

2 章

インテル® 64 プロセッサ・アーキテクチャーと IA-32 プロセッサ・アーキテクチャー

本章では、最新世代のインテル® 64 プロセッサと IA-32 プロセッサ（インテル® マイクロアーキテクチャー Haswell⁺、インテル® マイクロアーキテクチャー Ivy Bridge⁺、インテル® マイクロアーキテクチャー Sandy Bridge⁺ ベースのプロセッサと、インテル® Core™ マイクロアーキテクチャー・ベースのプロセッサ、拡張版インテル® Core™ マイクロアーキテクチャー、インテル® マイクロアーキテクチャー Nehalem⁺）におけるソフトウェア最適化に関連するプロセッサの機能について概説する。これらの機能には、以下のものが含まれる。

- 高クロックレートかつ高スループットでの命令実行が可能なマイクロアーキテクチャー、高速なキャッシュ階層、高速システムバス
- インテル® Core™ プロセッサとインテル® Xeon® プロセッサ・ファミリーで利用可能なマルチコア・アーキテクチャー
- ハイパースレッディング・テクノロジー¹ (HT テクノロジー) のサポート
- インテル® 64 プロセッサのインテル® 64 アーキテクチャー
- SIMD 拡張命令: インテル® MMX® テクノロジー、ストリーミングSIMD 拡張命令 (SSE)、ストリーミングSIMD 拡張命令2 (SSE2)、ストリーミングSIMD 拡張命令3 (SSE3)、ストリーミングSIMD 拡張命令3 補足命令 (SSSE3)、ストリーミングSIMD 拡張命令4.1 (SSE4.1)、ストリーミングSIMD 拡張命令4.2 (SSE4.2)
- インテル® アドバンスド・ベクトル・エクステンション (インテル® AVX)
- 半精度浮動小数点変換と RDRAND 命令
- 乗算加算融合 (FMA) 拡張
- インテル® AVX2

インテル® Core™ 2 プロセッサ・ファミリー、インテル® Core™ 2 Extreme プロセッサ・ファミリー、インテル® Core™ 2 Quad プロセッサ・ファミリー、インテル® Xeon® プロセッサ 3000/3200/5100/5300/7300 番台は、電力効率に優れた高性能のインテル® Core™ マイクロアーキテクチャーをベースにしている。インテル® Xeon® プロセッサ 3100/3300/5200/5400/7400番台、インテル® Core™ 2 Extremeプロセッサ QX9600/Q9700 番台、インテル® Core™ 2 Quad プロセッサ Q9000/Q8000 番台は拡張版インテル® Core™ マイクロアーキテクチャーをベースにしている。インテル® Core™ i7 プロセッサは、インテル® マイクロアーキテクチャー Nehalem をベースにしている。インテル® Xeon® プロセッサ 5600 番台、インテル® Xeon® E7 とインテル® Core™ i7、i5、i3 プロセッサは、インテル® マイクロアーキテクチャー Westmere⁺ をベースにしている。

インテル® Xeon® プロセッサ E5 ファミリー、インテル® Xeon® プロセッサ E3-1200 ファミリー、インテル® Xeon® プロセッサ E7-8800/4800/2800 製品ファミリー、インテル® Core™ i7-3930K プロセッサ、および第 2 世代インテル® Core™ i7-2xxx、インテル® Core™ i5-2xxx、インテル® Core™ i3-2xxx プロセッサ・シリーズは、インテル® マイクロアーキテクチャー Sandy Bridge をベースにしている。

インテル® Xeon® プロセッサ E3-1200 v2 製品ファミリーと第 3 世代インテル® Core™ プロセッサは、インテル® 64 アーキテクチャーをサポートするインテル® マイクロアーキテクチャー Ivy Bridge をベースにしている。

インテル® Xeon® プロセッサ E3-1200 v3 製品ファミリーと第 4 世代インテル® Core™ プロセッサは、インテル® 64 アーキテクチャーをサポートするインテル® マイクロアーキテクチャー Haswell をベースにしている。

