



UNIVERSIDADE DE SÃO PAULO

Escola de Artes, Ciências e Humanidades

Relatório Técnico PPgSI-002/2022

Estudo sobre a microarquitetura Nehalem e um paralelo com a microarquitetura Sandy Bridge

Andrei Clemente de Oliveira,
Juliana Witzke de Briton,
Victor Hugo Vieira Carlota,
Norton Trevisan Roman

Junho - 2022

O conteúdo do presente relatório é de única responsabilidade dos autores.

Série de Relatórios Técnicos

PPgSI-EACH-USP

Rua Arlindo Béttio, 1000 – Ermelino Matarazzo

03828-000 – São Paulo, SP.

TEL: (11) 3091-8197

<http://www.each.usp.br/ppgsi>

Estudo sobre a microarquitetura Nehalem e um paralelo com a microarquitetura Sandy Bridge

Andrei Clemente de Oliveira¹, Juliana Witzke de Briton¹,
Victor Hugo Vieira Carlota¹

¹Escola de Artes, Ciências e Humanidades – Universidade de São Paulo
São Paulo – SP, Brazil

andrei.oliveira@usp.br, jubrito@usp.br, victorviera@usp.br

Resumo. *A evolução da arquitetura do conjunto de instruções dentro do âmbito de arquitetura de computadores vem mostrando passos significativos ao longo desta última década. Neste relatório, trataremos de temas como o histórico, uso atual, uso pretendido, tipo, detalhes, desempenho e conjunto de instruções de uma arquitetura específica, a Intel Core Nehalem, tal como realizaremos um paralelo com outra arquitetura existente, sucessora e pertencente à mesma família de arquiteturas Intel Core, a Sandy Bridge.*

1. Introdução

O processador é um microchip especializado em realizar operações. Ao receber determinados volumes de dados binários (0 e 1), as instruções armazenadas em sua memória interna permitem o processamento das informações, a realização de instruções aritméticas e lógicas, bem como a resolução de entrada e saída de dados. (Pillou, 2020). Por ser a parte central - ou cérebro - do computador, é frequentemente a peça mais cara da máquina.

Os primeiros computadores, anteriores à década de 50, não eram capazes de armazenar programas, exigindo a manipulação da parte física da máquina - como cabos e chaves - para a execução de diferentes tarefas. (Harman e Gimblett, 2009). A partir da introdução do conceito de unidade central de processamento por John Von Neumann em 1945, reflexões sobre a organização interna dos computadores ganharam força, viabilizando o surgimento dos processadores como conhecemos hoje. Apesar da história dos processadores ser antiga, os processadores modernos tiveram seu início na década de 70 após o lançamento pela Intel do modelo 8088 para compor o primeiro computador em 1979. Após alguns anos, em 1985, foi lançado o modelo 386, o primeiro a introduzir as instruções básicas, que são utilizadas até hoje nos processadores atuais. (Harman e Gimblett, 2009).

Desde o modelo 386, muitas melhorias foram adotadas na construção dos processadores, tanto com relação ao hardware - como a inclusão de mais transistores e diferentes tipos de circuitos - quanto às técnicas de processamento - como pipeline e multi threading. (Harman e Gimblett, 2009). Este trabalho tem como objetivo analisar dois processadores da linha Intel Core, Nehalem e Sandy Bridge, buscando um entendimento dos principais avanços ao longo do tempo, através da comparação de diversos aspectos entre os dois processadores, a fim de servir para consultas por outros alunos e docentes e fomentar futuras pesquisas na área.

2. Microarquitetura Nehalem

A Intel, companhia norte-americana de manufatura de chips, inventou quase que por acidente o microprocessador. Os engenheiros da Intel tinham a missão de criar um conjunto

de 12 chips personalizados para uma companhia de calculadoras eletrônicas japonesa. Durante o desenvolvimento, descobriram que um chip simples, de propósito geral, seria capaz de realizar as mesmas ações por um preço mais baixo, dando surgimento ao processador de 4-bits 4004. (Harman e Gimblett, 2009).

Posteriormente, a Intel identificou o potencial do chip e, apesar das opiniões divididas sobre a relevância da descoberta, a empresa optou por comprar os direitos da fabricação do 4004 pelo mesmo preço cobrado pela realização da tarefa. Como resultado, desenvolveram o modelo 8008 em 1978, obtendo um grande sucesso ao introduzirem o primeiro microprocessador de 8 bits. Uma versão melhorada, o 8080, foi disponibilizada em 1974, e o seminal 8086 em 1978. (Harman e Gimblett, 2009).

Cada nova geração era mais sofisticada que suas antecessoras, e com isso, a arquitetura também evoluiu. O 8080 e outros processadores anteriores eram máquinas acumuladoras, ou seja, possuíam um único registrador utilizado como operando e destino para praticamente todas as operações aritméticas. Os 8086/88 tinham mais registradores mas não eram, em sua maior parte, de propósito geral. Esta última arquitetura foi chamada de arquitetura de acumuladores estendida. (Harman e Gimblett, 2009).

O 80386 foi lançado em 1985. Era uma extensão de 32-bits do 286, este já uma extensão de 24 bits do 8086. O 8086 introduziu novas instruções e um modelo de memória muito mais elaborado. As novas instruções reduziram o nível de dedicação às tarefas específicas dos registradores de dados da máquina, paralelamente à extensão destes para 32 bits. Neste momento, a arquitetura era similar a de uma máquina de registradores gerais — exceto pelos pontos flutuantes e outros poucos aspectos. (Harman e Gimblett, 2009).

Poucas mudanças foram executadas na arquitetura, com exceção às tecnologias MMX e SSE, porém a implementação mudou significativamente. É efetivamente a esta arquitetura que se faz referência quando se menciona a ‘Arquitetura Intel’, ‘x86’, ou precisamente a ‘IA32’. (Harman e Gimblett, 2009).

Para obter a mesma performance, um processador IA32 deve rodar mais rápido que um processador da série Power, SPARC ou Alpha. A decodificação, por exemplo, é muito mais difícil, sendo necessárias pipelines mais longas. Como consequência, tem-se um maior número de perdas em um desvio mal-previsto, levando a melhor (e mais cara) previsão de desvios. Economias de escala são suficientes para superar estas desvantagens, tendo como exemplo Intel e AMD, ambas utilizadoras desta arquitetura e com números de vendas maiores com relação aos processadores RISC. (Harman e Gimblett, 2009).

