INSIDE 6TH GEN INTEL® CORE™: NEW MICROARCHITECTURE CODE NAMED SKYLAKE

Jack Doweck, Wen-fu Kao
Intel Corporation, 2016

Authors: Ittai Anati, David Blythe, Jack Doweck, Hong Jiang, Wen-fu Kao, Julius Mandelblat, Lihu Rappoport, Efraim Rotem, Ahmad Yasin
SOME ARCHEOLOGY
What is Intel® Core™ Microarchitecture?

The Intel® Core™ Microarchitecture is a new foundation for Intel® architecture-based desktop, mobile, and mainstream server multi-core processors.

Designed for efficiency and optimized performance across a range of market segments and power envelopes.

2006 Intel Core Microarchitecture based processors:

**Server:**
Dual-core Intel® Xeon® 51xx Processors
Quad-core codenamed Clovertown

**Desktop:**
Dual-core Intel® Core™ 2 Duo Processors
Quad-core codenamed Kentsfield

**Mobile:**
Dual-core Intel® Core™ 2 Duo Processors
What is Intel® Core™ Microarchitecture?

The Intel® Core™ Microarchitecture is a new foundation for Intel® architecture-based desktop, mobile, and mainstream server multi-core processors

Designed for efficiency and optimized performance across a range of market segments and power envelopes

2006 Intel Core Microarchitecture based processors:

**DP Server:**
- Dual-core Intel® Xeon® 51xx Processors
- Quad-core codenamed Clovertown

**Desktop:**
- Dual-core Intel® Core™ 2 Duo Processors
- Quad-core codenamed Kentsfield

**Mobile:**
- Dual-core Intel® Core™ 2 Duo Processors
AN HISTORICAL PERSPECTIVE
WHAT CHANGED IN ~10 YEARS?

SKL (HOT CHIPS 2016) COMPARED TO MRM (HOT CHIPS 2006)

UP TO 10x MORE EFFICIENT
3-5x FASTER WITH 2x LOWER TDP¹

5.0x DENSER
1.5x SMALLER
6.0x HIGHER

PROCESS: 65NM
DIE SIZE: 143 MM²
TRANSISTOR COUNT: 291 M

¹ Based on estimated SPECint rate_base2006 and SPECfp rate_base2006. Range depends on benchmark, product and platform. Gains include advance on compiler and memory technologies.
WHAT CHANGED IN ~10 YEARS?

**SKL (HOT CHIPS 2016)**

**MRM (HOT CHIPS 2006)**

**UP TO 10x MORE EFFICIENT**

3-5x FASTER WITH 2x LOWER TDP\(^1\)

**1.5x SMALLER AREA**

**6.0x HIGHER TRANSISTOR COUNT**

\(^1\) Based on estimated SPECint_rate_base2006 and SPECfp_rate_base2006. Range depends on benchmark, product and platform. Gains include advance on compiler and memory technologies.
WHAT CHANGED IN ~10 YEARS?

**SKL (HOT CHIPS 2016)**

- 6 DIE VERSIONS, 2 OR 4 CORES
- 1 (OR 2 CHIP) SOLUTION
- TDP: 4.5W TO 91W
- GRAPHICS, MEMORY CONTROLLER, IMAGE PROCESSING, I/O S

**MRM (HOT CHIPS 2006)**

- 3 DIE VERSIONS, 1 OR 2 CORES
- 3 CHIPS SOLUTION
- TDP: 5.5W TO 75W
- CPU ONLY

**COMPARSED TO**

- UP TO 10x MORE EFFICIENT
- 3-5x FASTER WITH 2x LOWER TDP\(^1\)
- 1.5x SMALLER AREA
- 6.0x HIGHER TRANSISTOR COUNT

\(^1\) Based on estimated SPECint_rate_base2006 and SPECfp_rate_base2006. Range depends on benchmark, product and platform. Gains include advance on compiler and memory technologies.
SKYLAKE’S PERFORMANCE AND POWER
** System Configuration details:
5th generation Intel® Core™ system: Intel Reference Platform, Intel® Core™ M-5Y71 processor (Turbo up to 2.9GHz/2.6GHz, PL1=4.5W TDP, 2C4T), Memory: 4GB (2x2GB) LPDDR3-1600, HDD: Intel® Solid State Drive (Intel® SSD), Display: 1920x1080, OS: Windows® 10.

