Trans. Model. Perform. Eval. Comput. Syst. (TOMPECS)* 2018, 3, 1–26.
9. Rehman, Z. Survey—Branch Prediction Techniques. *Res. J. Innov. Ideas Thoughts* 2017, 5, 47–58.
10. Mehta, H.; Owens, R.M.; Irwin, M.J.; Chen, R.; Ghosh, D. Techniques for low energy software. In Proceedings of the 1997 International Symposium on Low Power Electronics and Design, Monterey, CA, USA, 18–20 August 1997; pp. 72–75.
11. Capra, E.; Francalanci, C.; Slaughter, S.A. Measuring application software energy efficiency. *IT Prof.* 2012, 14, 54–61.
12. Sushko, S.; Chemeris, A. The dependence of microprocessor system energy consumption on software optimization. In Proceedings of the 2017 IEEE 37th International Conference on Electronics and Nanotechnology (ELNANO'17), Kyiv, Ukraine, 18–20 April 2017; pp. 451–454.
13. Al-Hashimi, M.; Saleh, M.; Abulnaja, O.; Aljabri, N. Evaluation of control loop statements power efficiency: An experimental study. In Proceedings of the 2014 9th International Conference on Informatics and Systems (INFOS'14), Cairo, Egypt, 15–17 December 2014; pp. PDC–45–PDC–48.
14. Abulnaja, O.A.; Ikram, M.J.; Al-Hashimi, M.A.; Saleh, M.E. Analyzing power and energy efficiency of bitonic mergesort based on performance evaluation. *IEEE Access* 2018, 6, 42757–42774.
15. Aljabri, N.; Al-Hashimi, M.; Saleh, M.; Abulnaja, O. Investigating power efficiency of mergesort. *J. Supercomput.* 2019, 75, 6277–6302.
16. Al-Hashimi, M.; Saleh, M.; Abulnaja, O.; Aljabri, N. On the power characteristics of mergesort: An empirical study. In Proceedings of the 2017 Int'l Conf. on Advanced Control Circuits and Systems & 2017 Int'l Conf. on New Paradigms in Electronics & Information Technology (ACCS'17/PEIT'17), Alexandria, Egypt, 5–8 November 2017; pp. 172–178.
17. Oo, N.Z.; Chaikan, P. The Effect of Loop Unrolling in Energy Efficient Strassen's Algorithm on Shared Memory Architecture. In Proceedings of the 2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), Jeju, Republic of Korea, 27–30 June 2021; pp. 1–4.
18. Jammal, F.; Aljabri, N.; Al-Hashimi, M.; Saleh, M.; Abulnaja, O. A preliminary empirical study of the power efficiency of matrix multiplication. *Electronics* 2023, 12, 1599.