O primeiro x86 a receber pipelines foi o 80386. O 80486 tinha duas unidades funcionais e o Pentium foi o primeiro a utilizar uma arquitetura superescalar. ‘P6’ era o nome interno utilizado pela Intel para o sucessor do Pentium, aparecendo primeiro como Pentium Pro, em 1994/95. Seu diferencial consistia em uma embalagem de chip com duas cavidades e uma Cache de Nível 2 bem próxima ao processador, tornando-o rápido porém caro. Em 1997, o Pentium original teve a adição MMX — extensões multimídia que consistiam em um conjunto básico de operações vetoriais - e, mais tarde no mesmo ano, o Pentium II foi lançado. (Harman e Gimblett, 2009).

O Pentium II era um processador baseado no P6, como o Pentium Pro, mas com MMX adicionadas e maiores caches de instrução e dados Nível 1 (16KB vs 8KB cada).

Os caches maiores de Nível 1 sobrepunham-se ao fato de que a cache de Nível 2 próxima ao processador foi trocada por RAM estática comum fora do chip, o que era mais barato. (Harman e Gimblett, 2009).

A organização do P6 foi subsequentemente utilizada para o Pentium III, com as MMX sendo substituídas pela SSE (extensão de streaming SIMD). Os Pentium III tinham uma cache Nível 2 dentro ou fora (e maior) do chip, dependendo da versão. As várias versões de P6 são tipicamente conhecidas pelos seus nomes internos da Intel. Por exemplo, Katmai foi o primeiro Pentium III, que foi depois substituído pelo Coppermine. Esta é uma prática comum na indústria de microprocessadores; o mesmo aconteceu com o Pentium 3 e está acontecendo com a família Intel Core de processadores. (Harman e Gimblett, 2009).

Por volta dos anos 2000, a microarquitetura Intel NetBurst ganhou espaço ao formar a base dos Pentium 4, sendo substituída posteriormente pela arquitetura 'Intel Core' em 2006. (Harman e Gimblett, 2009).

Os processadores NetBurst rodavam mais quentes em geral, e portanto eram inviáveis para computação móvel em particular, como os laptops. Portanto, em paralelo com o NetBurst, a Intel reviveu a linha P6, fazendo um número de modificações para auxiliar o ambiente móvel - principalmente com o objetivo de melhorar a performance térmica e elétrica - nomeando-a Pentium M. Em 2006, a Intel lançou a microarquitetura Core, focada em múltiplos núcleos e virtualização de hardware como componentes chave, demonstrando melhores propriedades térmicas. (Harman e Gimblett, 2009).

Em 2007, um novo modelo de produção adotado pela Intel foi inaugurado, o Tick-tock. Sob este modelo, toda a mudança microarquitetural (tock) era seguida por uma diminuição de tamanho de componentes na tecnologia (tick). Sua substituta foi anunciada em 2016 e conhecida como "Processo, Arquitetura, e Otimização", e é similar ao ciclo tick-tock, com a adição de uma fase de otimização. (Intel, 2007).

Cada 'tick' representa um encolhimento da tecnologia de processamento da microarquitetura anterior, e cada 'tock' designava uma nova microarquitetura. Estas ocorreram em média entre cada 12 e 18 meses. Em 2016 a Intel anunciou em um relatório que estava deixando o ciclo tick-tock em favor de um processo de três passos sob o qual três gerações de processadores eram produzidos sob um único processo de manufatura, com a terceira geração focando em otimização. (Intel, 2007).

Neste relatório, exploraremos o ciclo tick-tock correspondente à geração Nehalem e Sandy Bridge.

2.1. Contexto Atual da Microarquitetura

A arquitetura Nehalem foi primeiro introduzida no mercado em Novembro de 2008, sendo a sucessora na família Core de microarquitecturas (Intel Core 2). Esta arquitetura pode ser encontrada tanto em plataformas móveis, com a série Clarkfield de processadores (como o i7-940XM para notebooks de alta performance), quanto em plataformas desktop, com a série Bloomfield (como o i7-965, o alvo principal desse relatório) e servidores, com a série Beckton (como o Xeon X7560, para servidores de alta performance). (Intel, 2008).

O alvo desta análise é o processador Intel Core i7-965 Extreme Edition, pertencente à microarquitetura Nehalem. Quando foi lançado, este processador de quatro núcleos era

um dos processadores mais rápidos do planeta, sendo destinado à computação entusiasta. (Intel, 2008).

A linha de processadores utilizando a microarquitetura Nehalem teve sua produção e utilização principal até o fim de 2010, com a microarquitetura Sandy Bridge substituindo-a em Janeiro de 2011. (Torres, 2011). Desde então, a produção dos processadores de microarquitetura Nehalem foi descontinuada e sua utilização e aquisição nos dias de hoje somente pode ser feita através de fornecedores com estoque e componentes da época. No entanto, devido à revolucionária introdução desta arquitetura no mercado, ainda hoje alguns desktops podem ser encontrados rodando estes chips com velocidades de até 5GHZ (overclocked), mostrando a robustez desta arquitetura.

2.2. Utilização da Microarquitetura

Esta arquitetura surgiu após o fracasso da Intel com a arquitetura NetBurst, onde a empresa procurou desenvolver a microarquitetura Core tomando de suma importância a escalabilidade ponta a ponta de servidores multi-socket a notebooks ultra-portáteis. A tarefa, com grande nível de dificuldade, procurou balancear economias de consumo de eletricidade, um dos mais importantes fatores em aplicações móveis com bateria limitada para servidores altamente escaláveis de alta performance. (Swinburne, 2008).

É neste âmbito que o desenho modular da Intel entra em cena: ainda que o Nehalem seja um chip maior e mais complexo, internamente este é dividido em partes que podem na teoria serem combinadas como blocos de construção. Ao reter este elemento de flexibilidade na produção, a empresa pode satisfazer novos mercados mais rapidamente e alcançar a competição mais cedo, ambos pontos bons para os consumidores. (Swinburne, 2008).

