

مبانی کاربردی استفاده از پردازنده گرافیکی برای محاسبات عددی در علوم مهندسی

تأليف

دکتر سیدعارضا ذوالفقاری

(عضو هیات علمی گروه مهندسی مکانیک دانشگاه بیرجند)

دکتر حمید سعادت

(عضو هیات علمی گروه مهندسی کامپیوتر دانشگاه بیرجند)

مهندس علی فوادالدینی

(دانشجوی دکتری مهندسی مکانیک دانشگاه بیرجند)



مبانی کاربردی استفاده از پردازنده گرافیکی برای محاسبات عددی در علوم مهندسی

سرشناسه: ذوالفقاری، سیدعلیرضا، ۱۳۶۱ -
عنوان و نام پدیدآور: مبانی کاربردی استفاده از پردازنده گرافیکی برای محاسبات عددی در علوم مهندسی / تألیف
سیدعلیرضا ذوالفقاری، حمید سعادتفر، علی فوادالدینی.
مشخصات نشر: بیرجند: دانشگاه بیرجند، انتشارات، ۱۳۹۸.
مشخصات فیزیکی: ۱۸۶ ص.: مصور (رنگی)، جدول (رنگی)، نمودار.
شابک: ۹۷۸-۶۰۲-۶۵۸۲-۱۰۰۰-۰
وضعیت فهرست نویسی: فیپا
موضوع: پردازش موازی
موضوع: Parallel processing (Electronic computers)
شناسه افزوده: سعادتفر، حمید، ۱۳۶۳ -
Saadatfar, Hamid
شناسه افزوده: فوادالدینی، علی، ۱۳۷۰ -
Afoadaddini, Ali
شناسه افزوده: دانشگاه بیرجند. انتشارات
رده بندی کنگره: QA۷۶/۵۸
رده بندی دیویی: ۰۰۴/۳۵
شماره کتباتشناسی ملی: ۵۹۱۱۲۶۶

