

AMD معماری توسط Intel Core i7

بررسی معماری Core i7

نوشتن مقاله‌ای مختصر در مورد معماری شبیه به Nehalem کار بسیار دشواری است. این معماری از مجموعه‌ای از بلوک‌های مشابه با اسباب بازی لگو تشکیل شده که مهندسان اینتل با در کنار قرار دادن آنها می‌توانند نسخه‌های متنوعی از این معماری را معرفی کنند. در حال حاضر نسخه‌های پیشرفته این معماری جدید، نظیر نسخه بسیار پیشرفته برای سرورها و نسخه‌هایی با کارایی بالا برای ایستگاه‌های کاری معرفی شده است. در نگاه اول به مشخصات این پردازنده‌ها بطور حتم به یاد معماری K10 (بارسلونا) شرکت "ای‌ام‌دی" خواهید افتاد. پردازنده‌های چهار هسته‌ای مبتنی بر این معماری دارای سه سطح حافظه نهان، کنترل کننده حافظه مجتمع شده در پردازنده با پشتیبانی از حافظه‌های DDR3، یک اتصال نقطه به نقطه (Point-to-Point) برای ارتباط با وسایل جانبی و دیگر پردازنده‌ها در پیکره‌بندی چند پردازنده ای هستند. استفاده از گذرگاهی شبیه به Hyper Transport ثابت کرد که اینتل هر زمان متوجه شود که فناوری‌های رقبایش عملکرد بهتری دارد قطعاً از آنها در محصولاتش استفاده خواهد کرد! اما معماری جدید اینتل تغییرات عمده‌ای نسبت به معماری رقیبش داشته است. با تعبیه بیش از ۷۰۰ میلیون ترانزیستور (بطور دقیق ۷۳۱ میلیون) مهندسان اینتل قادر بودند مشخصه‌های هسته‌های اجرایی را بطور چشمگیری بهبود بخشند در حالیکه عملکردهای جدیدی نظیر چند پردازش همزمان (Simultaneous Multi-Threading) که پیش از این در پردازنده‌های پنتیوم چهار مبتنی بر "Northwood" تحت نام Hyper-Threading شناخته می‌شد، به این هسته‌ها اضافه شده است. برای مثال پردازنده‌های چهار هسته‌ای Nehalem قادرند حداکثر شانزده رشته را بطور همزمان اجرا کنند. این موضوع سبب فشار زیادی روی سطوح مختلف خط لوله‌ها می‌شود که در نتیجه اینتل را مجبور به تغییر در حجم بسیاری از بافرها کرده است. مانند سال‌های گذشته اینتل به معماری جدید خود دستورالعمل‌های SSE جدیدی نیز اضافه کرده است. این معماری از SSE 4.2 پشتیبانی می‌کند که مشخصات آن از ریز معماری K10 شرکت AMD اقتباس شده است. همانطور که گفته شد پردازنده‌های چهار هسته‌ای Nehalem شامل ۷۳۱ میلیون ترانزیستور هستند که مقدار آنها نسبت به هسته Yorkfield، هسته رایج در پردازنده‌های چهار هسته‌ای مبتنی بر ریزمعماری Penryn با ۸۲۰ میلیون ترانزیستور، کاهش پیدا کرده است. اما اندازه سطح هسته از ۲۱۴ میلی‌متر مربع به ۲۶۳ میلی‌متر مربع افزایش یافته است (ترانزیستورهای کمتر اما ابعاد هسته بزرگ‌تر!!!). کاهش ترانزیستورها ناشی از کاهش اندازه حافظه نهان و افزایش ابعاد هسته ناشی از تغییرات بنیادی در

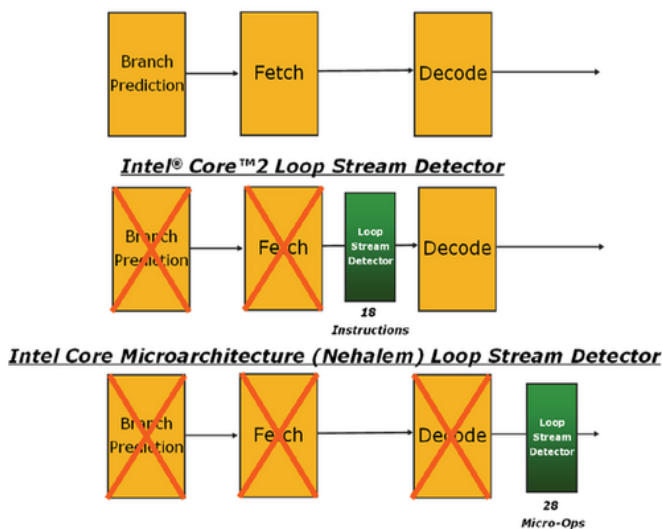
زیرمعماری پردازنده است. در پردازنده‌های چهار هسته‌ای مبتنی بر این معماری هر هسته شامل یک حافظه نهان سطح یک و دو مستقل و تمام چهار هسته شامل یک حافظه سطح سه بزرگ با ظرفیت هشت مگابایت هستند. حافظه نهان سطح یک نسبت به هسته Penryn (پردازنده‌های ۴۵ نانومتری رایج) هیچ‌گونه تغییری نداشته اگرچه اندکی کندتر (چهار سیکل در عوض سه سیکل) شده است. حافظه نهان سطح دو با وجود آن که سریع‌تر شده اما اندازه آن به ۲۵۶ کیلوبایت برای هر هسته کاهش پیدا کرده است. پردازنده‌های دو هسته‌ای Penryn دارای یک حافظه نهان اشتراکی سطح دو با ظرفیت حداقل ۳ مگابایت هستند. حافظه نهان سطح سه خصوصیت جدیدی است که بین تمام چهار هسته به اشتراک گذاشته شده و به برنامه‌های چند وظیفه‌ای (Multi threaded) کمک می‌کند تا روان‌تر اجرا شوند (همانند آنچه در سرورها با آن مواجه هستید). اکنون که بطور خلاصه با معماری جدید اینتل آشنا شدید زمان آن رسیده که با جزئیات بیشتر این معماری آشنا شوید. کار را با خط لوله آغاز می‌کنیم که وظیفه آن خواندن دستورالعمل‌ها از حافظه و آماده کردن آنها برای اجرا است.

