ISSN:0975-9646

The Haswell Microarchitecture - 4th Generation Processor

Tarush Jain[#], Tanmay Agrawal[*]

[#]*IT Department,*
MIT College of Engineering ,Pune, India
[*]*IT Department*
Maharashtra Institute of Technology ,Pune, India

Abstract: **Haswell is the codename for processors and processor microarchitectures which will replace Sandy Bridge and Ivy Bridge. The Haswell family features a new CPU core, new graphics and substantial changes to the platform in terms of memory and power delivery and power management. The new microarchitecture is expected to improve performance and power consumption, featuring new AVX2 instructions and taking advantage of Intel's 22nm FinFET process technology, while simultaneously introducing new integrated graphics named Iris. The improvements in Haswell are concentrated in the out-of-order scheduling, execution units and especially the memory hierarchy. 4th generation Intel Core processor family based on Haswell microarchitecture will bring faster, thinner, lighter, cooler, more secure systems with built-in graphics to mainstream.**

I.INTRODUCTION

Haswell is the processor microarchitecture as a successor to the Sandy Bridge and Ivy Bridge architecture.

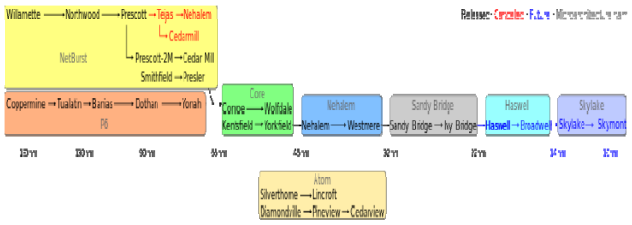

Figure 1.1: Intel CPU core roadmaps from NetBurst and P6 to Skymont

The Haswell architecture is specifically designed to optimize the power savings and performance benefits on the improved 22nm process node.
Haswell represents the "tock" in Intel's CPU development program.

II.INTEL TICK TOCK

"Tick-Tock" is a model adopted by chip manufacturer Intel Corporation since 2007 to follow every microarchitectural change with a die shrink of the process technology. Every "tick" is a shrinking of process technology of the previous microarchitecture and every "tock" is a new microarchitecture. Every year, there is expected to be one tick or tock.

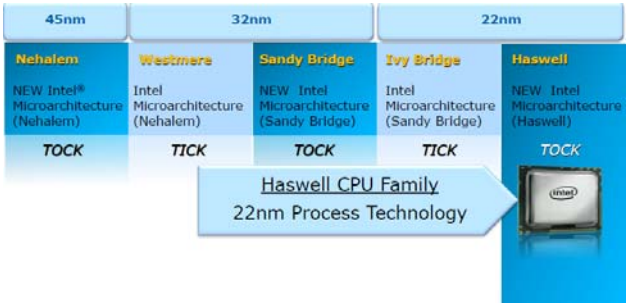

Figure 2.1: Intel Tick-Tock (Haswell is a Tock).

Intel's "tick-tock" model inspires confidence in the future of microprocessors and the devices that depend on them. A "tick" advances manufacturing technology and a "tock" delivers a new microarchitecture. With a "tick" cycle every couple of years, look for Intel to advance manufacturing process technology and continue to deliver the expected benefits of Moore's Law to users. The typical increase in transistor density enables new capabilities, higher performance levels, and greater energy efficiency—all within a smaller, more capable version of the previous "tock" microarchitecture. In alternating "tock" cycles, expect Intel to use the previous "tick" cycle's manufacturing process technologies to introduce the next big innovation in processor microarchitecture. Intel microarchitecture advancements seek to improve energy efficiency and performance as well as functionality and density of features such as hardware-supported video transcoding, encryption/decryption, and other integrated capabilities.

III. POWER EFFICIENCY

Haswell processors are much less power hungry than the ivy bridge processors. This power efficiency is the result of Intel's effort in trying to optimize the power consumption of every component on the motherboard. Intel has also equipped the haswell architecture with more power ratings and lower power modes. The haswell based processors can also switch power modes 25% faster than the ivy bridge. With Haswell, Intel has dropped the energy usage of the chip to 10 watts, down from 17 watts used by Ivy Bridge.

