
بسم الله الرحمن الرحيم

دانشگاه اصفهان

دانشکده فنی و مهندسی

گروه مهندسی کامپیوتر

پایان نامه‌ی کارشناسی ارشد رشته‌ی مهندسی کامپیوتر گرایش

معماری سیستم

طراحی یک سیستم حافظه نهان با مصرف انرژی بهینه با استفاده از تکنیک های ویژه سخت افزاری و نرم افزاری

استاد راهنما

دکتر عباس وفایی

پژوهشگر

نوید خیبر

خرداد ماه ۱۳۸۸

دانشگاه اصفهان

دانشکده فنی و مهندسی

گروه مهندسی کامپیوتر

پایان نامه ی کارشناسی ارشد رشته ی مهندسی کامپیوتر گرایش معماری سیستم آقای نوید خیبر تحت عنوان

طراحی یک سیستم حافظه نهان با مصرف انرژی بهینه با استفاده از تکنیک های ویژه سخت افزاری و نرم افزاری

در تاریخ ۱۳۸۸/۳/۳ توسط هیأت داوران زیر بررسی و با درجه عالی به تصویب نهایی رسید.

۱- استاد راهنمای پایان نامه دکتر عباس وفایی با مرتبه ی علمی استادیار امضا

۲- استاد داور داخل گروه دکتر محمدرضا خیام باشی با مرتبه ی علمی استادیار امضا

۳- استاد داور خارج از گروه دکتر سید مسعود سیدی با مرتبه ی علمی استادیار امضا

از طرف

امضای مدیر گروه

با سپاس فراوان از استاد گرامی و ارجمند

جناب آقای دکتر وفایی

تقدیم به

پدر و مادرم

چکیده

سرعت اجرای پردازنده های مدرن امروزی با سرعت چشمگیری در حال افزایش است که این امر موجب شده تا فاصله بین زمان دستیابی حافظه اصلی و سرعت اجرای پردازنده با نرخ بالاتری افزایش پیدا کند. یک راه حل مناسب جهت کاهش این فاصله استفاده از حافظه های نهان روی تراشه می باشد، بطوریکه هر روزه شاهد عرضه پردازنده هایی با مقادیر بیشتر حافظه نهان روی تراشه هستیم. اما با توجه به اینکه حدود ۴۰ درصد انرژی مصرفی پردازنده مربوط به حافظه نهان روی تراشه می باشد، افزایش میزان این حافظه موجب می شود تا توان مصرفی پردازنده به مقدار زیادی افزایش پیدا کند که این افزایش توان مصرفی پردازنده موجب می شود تا قابلیت اطمینان و چگالی تراشه نیز کاهش پیدا کند. علاوه بر این در سیستمهای قابل حمل نیز این افزایش توان مصرفی موجب کاهش عمر باتری می شود. اگرچه در پردازنده های مدرن امروزی اغلب از حافظه های نهان مجازی در سطح ۱ استفاده می شود، اما این نوع حافظه های نهان دارای یک مشکل عمده بنام مترادف می باشند. در این پایان نامه یک راهکار برای حل مشکل مترادف در حافظه های نهان مجازی و بهبود زمان دستیابی اصابت و توان مصرفی این نوع حافظه ها ارائه کرده ایم. سازمان حافظه نهان پیشنهادی ما، مجازی بوده و دارای مشکل مترادف نمی باشد. علاوه بر این، سازمان حافظه نهان پیشنهادی ما دارای زمان دستیابی پائین تر و توان مصرف کمتری نسبت حافظه های نهان نگاشت مستقیم و انجمنی- گروهی معادلش می باشد.