6th generation Intel® Core™ system: Intel Reference Platform, Intel® Core™ m7-6Y75 processor (Turbo up to 3.1GHz/2.9GHz, PL1=4.5W TDP 1 PL1=4.5W TDP 1 PL1=4.5W TDP 1 PL1=4.5W TDP 1, Memory: 4GB (2x2GB) LPDDR3-1600, HDD: Intel® Solid State Drive (Intel® SSD), Display: 1920x1080, OS: Windows® 10.

** SPEC CPU 2006 estimates based on measurements on reference boards.
**System Configuration details:**

5th generation Intel® Core™ system: Intel Reference Platform, Intel® Core™ i7-5550U processor (Turbo up to 2.8GHz, PL1=3W TDP, 2C4T), Memory: 4GB (2x2GB) LPDDR3-1600, HDD: Intel® Solid State Drive (Intel® SSD), Display: 1920x1080, OS: Windows® 10.

6th generation Intel® Core™ system: Intel Reference Platform, Intel® Core™ i7-6650U processor (Turbo up to 3.1GHz, PL1=4.5W TDP, 2C4T), Memory: 4GB (2x2GB) LPDDR3-1600, HDD: Intel® Solid State Drive (Intel® SSD), Display: 1920x1080, OS: Windows® 10.

**SPEC CPU 2006 estimates based on measurements on reference boards.**

---

**9 HOURS BATTERY LIFE**

Up to 9 hours of 1080p video playback

In a 15W chassis:

**UP TO 28% MORE COMPUTE**

**70% BETTER 3D GAMING**

Vs. prior gen

---

**Inside 6th generation Intel Core Code Name Skylake - HOT CHIPS 2016**
INSIDE SKYLAKE
INSIDE SKYLAKE: OUTLINE

• High level view
• Cache and Memory subsystem
• Power management and optimizations
• The core
• Processor Graphics
• Media Engine
INSIDE SKYLAKE: ORIENTATION MAP

- High level view
- Cache and Memory subsystem
  - Power management and optimizations
  - The core
  - Processor Graphics
  - Media Engine
Inside 6th generation Intel Core Code Name Skylake - HOT CHIPS 2016

**INTEL’S SKYLAKE MICROARCHITECTURE**

**High Level View**

- Increased chipset I/O throughput, Tablet I/Os, Audio DSP Upgrade, Sensor Hub
- Higher resolution display
- Bigger/wider core, better instruction per clock, improved power efficiency
- Enhanced ring/LLC for improved throughput

**Features**

- Integrated camera ISP
- Extended overclocking capabilities
- Faster DDR Memory
- Advanced Processor Graphics GT3 + eDRAM, GT4 + eDRAM; OpenCL™ 2.0 API, DirectX® 12, OpenGL® 4.4
CACHE AND MEMORY SOLUTIONS

• Improved coherent fabric efficiency
  – Double throughput of the last level cache (LLC) miss handling
  – About 50% improvement in bandwidth per watt

• New eDRAM cache architecture, higher bandwidth, more products with eDRAM

• Support of faster DDR memory (up to DDR4 2400)

• Memory Encryption Engine for Intel® Software Guard Extension (Intel® SGX) and other security-critical data

• Improved memory QoS for high resolution displays and integrated image signal processor (ISP)
  – Higher concurrent bandwidth
EDRAM AS MEMORY SIDE CACHE

- **Observed by all memory accesses, i.e. fully coherent**
- **Not architectural and thus can cache any data including “uncacheable memory”**
- **No need to flush it for coherency maintenance**
- **Available for use by I/O devices and Display Engine**
- **For optimal performance GFX driver may choose caching of particular data in the eDRAM only and not in the LLC**