خواندن و رمزگشایی دستورالعمل‌ها

برخلاف تغییرات ایجاد شده در حرکت از Core 2 به Core Nehalem اعمال نکرده است. این بخش دارای چهار رمزگشای یکسان با Conroe است که شامل سه رمزگشایی ساده و یک رمزگشایی پیچیده می‌شود. یکی از وظایف این رمزگشاهای تبدیل دستورالعمل‌های x86 به دستورالعمل‌های با طول ثابت است که Micro-op نامیده می‌شود. اینتل برای اولین بار در پردازنده‌های مبتنی بر ریزمعماری Core از نوآوری بنام "همجوشی Macro-op" استفاده کرد. "همجوشی Macro-op" این امکان را فراهم می‌کند که دو دستورالعمل وابسته x86 در عوض تبدیل به دو Micro-op به یک Micro-op تبدیل شوند. به عبارت ساده‌تر دو دستورالعمل وابسته را با یکدیگر ادغام کنند. Nehalem نیز از این نوآوری استفاده می‌کند و از نظر توان عملیاتی هنوز مشابه با گذشته است یعنی هر پنج دستورالعمل در هر سیکل را به چهار دستورالعمل تبدیل می‌کند. اگرچه تحولات انقلابی در بخش رمزگشاهای ایجاد نشده اما در ریز جزئیات آن تغییراتی اعمال شده است. همانطور که می‌دانید افزایش تعداد واحدهای پردازشی روش موثری برای بالابردن کارایی نیست زیرا تنها موجب افزایش قیمت پردازنده و در برخی مواقع نیز موجب افزایش اندک کارایی می‌شود. بنابراین مهندسين اينتل در عوض اضافه کردن یک کدگشای جدید تصمیم گرفتند تا کدگشاهای

رایج را موثرتر و کارآمدتر کنند. نوآوری "همجوشی Macro-op" در پردازنده‌های Conroe تنها در وضعیت ۳۲ بیت عمل می‌کرد. به عبارت ساده‌تر زمانی که پردازنده در وضعیت ۶۴ بیت قرار می‌گرفت (زمانی که از سیستم عامل ۶۴ بیتی استفاده می‌شد) این نوآوری بلا استفاده می‌شد. اما اکنون "همجوشی Macro-op" از وضعیت ۶۴ بیت نیز پشتیبانی می‌کنند. این ویژگی برای معماری شبیه به Nehalem کاملاً قانع کننده است زیرا که پردازنده‌های سرور آینده اینتل نیز از آن مشتق می‌شوند. اما همه تغییرات تنها به این موضوع خلاصه نمی‌شود. معماری Conroe توسط "همجوشی Macro-op" می‌تواند تعداد دستورالعمل‌های محدودی را با یکدیگر ترکیب کند درحالیکه معماری Nehalem از تعداد دستورالعمل‌های بیشتری پشتیبانی می‌کند بطوریکه استفاده از این نوآوری خیلی اوقات فراهم می‌شود. خصوصیت جدید دیگری که توسط Conroe معرفی شد و اکنون بهبود یافته است: واحد کشف جریان حلقه (Loop Stream Detector) است. واحد کشف جریان حلقه در واقع یک بافر داده است که تعداد اندکی از دستورالعمل‌ها را درون خودش نگه می‌دارد (پردازنده‌های Core 2 هجده دستورالعمل در خود نگه می‌داشتند). این ویژگی موجب می‌شود تا زمانیکه پردازنده یک حلقه را شناسایی می‌کند بخش‌های معینی از خط لوله را غیرفعال می‌کند. از آنجاییکه یک حلقه شامل اجرای دستورالعمل‌های یکسان در یک تعداد زمان مشخص است، لازم نیست که واحد پشتیبانی انشعاب برای دریافت دستورالعمل‌ها از حافظه نهان سطح اول در هر تکرار حلقه کار کند. بنابراین واحد کشف جریان حلقه بعنوان یک حافظه نهان کوچک عمل می‌کند که اولین مرحله از خط لوله را در چنین وضعیتی اتصال کوتاه می‌کند.

این شیوه دو مزیت دارد: کاهش مصرف توان بوسیله غیرفعال کردن واحدهای بلا استفاده و بهبود کارایی بوسیله کاهش فشار روی حافظه نهان دستورالعمل سطح یک. در معماری Nehalem اینتل عملکرد واحد کشف جریان حلقه را بهبود بخشیده است. اولاً این بافر بزرگتر شده و اکنون می‌تواند بیست و هشت Micro-op را در خود ذخیره کند و دوم اینکه موقعیت آن در خط لوله تغییر کرده



است. در Conroe این واحد بعد از مرحله واکشی دستورالعمل واقع شده بود. اما اکنون این واحد

بعد از رمزگشاها واقع شده است. این موقعیت جدید اجازه می‌دهد تا بخش وسیعی از خط لوله غیرفعال شود.

سوم اینکه واحد کشف جریان حلقه در Nehalem با توجه به اینکه بعد از رمزگشاها واقع شده دیگر تنها دستورالعمل‌های x86 بیشتری را ذخیره نمی‌کند بلکه در حقیقت این واحد Micro-opها را ذخیره می‌کند. از آنجاییکه ممکن است توسط "همجوشی Macro-op" دو دستورالعمل به یک Micro-op ترجمه شود بنابراین این واحد افزایش کارایی چشمگیری خواهد داشت. واحد کشف جریان حلقه تقریباً مشابه با حافظه نهان Trace در Pentium 4 است. شاید این مسئله جای تعجب نداشته باشد زیرا تیم Hillsboro که اکنون متصدی پروژه Nehalem است مسئولیت پروژه Pentium 4 را نیز به عهده داشته است.