خرید اینترنتی این اثر در سایت: www.fekrebukr.ir

ناشر: دانشگاه بیرجند

تیراژ: ۵۰۰ جلد

چاپ اول: پاییز ۱۳۹۸

امور فنی: انتشارات فکربرکر

صفحه آرای: فائزه دستگردی

طراح جلد: نفیسه شکرانی

لیتوگرافی و چاپ: واصف

قیمت: ۴۵۰۰۰ ریال

شابک: ۹۷۸-۶۰۲-۶۵۸۲-۱۰۰۰-۰

فهرست مطالب

پیشگفتار ۹

فصل ۱: پردازش موازی و پردازنده گرافیکی ۱۱

۱-۱. واحد پردازش مرکزی ۱۴

۲-۱. پردازنده گرافیکی ۱۴

۳-۱. کودا ۱۷

۴-۱. نمونه‌ای از کاربردهای پردازش موازی با پردازنده گرافیک ۲۰

۴-۱-۱. تصویر برداری پزشکی ۲۰

۴-۱-۲. محاسبات دینامیک سیال ۲۱

۴-۱-۳. علوه ریت حیطی ۲۱

۵-۱. نتیجه گیری ۲۳

فصل ۲: منابع و پیش‌نماهی ک ۲۵

۱-۲. منابع کودا ۲۸

۱-۱-۲. منابع اینترنتی ۲۸

۲-۱-۲. کتاب‌ها ۲۹

۲-۲. محیط توسعه کودا ۳۲

۳-۲. پردازنده‌های گرافیکی پشتیبانی کننده کودا ۳۲

۴-۲. درایور دستگاه انویدیا ۳۲

۵-۲. جعبه ابزار توسعه کودا ۳۳

۶-۲. کامپایلر استاندارد C ۳۳

۱-۶-۲. سیستم عامل ویندوز ۳۴

۲-۶-۲. سیستم عامل لینوکس ۳۴

۳-۶-۲. سیستم عامل مکیتاش ایکس ۳۴

۷-۲. نتیجه گیری ۳۵

فصل ۳: مقدمات کدنویسی به زبان CUDA C ۳۷

۱-۳. اولین برنامه ۴۰

۲-۳. فراخوانی یک تابع ۴۰

۳-۳. ارسال پارامترها به توابع ۴۱

۴-۳	استخراج مشخصات دستگاه	۴۵
۵-۳	استفاده از خصوصیات دستگاه	۴۹
۶-۳	نتیجه گیری	۵۲
۵۳	فصل ۴: محاسبات عددی در CUDA C	
۱-۴	توزیع محاسبات روی بلاک های موازی	۵۶
۱-۱-۴	جمع برداری	۵۶
۱-۱-۱-۴	جمع برداری در پردازنده مرکزی	۵۷
۱-۱-۲-۴	جمع دو بردار در پردازنده گرافیکی	۵۸
۲-۴	وزن محاسبات روی نخ های موازی	۶۴
۱-۲-۴	توسعه جمع برداری	۶۴
۱-۲-۱-۴	جمع برداری با بهره گیری از نخ ها	۶۵
۲-۲-۴	جمع برداری بردارهای بلند	۶۷
۳-۲-۴	جمع برداری با اندازه دلخواه	۶۹
۳-۴	همکاری نخ ها، حافظه اشتراکی و هم گام سازی	۷۱
۱-۳-۴	ضرب نقطه ای	۷۲
۴-۴	شبکه محاسباتی ۲ بعدی	۸۵
۵-۴	حافظه ثابت	۹۳
۶-۴	روش های عددی	۹۶
۱-۶-۴	روش حذفی گوس	۹۷
۲-۶-۴	الگوریتم حل دستگاه معادلات سه قطری	۹۹
۱-۲-۶-۴	الگوریتم توماس	۹۹
۲-۲-۶-۴	الگوریتم کاهش متناوب (CR)	۱۰۱
۳-۲-۶-۴	الگوریتم کاهش متناوب موازی (PCR)	۱۰۲
۴-۲-۶-۴	مقایسه الگوریتم های سری و موازی	۱۰۳
۷-۴	نتیجه گیری	۱۰۴
۸-۴	پروژه پیشنهادی	۱۰۵

فصل ۵: کد نویسی بهینه در CUDA C

۱-۵	روند پردازش در پردازنده گرافیکی	۱۱۰
۲-۵	راهکارهای پردازش بهینه روی پردازنده گرافیکی	۱۱۳

۱۱۴	۱-۲-۵. روش محاسباتی
۱۱۴	۲-۲-۵. کاربرد و دسترسی بهینه به حافظه‌ها
۱۱۴	۱-۲-۲-۵. حافظه سراسری
۱۱۷	۲-۲-۲-۵. حافظه اشتراکی
۱۲۲	۳-۲-۲-۵. حافظه ثابت
۱۲۴	۴-۲-۲-۵. حافظه ثابت
۱۲۴	۳-۲-۵. مشغولیت پردازنده گرافیکی و پنهان سازی تاخیرات حافظه
۱۲۵	۱-۳-۲-۵. روش محاسباتی
۱۲۵	۲-۳-۲-۵. اندازه بلاک
۱۲۵	۳-۲-۵. حافظه ثابت
۱۲۵	۴-۳-۲-۵. حافظه اشتراکی
۱۲۶	۴-۲-۵. پرهیز از عملیات انشعابی
۱۳۰	۳-۵. مطالعه بیشتر
۱۳۰	۴-۵. نتیجه گیری
۱۳۱	۵-۵. پروژه پیشنهادی