کلمات کلیدی : سیستم حافظه، حافظه نهان کم مصرف، حافظه نهان مجازی، مشکل مترادف

فهرست مطالب

عنوان صفحه

عنوان | صفحه

عنوان | صفحه

فهرست شکل ها

عنوان صفحه

عنوان | صفحه

عنوان | صفحه

فهرست جدول ها

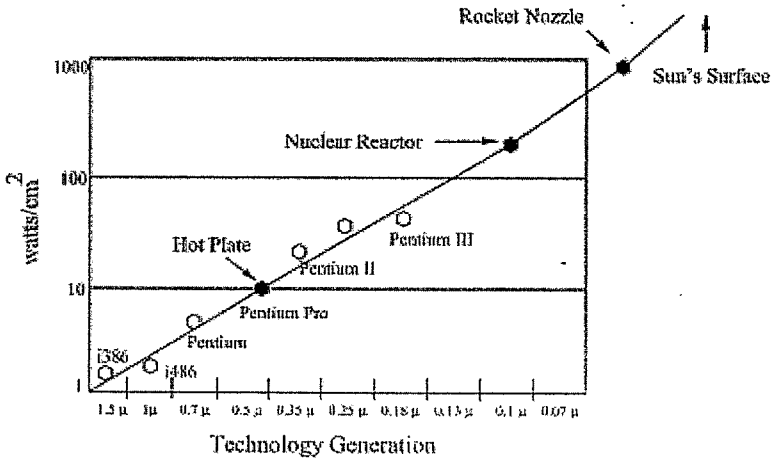
عنوان **صفحه**

فصل اول مقدمه

۱-۱- مقدمه

در گذشته وسایل الکترونیکی قابل حمل ، شامل وسایلی مانند ساعت مچی و ماشین حساب بودند که امروزه این وسایل جای خود را به وسایل قابل حمل کاراتری مانند کامپیوترهای قابل حمل، گوشی های موبایل و دستگاه های پخش موزیک و پخش فیلم قابل حمل داده اند. از طرفی دیگر روز به روز شاهد تقاضاهای روز افزونی جهت ارائه خدمات چند رسانه ای بر روی این وسایل الکترونیکی قابل حمل هستیم، بنابراین این سیستمها باید دارای کارایی و سرعت پردازش بالایی باشند تا بتوانند خدمات مطلوب را به کاربر ارائه دهند. اما بالاتر رفتن سرعت پردازش، موجب می شود تا توان مصرفی نیز افزایش پیدا کند (بدلیل افزایش فرکانس سیگنال ساعت سیستم) و با توجه به اینکه در سیستمهای قابل حمل انرژی مورد نیاز سیستم بوسیله باتری مهیا می شود بنابراین توان مصرفی سیستم تاثیر مستقیم بر روی عمر باتری سیستم، و در نتیجه آن مدت زمانی که از سیستم می توان استفاده کرد، دارد. به همین دلیل عمر باتری، یکی از معیارهای مهم سنجش کیفیت چنین سیستمهایی بشمار می رود. از طرفی دیگر، توان مصرفی سیستم تاثیر مستقیم بر روی گرمای تولید شده توسط سیستم دارد بنابراین در سیستمهایی که توان مصرفی بالایی دارند باید تجهیزات خنک کننده زیادی جهت پائین نگه داشتن دمای سیستم به آن اضافه شود که این امر موجب افزایش حجم و هزینه ساخت سیستم می شود. علاوه بر این بدلیل مسائل

خنک کنندگی، قابلیت اطمینان و بسته بندی[1]، میزان توان مصرفی تاثیر تعیین کننده ای بر روی طراحی تراشه دارد و یک عامل محدود کننده برای تعداد ترانزیستوری است که می توان بر روی یک تراشه قرار داد. با پیشرفت تکنولوژی ساخت مدارات مجتمع و کوچکتر شدن اندازه ترانزیستورها، چگالی ترانزیستورهای روی تراشه افزایش پیدا کرده است و همانطوریکه در شکل ۱-۱ نشان داده شده هر چه چگالی ترانزیستورهای روی تراشه افزایش پیدا می کند، چگالی توان مصرفی و در نتیجه آن گرمای تولید شده توسط تراشه نیز افزایش پیدا خواهد کرد و در صورتیکه راه حل مناسبی برای کاهش توان مصرفی تراشه اندیشیده نشود، این عامل محدود کننده (بدلیل گرم شدن بیش از حد تراشه) باعث متوقف شدن روند پیشرفت تکنولوژی ساخت مدارات مجتمع می شود.



