

# دانلود مقاله ارائه ساختاری جدید برای حافظه نهان در پردازنده های بسا هسته ای

جهت مشاهده [دانلود مقاله ارائه ساختاری جدید برای حافظه نهان در پردازنده های بسا هسته ای](#) به پایین

همین صفحه مراجعه نمایید

تعداد صفحات : 9 صفحه

برای دریافت اینجا کلیک کنید

فرمت WORD قابل ویرایش



چکیده

با پیشرفت تکنولوژی و کوچکتر شدن روز افزون تعداد ترانزیستورها، پیچیدگی سیستم های چند پردازنده ای بر روی تراشه ها در حال افزایش است به دلیل افزایش عناصر پردازشی بر روی تراشه ما با مشکل منابع حافظه به دلیل تأخیر سیم و همچنین تأخیر کش برای منابع بر روی تراشه و تراکم ترافیک شبکه مواجه هستیم. شبکه بر تراشه (NOC) در حال حاضر به عنوان یک الگوی امیدوارکننده از ارتباط بین هسته ها در پردازنده های چند هسته ای در نظر گرفته شده است. در این گزارش ما به بررسی اینکه چگونه سازمان منطقه ای کش میانگین تأخیر شبکه را کاهش می دهد، می پردازیم و یک ساختار کش محلی با نماینده مدیریت واحدها برای پردازنده های پایه NOC ارائه می دهیم. آزمایشات نشان داده است که تأخیر دسترسی کش L2 تا حدود زیادی توسط سازمان الگوی اتصال در شبکه بر روی تراشه معلوم می شود، و سازمان کش منطقه ای بسیار مهم در کارایی بهتر کش NOC است.

کلمات کلیدی: پردازنده های بساهسته ای، بهبود سرعت ، افزایش کارایی، ساختار حافظه نهان

مقدمه

به دلیل افزایش تعداد ترانزیستورها در سال ۲۰۰۲ تولیدکنندگان تراشه متوجه شدند، دیگر نمی توانند ولتاژ عملیاتی ترانزیستورها را مانند قبل کاهش دهند و در نتیجه نمی توانند با همان نرخ، اندازه ترانزیستورها را کاهش و سرعت تراشه ها را افزایش دهند. این وضعیت به افزایش گرمایی منجر شد که در هر مترمربع سیلیکون تولید می شد. در نتیجه طراحان با مشکلی مواجه شدند که از آن با عنوان دیوار توان یاد می شود؛ یعنی حداکثر توان قابل دستیابی برای هر پردازنده ، به دلیل افزایش بسیار زیاد گرما با محدودیت چشمگیری مواجه شد. اکنون طراحان این واقعیت را پذیرفته اند که با وجود کوچک تر شدن ترانزیستورها، امکان افزایش سرعت پردازنده ها به واسطه افزایش ترانزیستورهای یک تراشه وجود ندارد. بنابراین اگر تعداد بسیار زیادی ترانزیستور را درون یک پردازنده بزرگ قرار دهید، سرعت پردازش آن بیش از یک تراشه امروزی نخواهد بود. این واقعیت رویکرد جدید را که در آن ترانزیستورهای مذکور را به جای یک پردازنده در چندین هسته پردازشی نصب

می کنند به خوبی توجیه می کند. پردازشی که در آن بیش از یک هسته استفاده شود را پردازش چند هسته ای (multi cores) گویند، یکی از دلایل دیگر استفاده از پردازنده های چند هسته ای افزایش قابلیت اعتماد سیستم می باشد. همچنین با استفاده از این سیستم ها بار موجود بر روی هسته پخش شده و صف انتظار کاهش پیدا می کند و توان عملیاتی زیاد می شود. یکی دیگر از قابلیت های این سیستم ها استفاده از حافظه نهان (cache) در دو لایه یا بیشتر می باشد که وجود چند لایه کش برای هر کدام از هسته ها باعث میشود علاوه بر آنکه نوع پردازنده به واسطه چند هسته، پردازش سریعتری داشته باشد بتواند اطلاعات مورد نیاز در پردازش را سریعتر بدست آورند و به طور کلی فرآیندها سریعتر به پایان برسند. در این پردازنده ها از حافظه نهان L2 به صورت مشترک استفاده می شود.

