CPU چو بیرون رود، GPU در آید!

هومن سیاری
Sayyari@ComputerNews.ir

استفاده از کارت گرافیکی برای پردازش به جای پردازنده! راحت‌تر بگوییم، استفاده از GPU به جای CPU! آیا کارت‌های گرافیکی، پردازنده‌ها را حذف خواهند کرد؟ مطالب و آزمایشات ارایه شده در این مقاله، حول محور فوق دور می‌زند

مدتی است که زمزمه استفاده از GPU به جای CPU نقل بسیاری از سایت‌ها و پیش‌قراولان سخت‌افزار شده است. دلیل آن هم بدیهی است. امروزه پیشرفته‌ترین پردازنده‌های دسکتاپ اینتل یا AMD فقط ۴ هسته دارند در حالی که کارت‌های گرافیکی موجود صدها هسته دارند. پس اگر بشود پردازش‌های رایج یا حتی تنها بخشی از آنها را توسط GPU انجام داد، اولاً راندمان بسیار بالایی به دست می‌آید و ثانیاً بار پردازشی از دوش CPU برداشته شده و امکان انجام پردازش‌های دیگر توسط آن فراهم می‌شود. جنگ GPU بین ATI و nVIDIA که از اوایل سال ۲۰۰۰ شروع شده، منجر به افزایش قدرت پردازشی GPUها شده است. به موازات این افزایش قدرت و سرعت، هر دو شرکت مذکور نیاز به افزایش کیفیت ویدیویی مخصوصاً در برنامه‌های گرافیکی سه‌بعدی نیز داشتند که این امر در بحث AntiAliasing و AntiSotropic بروز کرد. اگرچه فناوری‌های عرضه شده توسط آنها در مقاطع مختلف، باعث افزایش کیفیت شده است، اما تحولی اساسی ایجاد نکرده است. اینک هر دو با فناوری جدیدی (هر چند با اسامی متفاوت) قصد انقلاب در حیطه پردازش‌های سنگین را دارند. هر دو غول گرافیکی دنیا یعنی nVIDIA و ATI (یا بهتر بگوییم AMD) این ایده را در مرحله عمل آورده‌اند و امکان انجام برخی از پردازش‌های CPU را توسط کارت‌های گرافیکی خود میسر کرده‌اند. nVIDIA نام فناوری خود را CUDA گذاشته است و ATI هم نام Stream را انتخاب کرده است. باید دید که آیا GPUها موفق می‌شوند CPUها را از دور خارج کنند؟

nVIDIA CUDA

CUDA که مخفف عبارت Compute Unified Device Architecture است، یک معماری جدید پردازش موازی است که توسط nVIDIA ابداع شده است.

CUDA موتور پردازشی در GPUهای nVIDIA است که توسط برنامه‌نویسان و در زبان برنامه‌نویسی خاص خود قابل دسترسی است. برنامه‌نویسان در نسخه خاصی از زبان قدرتمند C که دارای ضمیمه nVIDIA (C with nVIDIA extensions) است، برنامه‌های خود را می‌نویسند و بعد از کامپایل قادرند آنها را روی GPU اجرا کنند. البته به زودی با سایر زبان‌های برنامه‌نویسی مثل C++ و Fortran هم می‌توان روی GPU برنامه نوشت.

آخرین نسخه درایورها، همگی دارای تمامی اجزای مورد نیاز برای CUDA هستند. CUDA با تمامی GPUهای سری ۸ به بعد شرکت nVIDIA در هر سه خط تولید Quadro، GeForce و Tesla کار می‌کند. nVIDIA تصریح می‌کند که تمامی برنامه‌هایی که برای GeForce سری ۸ نوشته شده‌اند، همچنان بدون نیاز به تغییر روی کارت‌های جدیدتر اجرا خواهند شد. CUDA همچنین امکان دسترسی برنامه‌نویسان به مجموعه دستورالعمل‌های Native و اجزای حافظه در پردازش موازی را نیز فراهم می‌کند که این امر توانایی برنامه‌نویس را در کنترل سخت‌افزار بسیار بالا می‌برد. بنابراین با استفاده از CUDA، معماری GPUها هم مثل CPU ها باز می‌شود، گرچه GPUها برخلاف CPUها دارای یک معماری "بسیار-هسته‌ای" موازی هستند که طی آن هسته‌ها توانایی پردازش هزاران Thread را به طور همزمان دارند. بنابراین اگر برنامه‌ای منطبق با این معماری باشد، می‌تواند راندمانی بسیار بالاتر از اجرا بر روی CPU را به ارمغان بیاورد.