توضیحات تصویر : واحد کشف جریان حلقه

پیش بینی انشعاب

آخرین پیشرفت در بخش جلویی خط لوله‌های Nehalem، بهبود پیش‌بینی انشعاب است. کارایی الگوریتم‌های پیش‌بینی انشعاب در معماری‌های که نیاز به سطح بالایی از دستورالعمل دارند بسیار تعیین کننده و مهم است. پیش‌بینی انشعاب تعیین می‌کند که یک انشعاب پذیرفته شود یا خیر. اگر واحد پیش‌بینی انشعاب با یک انشعاب موافقت کند، به سرعت آدرس هدف را برای اجرای مداوم تعیین می‌کند. یک انشعاب زمانی شکست می‌خورد که مجبور شود هنگام اجرای یک دستورالعمل منتظر نتایج دستورالعمل‌های قبل از خود بماند. اگر چنین اتفاقی در خط لوله بیفتد کلیه دستورالعمل‌های جاری در خط لوله باید از ابتدا واکنشی شوند. این موضوع نیاز به صرف زمان دارد و سبب کاهش کارایی پردازنده می‌شود. شگردهای مورد نیاز برای انجام این کار پیچیده نیستند. این کار تنها نیازمند یک آرایش انشعاب است که "بافر هدف انشعاب" (BTB) نامیده می‌شود و نتایج انشعابات را بعنوان گردش اجرایی و یک الگوریتم برای تعیین نتایج انشعابات بعدی ذخیره می‌کند. اینتل هیچ‌گونه جزئیاتی در مورد الگوریتم‌های که در پیش‌بینی کننده‌های جدیدش بکار گرفته مشخص نکرده است. اما این موضوع مشخص است که این پیش‌بینی کننده‌های جدید اکنون دو سطح دارند. اولین سطح نسبت به معماری Conroe عوض نشده است اما سطح جدید با دسترسی آهسته‌تر می‌تواند تاریخچه پیش‌بینی بیشتری را ذخیره کند. مطابق با گفته‌های اینتل، این پیکره‌بندی موجب بهبود پیش‌بینی انشعاب برای برنامه‌های کاربردی مشخصی که از مقادیر وسیعی کد استفاده می‌کنند، نظیر پایگاه داده‌ها می‌شود. منظور اینتل از برنامه‌های کاربردی مشخص نظیر پایگاه داده‌ها، برنامه‌های است که بیشتر در سرورها استفاده می‌شود. بنابراین یکبار

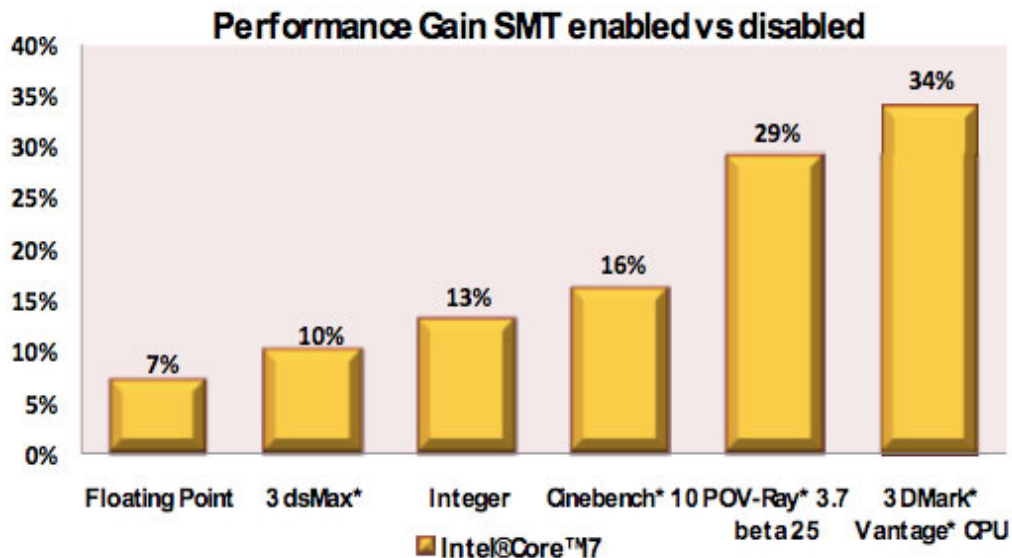
دیگر به این حقیقت می‌رسیم که Nehalem گرایش به سیستم‌های سرور دارد. یکی دیگر از پیشرفت‌ها Return Stack Buffer است که آدرس‌های برگشتی وظایف را زمانیکه آنها فراخوانده می‌شوند، ذخیره می‌کند.

بازگشت Hyper-Threading

همانطور که گفته شد بخش جلوی خط لوله‌ها تغییرات عمده‌ی نداشته‌است. این موضوع در مورد انتهای خط لوله‌ها نیز صادق است. این بخش واحدهای اجرایی مشابه‌ای با پردازنده‌های مبتنی بر معماری Core رایج دارد. اما در این بخش نیز مهندسی اینتل یک بار دیگر روی کارآمدتر شدن خط لوله‌ها کار کرده‌اند. واحدهای اجرایی پردازنده‌های مدرن امروزی تشنه قدرت هستند و فضای زیادی از هسته پردازنده را به خود اختصاص می‌دهند. سازندگان پردازنده‌ها سعی دارند تا این واحدها در اکثر مواقع تا حد امکان تغذیه شوند و کار کنند. آنها اندازه حافظه‌های نهان را افزایش می‌دهند تا پردازنده هرگز منتظر حافظه اصلی سیستم نماند، کنترلر حافظه را درون پردازنده ادغام می‌کنند تا دسترسی به حافظه اصلی سیستم سریع‌تر انجام گیرد، واحدهای پیش‌بینی انشعاب خود را بهبود می‌بخشند تا داده‌ها و دستورالعمل‌های مورد نیاز برای اجرای کاربردهای آینده برای پردازنده فراهم شود و ...