**Diagram Details:**
- **L1:**
  - L1D
  - L1I
- **L2:**
  - Core L1
  - Core L1
  - L2
  - L2

**Core Architecture Diagram:**
- **Core L1:**
  - L1D
  - L1I
  - L2

**GFX Architecture Diagram:**
- **GFX Graphics Caches:**
  - Graphics
  - Caches

**System Agent Diagram:**
- **System Agent:**
  - Other Devices
  - PCIe
  - EDRAM
  - EDreamCTL
  - M$ Tags
  - MC
  - DDR

**System Architecture Overview:**
- **EDRAM** as a Memory Side Cache
- **EDRAM CTL**
- **M$ Tags**
- **MC**
- **DDR**

**Inside 6th generation Intel Core Code Name Skylake - HOT CHIPS 2016**
INSIDE SKYLAKE: ORIENTATION MAP

• High level view
• Cache and Memory subsystem
• **Power management and optimizations**
• The core
• Processor Graphics
• Media Engine
POWER MANAGEMENT

INTEL® SPEED SHIFT TECHNOLOGY

A New revolutionary approach to Power/performance management

- Hardware P-state control - A new power architecture and interface
- Fully autonomous hardware controlled P-state selection
- Operating system set minimum QoS, maximum frequency and power/performance hint
- Better observation of micro architectural behavior of workloads
- Fine grain run time control

Higher performance and responsiveness at power constrained form factors
ARCHITECTURAL INTERFACE

DVFS – Intel SpeedStep® Technology

- $P \sim V^2 \cdot f \cdot C_{dyn} + \text{leakage}(V) \sim f^3$

Legacy: OS controls P-state

- P1-Pn enumerated via ACPI tables
- Explicit P-state selection to P1
- Autonomous control in turbo range

New: guided autonomous control

- OS provides min, max and preference
- Demand based HW control
• Most energy efficient frequency ($P_e$) is calculated at run time (EARTh algorithm\(^1\))
• No benefit in running lower than $P_e$ and lose energy $\rightarrow$ frequency clamped at $P_e$
  – Unless critical power saving is needed
  – Only if possible to enter package sleep state (Consumer Producer)
• Performance = energy: Preference ($\alpha$) allows limiting energy/performance
  – Semantics: frequency that meets $\Delta$Power/$\Delta$performance $\leq \alpha$


Inside 6th generation Intel Core Code Name Skylake - HOT CHIPS 2016
RESPONSIVENESS

Fast burst response while performing interactive work

- Filter out short interrupts and repeated work such as Video playback
- Filter cyclic workloads e.g. video playback
POWER OPTIMIZATIONS IN THE SKL CORE

Dynamic consumption based resource configuration
- Power Gating of Intel® AVX2 (Intel® Advanced Vector Extensions 2) hardware when it is not used
- Downscaling of underused resources

Improved scenario based power (e.g. media playback) for great mobile experience
- Idle power reduction
- C1 state power reduction: improved dynamic capacitance ($C_{\text{dyn}}$)

Better performance/Watt for the core, including focus on power at low utilization

Inside 6th generation Intel Core Code Name Skylake - HOT CHIPS 2016
INSIDE SKYLAKE: ORIENTATION MAP

- High level view
- Cache and Memory subsystem
- Power management and optimizations

- The core
  - Processor Graphics
  - Media Engine
SKYLAKE CORE: FRONT-END

IMPROVED FRONT-END

- Increased bandwidth of Instruction Decoders and \( \mu \)op-cache
- Higher capacity, improved Branch Predictor
- Reduced penalty for wrong direct jump target prediction
- Faster instruction prefetch
- Increased capacity of the \( \mu \)op queue / Loop Stream Detector
SKYLAKE CORE: OUT-OF-ORDER EXECUTION