شکل ۱-۱ : افزایش چگالی توان مصرفی[۱]

بنابراین با توجه به توضیحات فوق، جهت کاهش وزن و اندازه یک سیستم قابل حمل و افزایش عمر باتری آن، که از معیارهای مهم سنجش کارایی این نوع سیستمها می باشند، باید توان مصرفی این نوع سیستمها را کاهش دهیم. در سیستم های غیرقابل حمل نیز توان مصرفی معیار مهمی بشمار می‌آید زیرا هرچه توان مصرفی بیشتر باشد گرمای تولید شده توسط آن سیستم نیز بیشتر خواهد بود و سیستم به تجهیزات خنک کننده بیشتری نیاز خواهد داشت. بنابراین اگر در این سیستم ها توان مصرفی را پائین بیاوریم می توانیم بسیاری از تجهیزات خنک کننده را حذف کرده و قیمت سیستم را پائین بیاوریم و علاوه بر آن اندازه سیستم نیز می‌تواند کوچکتر شود.

[1] Packaging

برای کاهش توان مصرفی یک سیستم لازم است تا مولفه های آن سیستم بطور دقیق بررسی شوند تا مشخص شود زمانیکه سیستم در حال فعالیت است هر مولفه آن چه مدت زمانی فعال است و هر مولفه در زمان فعالیتش چه مقدار انرژی در واحد زمان مصرف می کند، بدلیل اینکه مولفه های یک سیستم زمانی انرژی مصرف می کنند که فعال شده و مورد دستیابی قرار بگیرند این نوع انرژی مصرفی تحت عنوان انرژی مصرفی پویا[1] شناخته می شود. علاوه بر این، نوعی دیگر از انرژی مصرفی بنام انرژی مصرفی نشتی[2] نیز وجود دارد که این نوع انرژی بیانگر آن میزان انرژی می باشد که یک مولفه زمانیکه غیرفعال است مصرف می کند. امروزه تلاشهایی که برای کاهش انرژی مصرفی صورت می گیرد در جهت کاهش انرژی پویای یک سیستم می باشند[۲] که میزان این انرژی متناسب با تعداد باری است که مولفه های یک زیر سیستم مورد دستیابی قرار می گیرند. بنابراین اگر مدت زمان فعال بودن یا توان مصرفی دینامیکی مولفه های سیستم را کاهش دهیم، مصرف توان سیستم کاهش پیدا می کند. لذا در اولین گام لازم است تا بررسی کنیم کدامیک از مولفه های سیستم بیشتر مورد دستیابی قرار می گیرند و کدامیک دارای بیشترین توان مصرفی دینامیکی می باشند سپس در گام بعدی با کاهش توان مصرفی دینامیکی این مولفه ها یا کاهش تعداد باری که مورد دستیابی قرار می گیرند، می توان مصرف انرژی سیستم را تا حد زیادی کاهش داد.

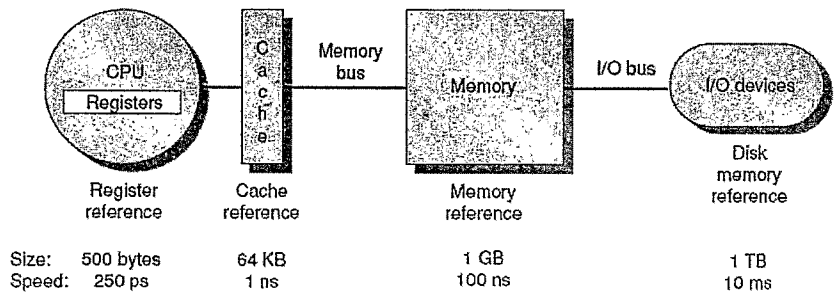
یکی از مهمترین اجزاء مصرف کننده انرژی در یک سیستم کامپیوتری پردازنده می‌باشد بنابراین اگر بتوان پردازنده‌ای کم مصرف طراحی کرد می توان میزان توان مصرفی کل سیستم را تا حدود زیادی کاهش داد. تحقیقات زیادی بر روی طراحی پردازنده های کم مصرف انجام شده است و در حال حاضر نیز یکی از شاخه های فعال در تحقیقات می‌باشد. بدلیل اینکه اغلب دستیابی های پردازنده به حافظه نهان[3] می باشد، برای کاهش توان مصرفی سیستم کامپیوتری استفاده از حافظه های نهان با مصرف انرژی پائین تبدیل به یک موضوع مهم در طراحی سیستمهای کامپیوتری مدرن شده است[۳].