با بیش از ۱۰۰ میلیون GPU فروخته شده با قابلیت CUDA در سراسر دنیا، بسیاری از برنامه‌نویسان با استفاده از کیت SDK (Software Development Kit) برنامه‌نویسی روی GPU که توسط nVIDIA فراهم شده است، در حال نوشتن برنامه‌های متفاوتی از سطح خانگی تا حرفه‌ای از پردازش‌های صوتی و تصویری گرفته تا شبیه‌سازی‌های فیزیکی، اکتشاف نفت و گاز، طراحی محصول، تصاویر پزشکی و تحقیقات علمی هستند.

بسیاری از کاربردهای فوق حداقل ۲۵۶ مگابایت حافظه اختصاصی گرافیکی نیاز دارند که امروزه این مقدار حافظه را روی تمامی کارت‌های گرافیکی می‌توان یافت.

در دنیای بازی‌های کامپیوتری، کارت‌های گرافیکی علاوه بر رندر تصاویر گرافیکی، برای محاسبات فیزیکی بازی (افکت‌های فیزیکی مثل دود، آتش، حرکت آب، شن و ...)

که اصطلاحاً PhysX نامیده می‌شوند، نیز مورد استفاده قرار می‌گیرند. CUDA همچنین برای افزایش سرعت کاربردهای غیر گرافیکی مثل محاسبات بیولوژیکی و رمزنگاری هم استفاده می‌شود.

CUDA هم APIهای سطح پایین و هم APIهای سطح بالا را فراهم کرده است. در ۱۵ فوریه ۲۰۰۷ کیت SDK مربوط به CUDA در محیط ویندوز و لینوکس ارایه شد و سپس در نسخه ۲,۰ که در تاریخ ۱۴ فوریه ۲۰۰۸ ارایه شد، MacOS هم مورد پشتیبانی قرار گرفت (شکل ۱).

مراحل پردازش CUDA در شکل ۱ به ترتیب زیر است:

۱ـ کپی داده‌ها از حافظه اصلی به حافظه GPU

۲ـ دستور انجام پردازش توسط CPU به GPU

۳ـ انجام پردازش موازی در هسته‌های GPU

۴ـ کپی نتایج از حافظه GPU به حافظه اصلی

ویژگی‌های CUDA

- استفاده از زبان C استاندارد برای نوشتن برنامه‌های پردازش موازی بر روی GPU
- کتابخانه‌های عددی استاندارد برای انجام تبدیل فوریه سریع و ساب‌روتین‌های اصلی جبر خطی
- درایور CUDA اختصاصی برای انتقال سریع داده‌ها بین GPU و CPU
- درایور CUDA اختصاصی برای عمل متقابل با OpenGL و DirectX
- پشتیبانی از ویندوز و لینوکس و Mac OS

مزایای CUDA

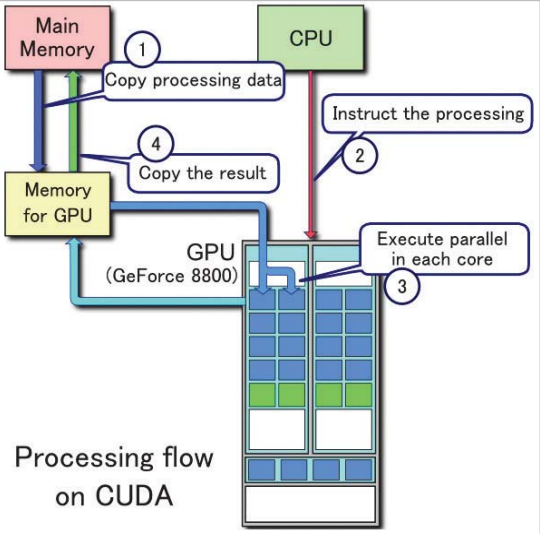
CUDA دارای چندین مزیت نسبت به روش‌های معمول پردازش توسط GPU است که در ادامه ذکر شده‌اند:

- خواندن پراکنده: کدها می‌توانند از چندین آدرس مختلف از حافظه خوانده شوند.
- حافظه اشتراکی: CUDA می‌تواند بخشی از حافظه را بین Threadها با سرعت بالا به اشتراک بگذارد که در واقع مثل کشی عمل می‌کند که توسط کاربر کنترل می‌شود و پهنای باند بالاتری را ایجاد می‌کند.
- دانلود و بازخوانی سریع‌تر به و از GPU
- پشتیبانی کامل از عملیات صحیح و بیتی

محدودیت‌های CUDA

- CUDA از یک زبان C، بدون توابع بازگشتی و بدون اشاره‌گر استفاده می‌کند و یک پردازش ساده در بخش‌های پراکنده حافظه پخش می‌شود، بر خلاف حالت عادی که از یک بخش یکپارچه از حافظه استفاده می‌گردد.
- رندر بافت‌ها (Texture Rendering) پشتیبانی نمی‌شود.
- پهنای باند و تاخیر گذرگاه بین CPU و GPU می‌تواند گلوگاه سیستم شود.
- در دقت مضاعف (Double Precision) هیچ مشکلی وجود ندارد اما در دقت تکی (Single Precision) به دلیل مشکلات ساختاری کمی عدم دقت پیش می‌آید.
- برای راندمان بالا Threadها باید در گروه‌های حداقل ۳۲تایی اجرا شوند در حالی که نیاز به هزارها thread است. انشعاب‌ها در کد برنامه باعث افت راندمان می‌شوند و هر ۳۲ تا thread یک مسیر اجرایی را طلب می‌کند. مدل اجرایی SIMD در زمان اجرای یک برنامه ذاتاً انشعاب‌پذیر دچار محدودیت‌های قابل توجهی می‌شود.
- GPUهای دارای CUDA فقط در تولیدات سری ۸ به بعد nVIDIA قابل پشتیبانی هستند.

شکل ۱: جریان پردازش در CUDA



تبدیل‌کننده Badaboom

nVIDIA در تولید نرم‌افزارهای مبتنی بر GPU پیش‌قدم شده و ابزاری به نام Badaboom برای تبدیل فرمت‌های مختلف فایل‌های ویدیویی به یکدیگر عرضه کرده است. nVIDIA می‌گوید Badaboom تا ۲۰ برابر سریع‌تر از سایر نرم‌افزارهای رایج تبدیل‌کننده عمل می‌کند. البته دوستان برنامه‌نویس می‌دانند که چرا nVIDIA تبدیل فایل‌های ویدیویی را به عنوان نمونه‌ای از کاربردهای CUDA عرضه کرده است، چرا که عمل تبدیل فرمت‌های ویدویی بسیار مناسب برنامه‌نویسی موازی است و هیچ انشعابی در برنامه پیش نمی‌آید. البته کار عجیبی که انجام داده، این است که قیمت ۳۰ دلار را برای این نرم‌افزار تعیین کرده است. معمولاً شرکت‌ها برای توسعه یک فناوری جدید، از سودهای ناچیز چشم‌پوشی کرده و شرایط را برای معرفی بیشتر فناوری مورد نظر مساعد می‌کنند.

ATI Stream Acceleration

ATI هم با همان ایده nVIDIA اقدام به پردازش موازی توسط GPU کرده است و آن را ATI Stream Acceleration نامیده است. Stream مجموعه‌ای از فناوری‌های پیشرفته سخت‌افزاری و نرم‌افزاری است که امکان همکاری پایاپای پردازنده‌های گرافیکی AMD ATI سابق را با CPU فراهم می‌کند تا بسیاری از کاربردها و مهم‌تر از همه، پردازش‌های گرافیکی با سرعت بالایی انجام شوند. در واقع AMD اکوسیستمی طراحی کرده که کارایی بالا، کاربردهای متنوع، نرم‌افزارهای خاص و ابزارهای ویژه از جمله ویژگی‌های آن است.