یکی دیگر از روش‌های تغذیه بیشتر واحدهای اجرایی و افزایش کارایی پردازنده‌ها اجرای چندین Thread بطور همزمان (SMT) است. این روش بدلیل آنکه موجب افزایش اندک اندازه هسته پردازنده و تعداد ترانزیستورهای آن می‌شود از روش‌های دیگر مفیدتر است. بوسیله SMT به آسانی می‌توان استفاده از یک واحد اجرایی پردازنده را به میزان دو برابر افزایش داد. Hyper Threading نامی است که اینتل برای SMT در نظر گرفته و این کار را با ارسال دو Thread از دستورالعمل‌ها به طور همزمان به هر یک از هسته‌های پردازنده‌ها و بدون نیاز به افزایش منابع اجرایی انجام می‌دهد. در این روش بوسیله افزایش تعداد دستورالعمل‌ها به میزان دو برابر این احتمال که پردازنده منتظر دستورالعمل خاصی بماند و یا منتظر دریافت داده‌ها از حافظه شود کاهش می‌یابد. با Nehalem یکبار دیگر فناوری Hyper Threading به پردازنده‌های اینتل بازگشته است. این فناوری برای اولین بار در نسخه Northwood معماری NetBurst اینتل معرفی شد. در روزهای که Hyper Threading برای اولین بار معرفی شد به ندرت شاهد افزایش کارایی وسیعی در کامپیوترهای بودیم. این نوآوری در آن زمان تنها پنج درصد کارایی را افزایش می‌داد و همچنین تعداد برنامه‌های کاربردی که از آن پشتیبانی می‌کردند بسیار اندک بود. اما در

بازگشت این نوآوری کارایی به میزان ۰ الی ۳۵ درصد افزایش پیدا می‌کند. زیرا این روزها برنامه‌های کاربردی چند وظیفه‌ای بسیار عمومی‌تر نسبت به شش سال گذشته شده‌اند.



زمانی که Pentium 4 برای اولین بار از این نوآوری استفاده کرده همه نگران اندازه سطح هسته و مصرف توان که در حال تبدیل شدن به یک معضل بود، بودند. در آن روزها مصرف توان، اندازه هسته و کارایی دست به دست هم دادن تا به کارایی حاصل از Hyper Threading از منظر توان نگاه شود (در جدول زیر به کارایی ناشی از فعال و غیر فعال شدن Hyper Threading همراه با میزان مصرف توان توجه کنید).

Intel Core i7-965 (Nehalem 3.2GHz)	POV-Ray 3.7 Beta 29	نرم‌افزار Cinebench	بازی Grid
HT غیرفعال	3239 PPS ۲۰۷ وات	4671 CBMarks	۱۰۳ فریم در ثانیه ۱۶۱,۸ وات
HT فعال	4202 PPS ۲۳۳ وات	4452 CBMarks	۱۰۲,۹ فریم در ثانیه ۱۵۹,۵ وات

کاربرد SMT

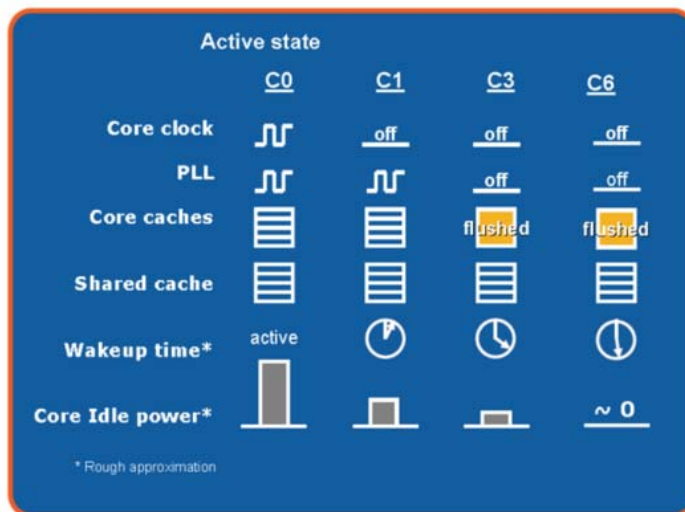
تاثیر SMT روی کارایی در اکثر مواقع مثبت است درحالیکه همه برنامه‌های کاربردی از این ویژگی استفاده نمی‌کنند. از آنجاییکه در Nehalem همه Threadها یکسان ایجاد نمی‌شوند، در استفاده از SMT برنامه نویسان بطورقطع با مشکل مواجه خواهند شد. برای حل این مشکل اینتل به برنامه نویسان کمک کرده و توسط روش تصمیم‌گیری دقیق که توپولوژی صحیح پردازنده را تشخیص می‌دهد (تعداد پردازنده‌های فیزیکی و منطقی) به برنامه نویسان اجازه می‌دهد تا از مکانیزم پیوستگی سیستم عامل برای واگذار کردن هر یک از Threadها به پردازنده استفاده کنند. این موضوع نباید مشکلی را برای برنامه نویسان بازی‌ها ایجاد کند زیرا آنها پیش از این با این روش در پردازنده‌های Xenon (پردازنده‌ای که در Xbox 360 استفاده شده است) آشنا شده‌اند. از آنجاییکه SMT بار سنگین‌تری را روی موتور اجرایی خارج از ترتیب اعمال می‌کند، اینتل اندازه بافرهای داخلی‌اش را برای آنکه آنها تبدیل به گلوگاه نشوند، افزایش داده است. بنابراین بافر سفارش مجدد (reorder) که اطلاعات مربوط به دستورالعمل‌های که پیش از این اجرا شدند را در خود نگهداری می‌کند، از ۹۶ ثابت در معماری Core به ۱۲۸ ثابت در Nehalem افزایش یافته است. در عمل این بافر تقسیم‌بندی شده تا هیچ یک از Threadها همه منابع را در انحصار خود نگه ندارد. یا به عبارت دیگر هر Thread تنها به یک بخش از بافر دسترسی دارد. با SMT اندازه هر یک از این بخش‌ها به ۶۴ ثابت کاهش پیدا کرده است. اما اگر یک Thread بصورت تکی اجرا شود، به همه ثابت‌ها دسترسی خواهد داشت. این موضوع بدین معنی است که Nehalem هیچ‌گاه کارایی پایین‌تری نسبت به نسل قبلی‌اش نخواهد داشت. درحالیکه اگر بخاطر داشته باشید در پردازنده‌های Pentium 4 در برخی مواقع زمانی که Hyper Threading فعال می‌شد کارایی پردازنده افت پیدا می‌کرد. ایستگاه رزرو کردن، واحدی است که وظیفه واگذاری دستورالعمل‌ها را به واحدهای اجرایی مختلف دارد. اندازه این بافر از ۳۲ ثابت به ۳۶ ثابت افزایش پیدا کرده است. اما برخلاف بافر سفارش مجدد، اینجا تقسیم‌بندی بصورت دینامیکی است بطوریکه Threadها می‌توانند ثابت‌های بیشتر و یا کمتری را بستگی به عملیات مورد نیازشان دریافت کنند. دو تا از بافرهای دیگر که اندازه‌شان تغییر کرده، بافر بارگذاری و بافر ذخیره سازی است. این بافر بارگذاری از ۳۲ ثابت به ۴۸ ثابت و بافر ذخیره‌سازی از ۱۶ ثابت به ۳۲ ثابت افزایش یافته است. در این دو بافر نیز تقسیم‌بندی بین Threadها استاتیکی است.