¹ ハイパースレッディング・テクノロジーを利用するには、ハイパースレッディング・テクノロジーに対応したインテル® プロセッサを搭載したコンピューター・システム、および同技術に対応したチップセットと BIOS、OS が必要である。性能は使用するハードウェアやソフトウェアによって異なる。

2.1 インテル® マイクロアーキテクチャー Haswell†

インテル® マイクロアーキテクチャー Haswell は、インテル® マイクロアーキテクチャー Sandy Bridge とインテル® マイクロアーキテクチャー Ivy Bridge の成功を受けて開発された。この新しいマイクロアーキテクチャーの基本パイプライン (図 2-1 を参照) は、以下の革新的な機能を提供している。

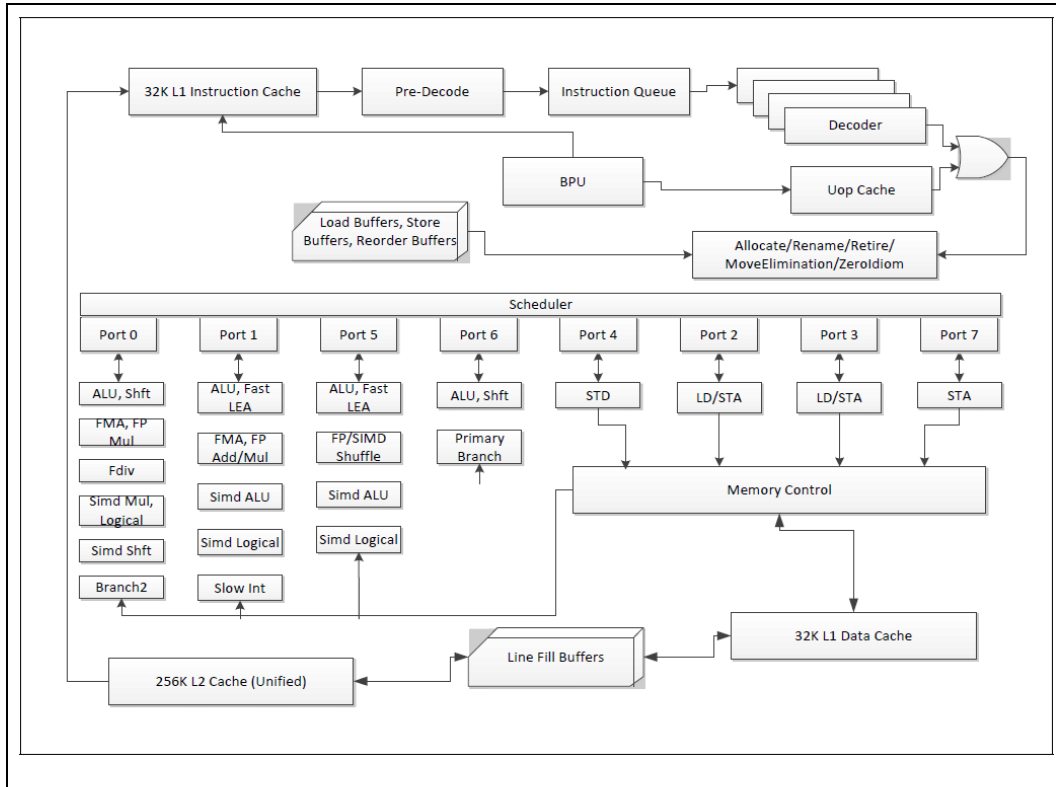


図 2-1 インテル® マイクロアーキテクチャー Haswell の CPU コア・パイプライン

- インテル® アドバンスド・ベクトル・エクステンション 2 (インテル® AVX2)、FMA のサポート
- 整数値演算と暗号化を高速化する新しい汎用命令
- インテル® トランザクショナル・シンクロナイゼーション・エクステンション (インテル® TSX) のサポート
- 各コアでサイクルごとに最大 8 マイクロオペレーション (uOP) をディスパッチ可能
- メモリー操作、FMA、インテル® AVX 浮動小数点実行ユニット、インテル® AVX2 整数実行ユニット用の 256 ビット・データ・パス
- L1 データキャッシュと L2 キャッシュの帯域幅が増加
- 2 つの FMA 実行パイプライン
- 4 つの数値演算ユニット (ALU)
- 3 つのストア・アドレス・ポート
- 2 つの分岐実行ユニット
- IA プロセッサ・コアおよびアンコア・サブシステム向けの高度な電力管理機能
- オプションの L4 キャッシュをサポート