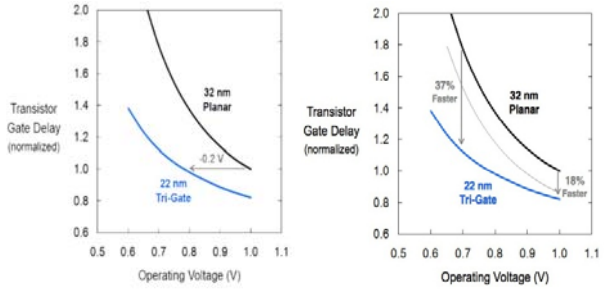

Figure 3.1: Transistor Gate Delay

With the lower energy consumption comes the benefit that ultrabooks and laptops will have a longer battery life on a yet thinner form. Haswell's promise could translate into lighter-weight ultrabooks with practically all-day better life. Reports say Haswell-powered devices will likely carry less obtrusive fans along with the thinner form factor compared to current ultrabooks.

IV. INTEGRATED GRAPHICS

The top two levels of integrated graphics in Intel's Haswell microarchitecture are codenamed 'Iris' and 'Iris Pro'. These graphics will be able to provide upto 2 or 3 times the performance of Intel HD Graphics 4K that comes with current Ivy bridge processors.

This is one of the most hyped performance improvement in the Haswell architecture. Intel has scrapped the nomenclature of HD Graphics xxxx in the favour of GT1, GT2 and GT3. These onboard graphics are meant to run modern games. The number of EUs (execution units) in GT2 and GT3 will be 20 and 40 respectively. To prevent malfunctioning due to some faulty EU, Intel has decided to provide an extra EU in the array. It is even stated that the GT3's performance will be comparable to the older gen. mainstream dedicated GPU like Nvidia GT 650m!

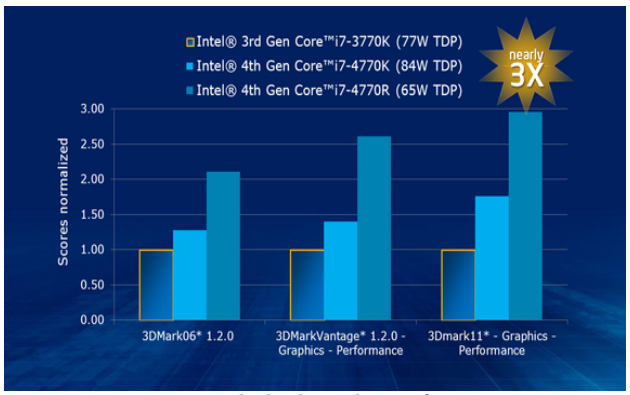

Figure 4.1: Ultrabook Graphics Performance

V. HASWELL EXECUTION UNIT

The execution units in Haswell are tremendously improved over Sandy Bridge, particularly to support AVX2 and the new FMA. Haswell adds an integer dispatch port and a new memory port, bringing the execution to eight uops/cycle. But the biggest changes are to the vector execution units. On the integer SIMD side, the hardware has been extended to single cycle 256-bit execution. For floating point vectors, the big change is 256-bit fused multiply add units for two of the execution ports. As a result, the theoretical peak performance for Haswell is more than double that of Sandy Bridge.

Every cycle, up to eight uops are sent from the unified scheduler to the dispatch ports. As shown in Figure 5.1, computational uops are dispatched to ports 0, 1, 5, and 6 and executed on the associated execution units. The execution units include three types: integer, SIMD integer, and FP (both scalar and SIMD).

Port 6 on Haswell is a new scalar integer port. It only accesses the integer registers and handles standard ALU (Arithmetic Logic Unit) operations, including shifts and branches that were previously on port 5 (in Sandy Bridge). One of the advantages of the new integer port is that it can handle many instructions while the SIMD dispatch ports are fully utilized.

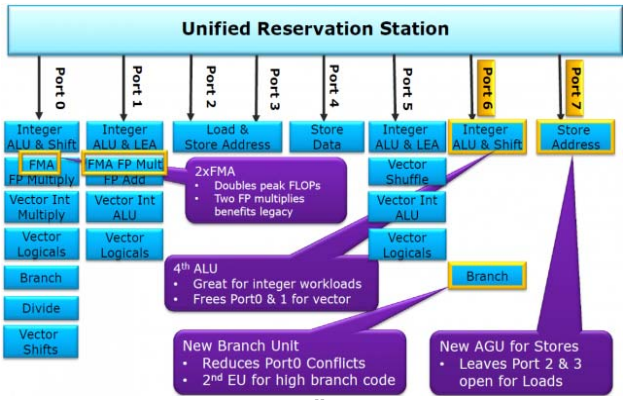

Figure 5.1: Haswell Execution Unit

Turning to the SIMD execution units, Haswell boasts a wide variety of improvements for AVX2 that roughly double the throughput. The SIMD ALU, multiplier, and shifter on port 0 have been extended to 256-bits, along with the vector ALU and blending on Port 1. The ALU, shuffle, and blend units on port 5 were also extended to 256-bits. Generally, the SIMD integer performance doubled due to wider 256-bit AVX2 instructions.