امروزه با زیادتر شدن نیاز به افزایش سرعت در پردازش ها سیستم های چند هسته ای کم کم جای خود را به سیستم های بسا هسته ای (many cores) می دهند. یکی از جریانات مهم در ساخت و طراحی سیستم های دیجیتال، سیستم های روی تراشه هستند. در این سیستم ها، چندین و در آینده نزدیک ده ها و صدها هسته پردازشی و حافظه ای شامل پردازنده های سیگنال، سخت افزارهای خاص منظوره، مدارهای منطقی برنامه پذیر، پردازنده های همه منظوره و انواع حافظه و مدارات جانبی در داخل یک تراشه مجتمع می شوند. [۱]

در ادامه به بررسی کاربرد حافظه نهان و چند نمونه ساختار حافظه نهان می پردازیم و در ادامه یک سطح جدید برای آن ارائه می دهیم و با استفاده از نرم افزار Noxim به بررسی و مقایسه تأخیرها می پردازیم.

### ساختار کش و تأخیر شبکه

شبکه بر روی هسته، معمولاً مش های سوئیچ و منابع  $n \times m$  می باشد که منابع با استفاده بسته های داده ارتباط برقرار می کنند. منبع می تواند یک المان پردازشی باشد یا بلوک حافظه یا هر سخت افزار سفارش شده، درست مثل ارتباط داخلی هسته. حافظه نهان های بزرگ امروزه به بانک های بسیار زیادی تقسیم شده اند که بدین ترتیب سرعت دسترسی بالا را حفظ می کنند. برخی بانک های شدید معماری کش که تأخیر سیم را برای بانک های نزدیک پردازنده کاهش می دهد، توسط کیم نشان داده شده است. [۲] اگر چه این ساختار کش برای شبکه بر روی قطعه پردازنده های بسا هسته ای مناسب نیست. هر دو معماری کش مرکزی و کش ناحیه ای هر دو از تأخیر شبکه رنج می برند، مخصوصاً

زمانی که ترافیک شبکه بالاست. آزمایشات نشان می دهد که میانگین تأخیر شبکه با افزایش سایز شبکه به سرعت افزایش می یابد. در شبکه  $8 \times 8$  اگر یک پردازنده بخواهد به کش L2 دسترسی پیدا کند که در مرکز ۱۶ بانک قرار گرفته است تأخیر دسترسی شبکه هزاران سایکل می شود ولی اگر آن را به ۴ ناحیه  $4 \times 4$  تقسیم کنیم تأخیر حدود ۸ سایکل می شود.

با توجه به این، ما استدلال می کنیم که کش پردازنده های بسا هسته ای مبتنی بر شبکه NOC باید به طور ناحیه ای سازمان یافته شود المان های پردازنده ی شبکه که به ناحیه هایی تقسیم می شوند، در داخل برخی از آن ها کش را که ناحیه تخصیص داده شده است را به اشتراک می گذارد، دسترسی کش از طریق ناحیه ها ممنوع است. در این روش میانگین تأخیر دسترسی کش به شدت کاهش یافته است، همچنین بار ترافیک شبکه کاهش یافته است.

سازماندهی ناحیه ای شهودی و طبیعی است، اما آیا اینکه واقعا کار می کند، هنوز اثبات نشده است، و برخی از ساختار های کش به ظاهر مجزا مانند: کش چند سطحی در مقابل NOC یک سطحی می توانند سازماندهی ناحیه ای را به اشتراک بگذارند.

براساس این الگو مایک ساختار کش L2 با نماینده مدیریت حافظه واحد (D\_MMU) را بررسی می کنیم. پردازنده های بسا هسته ای بر پایه NOC با کامپیوترهای با هسته کم متفاوت است که تأخیر ارتباط داخلی تقریباً غیر قابل پیش بینی است ما می خواهیم ساختار کش را در NOC مدل کنیم و بهترین راه حل برای طراحی آن را بررسی کنیم. فرض کنیم که هر المان پردازشی (PE) در NOC کش L1 مخصوص خود را دارد و کش L2 فقط در صورت دیده نشدن در L1 ملاقات می شود.