---

[1] Dynamic Power
[2] Leakage Power
[3] Cache

۱-۲- حافظه نهان و جایگاه آن در سیستم کامپیوتری

پیشگامان علم کامپیوتر بدرستی پیش بینی کردند که برنامه نویسان در آینده، مقادیر زیادی از حافظه های سریع را طلب خواهند کرد. یک راه حل اقتصادی برای پاسخ به این خواسته استفاده از سلسله مراتب[۱] حافظه است، که از مزایای اصل محلی بودن ارجاعات[۲] و هزینه – کارایی[۳] تکنولوژیهای حافظه بهره برده و یک سیستم حافظه ارائه می دهد که سرعت آن تقریباً معادل سریعترین و گرانترین نوع حافظه و هزینه آن تقریباً معادل ارزانترین و کندترین نوع حافظه می باشد. اصل محلی بودن ارجاعات بیان می کند که اکثر برنامه ها کد و داده هایشان را به یک شکل مورد دستیابی قرار نمی دهند. محلی بودن در زمان[۴] و در فضای کد برنامه[۵] رخ می دهد. این مفهوم بعلاوه این نکته که سخت افزارهای کوچکتر سریعتر می باشند منجر به ایجاد یک سلسله مراتب حافظه براساس سرعت و اندازه متفاوت شد. در شکل ۱-۲ یک سلسله مراتب حافظه چند سطحی همراه با اندازه و سرعت دستیابی معمول برای هر سطح، نمایش داده شده است.



شکل ۱-۲ : مثالی از یک سلسله مراتب حافظه چند سطحی[۵]

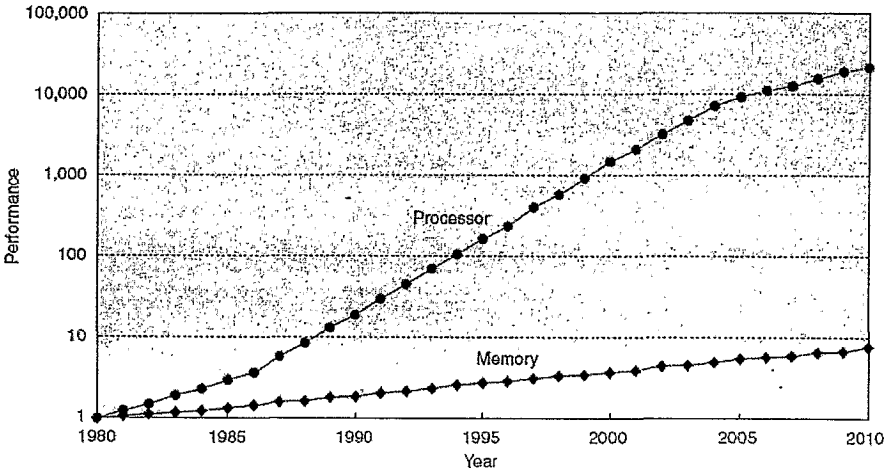
بدلیل اینکه حافظه های سریع گران هستند سلسله مراتب حافظه در سطوح مختلفی سازمان دهی شده است بطوریکه هر سطح آن، از سطح بعدی اش کوچکتر و سریعتر بوده و هزینه هر بایت آن گرانتر است.

بدلیل بهبود در کارایی پردازنده ها، سلسله مراتب حافظه اهمیت خاصی پیدا کرده است. شکل ۱-۳ مقایسه ای بین روند رشد کارایی پردازنده ها (سرعت اجرای پردازنده) و زمان دستیابی[۶] حافظه اصلی در زمان را

---

[1] Memory Hierarchy
[2] Locality
[3] Cost-Performance
[4] Temporal Locality
[5] Spatial Locality
[6] Access Time

نشان می دهد[۵]. بدلیل افزایش فرکانس سیگنال ساعت و استفاده از موازی سازی در سطح دستورالعمل[1] (ILP)، پردازنده ها دارای نرخ بهبود کارایی ۶۰ درصد در سال هستند در حالیکه نرخ بهبود زمان دستیابی حافظه ها ۱۰ درصد در سال می باشد[۶].