فصل ۶: کتابخانه‌های کودا..... ۱۳۳

۱۳۶	۱-۶. مزایای استفاده از کتابخانه‌ها
۱۳۶	۲-۶. انواع کتابخانه‌های CUDA C
۱۳۷	۳-۶. روند کاری در یک کتابخانه
۱۳۸	۱-۳-۶. ساخت گیره کتابخانه
۱۳۸	۲-۳-۶. تخصیص حافظه دستگاه
۱۳۸	۳-۳-۶. تبدیل ورودی‌ها به قالب‌های پشتیبانی شده توسط کتابخانه
۱۳۸	۴-۳-۶. پیکربندی کتابخانه
۱۳۸	۵-۳-۶. اجرا
۱۳۹	۶-۳-۶. بازیابی اطلاعات از حافظه دستگاه
۱۳۹	۷-۳-۶. تبدیل داده‌ها به قالب برنامه
۱۳۹	۸-۳-۶. آزاد نمودن منابع کودا
۱۳۹	۹-۳-۶. ادامه محاسبات مربوط به برنامه
۱۳۹	۴-۶. آشنایی با چند کتابخانه پرکاربرد

۱۳۹	۱-۴-۶. کتابخانه cuSPARSE
۱۴۰	۱-۱-۴-۶. قالب‌های ذخیره اطلاعات در cuSPARSE
۱۴۳	۲-۱-۴-۶. کاربرد انواع مختلف قالب‌ها
۱۴۴	۳-۱-۴-۶. تبدیل قالب‌ها در cuSPARSE
۱۴۵	۴-۱-۴-۶. مثالی از کاربرد cuSPARSE
۱۴۷	۲-۴-۶. کتابخانه cuBLAS
۱۴۸	۱-۲-۴-۶. مدیریت داده‌ها در cuBLAS
۱۴۹	۲-۲-۴-۶. مثالی از کاربرد cuBLAS
۱۵۰	۳-۴-۶. کتابخانه cuRAND
۱۵۱	۱-۳-۴-۶. مفاهیم اولیه پیرامون اعداد تصادفی
۱۵۲	۲-۳-۴-۶. بکارگیری کتابخانه cuRAND
۱۵۵	۳-۳-۴-۶. مثالی از کاربرد cuRAND
۱۵۷	۵-۶. نتیجه‌گیری
۱۵۹	فصل ۷: عملیات بی‌وقفه
۱۶۲	۱-۷. کاربرد عملیات بی‌وقفه
۱۶۴	۲-۷. روش استفاده از عملیات بی‌وقفه
۱۶۴	۱-۲-۷. ضرب برداری
۱۶۵	۲-۲-۷. پیاده‌سازی بهینه عملیات ضرب بردار
۱۶۷	۳-۷. نحوه تعریف توابع بی‌وقفه
۱۶۹	۴-۷. توابع بی‌وقفه پشتیبانی شده در کودا
۱۷۱	۵-۷. نتیجه‌گیری
۱۷۱	۶-۷. پروژه پیشنهادی
۱۷۲	مراجع
۱۷۳	پیوست الف: پیش‌نیازهای کدنویسی در CUDA C
۱۸۱	پیوست ب: محاسبه عملکرد
۱۸۴	پیوست ت: تغییر تنظیمات برای استفاده از توابع اتمی

پیشگفتار

نوشتن یک کتاب در مورد پردازش موازی، کاری پیچیده است. خصوصاً اگر به عرصه نوینی نظیر پردازش موازی روی پردازنده گرافیکی وارد شده و مخاطبان کتاب هم تا حدی با مفاهیم عمیق حوزه نرم افزار و علوم کامپیوتر بیگانه باشند. در این صورت باید شیوه نگاه به موضوع با اغلب کتاب های تخصصی تدوین شده در این زمینه متفاوت باشد. در نگاه اول، برای مهندسان ممکن است پردازش موازی به سادگی تقسیم یک کار بین چند نفر به نظر برسد. پردازش موازی یک تقسیم کار است؛ اما این توصیف هرگز روشن کننده عمق چالش های این حوزه نیست. در حقیقت، همکاری بین چند نفر هم می تواند به موضوع چالش برانگیزی تبدیل شود. این که یک کار برای رسان به بهترین بازده باید بین چند نفر تقسیم شود و هر نفر چه کاری را انجام دهد ساده ترین بخش مسئله است. موضوع پیچیده تر می تواند نحوه تعامل بین این افراد در بخش های خاصی از کار باشد. این تعامل باید به موقع و سریع انجام شود. شاید همین تعامل، پرهزینه ترین و چالش برانگیزترین قسمت کار باشد. چالش ها در پردازش موازی با استفاده از پردازنده گرافیکی نیز از همین جنس است. این چالش ها در حدی است که با وجود گذشت بیش از یک دهه از اولین تلاش ها، توسعه دهندگان برنامه های کاربردی همچنان زمان زیادی را صرف نوشتن کدهای سریع تر برای مسائلی قدیمی می کنند.