2.3. Detalhes da Microarquitetura

O pipeline de Nehalem, ilustrado na Figura 1, foi projetado de maneira a maximizar o fluxo de macro instruções por meio de múltiplas UFs. Ele segue o modelo de pipeline de micro-arquitetura de quatro larguras com comprimento total de 16 ciclos, sendo medido pelo atraso de erros de previsão de ramificação. (Thomadakis, 2011).

Os núcleos da Nehalem são modernos micro-processadores com emissão de instruções em ordem e caminhos de dados de execução superescalares fora de ordem, acoplados a uma hierarquia de armazenamento multinível (Thomadakis, 2011). Os núcleos da Nehalem têm amplo suporte para previsão de ramificação, execução de instruções especulativas, pré-busca de dados e várias UFs (Uncore Frequency Scaling) em pipeline. Seu pipeline foi projetado de maneira a maximizar o fluxo de macro instruções utilizando múltiplas UFs. (Thomadakis, 2011). A Figura 1 ilustra uma visão a nível funcional da pipeline de instruções do Nehalem. O comprimento total do pipeline, medido pela latência do erro de previsão de branch, são 16 ciclos, que são dois ciclos a mais que seu predecessor (Thomadakis, 2011). Referente a Figura 1, o núcleo consiste de:

- uma Pipeline Front-End (FEP) em ordem que recolhe instruções Intel64 da memória, utiliza quatro decodificadores para decodificá-las em micro-operações e armazená-las em um buffer para os estágios seguintes;
- um Motor de Execução (EE) superescalar fora de ordem que pode dinamicamente escalar e enviar até seis micro-operações por ciclo para as unidades de execução, assim que os operandos e recursos fonte estiverem prontos;

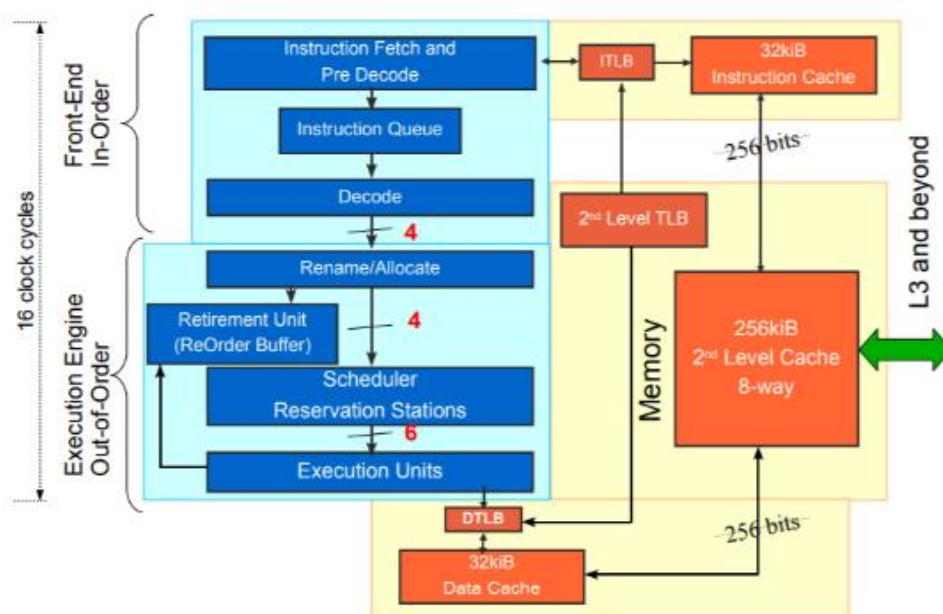


Figura 1. Diagrama de visão geral do nível funcional de um pipeline de instruções da Nehalem. Fonte: The Architecture of the Nehalem Processor and Nehalem-EP SMP Platforms, 2011

- uma Unidade de Descanso (RU) em ordem que garante que os resultados da execução de micro-operações são processadas e o estado “arquitetado“ é atualizado de acordo com a ordem original do programa, e;
- uma hierarquia de cache multinível e recursos de tradução de endereços;

Cada núcleo contém dois contextos de hardware e um pipeline 4-wide superescalar fora de ordem. A arquitetura Nehalem-EX se estende até oito núcleos. A descrição da arquitetura distingue entre os recursos que são totalmente replicados e privados para cada núcleo - estado do registro, a pilha de retorno e o ITLB (do inglês Instruction Translation Look-aside Buffer) de página grande - e recursos que são particionados quando duas threads estão ativas, completamente compartilhados e completamente inconscientes da threading. Embora muitos recursos ainda estejam particionados nessa arquitetura, ela ainda é muito mais flexível que as anteriores, por apresentar estações de reserva totalmente compartilhadas. Além disso, os designers permitem que a maioria dos recursos particionados fique totalmente disponível para uma única thread, mesmo no modo SMT (Tecnologia de Montagem em Superfície) quando outra thread estiver inativa. (Nemirovsky e Tullsen, 2013).

As instruções do SIMD (do inglês Single Instruction, Multiple Data, relacionado ao método de operação de computadores com várias unidades operacionais em computação paralela) usam conjuntos de registradores principais separados chamados registradores MMX e XMM. Estes têm tamanho de 64 bits e são relacionados à pilha de operandos para obter as instruções x87 herdadas. Os registros XMM têm tamanho de 128 bits e cada um pode armazenar 4 operandos de ponto flutuante de precisão simples (SP) ou 2 de precisão dupla (DP). As unidades de carregamento e armazenamento podem recuperar e salvar operandos de 128 bits do cache ou da memória principal. O subsistema de memória da Nehalem permite que áreas da memória possam ser usadas como buffers para

dados vetoriais entrando e saindo do núcleo, dispensando o armazenamento temporário em cache e permitindo com eficiência a recuperação do fluxo de operandos de sub-vetores da memória para registros XMM, a execução da computação SIMD e a transmissão dos resultados diretamente a memória. (Thomadakis, 2011).

A busca na cache e carregamento dos blocos de instruções alinhadas é realizado pela Unidade de Busca de instruções, que realiza a busca de instruções especulativa antecipando o carregamento de instruções sem precisar previamente definir o caminho do programa. (Thomadakis, 2011).