インテル® マイクロアーキテクチャー Haswell は、L3 (オプションでオフダイの L4 も) の複数のスライスへのリング・インターコネクト、プロセッサ・グラフィックス、統合型メモリー・コントローラー、インターコネクト・ファブリックなどを含むいくつかの要素で構

成される共有アンコア・サブシステムと、複数のプロセッサ・コアとの柔軟な統合をサポートしている。図 2-2 に、4 CPU コアとアンコア要素で構成されるシステム統合の例を示す。

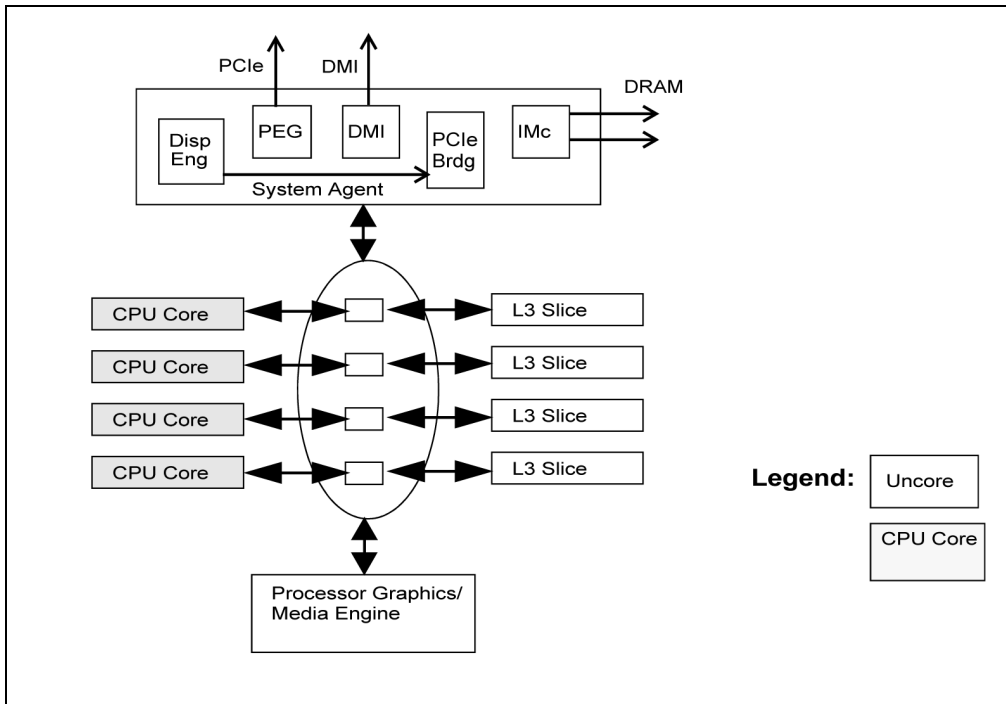


図 2-2 インテル® マイクロアーキテクチャー Haswell の 4 コアのシステム統合

2.1.1 フロントエンド

インテル® マイクロアーキテクチャー Haswell のフロントエンドは、インテル® マイクロアーキテクチャー Sandy Bridge (2.2.2 節) とインテル® マイクロアーキテクチャー Ivy Bridge (2.2.7 節) をベースに開発され、次の点が拡張されている。

- マイクロオペレーション (uOP) キャッシュ (またはデコード済み命令キャッシュ) は、2 つの論理プロセッサ間で均等に分割される。
- 命令デコーダーは、アクティブな論理プロセッサ間で交互に使用される。1 つの論理プロセッサがアイドル状態の場合は、もう一方のアクティブな論理プロセッサがデコーダーを続けて使用する。
- ループストリーム検出器 (LSD)/マイクロオペレーション (uOP) は、56 マイクロオペレーション (uOP) までの小さなループを検出できる。56 エントリーのマイクロオペレーション (uOP) キューは、ハイパースレディング・テクノロジーが有効な場合、2 つの論理プロセッサによって共有される (インテル® マイクロアーキテクチャー Sandy Bridge では、各コアに 28 エントリーのマイクロオペレーション (uOP) キューの複製が提供される)。