مصرف توان و SSE 4.2

در Nehalem نیز اینتل خصوصیات جدیدی را به لیست طولانی دستورالعمل‌های SSE اضافه کرده است. Nehalem از SSE 4.2 پشتیبانی می‌کند و شامل همه دستورالعمل‌های SSE 4.1 (که توسط Penryn پشتیبانی شده بود) به علاوه ۷ دستورالعمل جدید شده است. پنج عدد از این دستورالعمل‌ها برای دستکاری کارکترها (شخصیت‌های داستانی) هستند. اینتل ادعا می‌کند که این دستورالعمل‌ها موجب تسریع در پردازش فایل‌های XML نیز می‌شود. دو دستورالعمل دیگر نیز در برنامه‌های کاربردی مشخص مورد استفاده قرار می‌گیرد. یکی از این دستورالعمل‌ها POPCNT است که برای اولین بار همراه با بارسلونا معرفی شد و برای شمردن اعداد بدون بیت صفر در یک ثبات استفاده می‌کند. بر طبق گفته‌های اینتل این دستورالعمل مخصوصاً برای تشخیص صدا و ترتیب‌دهی DNA مفید است. آخرین دستورالعمل CRC32 است و برای شتاب‌دهی در محاسبات کدهای تشخیص خطا استفاده می‌شود.

مصرف توان تحت کنترل

اینتل درس خوبی از مشکلاتی که در زمینه مصرف توان در انتهای زمان عرضه پردازنده‌های Pentium 4 با آن مواجه شد، گرفته است. این شرکت از آن تاریخ به بعد هر زمانی که معماری جدیدی معرفی می‌کند در زمینه کاهش مصرف توان آن نیز می‌کوشد. در معماری Nehalem نیز مهندسان اینتل از شگردهای برای کاهش مصرف توان استفاده کرده‌اند. در این معماری اکنون یک میکروکنترلر (واحد کنترل توان) مجتمع شده قرار گرفته است که دائماً بر توان و درجه حرارت مورد استفاده هر یک از هسته‌ها نظارت می‌کند و می‌تواند بطور کامل آنها را در زمان‌های که استفاده نمی‌شوند غیرفعال کند.



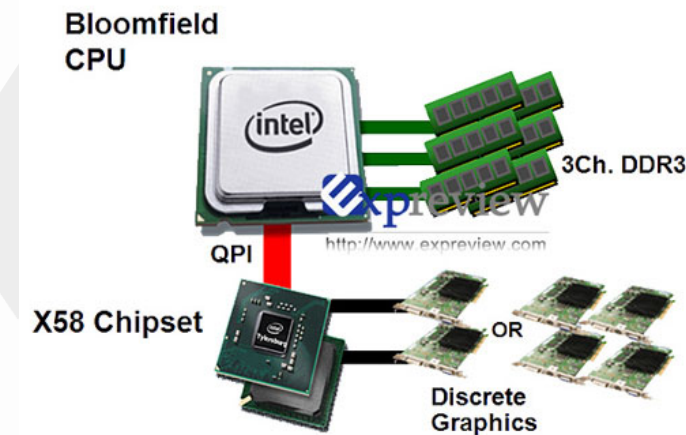
به واسطه این فناوری مصرف توان یک هسته بلااستفاده نزدیک به صفر است. دلیل اینکه مصرف توان کاملاً صفر نمی‌شود جریان نشتی ترانزیستورهای موجود در هسته‌ها است که تنها مختص Nehalem نیست و بطور کلی ترانزیستورهای همه پردازنده‌ها دارای چنین جریان نشتی هستند. اینتل به این

روش کاهش مصرف توان، وضعیت Turbo می‌گویند. زمانی که پردازنده پایین‌تر از استاندارد TDP خود عمل می‌کند، وضعیت Turbo فرکانس هسته در حال انجام کار را افزایش می‌دهد. لازم بذکر است که همانند پردازنده‌های Atom، حافظه نهان سطح یک و دو در پردازنده‌های مبتنی بر معماری Nehalem در عوض شش ترانزیستور از هشت ترانزیستور استفاده کرده‌اند. دلیل افزایش اندازه سطح پردازنده نیز بخاطر آرایش حافظه نهان پردازنده‌های جدید است.

QuickPath Interconnect

تمامی پردازنده‌های که تا کنون اینتل معرفی کرده برای ارتباط بین پل شمالی و پردازنده از گذرگاهی به نام "باس جلویی سیستم" (FSB) استفاده کرده‌اند. این گذرگاه با توجه به معایبی که دارد برای سیستم‌های پیشرفته امروزی مناسب نبود. اکنون اینتل در Nehalem از گذرگاه جدیدی مشابه با HyperTransport شرکت "ای‌ام‌دی" استفاده کرده است. این گذرگاه که QuickPath Interconnect یا QPI نام دارد، دارای دو مسیر مجزا برای ارتباط بین چیپست و پردازنده است. عرض هر یک از این مسیرها بیست بیت است که تنها شانزده بیت آن برای انتقال داده‌ها استفاده می‌شود و چهار بیت باقی‌مانده برای کدهای کشف خطا است. اولین نسخه QPI دارای فرکانس ۳/۲ گیگاهرتز است و از روش DDR برای انتقال اطلاعات استفاده می‌کند. بنابراین این رابط دارای نرخ انتقال ۶٫۴ گیگاتکسل در ثانیه (۶/۴ میلیون انتقال در هر ثانیه) است که در هر