Esta unidade apoia-se na Unidade de Previsão de desvios - que busca antecipar com precisão desvios na execução do programa antes das chamadas ou do resultado de uma instrução serem conhecidos. Isso permite que a busca especulativa tenha como resultado instruções no caminho correto do programa em questão. (Thomadakis, 2011).

As instruções são identificadas e separadas individualmente do bloco e colocadas em um Buffer de Fila de Instruções, que possui entradas cujas instruções esperam em sua ordem inicial de leitura para serem recebidas pela Unidade de Decodificação, equipada com decodificadores capazes de enviar para o Buffer de Fila de μ ops até 4μ ops em cada ciclo de clock. (Thomadakis, 2011).

As μ ops depois de ordenadas podem ser executadas. Porém, para driblar a espera de todos os elementos necessários à execução de cada uma na sua sequência natural (pipeline com 14 estágios), a arquitetura Nehalem prevê a execução de μ ops fora de ordem, conseguindo assim preencher vários estágios do pipeline com μ ops sem que estes precisem esperar aqueles já em execução, preparando os dados necessários numa fase mais adiantada. (Thomadakis, 2011).

A Nehalem tem como principal característica sua evolução com relação à paralelização de execução de instruções - como foi demonstrado previamente - dentro de cada núcleo, devido à Unidade de Previsão de Desvios, sua maior profundidade de buffers de retenção de μ ops nas fases de busca e execução fora de ordem, sua execução de μ ops em paralelo a partir da criação de diversas unidades funcionais, e sua consistência de todas as μ ops executadas fora de ordem através da lógica dos buffers. (Thomadakis, 2011).

Duas características de evolução com relação às microarquiteturas anteriores que merecem destaque são:

- Acesso direto à Memória Principal incorporada ao Controlador de Memória no chip da CPU;
- Ausência de limite no número de páginas, permitindo a descrição pormenorizada de implementações, experimentos e demais trabalhos;
- Aumento na capacidade da Memória Cache: três níveis, com os dois primeiros sendo exclusivos a cada núcleo e o terceiro atuando como um nível global para os 4 núcleos.

3. Microarquitetura Sandy Bridge

A Microarquitetura Sandy Bridge oferece como características inovadoras, dentre outras, um maior front-end e mecanismo de execução somado às melhorias na hierarquia de cache com caminho mais amplo. (Intel, 2011).

A existência de quatro decodificadores de instruções permite a decodificação por parte do processador de até quatro instruções por pulso de clock, sendo responsáveis por decodificar instruções em micro instruções RISC (μ ops), usadas internamente pelas unidades de execução, suportando tanto fusão de instruções quanto fusão de microinstruções. (Torres, 2011).

A arquitetura apresenta como novidade um cache de micro instruções decodificadas capaz de armazenar 1.536 microinstruções (aproximadamente 6 kB), chamada pela Intel de “cache L0”, fazendo com que as instruções estejam decodificadas no cache quando o programa precisa repetir a execução de um grupo de instruções várias vezes. Desta forma, evita-se a decodificação repetida das instruções pelo processador, aumentando o desempenho e economizando tempo. Este cache apresenta uma taxa de acerto de 80%, ou seja, o cache é usado pelo menos oitenta por cento do tempo. (Torres, 2011).

Na Sandy Bridge, a unidade de previsão de desvio foi redesenhada e o tamanho do buffer de destino de desvio (Branch Target Buffer ou BTB) dobrou em relação à Nehalem, além de utilizar uma técnica de compressão que permite um maior armazenamento de dados. A previsão de desvios - um circuito que carrega para dentro do processador as instruções possivelmente carregadas pelo processador - tenta adivinhar os próximos passos do programa. Caso acerte, economiza o tempo de carregamento das instruções na memória por parte do processador, visto que elas já estarão nele. O aumento do tamanho desse buffer viabiliza o carregamento antecipado de um maior número de instruções, aumentando o desempenho do processador. (Torres, 2011).

O agendador usado na microarquitetura Sandy Bridge é similar ao usado na microarquitetura Nehalem, com seis portas de despacho, três delas usadas pelas unidades de execução e três usadas pelas unidades de memória. Embora a configuração seja a mesma, a microarquitetura Sandy Bridge apresenta um maior número de unidades de execução: 15, em comparação às 12 presentes na Nehalem. Segundo a Intel, elas foram redesenhadas visando a um aumento de desempenho de operações envolvendo números de ponto flutuante. (Torres, 2011).

Diferente do que acontecia na Nehalem, após a execução de uma instrução o processador apenas indica que a instrução terminou de ser executada em uma lista, sem copiá-la de volta para o buffer de reordenamento, economizando bits e aumentando a eficiência do processador. (Torres, 2011).

Com relação às portas de memória, a microarquitetura Nehalem apresenta: uma unidade de carga, uma unidade de armazenamento de endereços e uma unidade de armazenamento de dados, conectadas a uma porta de despacho individual, podendo carregar do cache L1 de dados 128 bits de dados por pulso de clock. Na microarquitetura Sandy Bridge, as unidades de carga e armazenamento de endereços podem ser usadas como uma unidade de carga ou como uma unidade de armazenamento de endereços. Isso permite o carregamento do dobro de dados do cache L1 por pulso de clock (utilizando duas unidades de 128 bits ao mesmo tempo ao invés de vez de apenas uma). Esta melhoria faz com que processadores Sandy Bridge possam carregar 256 bits de dados do cache L1 por pulso de clock, garantindo um aumento no desempenho. (NASA, 2019).

Os processadores Sandy Bridge viabilizam uma comunicação dos componentes internos do processador, através de uma arquitetura em anel. Para realizar tal comunicação,

um componente coloca a informação no anel, fazendo com que chegue até o destinatário (comunicação indireta através do anel). Os componentes que utilizam o anel são os núcleos de processamento, cada cache de memória L3 (Cache de Último Nível, não é unificado), controlador de memória integrado, controlador do barramento PCI Express, unidade de controle de alimentação e vídeo 2D e processador de vídeo 3D. (Torres, 2011).