هدف این کتاب، کمک به مهندسان است تا با استفاده از یک تکنولوژی ارزان و در دسترس پاسخ محاسبه عددی خود را بسیار سریع تر از قبل دریافت کنند. برای رسیدن به این مقصود، تمرکز کتاب در اغلب بخش ها معطوف به محاسبات ریاضی پایه همراه با اشاراتی به مسائل پیچیده تر بوده است. لغات تخصصی رایانه تنها در مواردی که به خواننده برای کسب اطلاعات بیشتر یاری می رساند به کار رفته است. ترجمه لغات نیز تا حد امکان بر اساس واژه نامه تخصصی رایانه انجام شده است اما با توجه به نبودن این حوزه از علم رایانه در کشورمان، ترجمه برخی از لغات برای اولین بار در این کتاب صورت گرفته است.

ما به عنوان نویسندگان کتابی که قرار است پردازش موازی را به مهندسان رشته های مکانیک، برق، عمران و... به عنوان کاربران غیرحرفه ای "کودا" آموزش دهد، یک مسئولیت مهم خود را ساده سازی مفاهیم و همراهی کامل ایشان برای برداشتن اولین گام ها قرار داده ایم. در فصل ۱ به بررسی روند تاریخی و فنی ظهور فناوری کودا پرداخته ایم و ویژگی ها و دستاوردهای آن را معرفی کرده ایم. در فصل ۲ به معرفی منابع کودا پرداخته و ابزارهای مورد نیاز برای پردازش موازی روی پردازنده گرافیکی معرفی شده است. در فصل های ۳ و ۴ مقدمات کدنویسی و

محاسبه روی پردازنده گرافیکی به صورت گام به گام ارائه شده است. در این دو فصل از توضیح هیچ نکته مهمی عبور نکرده ایم و تمامی مثال های ارائه شده نیز به صورت بخش به بخش تشریح شده اند. این مثال ها شامل تمامی بخش های ضروری یک کد هستند تا برای اجرای آن ها هیچ کاری به خواننده واگذار نشود. در فصل ۵ مباحثی پیشرفته پیرامون بهینه سازی عملکرد کد با زبانی تا حد امکان ساده ارائه شده است. این فصل را می توان یک راهنمای جامع برای افزایش سرعت برنامه ها دانست که حاصل پژوهش ها و تجربیات نویسندگان کتاب در این حوزه است. مثال های بخش ۵ تنها حاوی بخش های چالش برانگیز کد است و برای پیشروی سریع خواننده از آوردن مقدمات خسته کننده پرهیز شده است. فصل ۶ مطلوب خوانندگانی است که می خواهند در کمترین زمان نتیجه مطلوبی بگیرند. در این فصل کتابخانه هایی معرفی شده اند که به کمک آن ها می توانید سرعت اجرای برنامه های خود را با کدنویسی اندک افزایش دهید. در فصل ۷ به عنوان فصل پایانی به یکی از حالت های اساسی کدنویسی موازی پرداخته ایم و نوع خاصی از عملیات را با نام عملیات وابسته معرفی کرده ایم.

آشنایی با ریاضیات پایه و زبان برنامه نویسی C در سطح مقدماتی، پیشنیاز مطالعه این کتاب است. برخی از موارد مربوط به برنامه نویسی پیشرفته نیز که دانستن آن برای فهم بخشی از مطالب کتاب ضرورت دارد، در پیوست ارائه شده است.

طبیعتاً متن حاضر خالی از اشکال نیست و مایه انتان خواهد بود اگر خوانندگان محترم اشکالات مفهومی و چاپی را برای اصلاح در چاپ بعدی به نویسندگان اطلاع دهند.