

تقاضانامه همکاری پژوهشی

متقاضی: نگار اکبرزاده
نوع همکاری: پسادکتری

عنوان طرح: ارائه‌ی یک سیستم حافظه‌ی مقیاس پذیر برای پردازنده‌های گرافیکی

بیان مسئله

بررسی سیر تکامل پردازنده‌های گرافیکی نشان می‌دهد که سه عامل ظرفیت، پهنای باند و توان مصرفی حافظه همواره مانعی برای بهبود کارایی و رشد بیشتر این پردازنده‌ها با پیشرفت تکنولوژی بوده است. با این که طبق پیش‌بینی قانون مور، تعداد ترانزیستورها در پردازنده‌های گرافیکی هر دو سال دو برابر شده است، اما ظرفیت و پهنای باند حافظه با در نظر گرفتن محدودیت توان مصرفی، رشد بسیار کمتری داشته است. از طرفی به دلیل این که امروزه پردازنده‌های گرافیکی با محدودیت جدی از نظر اندازه‌ی تراشه‌ی قابل ساخت مواجه شده‌اند، فراهم کردن پهنای باند و ظرفیت بیشتر در این فضای محدود چالش‌برانگیز است. به همین دلیل، نیاز است تا با بهره‌گیری از تکنیک‌های مناسب و جدید در سطح معماری، سیستم حافظه‌ی پردازنده‌های گرافیکی بهبود یابد.

در پژوهش پیشنهادی قصد داریم به دو چالش پهنای باند و ظرفیت محدود پردازنده‌های گرافیکی بپردازیم و با بهره‌گیری از پشته‌سازی سه بعدی و حافظه‌های غیرفرار این محدودیت‌ها را بهبود دهیم.

پیشینه‌ی پژوهش

با توجه به این که امروزه با کاربردهایی مانند شبکه‌های عصبی عمیق، بینایی ماشین و پردازش گراف رو به رو هستیم که در آن‌ها به موازات و قدرت پردازشی بالا مورد نیاز است، بکارگیری پردازنده‌های گرافیکی که در گذشته تنها برای کاربردهایی مانند پردازش تصویر مورد استفاده قرار می‌گرفتند، بیش از پیش رونق یافته است. از طرفی، به دلیل این که حجم داده‌ی ورودی در این کاربردها بسیار بزرگ بوده و روز به روز نیز بیشتر رشد می‌کند، سه چالش اساسی طراحی پردازنده‌های گرافیکی یعنی محدودیت ظرفیت، پهنای باند و توان مصرفی بیش از پیش اهمیت پیدا می‌کنند. بدیهی است که برای پاسخگویی به نیاز این کاربردها و ایجاد جهش در روند رو به رشد اما کند سیستم حافظه‌ی پردازنده‌های گرافیکی، بهره‌گیری از فناوری‌ها و ساختارهای نوظهور حافظه مانند حافظه‌های مقاوم‌تی و حافظه‌های پشته‌ای اجتناب ناپذیر خواهد بود.

کارهای اخیر که استفاده از حافظه‌های غیرفرار و پشته‌سازی سه بعدی لایه‌های حافظه در ساختار پردازنده‌ی گرافیکی را پیشنهاد کرده‌اند، عمدتاً بر روی یکی از این گلوگاه‌ها تمرکز داشته‌اند و راه حل مقیاس‌پذیری برای حل همزمان چالش‌های ظرفیت و پهنای باند برای نسل‌های آینده پردازنده‌ی گرافیکی ارائه نداده‌اند. در این پژوهش، امکان پشته‌سازی سه بعدی واقعی حافظه بر

روی پردازنده‌ی گرافیکی با بهره‌گیری از حافظه چگال و غیرفرار PCM را بررسی می‌کنیم. این ساختار امکان افزایش ظرفیت و پهنای باند حافظه‌ی پردازنده‌ی گرافیکی با توجه به محدودیت انرژی و دما را فراهم می‌کند.

اهداف و روش تحقیق

- پشته‌سازی سه بعدی: شامل بررسی دقیق تأثیر پشته‌سازی انواع مختلف حافظه بر کارایی، مصرف توان و مشخصه‌ی دمایی پردازنده‌های گرافیکی خواهد بود.

- حافظه‌های غیرفرار: برای حل چالش‌های حافظه‌های غیرفرار از جمله تاخیر و توان مصرفی نوشتن بیشتر و تحمل‌پذیری کم، از تکنیک‌های کارا و مناسب در سطح معماری استفاده خواهیم کرد تا با فراهم کردن امکان پشته‌سازی سه بعدی، محدودیت ظرفیت و پهنای باند در سیستم حافظه‌ی پردازنده‌ی گرافیکی برطرف شود.

- آزمایش‌ها: آزمایش‌های دقیق و ارزیابی‌ها برای اثبات کارایی تکنیک‌های پیشنهادی انجام خواهند شد.

ابزارهای مورد نیاز

دسترسی به سرور مناسب برای پیاده‌سازی و اجرای ایده‌ها ضروری است. همچنین، از دیگر ابزارهای مورد نیاز می‌توان به نرم‌افزارهای شبیه‌ساز Accel-sim، Accel-wattch و مجموعه داده‌هایی چون DeepBench اشاره کرد.

نتایج مورد انتظار

با بهره‌گیری از تکنیک‌هایی مانند استفاده از حافظه‌ی نهان بزرگتر و یا بکارگیری حافظه‌ی ترکیبی می‌توان چالش‌های حافظه‌های غیر فرار را برطرف کرده و کارایی پردازنده‌ی گرافیکی را با در نظر گرفتن محدودیت دمایی و توان مصرفی در ساختار سه بعدی به میزان قابل توجهی افزایش داد.

خلاصه

یک پروژه تحقیقاتی جهت طراحی یک سیستم حافظه‌ی مقیاس پذیر برای پردازنده‌ی گرافیکی پیشنهاد می‌شود. انتظار می‌رود با استفاده از تکنیک‌های مناسب در سطح معماری و پشته‌سازی سه بعدی لایه‌های حافظه روی پردازنده‌ی گرافیکی بتوان کارایی را با حفظ محدودیت توان مصرفی و مساحت افزایش داد.