4. Comparação de Desempenho

4.1. Desempenho da microarquitetura Nehalem

Com relação ao Intel i7-965 Extreme Edition, este processador é fabricado com uma tecnologia de 45 nanômetros - adotada pela Intel no fim de 2007 - e possui quatro núcleos com 64KB de memória cache L1 (sendo metade desta memória destinada à informação, e metade destinadas à instruções), 256KB de memória cache L2 e 8MB de memória cache L3, que é dividida pelos quatro núcleos do processador. Também utiliza de um canal QPI (QuickPath Interconnect) de 3200MHz, uma tecnologia de conectores ponta-a-ponta que permite que os processadores sejam ligados através da implementação de uma tecnologia de acesso à memória de maneira não uniforme (NUMA). (Intel, 2008).

Nas especificações voltadas ao desempenho, a arquitetura possui uma frequência baseada em processador de 3.20GHz, frequência turbo max de 3.46GHz, velocidade de barramento de 6.4 GigaTransfers por segundo, além de uma potência de design térmico (TDP) de 130W e um intervalo de voltagem VID entre 0.800V e 1.375V. Por fim, na arquitetura Nehalem cada núcleo do processador consegue enviar 4 pontos flutuantes de precisão dupla com a performance máxima de 44.8Gflop/s por chip. Sua capacidade de processar oito threads simultaneamente (duas por núcleo) gera grande melhoria na performance, por ser possível esconder a latência da memória ao mudar a thread ativa. (Barker et al., 2008).

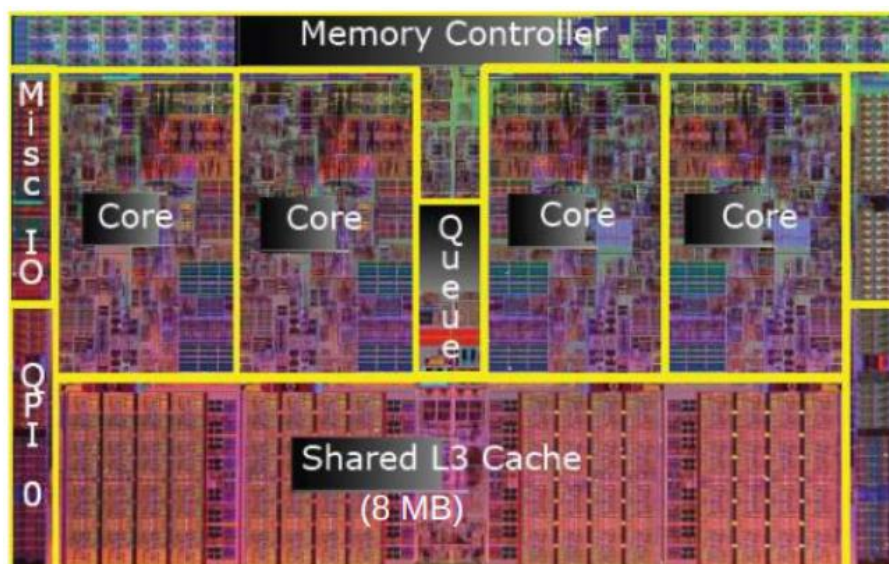


Figura 2. Estrutura do processador Nehalem. Fonte: Intel Core X-series (HED lines), Dezsó Sima, 2018.

A Figura 2 ilustra um chip Nehalem CMP (Chip Multi Processador) e suas partes principais, consistidas dos seguintes componentes:

- quatro núcleos de computação idênticos;
- UIU: Unidade de Interface Un-Core (um switch conectando os quatro núcleos aos quatro segmentos de cache L3, às portas IMC e às portas QPI);
- L3: controlador de cache nível 3 e bloco de dados de memória;
- IMC: Um controlador de memória integrado com três canais de memória DD3;
- QPI: Duas portas Quick-Path Interconnect, e;
- circuito auxiliar para coerência de cache, controle de força, gerenciamento de sistema e lógica de monitoramento de performance;

O processador Nehalem apresenta uma escala dinâmica de tensão e frequência (em inglês *Dynamic voltage and frequency scaling*, ou DVFS) agressiva que mede iterativamente tensão, consumo de energia, temperatura e outros fatores ambientais. Essas medições são usadas para ajustar a tensão e a frequência dos núcleos, usando o espaço térmico e de energia disponível para aumentar a frequência e o desempenho de rosca única. Em conjunto, a Nehalem usa portas de energia baseadas em PFET para eliminar vazamentos para núcleos ociosos, liberando assim mais espaço para aumentos de frequência. (Kanter, 2010).

4.2. Desempenho da Microarquitetura Sandy Bridge

A GPU Sandy Bridge compartilha o cache L3 (em inglês *Last Level Cache* ou LLC) — observado na Figura 3 abaixo — com as unidades de execução da CPU. Com a integração de gráficos e CPUs em um único chip, a PCU pode gerenciar com flexibilidade a energia e o recurso térmico entre os dois componentes com muito mais precisão e menor latência. O Sandy Bridge pode fornecer desempenho superior para a maioria dos aplicativos, compartilhando recursos térmicos e de energia entre os gráficos e as CPUs, em vez de alocar estaticamente um orçamento de energia e térmico. (Kanter, 2010).