ارتباط در NOC به صورت N به N است، یعنی انتقال از منابع مختلف و مقاصد سفر به طور همزمان است. این کار باعث استفاده خوب و مناسب از منابع موجود با کمک موازی سازی می شود، اگر چه موازی سازی خود باعث تراکم شبکه می شود در ادامه ما به طور خلاصه ۲ نمونه معمولی کش در پردازنده های ۸×۸ NOC را مورد بررسی قرار می دهیم، برای هر الگو ما یک نمودار نشان می دهیم.

شکل یک ساختار ی را نشان می دهد که ۶۴ المان پردازشی متصل شده توسط (۸×۸ NOC 8×۸ surround) وجود دارند. کش L2 به ۱۶ بانک تقسیم می شود و همه این ۱۶ کش به NOC متصل هستند و دور المان های پردازشی قرار گرفته اند. آزمایشات ما نشان داد که میانگین تأخیر این نمونه ها، به شدت با افزایش بار ترافیک زیاد می شود. دلیل آن این است که گاهی اوقات در خواست حافظه مرجع (و همچنین پاسخ متناظر) باید یک فاصله طولانی را بپیماید به عنوان مثال از پردازنده گوشه سمت چپ بالا به بانک کش در گوشه راست پایین.

شکل ۸×۸ : 1 surround

در شکل دو یک NOC با ساختار کش محلی وجود دارد، المان های پردازشی به ۴ ناحیه تقسیم شده اند که هر منطقه کش اختصاص خود را دارد ( توجه داشته باشید که هر بلوک کش ممکن است نشان دهنده ی بانک های متعدد باشد) قسمت چپ بالا که توسط خط چین جدا شده است یکی از این چهار ناحیه می باشد. اگر کش L1 دیده نشود المان پردازشی در خواست دسترسی به کش L2 را در ناحیه خودش را می دهد، دسترسی به کش های خارج از محدوده ممنوع می باشد. بلوک کش L2 یک ناحیه ممکن است با کانال شخصی خودش همگام شده باشد پس دسترسی به کش ناحیه دیگر امکان دارد ما را دچار مشکل کند. ساختار کش منطقه ای در شکل ۲ از تأخیر شبکه ای کمتری برخوردار است و این تأخیر کم به دلیل فاصله کم درخواست حافظه های ارجاعی می باشد.

شکل ۸×۸ : 2 Region ۴

در شکل (۳) ما یک ۴×۴ NOC را نشان دادیم ، این می تواند یک پردازنده مستقل بر اساس NOC باشد یا یک ناحیه ۴×۴ در یک شبکه ۸×۸ باشد، این ساختار با ساختار شکل ۲ متفاوت است زیرا که ۴ کش جداگانه در مرکز قرار دارد که هرکدام مسیر یاب خود را دارند. ممکن است فکر کنید که این ساختار کش باعث کاهش میانگین تأخیر شبکه شود چون ۴ بلوک کش می توانند به طور موازی از طریق چهار مسیر یاب دیده شوند، اما آزمایشات خلاف آن را نشان داد، همانطور که پیش تر گفته شد موازی سازی باعث افزایش تراکم شبکه می شود، جزئیات بیشتر این ساختار در بخش ۴ بررسی شده است.

شکل ۴×۴ : 3 Block ۴

طرح پردازنده و D-MMU

در این بخش ما یک معماری کش ناحیه ای برای پردازنده های بسا هسته ای بر اساس NOC را نشان می دهیم. همانطور که در شکل ۲ می بینیم زیر ناحیه ۴×۴ از ۸×۸ NOC و یک کش L2 به عنوان یک ناحیه در نظر گرفته شده است، در کش L1 در صورت فقدان یک المان پردازشی در وهله اول در ناحیه داخلی خود به کار

گرفته می شود بدون مراجعه به منابع خارجی (به عنوان مثال المان های پردازشی، مسیر یاب های و بلوک های کش ) بنابراین ما در این قسمت بحث خود را محدود می کنیم به داخل یک ناحیه .