Deeper Out-of-Order buffers extract more instruction parallelism

- 97 entry scheduler, 224 entry Reorder Buffer

- Improved throughput and latency for divide and SQRT
- Balanced throughput and latency of Floating point ADD, MUL and FMA
- Significantly reduced latency for AES instructions
SKYLAKE CORE: MEMORY SUB-SYSTEM

• Deeper load and store buffers, increasing loads and stores throughput
• Reduced penalty for page-split loads
• Reduced Store-to-load forwarding latency
• Stores that miss in the L1$ generate L2$ requests earlier
• Higher BW from L2$ and L3$
• Higher write BW to L3$
NEW SECURITY TECHNOLOGIES

Intel® Software Guard Extensions (Intel® SGX)

• A new infrastructure for trusted applications

• Instructions and flows to create, launch and operate “Enclaves” and protect them from malware and privileged software

• Provides for confidentiality, integrity, replay protection and attestation

Intel® Memory Protection Extensions (Intel® MPX)

• Memory buffer boundary testing prior to memory accesses, to ensure the memory access falls within the legitimate bounds of the object

• Helps improve SW robustness and prevent certain malware
NEW INSTRUCTIONS AND ENHANCEMENTS

XSAVE/RESTORE IMPROVEMENTS:

• Supports new Architectural state in SKL
• New instruction XSAVEC stores state in a compact layout

CLFLUSHOPT: A WEAKLY ORDERED VARIANT OF CLFLUSH
**PERFORMANCE MONITORING UNIT (PMU)**

**SKYLAKE PMU ENABLES NEW USAGES:**

Enhanced Top-Down Analysis Method \[1\]
- A structured approach for microarchitectural performance bottleneck detection
- Better quality and coverage for underlying PerfMon events

PMU virtualization and sharing are now easier

Precise attribution of front-end events to original code (e.g. i-cache misses)

Higher resolution event-based sampling

New class of usages with accurate Basic-Block timing in Last Branch Records
  - E.g.: Time function calls, Min/Max/Avg time per basic block

---

<table>
<thead>
<tr>
<th>FEATURE</th>
<th>PREVIOUS</th>
<th>SKYLAKE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARCH PERFMON VERSION</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>EVENTS QUANTITY (COVERAGE)</td>
<td></td>
<td>RICHER</td>
</tr>
<tr>
<td>EVENTS QUALITY</td>
<td></td>
<td>BETTER</td>
</tr>
<tr>
<td>TOP-DOWN ANALYSIS</td>
<td>BASIC</td>
<td>ACCURATE</td>
</tr>
<tr>
<td></td>
<td>SMT-OFF</td>
<td>SMT ON</td>
</tr>
<tr>
<td>FRONT-END EVENTS COVERAGE</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>TSC INCLUDED IN RECORD</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>TIMING INFO</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td># ENTRIES</td>
<td>16</td>
<td>32</td>
</tr>
</tbody>
</table>


Inside 6th generation Intel Core Code Name Skylake - HOT CHIPS 2016
INSIDE SKYLAKE: ORIENTATION MAP

• High level view
• Cache and Memory subsystem
• Power management and optimizations
• The core
  • Processor Graphics
  • Media Engine
SKYLAKE GRAPHICS

• Scalable performance from 24-72 EUs
  – 384-1152 SP FLOPS/clock
  – Geometry throughput increases
  – Pixel fill rate increase by 1.33-2x/clock
  – Lossless render compression

• New 3D Application Features
  – Conservative Rasterization
  – Bindless Textures
  – Render Target read

• API Support
  – DirectX® 12, Vulkan, Metal, OpenGL
  – OpenCL™ 2.0, extends CPU/GPU Programmability
Inside 6th generation Intel Core Code Name Skylake - HOT CHIPS 2016

**SKYLAKE GRAPHICS**

**MICROARCHITECTURE IMPROVEMENTS**

**LOSSLESS RENDER COMPRESSION**
- Texture, pixel read/write paths
- 2:1 peak compression ratio
- Saves bandwidth and memory power