The new FMA instructions also required significant changes for the floating point units in Haswell. Intel's architects opted for a fully pipelined fused multiply-add that keeps the latency at five cycles—the same as an FP multiply (FMUL)—so the extra add in the FMA is essentially free from a latency perspective. On Sandy Bridge, port 0 was for FP multiplies, while port 1 was used for FP addition. Haswell added 256-bit FMA units to both port 0 and 1 that double for executing FP multiplies. So FMAs and FMULs can issue on both ports, but FP addition (FADD) must go to port 1. For recompiled code, the floating point performance has basically doubled by virtue of the FMA instructions, yielding 16 DP FLOP/cycle for each Haswell core. Existing code that depends on FMUL throughput will also get substantially faster.

VI. HASWELL MEMORY HIERARCHY

The memory hierarchy for Haswell is probably the biggest departure from the previous generation. The cache bandwidth doubled in tandem with an increase in FLOP/s from the new FMA units. Moreover, the whole memory system has been enhanced to support gather instructions and transactional memory.

Memory accesses start by allocating entries in the load and store buffers, which can track more than 100 uops, statically split between two threads. For Sandy Bridge, ports 2 and 3 calculated addresses, with port 4 for writing data into the L1 data cache. The new port 7 on Haswell handles address generation for stores. As a result, Haswell can now sustain two loads and one store per cycle under nearly any circumstances.

Metric	Nehalem	Sandy Bridge	Haswell
L1 Instruction Cache	32K, 4-way	32K, 8-way	32K, 8-way
L1 Data Cache	32K, 8-way	32K, 8-way	32K, 8-way
Fastest Load-to-use	4 cycles	4 cycles	4 cycles
Load bandwidth	16 Bytes/cycle	32 Bytes/cycle (banked)	64 Bytes/cycle
Store bandwidth	16 Bytes/cycle	16 Bytes/cycle	32 Bytes/cycle
L2 Unified Cache	256K, 8-way	256K, 8-way	256K, 8-way
Fastest load-to-use	10 cycles	11 cycles	11 cycles
Bandwidth to L1	32 Bytes/cycle	32 Bytes/cycle	64 Bytes/cycle
L1 Instruction TLB	4K: 128, 4-way 2M/4M: 7/thread	4K: 128, 4-way 2M/4M: 8/thread	4K: 128, 4-way 2M/4M: 8/thread
L1 Data TLB	4K: 64, 4-way 2M/4M: 32, 4-way 1G: fractured	4K: 64, 4-way 2M/4M: 32, 4-way 1G: 4, 4-way	4K: 64, 4-way 2M/4M: 32, 4-way 1G: 4, 4-way
L2 Unified TLB	4K: 512, 4-way	4K: 512, 4-way	4K+2M shared: 1024, 8-way

All caches use 64-byte lines

Figure 6.1: Haswell Memory Hierarchy

Once an address has been calculated by the Address Generation Unit (AGU), the uop will probe the translation look-aside buffers (TLBs). The L1 DTLB in Haswell is the same organization as in Sandy Bridge. However, there is a third port on the DTLB to accommodate the new store AGU on port 7. Misses in the L1 DTLB are serviced by the unified L2 TLB, which has been substantially improved with support for 2MB pages and twice the number of entries.

Similarly, the L1 data cache in Haswell is the same size and latency (minimum of four cycles), but with a third more bandwidth. The data cache can sustain two 256-bit loads and a 256-bit store every cycle, for 96B/cycle compared with 48B/cycle for Sandy Bridge. Moreover, the data cache in Sandy Bridge was banked, meaning that conflicts could potentially reduce the actual bandwidth. Turning to Haswell's L2 cache, the capacity, organization, and latency is the same, but the bandwidth has also doubled. A full 64B cache line can be read each cycle.

