به نام خدا

برنامه نویسی حرفه ای

CUDA C

با مقدمه دکتر باربارا چاپمن

مرکز محاسبات و سیستم های پیشرفته دانشگاه هوستون

مترجمان:

دکتر صبا جودکی (استادیار – عضو هیات علمی دانشگاه آزاد اسلامی واحد خرم آباد)

مهندس ندا محمودی (کارشناس ارشد مهندسی کامپیوتر – نرم افزار)

مؤسسه فرهنگی هنری دیباگران تهران

عنوان کتاب: **برنامه نویسی حرفه ای CUDA C**

عنوان کتاب اصلی: **Professional CUDA C programming**

مترجمان : دکتر صبا جودکی

هندس ندا محمودی

ناشر: موسسه فرهنگی هنری دیباگران تهران

صفحه آرایی: نازنین صبوری

طراح جلد: داریوش فرسایی

نوبت چاپ: اول

تاریخ نشر: ۱۳۹۷

چاپ و صحافی: درج عقیق

تیراژ: ۱۰۰۰ جلد

قیمت: ۱۴۲۰۰۰۰ ریال

شابک: ۹-۰۳۷-۲۱۸-۶۲۲-۹۷۸

نشانی واحد فروش: تهران، میدان انقلاب، خ کارگر جنوبی، روبروی پاساژ مهستان، پلاک ۱۲۵۱

تلفن: ۶۶۴۱۰۰۴۶-۲۲۰۸۵۱۱۱

فروشگاههای اینترنتی دیباگران تهران :

WWW.MFTBOOK.IR

www.dibagarantehran.com

www.mftdibagaran.ir

نشانی تلگرام:@mftbook

سرشناسه:چنگ،جان cheng,jon

عنوان و نام پدیدآور:برنامه نویسی حرفه ای CUDA C / تالیف: جان چنگ،ماکس گروسمن،تای مکرچر،با مقدمه باربارا چپمن،مترجمان:صبا جودکی ،ندا محمودی

مشخصات نشر: تهران : دیباگران تهران :۱۳۹۷

مشخصات ظاهری:۷۳۶ ص:مصور.

شابک: ۹-۰۳۷-۲۱۸-۶۲۲-۹۷۸

وضعیت فهرست نویسی: فیپا

یادداشت : عنوان اصلی :۲۰۱۴ professional cuda c programming

موضوع:برنامه نویسی موازی (parallel programming(computer science

موضوع:واحدهای پردازش گرافیک-برنامه نویسی graphics processing units-programming

موضوع:نرم افزار کاربردی-طراحی وتوسعه application software-development

موضوع: کودا(ساختار کامپیوتر) cuda (computer architecture)

موضوع:واحدهای پردازش گرافیکی graphicss processing units

شناسه افزوده:گروسمن،ماکس-grossman,max

شناسه افزوده:مکرچر،تای-mc crer,ty

شناسه افزوده:چپمن،باربارا،مقدمه نویس chapman,bar ra

شناسه افزوده:جودکی،صبا،۱۳۶۰-مترجم

شناسه افزوده: محمودی،ندا،۱۳۶۸-مترجم

رده بندی کنگره:۱۳۹۷ ۴ ب۹چ/QA ۷۶/۶۴۲

رده بندی دیویی: ۰۰۵/۲۷۵

شماره کتابشناسی ملی: ۵۴۴۲۰۹۶

فهرست مطالب

فصل ۴ پارم / حافظه سراسری

فصل پنجم / حافظه اشتراکی و حافظه ثابت

فصل ششم / جریان ها و همزمانی

فصل هفتم / تنظیم اولویت های سطح دستورالعمل

فصل هشتم / OpenACC و کتابخانه های کودای تسریع دهنده GPU

فصل نهم / برنامه نویسی چند GPUای

فصل دهم / ملاحظات پیاده سازی

پیشگفتار

GPUها راه طولانی را طی کرده‌اند. در ابتدا به عنوان پردازنده‌های گرافیکی که می‌توانستند به سرعت تصاویر را به صورت خروجی برای واحد نمایش تولید کنند مطرح شدند، و رفته‌رفته به فناوری تبدیل شدند که برای پردازش‌های فراسریع نیاز می‌شوند. در چند سال گذشته، GPUها به طور فزاینده برای تسریع یک آرایه محاسباتی به نام محاسبات ناهمگن به CPU پیوستند. امروزه، GPUها بر روی بسیاری از سیستم‌های رومیزی، بر روی کلاسترهای محاسباتی، و حتی بر روی بسیاری از بزرگترین ابرکامپیوترها در جهان پیکربندی می‌شوند. در نقش توسعه‌یافته آن‌ها به عنوان فراهم‌کننده حجم بزرگی از قدرت محاسباتی برای محاسبات فنی، GPUها در علم و مهندسی در حوزه‌های گوناگون پیشرفت کرده‌اند. آن‌ها همچنین در عین حفظ قدرت، تعداد بزرگی هسته محاسباتی را برای کار به صورت موازی ایجاد کرده‌اند.