جهت پهنای باندی معادل ۱۲/۸ گیگابایت در ثانیه فراهم می‌کند. در مقایسه FSB، QPI در هر سیکل اطلاعات کمتری انتقال می‌دهد اما در فرکانس بالاتر عمل می‌کند. سریع‌ترین فرکانس FSB در پردازنده‌های رایج اینتل اکنون ۴۰۰ مگاهرتز است که حداکثر پهنای باندی معادل ۱۲/۸ گیگابایت در ثانیه دقیقاً مشابه با QPI فراهم



می‌کند. اما QPI قادر است اطلاعات مربوط به خواندن و نوشتن را در دو مسیر مجزا انتقال دهد و

در هر مسیر دارای پهنای باندی برابر با ۱۲/۸ گیگابایت در ثانیه است در حالیکه گذرگاه FSB ۱۶۰۰ مگاهرتزی کنونی اینتل همین میزان پهنای باند را برای هر دو عملیات خواندن و نوشتن فراهم می‌کند. علاوه بر این حجم اطلاعاتی که باید در FSB انتقال پیدا کند بیشتر از QPI است زیرا در FSB باید داده‌های مربوط به I/O و حافظه انتقال پیدا می‌کند اما در QPI با توجه به اینکه کنترل کننده حافظه در پردازنده مجتمع شده تنها اطلاعات مربوط به I/O انتقال پیدا می‌کند. بنابراین با توجه به این موضوع QPI کمتر مشغول خواهد شد و پهنای باند قابل دسترس بیشتری خواهد داشت.

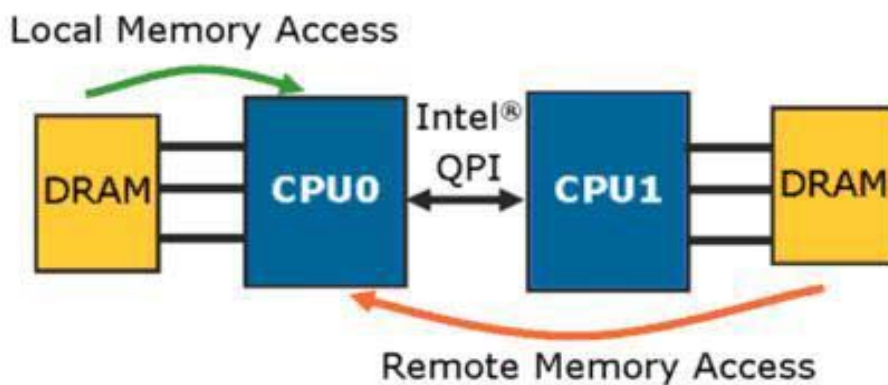
توضیحات تصویر : QPI دارای مسیرهای ورودی و خروجی اطلاعات بطور مجزا است.

زیر سیستم حافظه

کنترل کننده حافظه مجتمع

کنترلر حافظه مجتمع شده در پردازنده‌های شرکت "ای‌ام‌دی" موجب شد تا پردازنده‌های این شرکت کارایی بالاتری را نسبت به پردازنده‌های اینتل در بخش زیر ساخت حافظه فراهم کنند. اینتل اکنون این عقب‌ماندگی خود نسبت به "ای‌ام‌دی" را جبران کرده است. پردازنده‌های بارسلونا دارای دو کنترلر حافظه ۶۴ بیت بودند که از حافظه‌های DDR2 پشتیبانی می‌کردند. بالاترین محصول مبتنی بر معماری Nehalem شرکت اینتل دارای سه کنترلر ۶۴ بیتی با پشتیبانی از حافظه‌های DDR3 است. این کنترلر از حافظه‌های DDR3-1333 پشتیبانی می‌کند که حداکثر پهنای باند ۳۲ گیگابایت بر ثانیه را فراهم می‌کند. اما مزیت یک کنترلر حافظه مجتمع تنها موضوع پهنای باند نیست. کنترلر حافظه مجتمع موجب کاهش زمان‌های تاخیر دسترسی به حافظه می‌شود. با توجه به این که هر دسترسی چند صد سیکل ارزش دارد بنابراین کاهش زمان‌های تاخیر حاصل شده بوسیله یک کنترلر حافظه مجتمع محسوس خواهد بود. مزیت این معماری در سرورهای با پشتیبانی از چند پردازنده بیشتر به چشم خواهد آمد. در سرورهای حال حاضر زمانی که یک پردازنده به سیستم اضافه می‌شود پهنای باند حافظه هیچ‌گونه تغییری پیدا نمی‌کند در حالیکه با کنترل کننده حافظه مجتمع با توجه به اینکه هر پردازنده حافظه‌های مستقل به خود دارد با اضافه کردن یک پردازنده پهنای باند نیز افزایش پیدا می‌کند. در سرورهای که دارای بیش از یک پردازنده هستند این کاهش زمان‌های تاخیر متفاوت است. دستیابی به حافظه محلی بطور آشکار باعث کاهش زمان تاخیر و افزایش پهنای باند می‌شود و بالعکس دستیابی به یک حافظه دور دست نیازمند عبور از اتصال QPI است که موجب کاهش کارایی می‌شود. کنترلر حافظه مجتمع نسبت به پیش‌بینی انشعاب تاثیر کمتری روی کارایی دارد. زیرا این فناوری وابسته به برنامه‌های

کاربردی و سیستم عامل است. بنابر گفته اینتل کارایی در زمانی که پردازنده به یک حافظه دور دست دسترسی پیدا می کند کارایی تا حدود ۷۰ درصد بر حسب زمان تاخیر کاهش پیدا می کند و پهنای باند می تواند به نصف در مقایسه با حافظه محلی کاهش پیدا کند. بر طبق گفته های اینتل اما حتی در وضعیت دسترسی به یک حافظه دور دست از طریق مسیر QPI تاخیر نسبت به پردازنده های قبلی که کنترل کننده حافظه درون چیپست پل شمالی است، پایین تر خواهد بود.