ساختار داخل شکل ۴ در واقع همان شکل ۲ می باشد، ما فقط المان پردازشی کنار کش L2 را برجسته کرده ایم ، اما هیچ تغییر سخت افزاری ایجاد نکرده ایم.

شکل : ۴ یک ناحیه ۴×۴ دارای D-MMU

اگر چه المان پردازشی برجسته شده نقش را در ساختار ما بازی می کند و ما آن را به عنوان واحد مدیریت نماینده حافظه می نامیم. ما اسم این ساختار را ساختار D-MMU می نامیم.

مدیریت فقدان کش L1

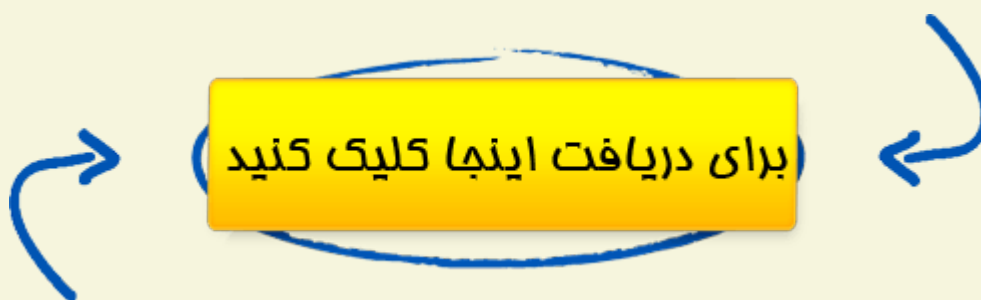
المان های پردازشی D-MMU کش های L1 شخصی خود را دارند. اگر فقدان L1 رخ دهد، المان پردازشی درخواست دسترسی حافظه D-MMU را به جای L2 می دهد، با دریافت درخواست ، D-MMU مسئول بازیابی داده های مورد نیاز از حافظه های نهان L2 (یا حافظه خارج از تراشه اگر از L2 نتواند) و ارسال داده ها به المان پردازشی خواهد بود. از آنجایی که D-MMU اساساً یک المان پردازشی عادی در سخت افزار استریال این مدیریت فقدان کش L1 با نرم افزار انجام می شود.

در شکل ۵ ما یک ساختار DMMU با ۳ المان پردازشی را نشان می دهیم که دو تا از آن ها المان پردازش ساده و یکی باقیمانده که در وسط قرار دارد D-MMU است . کش L1 اختصاصی المان های پردازشی نرمال به طور مجازی نمایه و برجسته گذاری شده است (V/V) است (به طور مجازی نمایه و فیزیکی برجسته گذاری شده است ، ما جزئیات و مزایای استفاده از این معماری را در ادامه این بخش توضیح خواهیم داد.[۵]

شکل : ۵ ساختار یک D-MMU

پردازنده ایمدرن معمولاً برای دسترسی به داده از آدرس حافظه مجازی استفاده می کنند، اما حافظه فیزیکی با آدرس فیزیکی نمایه می شود.

بنابراین برخی سطوح سلسله مراتب حافظه، ترجمه آدرس مورد نیاز است. بعضی مواقع آدرس قبل از دسترسی به کش L1 انجام می شود(بنابراین کش L1 با آدرس های مجازی نمایه می شود) همانطور که در شکل ۵ نشان داده شده است، ما از D-MMU برای بهبود بهره وری ترجمه آدرس در پردازنده های بسا هسته ای استفاده می کنیم.



مقالات مرتبط

- [دانلود مقاله بررسی کاربرد غشاهای در صنعت نفت و تصفیه آب و بساب](#)
- [دانلود مقاله ناویری دقیق رویات فضاپیما یا ردیابی یک هدف زمینی](#)
- [دانلود مقاله مروری بر الگوریتم های رمزنگاری تصویر مبتنی بر تئوری آشوب](#)

از این سایت ها نیز دیدن نمایید

- [ترنس لاین ، مرجع مقالات تخصصی فارسی ایران](#)
- [گت پیپر ، منبع مقالات انگلیسی و فارسی](#)
- [دانش رسان ، بیش از 1.5 میلیون مقاله فارسی](#)