2.1.2 アウトオブオーダー・エンジン

以下に、アウトオブオーダー・エンジンの主要構成要素と主な改善点を示す。

リネーマー: リネーマーは、マイクロオペレーション (uOP) キューからスケジューラーのディスパッチ・ポートへマイクロオペレーション (uOP) を移動し、実行リソースにバインドする。ゼロイディオム、1 イディオム、ゼロレイテンシーのレジスター移動命令はリネーマーによって実行され、スケジューラーと実行コアを解放することでパフォーマンスを向上できる。

スケジューラー: スケジューラーは、ディスパッチ・ポートへのマイクロオペレーション (uOP) のディスパッチを制御する。アウトオブオーダー実行コアをサポートするため 8 つのディスパッチ・ポートがあり、そのうち 4 つは計算処理用の実行リソースを提供し、残り 4 つは 1 サイクルで最大 2 つの 256 ビット・ロード操作と 1 つの 256 ビット・ストア操作をサポートする。

実行コア: スケジューラーは、各ポートで 1 つずつ、サイクルごとに最大 8 つのマイクロオペレーション (uOP) をディスパッチできる。計算リソースを提供する 4 つのポートには ALU が 1 つずつあり、実行パイプのうち 2 つは FMA ユニット専用である。除算/平方根を除き、STTNI (String and Text New Instructions) /AESNI (Advanced Encryption Standard New Instructions) ユニット、ほとんどの浮動小数点および整数 SIMD 実行ユニットは 256 ビット幅である。メモリー操作の 4 つのディスパッチ・ポートは、2 つのロード/ストアアドレス操作のデュアルユース・ポート、ストアアドレス専用のポート、1 つのストアデータ専用ポートで構成されており、すべてのポートで 256 ビットのメモリー・マイクロオペレーション (uOP) を処理できる。浮動小数点のピーク・スループットは、FMA を使用した場合、単精度では 1 サイクルあたり 32 マイクロオペレーション (uOP)、倍精度では 16 マイクロオペレーション (uOP) であり、インテル® マイクロアーキテクチャー Sandy Bridge の 2 倍である。アウトオブオーダー・エンジンは、同時に 192 マイクロオペレーション (uOP) を処理できる (インテル® マイクロアーキテクチャー Sandy Bridge では 168 マイクロオペレーション (uOP) である)。

2.1.3 実行エンジン

次の表に、各ポートでディスパッチ可能なマイクロオペレーション (uOP) を示す。

表 2-1 ディスパッチ・ポートと実行スタック

ポート 0	ポート 1	ポート 2,3	ポート 4	ポート 5	ポート 6	ポート 7
ALU, Shift	ALU, Fast LEA,	Load_Addr, Store_addr	Store_data	ALU, Fast LEA	ALU, Shift, JEU	Store_addr, Simple_AGU
SIMD_Log, STTNI, SIMD_Shifts	SIMD_ALU, SIMD_Log			SIMD_ALU, SIMD_Log		
FMA/FP_mul, Div	FMA/FP_mul, FP_add			FP/Int Shuffle		
2nd_Jeu	slow_int					

リザベーション・ステーション (RS) が 60 エントリーに拡大され (インテル® マイクロアーキテクチャー Sandy Bridge では 54 エントリー)、マイクロオペレーション (uOP) の実行準備ができている場合、サイクルごとに最大 8 つのマイクロオペレーション (uOP) をディスパッチできる。RS でマイクロオペレーション (uOP) は特定のデータ型やデータの粒度を処理するスタックに分けられ、発行ポートから特定の実行クラスターにディスパッチされる。

あるスタックで実行されるマイクロオペレーション (uOP) のソースが、別のスタックで実行されるマイクロオペレーション (uOP) から取得される場合、遅延が生じる可能性がある。インテル® SSE 整数操作とインテル® SSE 浮動小数点操作の間の遷移でも遅延が発生する。これは、命令フローに追加されるマイクロオペレーション (uOP) によって、データ遷移が行われるためである。実行後にライトバックされるデータを、後続のマイクロオペレーション (uOP) 実行にバイパスする方法とその遅延サイクル数を表 2-2 に示す。