در واقع Stream به یک گروه از مسایل، کاربردها یا پردازش‌ها اشاره می‌کند که می‌توانند به عملیات‌های موازی و مجزا شکسته شوند و به طور همزمان روی یک پردازنده اجرا شوند. این جریان‌های داده موازی وارد پردازنده می‌شوند و به صورت موازی در هسته‌های آن پردازنده اجرا شده و نتایج به صورتی قابل اتصال به هم از پردازنده خارج می‌شوند. در واقع مزیت اصلی Stream در پردازش‌های (Single Instruction Multiple Data)SIMD

بـروز مـی‌کنـد. چـــون CPU بـــر اســـاس روش SISD(Single Instruction Single Data) کـــار می‌کنـد، بنابراین در پردازش‌های مـوازی Stream عالی ظاهر می‌شود (شکل ۲).

برنامه‌هایـی کـه مـی‌خواهنـد بـرای Stream طراحی شوندباید دو مشخصه اصلی داشته باشند:

- درجـه بالایی از محاسـبات ریاضی در هر واکشـی (Fetch) از حافظه
- محاسبات مسـتقل: انجام محاسبات روی هر واحد پردازشی بدون نیاز به بررسی یا تایید هر واحد پردازشی دیگر.

ویژگی‌های Stream عبارتند از:

- امکان اجرای برنامه‌های جدید بر روی معماری جدید
- امکان اجـرای پردازش‌های موازی کـه با معماری GPUهای جدید مطابقت دارند
- انتقال از توابع ثابت به پایپ‌لاین قابل برنامه‌نویسی کاربردهـای وسـیع در تحقیقـات و صنایـع تحـت عنوان GPGPU
- پشتیبانی از ۳۲۰ هسـته پردازشی (واحد محاسباتی یا ALU)
- پشتیبانی از GPUهای سری R600 به بعد

مهم‌ترین مزایای Stream

مزایای سخت‌افزاری:

- انجام محاسبات ممیز شناور با دقت مضاعف
- پشـتیبانی از ۲ گیگابایـت حافظـه اختصاصـی GDDR3
- مصرف توان پایین در پردازش‌های سنگین (پردازش بیش از ۵ میلیارد عملیات ممیز شناور با یک وات توان) نیاز به یک کانکتور برق برای Stream
- DMA غیر همزمان (انتقـال داده‌ها بدون نیاز به وقفه پردازنده)
- امکان استفاده از حافظه‌هایی با اندازه متفاوت برای نگهداری نتایج میانی
- استفاده از یک بورد کوچک (فقط ۲۳٫۵ سانتی‌متر)
- پشتیبانی از اینترفیس PCI Express 2.0 x16

مزایای نرم‌افزاری:

- قابـل برنامه‌نویسـی بـا محیطی شـبیه C با یک کامپایلر سطح بالا
- پشتیبانی از سیستم عامل‌های ویندوز و لینوکس ۳۲ و ۶۴ بیتی

در دسـامبر ۲۰۰۸ شـرکت AMD بـا عرضـه ATI Catalyst نسخه ۸,۱۲ امکان استفاده از قابلیت Stream را در کارت‌های خود فعال کرد. شرکت AMD برای تمرکز بر روی قابلیت Stream دو خط تولید جدید کارت گرافیکی به نام‌هـای ATI FirePro و AMD FireStream را راه‌اندازی کـرد تا به صورت کاملاً حرفـه‌ای کارت‌هایی طراحی کند کـه نهایت راندمان را براسـاس Stream به ارمغان بیاورد.

ATI Avivo

Avivo هـم نمونـه‌ای از کاربردهـای قابلیت Stream اسـت که توسط ATI بـرای تبدیـل فرمت‌های مختلـف فایل‌هـای ویدیویی به یکدیگر عرضه شده است. Avivo معادل Badaboom شرکت nVIDIA است.

Avivo در GPUهای سـری Radeon R520 به بعد ارایه شـده اسـت. Avivo برای برداشـتن بار پردازشی از دوش پردازنـده اصلی طراحی شـده اسـت. از جمله این بار پردازشی می‌توان عملیات رمزنگاری یا رمزگشایی فایل‌های ویـــدیویـی (Video Decoding/Encoding) و پردازش‌های محاسباتی سـنگین را نام برد. اگر با سیستمی کـه دارای یک GPU با قابلیت Avivo اسـت، اقدام به پخش و کدگشایی یک فایل ویدیویی توسط نرم‌افزار خاص Avivo کنید، می‌بینید که درصد استفاده از پردازنده اصلی بسیار کمتر از حالتی اسـت که از نرم‌افزارهای رایج استفاده می‌کنیـد. نکته جالب آنکه ATI ایـن نرم‌افزار را برخلاف nVIDIA به صورت رایگان عرضه می‌کند.