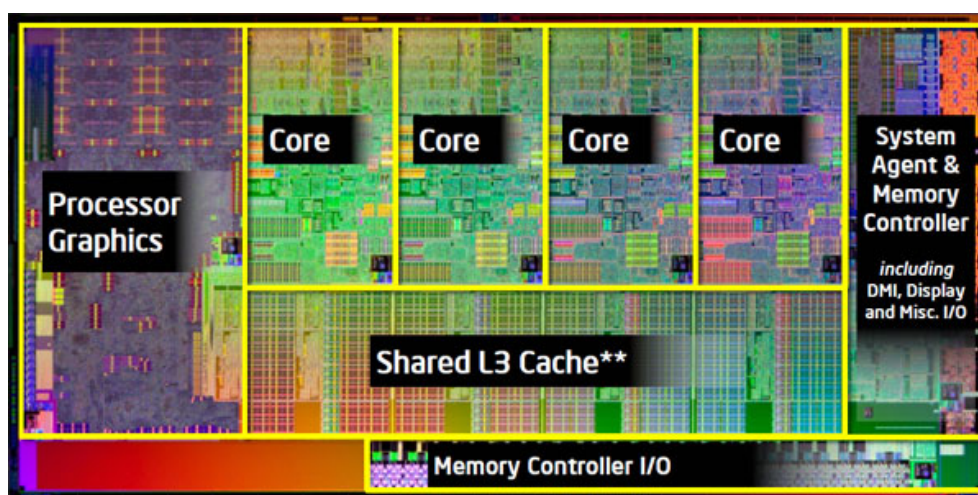


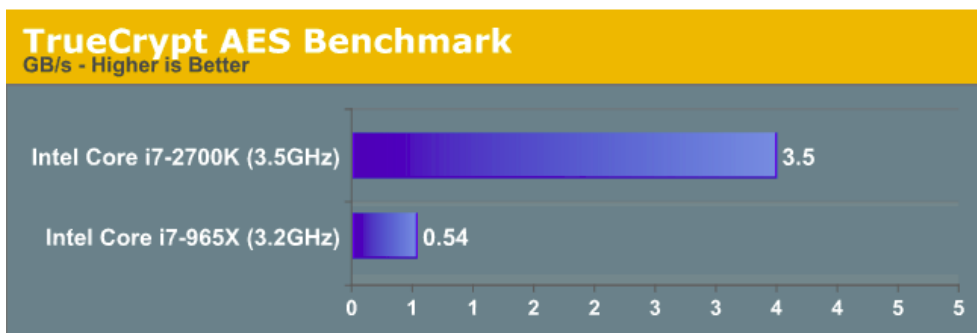
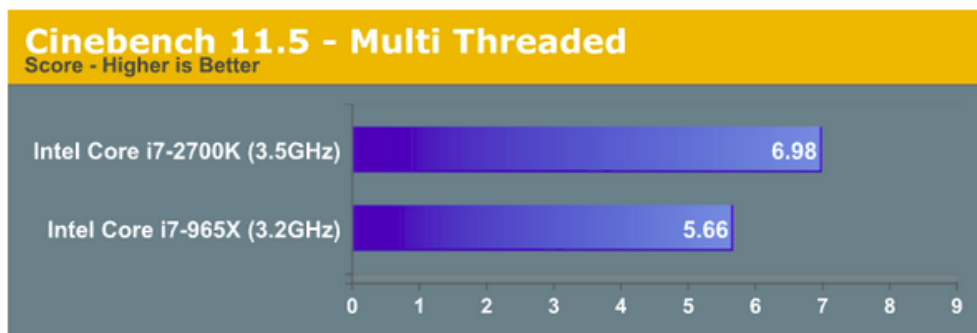
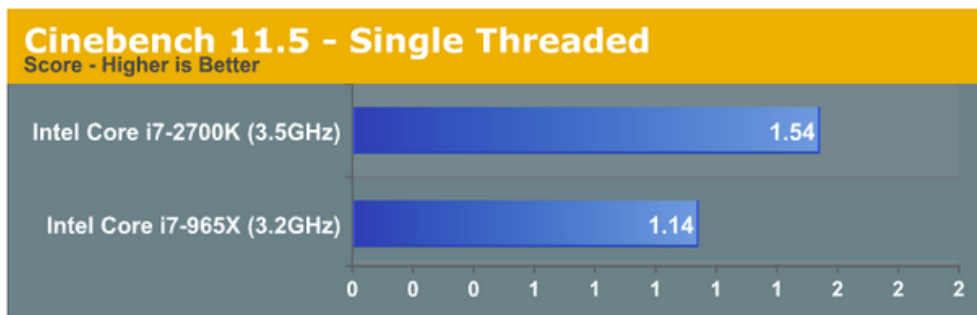
Figura 3. Estrutura do Processador Sandy Bridge. Fonte: Intel Core X-series (HED lines), Dezső Sima, 2018.

Uma das partes mais importantes do agente do sistema é a Unidade de Controle de Energia (PCU), um microcontrolador patchable responsável pelo gerenciamento térmico e de energia em todo o chip. A PCU Sandy Bridge utiliza um modelo mais sofisticado do chip que leva em consideração a capacitância térmica dinâmica, em vez de assumir

uma mudança instantânea de temperatura, como faz a PCU do Nehalem). Durante esse período de tempo em que a temperatura do dissipador de calor está baixa e crescente, ele absorverá o calor mais rapidamente do chip. O Sandy Bridge consegue aproveitar a transferência de calor extra para dissipar com segurança mais energia do que o limite TDP sustentado. (Kanter, 2010).

Em relação às questões técnicas do processador Intel I7 2700k extreme edition, a arquitetura Sandy Bridge teve algumas alterações comparada à Nehalem. Fabricada com uma tecnologia de 32 nanômetros, teve seu número de núcleos mantido em quatro ainda mantendo a relação de 2 threads por núcleo, deixando-a com um total de oito threads. Sua frequência baseada em processadores aumentou em 300MHz, e sua frequência turbo max em 440MHz. Assim como seu antecessor, também possui um cache L3 de 8MB compartilhado entre todos seus seis núcleos. Apresenta uma velocidade de barramento de 5GT/s e um TDP de 95W. (Intel, 2011).

Considerando o desempenho, a arquitetura Sandy Bridge (representada pela linha Intel I7 2700k), supera a Nehalem em todos os benchmarks testados, como pode ser observado nas Figuras 4 a 7.



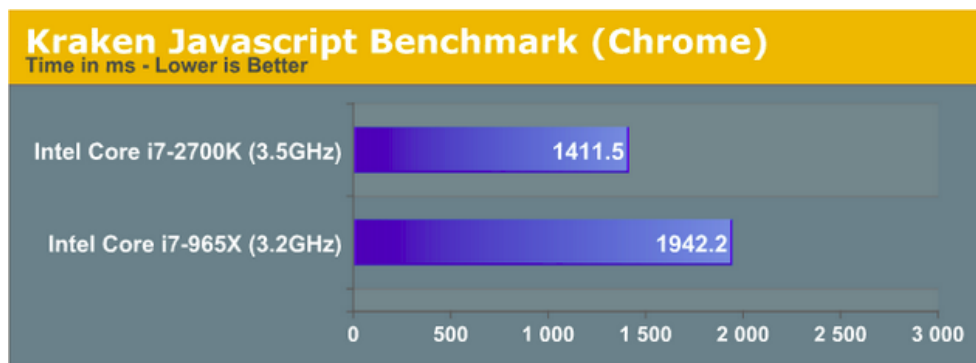


Figura 4. a 7: Testes de comparação dos processadores. The Haswell Review: Intel Core i7-4770K I& i5-4670K Tested - Anand Lal Shimpi, 2013