While the organization of the caches was largely unchanged, the capabilities are substantially greater in Haswell since the caches have been designed for TSX. As speculated, Haswell's transactional memory uses the L1 to store transaction data (either for Hardware Lock Elision or Restricted Transactional Memory). Transactions where the data fits in the L1D cache should be able to execute successfully. From a practical standpoint, this means that the L1D cache contains a bit of extra meta-data to track whether cache lines have been read or written to detect any conflicts.

While the closest levels of the memory hierarchy have been significantly improved, Haswell's system architecture has also been enhanced. The tags for the Last Level Cache (LLC) have been replicated, with one copy for reading data (at the same 32B/cycle) and another for prefetching and coherency requests. The write throughput for the memory controller is also significantly better due to larger write buffers for DRAM accesses and better scheduling algorithms.

To reduce power, the ring and LLC are on a separate frequency domain from the CPU cores. This means that the CPUs can enter in a low-power state, while the ring and LLC run at full throttle to feed the GPU. For many graphically intense workloads, this can reduce the power consumption substantially.

VII. CONCLUSION

Haswell, Intel's fourth-generation core microprocessor family, will offer better performance with lower power consumption. Haswell is a more powerful microarchitecture and has a bevy of new instructions. The Haswell microarchitecture has a modestly larger out-of-order window, with a 33 percent increase in dispatch ports and execution resources. Compared to previous generations, the theoretical FLOPs and integer operations have doubled for each core, primarily due to wider vectors. More significantly, the cache hierarchy can sustain twice the bandwidth, and it has fewer utilization bottlenecks. Because of these attributes, Haswell will serve as the main core microprocessor for ultrabooks. The microprocessor will provide Intel Identity Protection Technology to improve security, and will also support multiple displays and high-definition 4K monitors.

VIII. REFERENCES:

[1] Rotem, E. ; Naveh, A. ; Rajwan, D., Ananthakrishnan, A.,Weissmann, E., "Power-Management Architecture of the Intel Microarchitecture Code-Named Sandy Bridge", Micro, IEEE.

[2] Molka, D. ; Hackenberg, D., Schone, R. ; Muller, M.S., "Memory Performance and Cache Coherency Effects on an Intel Nehalem Multiprocessor System", Parallel Architectures and Compilation Techniques, 2009. PACT '09. 18th International Conference.

[3] Doyle, B. ; Boyanov, B. ; Datta, S. , Doczy, M. , Hareland, S. , Jin, B. , Kavalieros, J. , Linton, T. , Rios, R. , Chau, R., "Tri-Gate fully-depleted CMOS transistors: fabrication, design and layout", VLSI Technology, 2003. Digest of Technical Papers. 2003 Symposium.

[4] Auth, C., "22-nm fully-depleted tri-gate CMOS transistors", Custom Integrated Circuits Conference (CICC), 2012 IEEE.

[5] Jan, C.-H. ; Bhattacharya, U. , Brain, R. , Choi, S.-J. , Curello, G. , Gupta, G. ,Hafez, W. , Jang, M. , Kang, M. , Komeyli, K. , Leo, T. , Nidhi, N. , Pan, L. , Park, J. , Phoa, K. , Rahman, A. , Staus, C. , Tashiro, H. , Tsai, C. , Vandervoorn, P. , Yang, L. , Yeh, J.-Y. , Bai, P., "A 22nm SoC platform technology featuring 3-D tri-gate and high-k/metal gate, optimized for ultra-low power, high performance and high density SoC applications", Electron Devices Meeting (IEDM), 2012 IEEE International.

[6] Kavalieros, J. ; Doyle, B. , Datta, S. , Dewey, G. , Doczy, M. , Jin, B., Lionberger, D. , Metz, M. , Rachmady, W. , Radosavljevic, M. , Shah, U. , Zelick, N. , Chau, R., "Tri-Gate Transistor Architecture with

High-k Gate Dielectrics, Metal Gates and Strain Engineering", VLSI Technology, 2006. Digest of Technical Papers. 2006 Symposium.
[7] Seifert, N. ; Gill, B. , Jahinuzzaman, S. , Basile, J. , Ambrose, V. , Quan Shi , Allmon, R. , Bramnik, A., "Soft Error Susceptibilities of 22 nm Tri-Gate Devices" , Nuclear Science, IEEE Transactions.
[8] White Paper, "Inside Intel® Core™ Microarchitecture and Smart Memory Access".
[9] White Paper, "First the Tick, Now the Tock: Intel® Microarchitecture (Nehalem)".
[10] White Paper, "Going Under the Hood with Intel's Next Generation Microarchitecture Code Name Haswell".