شکل ۱-۳ : مقایسه بین روند رشد سرعت دستیابی به حافظه با رشد کارائی پردازنده[۵]

واضح است که معماران کامپیوتر باید سعی کنند فاصله[2] بین این دو منحنی را کاهش دهند. برای کاهش این فاصله سه راه حل وجود دارد:

۱. استفاده از حافظه های بزرگ، سریع و گرانقیمت

۲. استفاده از پردازنده هایی با کارایی در حد حافظه های کُند و ارزان

۳. استفاده از سلسله مراتب حافظه و حافظه نهان سریع، کوچک و گرانقیمت

استفاده از روش اول باعث می شود تا سیستم، قیمت، مصرف انرژی و کارایی بالایی داشته باشد، استفاده از روش دوم باعث می شود تا سیستم، قیمت، مصرف انرژی و کارایی پائینی داشته باشد و روش سوم از مزایای هر دو روش بهره برده و معایب آنها را ندارد و باعث می شود تا سیستم قیمت و مصرف انرژی پائین، و کارایی بالایی داشته باشد.

حافظه نهان نامی است که به بالاترین (اولین) سطح از سلسله مراتب حافظه اختصاص داده می شود و آدرس صادر شده توسط پردازنده ابتدا با این سطح حافظه مواجه می شود. بدلیل اینکه استفاده از اصل محلی

[1] Instruction Level Parallelism
[2] Gap

بودن موجب بهبود کارایی می شود، از این اصل در سطوح مختلف بهره می برند بنابراین هر سیستم حافظه ای که داده های اخیراً استفاده شده را به قصد استفاده مجدد ذخیره می کند، تحت عنوان حافظه نهان شناخته می شود، بعنوان مثال حافظه های نهان فایل[1] و حافظه های نهان نام[2] نمونه هایی از این نوع حافظه ها می باشند.

زمانیکه پردازنده داده درخواست شده را در حافظه نهان پیدا می کند اصطلاحاً گفته می شود یک اصابت[3] رخ داده است و داده موردنظر از حافظه نهان خوانده شده و در اختیار پردازنده قرار می گیرد. زمانیکه داده در حافظه نهان پیدا نشود اصطلاحاً گفته می شود یک عدم اصابت[4] رخ داده است و در چنین حالتی، یک مجموعه داده با اندازه ثابت[5] بنام بلوک[6] که حاوی داده مورد نیاز نیز است، از حافظه اصلی خوانده شده و در حافظه نهان قرار می گیرد. بدلیل اینکه بنابر اصل محلی بودن در زمان، این احتمال وجود دارد که در آینده نزدیک، دوباره به این داده نیاز داشته باشیم و بنابر اصل محلی بودن در فضای کد برنامه، این احتمال نیز وجود دارد که در آینده نزدیک به داده های مجاور آن نیاز داشته باشم، بنابراین در حالاتی که عدم اصابت رخ میدهد، قرار دادن یک بلوک شامل داده و همسایه های آن از حافظه اصلی در حافظه نهان، برای بهبود کارایی پردازنده بسیار مفید می باشد. با توجه به اینکه اغلب دستیابی های پردازنده به حافظه نهان می باشد، بنابراین حافظه نهان نقش تعیین کننده ای در کارایی پردازنده دارد و بهبود کارایی حافظه نهان تاثیر مستقیم بر روی کارایی پردازنده دارد. به همین دلیل حافظه نهان از دیر باز مورد توجه معماران کامپیوتر بوده و سعی شده تا کارایی آن بهبود داده شود.

مدت زمان لازم جهت سرویس دهی به یک عدم اصابت، به تاخیر[7] و پهنای باند[8] حافظه وابسته است. تاخیر حافظه، بیانگر مدت زمان لازم برای بازیابی اولین کلمه از بلوک می باشد و پهنای باند حافظه، بیانگر مدت زمان لازم برای بازیابی بقیه بلوک را می باشد. عدم اصابت توسط سخت افزار سرویس دهی می شود و تا زمانیکه داده در دسترس قرار بگیرد، موجب توقف[9] پردازنده هایی می شود که بصورت In-Order برنامه ها را اجرا می کنند. اما در پردازنده هایی که برنامه ها را بصورت Out-of-Order اجرا می کنند، عدم اصابت موجب متوقف شدن اجرای دستوراتی می شود که به داده درخواست شده نیاز دارند اما دستورات دیگر اجرا خواهند شد.

---

[1] File Cache
[2] Name Cache
[3] Hit
[4] Miss
[5] Fixed-Size
[6] Block
[7] Latency
[8] Band Width
[9] Stall