حافظه نهان سطح سه

سلسله مراتب حافظه در Conroe بسیار ساده بود و اینتل با تغییر در حافظه نهان سطح دوم محصولات متنوعی تولید می کرد. اما در Nehalem اینتل از رقیبش "ایامدی" تقلید کرده است زیرا حافظه نهان سطح دوم اشتراکی برای معماری پردازنده های چهار هسته ای مناسب نبود. در پردازنده های مبتنی بر معماری Nehalem همانند پردازنده های Phenom هر هسته دارای حافظه نهان سطح دو مستقل است. از آنجاییکه هر هسته دارای حافظه نهان سطح دو مستقل و نسبتاً کوچک با حجم ۲۵۶ کیلوبایت است، مهندسان توانسته اند زمان های تاخیر آنها را نیز کاهش و کارایی را افزایش دهند. بنابراین حافظه های نهان سطح دو جدید نسبت به حافظه های موجود در Penryn دارای زمان تاخیر کمتری هستند (تقریباً ده سیکل در عوض پانزده سیکل) اکنون اینتل نیز مانند "ایامدی" از یک حافظه نهان سطح سه بزرگ با ظرفیت هشت مگابایت برای ارتباط بین هسته ها استفاده کرده است، در حالی که پردازنده های Phenom دارای دو مگابایت حافظه نهان سطح سه هستند. اگرچه در نگاه اول سلسله مراتب حافظه نهان در Nehalem ما را به یاد "بارسلونا" می اندازد اما عملکرد حافظه نهان سطح سوم این پردازنده بسیار متفاوت از "ایامدی" است. در معماری اینتل اگر یک هسته به یک داده نیاز داشته باشد و داده مورد نظر در حافظه نهان سطح سه موجود نباشد، پردازنده لازم نیست که حافظه های نهان اختصاصی هسته های دیگر را

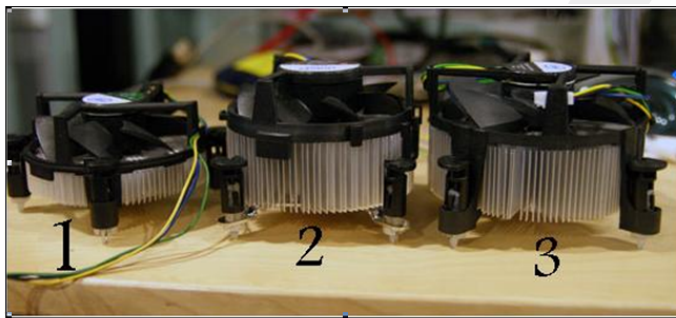
جستجو کند. زیرا داده‌ها در هیچ‌یک از آنها نخواهد بود. برعکس اگر اطلاعات در حافظه نهان سطح سه وجود داشته باشد توسط چهار مسیر که هر مسیر مختص به یک هسته است، قادر به تشخیص داده در حافظه نهان سطح سه خواهند بود. بطور کلی هسته‌ها توسط این چهار مسیر می‌توانند تشخیص دهند که داده‌های مورد نیازشان در حافظه نهان سطح سه موجود است یا خیر. این روش مبادلات بین هسته‌های مختلف را کاهش می‌دهد و برای عدم وابستگی حافظه‌های نهان اختصاصی به دیگر حافظه‌های نهان موثر است. همانطور که می‌دانید هر هسته برای اجرای یک وظیفه نیاز به تعدادی دستورالعمل و داده دارد که در حافظه نهان نگهداری می‌شود. در معماری حافظه نهان اشتراکی برای اجرای یک وظیفه، یک دستورالعمل به فضای حافظه نهان فراخوانده می‌شود و همه هسته‌ها از آن استفاده می‌کنند. اما در معماری جدید لازم است که هر حافظه نهان بطور مستقل دستورالعمل و داده را در خود ذخیره کند. اشکال این روش این است که حافظه نهان با داده‌های که در دیگر سطوح حافظه نهان نیز وجود دارد، بیهوده تلف خواهد شد. البته این موضوع تا حدی قابل چشم‌پوشی است، با توجه به این واقعیت که حافظه نهان سطح یک و دو در مقایسه با حافظه نهان سطح سه نسبتاً کوچک است (کل ظرفیت ذخیره‌سازی در این حافظه‌ها برابر با ۱/۲۵ مگابایت است). حافظه نهان سطح سه مانند "بارسلونا" در فرکانس یکسان با هسته و دیگر قسمت‌های پردازنده عمل نمی‌کند. در نتیجه تاخیر دسترسی به این نوع حافظه اندکی بیشتر از دو سطح دیگر است. در Nehalem این زمان برابر با چهل سیکل است. مشکل اصلی سلسله مراتب حافظه نهان در Nehalem، حافظه نهان سطح یک است. پهنای باند حافظه نهان دستورالعمل در این معماری افزایش پیدا نکرده و هنوز برابر با شانزده بایت در هر سیکل است در حالیکه "بارسلونا" دارای پهنای باند ۳۲ بایت در هر سیکل است. این موضوع می‌تواند تبدیل به یک کلوگاه خصوصاً در سرورها شود. از آنجاییکه دستورالعمل‌های ۶۴ بیتی بزرگ‌تر از دستورالعمل‌های ۳۲ بیتی هستند بخصوص اینکه Nehalem دارای یک کدگشایی بیشتر نسبت به "بارسلونا" است، فشار بیشتری روی حافظه نهان سطح یک اعمال شود. از طرفی دیگر زمان‌های تاخیر حافظه نهان داده‌ها به چهار سیکل در مقایسه با سه سیکل در Conroe افزایش پیدا کرده است اما این مشکلات با افزایش فرکانس و معرفی پردازنده‌های با فرکانس‌های بالاتر قابل جبران است.