**LOSSLESS COMPRESSION PERFORMANCE**

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Benefit</th>
<th>Benefit</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>I7-SKL-GT2</td>
<td>11%</td>
<td>4%</td>
<td>5%</td>
</tr>
<tr>
<td>BATMAN ARKHAM ORIGINS</td>
<td>4%</td>
<td>5%</td>
<td>3%</td>
</tr>
<tr>
<td>BIOSHOCK INFINITE</td>
<td>5%</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td>UNIGINE VALLEY</td>
<td>3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOMB RAIDER</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FINER GRAIN PREEMPTION**
- Increases responsiveness of system
- 3D: triangle boundaries
- Compute: Thread-group → Mid-thread

**FINER GRAIN PERFORMANCE**

<table>
<thead>
<tr>
<th>Thread Group</th>
<th>Benefit</th>
<th>Benefit</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-SERIES</td>
<td>4-6MS</td>
<td>17-22MS</td>
<td>300US</td>
</tr>
<tr>
<td>Y-SERIES</td>
<td>200-500MS</td>
<td>200-500MS</td>
<td>300US</td>
</tr>
<tr>
<td></td>
<td>17MS</td>
<td>24MS</td>
<td>240US</td>
</tr>
</tbody>
</table>

Adobe Photoshop*

Sample App1

Sample App2
SKYLAKE MEDIA SYSTEM-LEVEL IMPROVEMENTS

- Introducing Skylake with Gen9 Processor Graphics, built on 14nm
- Improved power management to optimize media usages
  - Ability to run only fixed function (FF) media in the Unslice without powering on the Slices for key workloads
  - Ability to run selective slices instead of full slices
  - Ability to use selective pair of EUs in the subslice
  - Inter-media engine memory compression to reduce bandwidth
- New Intel® Quick Sync Video mode
  - Now support Processor Graphics (PG) and Fixed Function (FF) modes
- Broad Enabling of Applications
  - Support for DirectX® 11.2, DirectX-next and OpenCL™ 2.0, extends CPU/GPU Programmability
**SKYLAKE MEDIA ENGINE IMPROVEMENTS**

### MULTI-FORMAT CODEC (MFX)

<table>
<thead>
<tr>
<th>DECODE</th>
<th>PAK</th>
<th>MULTI-FORMAT LEGACY</th>
<th>HEVC DECODE</th>
<th>FF ENC</th>
</tr>
</thead>
</table>

- **MFX**
  - HEVC 8b decode/encode for real-time 4Kp60 usages
  - VP8 and JPEG/MJPEG decode/encode
  - Fixed function low power low latency AVC encoder latency real-time usages

### VIDEO QUALITY ENGINE (VQE)

<table>
<thead>
<tr>
<th>DN &amp; DI</th>
<th>IEC</th>
<th>DM</th>
<th>WB</th>
</tr>
</thead>
</table>

- **VQE**
  - 16b processing path
  - Spatial 5x5 denoise filter
  - Local Adaptive Contrast Enhancement (LACE)
  - Camera processing features
  - Improved camera/image RAW processing capability

### SCALER AND FORMAT CONVERSION (SFC)

<table>
<thead>
<tr>
<th>AVS</th>
<th>IEF</th>
<th>CSC</th>
</tr>
</thead>
</table>

- **SFC**
  - New in-line scaling & format conversion engine
  - Includes IEF sharpening filter

Inside 6th generation Intel Core Code Name Skylake - HOT CHIPS 2016
SUMMARY

• Skylake delivers record levels of performance and battery life in many personal computing use cases and form factors

• Intel® Speed Shift Technology provides higher performance, responsiveness and efficiency at power constrained form factors

• Skylake Processor Graphics delivers scalable performance, >1TFLOPS compute, enhanced low power media engines, flexible power management, and end-to-end 4K experience

• Skylake family of products allows developers to:
  • Choose from wide range of platform capabilities
  • Innovate with products for wide range of thermal envelopes and I/O solutions
  • Optimize the system performance using the advanced PMU capabilities

• Skylake introduces Intel® SGX: a revolutionary game changer to trusted application security in the main stream SW environment