表 2-2 マイクロオペレーション (uOP) 間のバイパスによる遅延 (サイクル数)

遷移元/遷移先	整数	SSE-INT/ AVX-INT	SSE-FP/ AVX-FP_LOW	X87/ AVX-FP_High
整数		<ul style="list-style-type: none"> • uOP (ポート 5) • uOP (ポート 6) + 1 サイクル遅延 	<ul style="list-style-type: none"> • uOP (ポート 5) • uOP (ポート 6) + 1 サイクル遅延 	uOP (ポート 5) + 3 サイクル遅延
SSE-INT/ AVX-INT	uOP (ポート 1)		1 サイクル遅延	
SSE-FP/ AVX-FP_LOW	uOP (ポート 1)	1 サイクル遅延		uOP (ポート 5) + 1 サイクル遅延
X87/ AVX-FP_High	uOP (ポート 1) + 3 サイクル遅延		uOP (ポート 5) + 1 サイクル遅延	
ロード		1 サイクル遅延	1 サイクル遅延	2 サイクル遅延

2.1.4 キャッシュとメモリーサブシステム

キャッシュ階層は前世代と類似しており、各コアに L1 命令キャッシュ、L1 データキャッシュ、L2 ユニファイド・キャッシュがある。さらに、L3 ユニファイド・キャッシュもあり、そのサイズは製品構成に依存する。L3 キャッシュは複数のキャッシュスライスで構成されており、各スライスのサイズはリング・インターコネクで接続される製品構成に依存する。キャッシュトポロジーの詳細は、CPUID leaf 4 で確認できる。L3 キャッシュは、すべてのプロセッサ・コアで共有される「アンコア」サブシステムにある。一部の製品構成では L4 キャッシュもサポートされている。表 2-20 にキャッシュ階層の詳細を示す。

表 2-3 インテル® マイクロアーキテクチャー Haswell のキャッシュ・パラメーター

レベル	容量/アソシアティブ (ウェイ)	ラインサイズ (バイト)	最小レイテンシー ¹	スループット (クロック数)	ピーク帯域幅 (バイト/サイクル数)	アップデート方式
L1 データ	32KB/8	64	4 サイクル	0.5 ²	64 (ロード) + 32 (ストア)	ライトバック
命令	32KB/8	64	なし	なし	なし	なし
L2	256KB/8	64	11 サイクル	それぞれ異なる	64	ライトバック
L3 (共有)	それぞれ異なる	64		それぞれ異なる		ライトバック

¹ ソフトウェアから検知できるレイテンシーは、アクセスパターンやその他の要因により異なる。

² L1 データキャッシュは、最大 32 バイトのデータをフェッチ可能なロード操作を各サイクルで 2 つ処理できる。

TLB (Translation Lookaside Buffer) 階層は、L1 命令キャッシュ用の TLB、L1 データキャッシュ用の TLB、L2 ユニファイド・キャッシュ用の TLB で構成される。

表 2-4 インテル® マイクロアーキテクチャ Haswell の TLB パラメーター

レベル	ページサイズ	エントリー	アソシアティブ (ウェイ)	パーティション
命令	4KB	128	4 ウェイ	動的
命令	2MB/4MB	スレッドあたり 8		固定
L1 データ	4KB	64	4	固定
L1 データ	2MB/4MB	32	4	固定
L1 データ	1GB	4	4	固定
L2	4KB、2MB/4MB ページで共有	1024	8	固定

2.1.4.1 ロード操作とストア操作の拡張

L1 データキャッシュは各サイクルで 2 つの 256 ビット・ロード操作と 1 つの 256 ビット・ストア操作を処理でき、L2 ユニファイド・キャッシュは各サイクルで 1 つのキャッシュライン (64 バイト) を処理できる。さらに、マイクロオペレーション (uOP) の同時実行をサポートするため、72 のロードバッファと 42 のストアバッファが装備されている。