TLB

در چند سال گذشته پردازنده‌ها با آدرس‌های فیزیکی حافظه کار نمی‌کردند آنها تنها با آدرس‌های مجازی سر و کار داشتند. استفاده از آدرس‌های مجازی اجازه می‌دهد تا حافظه بیشتری به یک برنامه نسبت به زمانیکه از آدرس‌های فیزیکی استفاده می‌شود اختصاص پیدا کند. استفاده از آدرس‌های فیزیکی موجب می‌شود تا داده‌های مورد نیاز تنها در یک زمان مشخص در حافظه سیستم نگهداری شوند و بقیه داده‌ها روی هارددیسک باقی بماند. در سیستم‌های مبتنی بر اساس آدرس‌های مجازی در هریک از دسترسی‌های حافظه یک آدرس مجازی باید به یک آدرس فیزیکی ترجمه شود. در این سیستم‌ها جدولی بزرگ مسئولیت نگهداری از این پیگیری ارتباطاتی را برعهده دارد. مشکل اینجاست که این جدول خیلی بزرگ است و نمی‌توان روی یک چیپ ست در روی حافظه اصلی آن را ذخیره کرد. اگر در هر دسترسی به حافظه لازم بود که آدرس‌های مجازی به آدرس فیزیکی ترجمه شود، دسترسی به حافظه خیلی آهسته‌تر می‌شد. در نتیجه مهندسی از روش آدرس‌دهی فیزیکی بوسیله اضافه کردن یک حافظه نهان کوچک درون پردازنده که تعداد اندکی از آدرس‌های که به تازگی دسترسی یافته را ذخیره می‌کند، استفاده کرده‌اند. این حافظه نهان "Translation Lookaside Buffer" یا TLB نامیده می‌شود. اینتل بطور کامل در عملکرد TLB در معماری جدیدش تجدید نظر کرده است. پردازنده‌های Core 2 از TLB سطح یک که بی‌نهایت کوچک (۱۶ ثبات) اما برای بارگذاری‌ها بسیار سریع است و TLB سطح دو بزرگتر (۲۵۶ ثبات) که با بار از دست رفته TLB سطح یک بعلاوه ذخیره‌سازی سروکار دارد، استفاده کرده‌اند. Nehalem در حال حاضر یک TLB دو سطحی واقعی دارد. اولین سطح از TLB بین داده‌ها و دستورالعمل‌ها به اشتراک گذاشته شده است. TLB داده سطح یک اکنون ۶۴ ثبات را برای صفحات کوچک یا ۳۲ ثبات را برای صفحات بزرگ ذخیره می‌کند. درحالیکه TLB دستورالعمل سطح یک، ۱۲۸ ثبات را برای صفحات کوچک و هفت ثبات را برای صفحات بزرگ ذخیره می‌کند. TLB سطح دو یک کش متحد شده است که می‌تواند حداکثر تا ۵۱۲ ثبات را ذخیره کند که فقط با صفحات کوچک سروکار دارد. هدف این پیشرفت افزایش کارایی برنامه‌ها کاربردی است که از مجموعه داده‌های بزرگتری استفاده می‌کند. این مسئله همراه با پیش‌بینی انشعاب دو سطحی گواه این حقیقت است که Nehalem طراحی براساس سیستم‌های سرور دارد.

سوکت پردازنده

در پردازنده‌های جدید بدلیل آنکه کنترل کننده حافظه درون پردازنده گنجانده شده است تعداد پایه‌های پردازنده بطور چشمگیری افزایش پیدا کرده است. همانطور که می‌دانید پردازنده‌های Phenom شرکت "ای‌ام‌دی" دارای ۹۴۰ پایه و پردازنده‌های Core 2 شرکت اینتل دارای ۷۷۵ پایه هستند. اما پردازنده Core i7 با کنترلر حافظه سه کاناله DDR3 دارای ۱۳۶۶ پایه است که موجب شده سوکت و سطح پردازنده نسبت به گذشته از نظر فیزیکی بزرگ‌تر شود. زمانیکه کنترلر حافظه درون پردازنده گنجانده شود هرگونه تغییر در فناوری حافظه یا در تعداد کانال‌های حافظه موجب می‌شود تا پردازنده و سوکت جدیدی معرفی شود. در اواخر سال ۲۰۰۹ میلادی اینتل نسخه‌های ارزان قیمت پردازنده‌های مبتنی بر معماری Nehalem را معرفی خواهد کرد. این



پردازنده‌ها دارای کنترل حافظه دو کاناله هستند و از سوکت LGA با ۱۱۵۶ پایه استفاده خواهند کرد. پردازنده‌های آینده هشت هسته‌ای اینتل از سوکت LGA با ۱۳۶۶ پایه استفاده خواهند

کرد. بزرگ‌تر شدن اندازه سوکت موجب شده تا حرارت‌گیر نیز بزرگ‌تر شود.

توضیحات تصویر :

۱- خنک‌کننده Core 2 Duo ۴۵ نانومتری

۲- خنک‌کننده Core 2 Quad ۴۵ نانومتری

۳- خنک‌کننده Core i7 ۴۵ نانومتری

نتیجه‌گیری

اگر به ریز معماری Nehalem نگاه کنید یک موضوع کاملاً آشکار است: این پردازنده طراحی شده تا ضعف‌های اینتل در بازار سرور را جبران کند. البته این رویکرد به هیچ‌وجه اشتباه نیست اما این رویکرد کاملاً متفاوت با زمانی است که اینتل پردازنده‌های Conroe را معرفی کرد. در

پردازنده‌های مبتنی بر Conroe اینتل در حقیقت از معماری موبایل خود استفاده کرد و از مزیت‌های پردازنده‌های موبایل خود نظیر مصرف توان موثرتر و نسبت کارایی در هر وات بهتر برای پردازنده‌های خانگی‌اش استفاده کرد. بطور آشکار توان در حوزه سرورها نیز مورد توجه هست اما نه به اندازه‌ای که در حوزه موبایل به آن توجه می‌شود. در بازار پردازنده‌های سرور Conroe عملکرد خوبی داشت اما فاقد برخی از خصوصیات کلیدی بود. Nehalem در حقیقت معماری است که ضعف‌های بازار سرور اینتل را هدف قرار داده است. کنترلر حافظه مجتمع، Hyper Threading، TLB‌های بزرگ‌تر، بهبود کارایی Virtualization، دگرگونی سلسله مراتب حافظه نهان، دومین سطح پیش‌بینی انشعاب جدید، همه این خصوصیات بیشتر برای بخش سرور اینتل قابل رقابت خواهد بود. اما این خصوصیات برای کامپیوترهای خانگی نیاز به مصرف توان کمتر و کارایی بالاتری دارند.