خوشبختانه، رابط‌های برنامه‌نویسی GPUها با این تغییرات سریع حفظ می‌شود. در گذشته، تلاش اصلی برای استفاده از آن‌ها در خارج از محدوده برنامه‌ها نیازمند این بود که برنامه‌نویس GPU با بسیاری از مفاهیمی که برنامه‌نویس گرافیکی از آن‌ها آگاهی دارد آشنا باشد. سیستم‌های امروزی مفاهیم بسیار ساده‌تری برای ایجاد برنامه‌های نرم‌افزاری که بر روی آن‌ها ایجاد می‌شوند دارند. به طور خلاصه، ما کودا را داریم.

کودا یکی از محبوب‌ترین رابط‌های برنامه‌نویسی کاربردی برای تسریع حوزه محاسبات کرنل در GPU است. که می‌تواند کد نوشته شده در C یا C++ را برای اجرای موثر در GPU با تلاش برنامه‌نویسی خیلی مناسبی امکان‌پذیر کند. همچنین توازن بین نیاز به اشتراک‌گذاری به منظور بهره‌گیری کامل از آن، و نیاز به داشتن یک رابط برنامه‌نویسی که استفاده از آن آسان باشد و برنامه قابل خواندنی را نتیجه دهد بر هم می‌زند.

این کتاب منبع ارزشمندی برای افرادی است که می‌خواهند GPUها را برای برنامه‌نویسی علمی و فنی استفاده کنند . و معرفی جامعی از رابط برنامه‌نویسی کودا و بکارگیری آن ارائه می‌کند. برای شروع، اصول محاسبات موازی بر روی معماری‌های ناهمگن را شرح می‌دهد و ویژگی‌های کودا را بیان می‌کند. سپس نحوه اجرای برنامه‌های کودا را توضیح می‌دهد. کودا مدل اجرایی و حافظه را به برنامه‌نویس نمایش می‌دهد؛ در نتیجه برنامه‌نویس کودا کنترل مستقیمی بر روی محیط به شدت موازی دارد. علاوه بر ارائه جزئیات مدل حافظه کودا، متن کتاب اطلاعاتی در مورد نحوه بهینه‌سازی آن ارائه می‌دهد. فصل بعدی در مورد جریان‌ها بحث می‌کند، همچنین نحوه اجرای همزمان و همپوشانی کرنل‌ها را مورد بحث قرار می‌دهد. سپس اطلاعاتی در مورد تنظیم، در استفاده از کتابخانه‌های کودا، و استفاده از رهنمودهای OpenACC برای GPUها ارائه می‌شود. بعد از فصل برنامه‌نویسی چند GPUای، با بحث در مورد برخی ملاحظات پیکربندی کتاب به پایان می‌رسد. با این حال، گستره‌ای از مثال‌ها برای

کمک به خواننده به منظور شروع کار، که بسیاری از آن‌ها قابل دانلود و اجرا هستند ارائه می‌شود. کودا توازن خوبی بین صراحت و قابلیت برنامه‌نویسی ارائه می‌کند که در عمل اثبات شده است. با این حال، کسانی که آن را مامور ساده‌سازی توسعه برنامه‌شان کرده‌اند بایستی بدانند که این یک داستان در حال پیشرفت است. در چند سال گذشته، پژوهشگران کودا بر روی بهبود ابزارهای برنامه‌نویسی ناهمگن کار کرده‌اند . کودا ۶ چندین ویژگی جدید، از جمله حافظه متحد و کتابخانه‌های پلاگین، برای آسان کردن برنامه‌نویسی GPU معرفی می‌کند. آن‌ها همچنین مجموعه‌ای از دستورات به نام OpenACC را نیز ارائه کرده‌اند، که در این کتاب معرفی می‌شود. OpenACC تکمیل کودا را به واسطه ارائه مفاهیم ساده‌تر برای بهره‌گیری از قدرت برنامه‌نویسی کودا برای اجراهایی که نیاز به کنترل مستقیم کمتری دارند وعده می‌دهد. عجالتا نتایج نویدبخش‌تر هستند. OpenACC و کودا ۶، و دیگر عناوین پوشش داده شده در این کتاب به توسعه‌دهندگان کودا امکان می‌دهند تا برنامه‌هایشان را برای حداکثر نمودن کارایی تر یع دهند. داشتن این کتاب در یک محل ثابت در قفسه کتاب‌هایتان ضروری است.

شاد برنامه‌نویسی کودا!

BARBARA CHAPMAN
CACDS and Department of Computer Science
University of Houston