O Cinebench — um conjunto de testes cross-platform que mede o desempenho de recursos em relação ao software de animação Cinema 4D — realiza testes de uso comum do computador, como renderização, reprodução de vídeo e efeitos em 3D. Este benchmark foi utilizado para comparar o desempenho em cenários Single Threaded e Multi Threaded. No terceiro teste, o benchmark TrueCrypt AES utiliza um software que viabiliza a criptografia em tempo real e se relaciona ao padrão de criptografia avançada (em inglês Advanced Encryption Standard, ou AES). Considerando os benchmarks Cinebench 11.5 e TrueCrypt AES, quanto maior o número, melhor o resultado. Desta forma, podemos observar que o processador Intel Core i7-2700K (3.5GHz) apresenta melhores resultados em comparação ao Intel Core i7-965X (3.2GHz) em ambos os testes.

O Kraken Javascript benchmark (Chrome) mede a velocidade de vários casos de teste diferentes extraídos de aplicativos e bibliotecas do mundo real. Desta forma, a menor velocidade em milissegundos apresentada pelo Intel Core i7-2700K (3.5GHz) indica que ele é mais rápido.

5. Conjunto de Instruções e sua Arquitetura

Antes de tudo, é conveniente definir o que é uma instrução de máquina: consiste basicamente na operação mais elementar que um computador pode realizar. A estrutura de uma instrução e o seu tamanho variam de acordo com a arquitetura, podendo ir de 1 a 8 bytes. O formato segue um padrão de n bits para o código operacional, que representa a ação a ser tomada pelo processador, e x bits para o código do operando, que representa os parâmetros da operação. Os parâmetros da operação dependem de qual operação está sendo executada. Toda operação é executada por meio de unidades básicas de memória chamadas de registradores, que armazenam os dados a serem tratados nas instruções. (Pillou, 2020).

O código operacional (opcode) de uma instrução varia de tamanho dependendo da máquina, sendo que cada opcode tem uma sequência única de bits entendida pela máquina. Já as classes de operações não variam de máquina para máquina. São elas: Transferência de dados; Aritméticas; Lógicas; Conversão; E/S; Controle de sistema; Transferência de controle. O tamanho das instruções é definido pelo tipo de arquitetura: se for x86, as instruções deverão somente ter até 32 bits, o que limita a quantidade de memória possível de ser acessada a 3 GB dado que o número de bits para alocar o endereço de memória é

mais limitado. Se for x64, cada instrução poderá ter até 64 bits, sendo que a capacidade de endereçamento de memória aumenta e o computador pode ter mais do que 3 GB de memória RAM. (Pillou, 2020).

Definida o que é uma instrução de máquina, é possível introduzir o conceito de conjunto de instruções, para que posteriormente seja possível apresentar os conjuntos de instruções dos dois processadores analisados. Um conjunto de instruções nada mais é que o conjunto de operações básicas que um processador consegue realizar. O conjunto de instruções de um processador define qual a arquitetura trabalhada nesse computador. O conjunto de instruções é definido ao nível do circuito no processador, sendo que cada processador carrega em seu hardware as possíveis instruções a serem realizadas, definindo assim as funções do computador. Existem dois tipos de conjuntos de instruções: RISC e CISC. (Pillou, 2020).

- **Conjunto de instruções CISC:** Complex Instruction Set Computer - Computador com conjunto complexo de instruções é um conjunto de instruções que adiciona não somente as instruções (opcodes) mais simples, mas também instruções complexas no hardware do processador. Esse recurso tem suas vantagens e desvantagens: é vantajoso pois simplifica o desenvolvimento, dado que o desenvolvedor (compilador/montador) não precisará se preocupar em transformar certas instruções mais complexas em instruções básicas de máquina. Por outro lado, esse tipo de construção torna o processador mais caro, além de poder tornar a execução mais lenta, dado que uma instrução complexa pode levar mais de um ciclo para ser executada. Apesar da execução ser mais lenta devido à complexidade da instrução, o desempenho final pode ser mais rápido, pois o número de acessos à memória tende a diminuir devido ao formato complexo das instruções. (Pillou, 2020).
- **Conjunto de instruções RISC:** Reduced Instruction Set Computer - Computador com conjunto reduzido de instruções define que o conjunto de instruções (opcodes) básicos impressos em circuito sejam somente as instruções mais básicas, não registrando qualquer instrução complexa diretamente no circuito. Esse tipo de construção tem como vantagem um baixo custo de desenvolvimento e uma velocidade mais alta na execução do programa, dado que uma instrução simples sempre será executada em até um ciclo de clock. Por outro lado, RISC torna mais árdua a tarefa do desenvolvedor, pois todo tipo de instrução complexa deve ser transformada em um conjunto de instruções básicas. (Pillou, 2020).

5.1. Conjunto de instruções Nehalem

O conjunto de instruções da microarquitetura Nehalem IA-64 - resultado da implementação realizada no processador Intel Core i7-965 Extreme Edition debatido anteriormente - tem um formato de macro instruções CISC Intel64 (64 bits). Isso significa que as instruções exigem uma UCP mais cara e complexa de se construir, que possa entender as macro instruções e transformá-las em μ ops (micro ops) no formato RISC. Esse processo é feito dentro do pipeline do processador, que contém 14 estágios. (Thomadakis, 2011).

O fluxo de processamento de instruções e dados neste processador pode ser resumido da seguinte maneira: Uma série de macro instruções são buscadas, podendo ser um bloco de cache; Estes blocos são pré-decodificados nas diferentes instruções que o compõem;

Cada uma destas instruções é decodificada em micro ops, colocando cada uma delas em buffers próprios para a unidade funcional para serem tratadas; O despacho para a UF (Uncore Frequency Scaling) disponível é realizado a partir do momento em que os operandos de que necessitam estejam disponíveis; As micro ops completas são retiradas e o resultado é armazenado. (Thomadakis, 2011).

Como o pipeline tem muitos estágios para realizar o processo completo, o processador conta com execução fora de ordem de micro ops. Isso significa que, para que o pipeline não fique constantemente vazio, devido a questões relacionadas à necessidade de E/S, por exemplo, o processador executa instruções mais avançadas do programa que não tenham requisitos de instruções ainda não executadas. Dessa forma, os estágios do pipeline têm uma maior taxa de preenchimento. Utilizando todos esses recursos, o processador consegue executar 6 micro ops por ciclo de clock. (Thomadakis, 2011).

5.2. Comparação entre arquiteturas: Nehalem x Sandy Bridge

Fazendo uma comparação entre o conjunto de instruções do Nehalem com o do Sandy Bridge - arquitetura mais recente - é possível perceber semelhança e melhorias. Tomando como exemplo o processador Intel i7 Sandy Bridge x64 (Core I7-2700k Extreme Edition), podemos notar que o formato das instruções ainda é CISC, seguindo o padrão intel64 de 64 bits. (Torres, 2011). Porém, algumas tarefas estão mais fáceis de serem executadas nesse novo processador, pois o conjunto de instruções AVX foi adicionado, mudando substancialmente a forma como as operações de ponto flutuante são executadas. (Torres, 2011).

O AVX (Advanced Vector Extensions), é um conjunto de instruções para executar operações SIMD (Single Instruction Multiple Data). Ela foi implementada em processadores intel a partir da arquitetura Sandy Bridge. Essas extensões aumentam o registro de vetores de 128 bits para 256 bits possibilitando que o hardware de ponto flutuante possa suportar 16 operações de precisão única e 8 de precisão dupla por ciclo. O resultado disso é que mesmo que a velocidade do clock da CPU do processador Sandy Bridge (2,6 GHz) seja menor que a dos tipos de processadores mais antigos, como alguns da arquitetura Nehalem, o desempenho do ponto flutuante é maior para alguns aplicativos. (Torres, 2011).

6. Conclusão

A microarquitetura Sandy Bridge é uma das arquiteturas de microprocessador mais ambiciosas e agressivas já desenvolvidas pela Intel. A combinação de uma nova microarquitetura de CPU com uma nova microarquitetura gráfica traz como resultado uma significativa melhora com relação à geração anterior. (Thomadakis, 2011). Somada a isso, a integração no nível do chip deu um grande passo à frente ao permitir a disponibilização de um agente de sistema mais complexo, uma nova cache de nível 3 e um anel interconector compartilhado por todos os componentes. Lidar com todas essas diferentes visões do Sandy Bridge em uma única discussão é impossível, dado o escopo das mudanças, tornando a Sandy Bridge uma microarquitetura fundamentalmente nova para a Intel. Ainda que externamente pareça com a Nehalem e a P6, internamente é muito diferente, mostrando a filosofia de desenho dos engenheiros da Intel ao propor esta nova arquitetura.

Referências

- Barker, K. J., Davis, K., Hoisie, A., Kerbyson, D. J., Lang, M., Pakin, S., e Sancho, J. C. (2008). A performance evaluation of the nehalem quad-core processor for scientific computing. *World Scientific*. <https://www.worldscientific.com/doi/abs/10.1142/S012962640800351X>.
- com Lógica, S. (2016). Intel 17 nehalem. <https://sabercomlogica.com/pt/caso-de-estudo-intel-i7-nehalem/>.
- Harman, N. e Gimblett, A. (2009). Cs-323: High performance microprocessors. Technical report, University of Wales Swansea. Chapter 11 Intel: IA32 to P6 & Intel Core.
- Intel (2007). Intel tick-tock model. <https://www.intel.com/content/www/us/en/silicon-innovations/intel-tick-tock-model-general.html>.
- Intel (2008). *Processador Intel® Core™ i7-965 Extreme Edition*. <https://ark.intel.com/content/www/br/pt/ark/products/37149/intel-core-i7-965-processor-extreme-edition-8m-cache-3-20-ghz-6-40-gt-s-intel-qpi.html>.
- Intel (2011). *Processador Intel® Core™ i7-2700K*. <https://ark.intel.com/content/www/br/pt/ark/products/61275/intel-core-i7-2700k-processor-8m-cache-up-to-3-90-ghz.html>.
- Intel (2012). *Intel® Core™ i7-3820 Processor*. <https://ark.intel.com/content/www/us/en/ark/products/63698/intel-core-i7-3820-processor-10m-cache-up-to-3-80-ghz.html>.
- Kanter, D. (2010). Intel's sandy bridge microarchitecture. <https://www.realworldtech.com/sandy-bridge/9/>.
- Martins, L. G. A. (n.d). Organização de computadores 1. 5 – conjunto de instruções. <http://www.facom.ufu.br/gustavo/OC1/Apresentacoes/Conjunto%20de%20Instrucoes.pdf>.
- NASA (2019). Sandy bridge processors. https://www.nas.nasa.gov/hecc/support/kb/sandy-bridge-processors_301.html.
- Nemirovsky, M. e Tullsen, D. M. (2013). Multithreading architecture. *Morgan & Claypool Publishers*. Chapter 9.16.
- Pillou, J.-F. (2020). Processador. <https://br.ccm.net/contents/400-processador>.
- Shimpi, A. L. (2013). The haswell review: Intel core i7-4770k & i5-4670k tested. <https://www.anandtech.com/show/7003/the-haswell-review-intel-core-i74770k-i54560k-tested/6>.
- Sima, D. (2018). Intel core x-series (hed lines). https://users.nik.uni-obuda.hu/sima/letoltes/Processor_families_Knowledge_Base_2019/Intels_E-lines_2019_02_15.pdf.
- Swinburne, R. (2008). Intel core i7 - nehalem architecture dive. <https://www.bit-tech.net/reviews/tech/cpus/intel-core-i7-nehalem-architecture-dive/1/>.
- Thomadakis, M. E. (2011). The architecture of the nehalem processor and nehalem-ep smp platforms. <http://stellar.mit.edu/S/course/6/fa13/6.172/courseMaterial/topics/topic9/readings/nehalem/nehalem.pdf>.
- Torres, G. (2011). Por dentro da microarquitetura intel sandy bridge. <https://www.clubedohardware.com.br/artigos/processadores/por-dentro-da-microarquitetura-intel-sandy-bridge-r35529/?nbcpage=2>.