تست Avivo

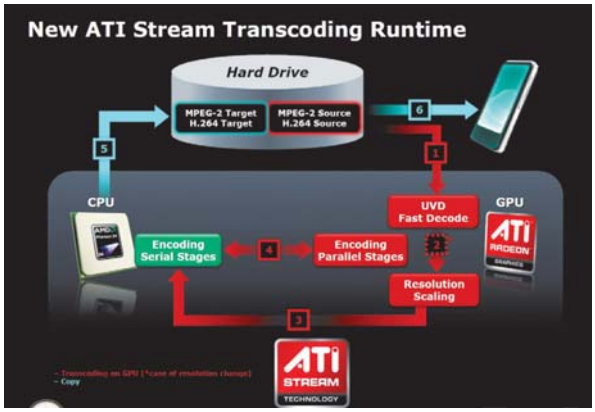
مـا در لابراتوار تصمیم گرفتیم که به ارزیابی قدرت GPU در کاربردهایی که تا به حال توسـط CPU انجام می‌شده است، بپردازیم.

معمولاً شـرکت‌های ارایـه کننده فناوری‌هـای جدید کمی در مـورد نوآوری‌هـای خـود اغـراق می‌کننـد و ATI و nVIDIA هم از این قائده مستثنی نیستند.

هـر دو شـرکت ادعـا می‌کنند که مثـلاً در مقولـه تبدیل فرمت‌های مختلـف فایل‌های ویدیویی بـه یکدیگر، حدود ۲۰ برابر نسـبت به نرم‌افزارهای رایج که از CPU استفاده می‌کنند، سـریع‌تر عمل می‌کنند. البته تنها قابلیتی که فعلاً می‌توان تسـت کرد، همان تبدیل فایل‌های ویدیویی است، چرا که نرم‌افزار آن توسط شرکت‌های فوق ارایه شده است.

تست با فرمت‌های مختلف

ما برای تست از یک فایل ۲۰۰ مگابایتی با فرمت MPEG2 و یک فایل یک گیگابایتی با فرمت VOB (فرمت DVD) استفاده کردیم و آنها را به فرمت‌های مختلف تبدیل کردیم. هـر تبدیل را یکبـار با نرم‌افـزار ATI Avivo و یکبار با نرم‌افزار Main Concept Reference تکرار کردیم و نهایت دقت را در یکسـان بودن تمامـی تنظیمات تبدیل انجام دادیم. بدیهی است که برای کدینگ ATI Avivo از GPU و Main Concept Reference از CPU استفاده می‌کنند. نتایج آزمایشات در شکل‌های ۳ تا ۶ نشان داده شده است.



شکل ۲: مراحل تبدیل فرمت ویدیویی در ATI Stream

مشخصات سخت‌افزاری	
CPU	Intel Core2Duo E8200 2.66 GHz 6MB cache
Graphic Card	SAPPHIRE Radeon HD4890
Motherboard	GIGABYTE GA-EX38-DS4
RAM	2×1GB DDR2 Corsair Dominator 1066
HARD	320GB Western Digital WD3200AAJS
DVD-RW	SAMSUNG SH-S203
Monitor	SAMSUNG SyncMaster 2032BW
Power	Green GP580B 580W
مشخصات نرم‌افزاری	
OS	Windows Vista Ultimate SP1 32bit
Driver	ATI Catalyst 9.7 Full
Applications	ATI Avivo 9.7 Main Concept Refrence 1.6.34425

جدول ۱: مشخصات سیستم تست

پس چرا راندمان، ۲۰ برابر نشد؟

همانطور که قبلاً اشاره شد، ATI و nVIDIA مدعی هستند که در تبدیل فرمت‌های مختلف فایل‌های ویدیویی به یکدیگر، حدود ۲۰ برابر نسبت به نرم‌افزارهای رایج که از CPU استفاده می‌کنند، سریع‌تر عمل می‌کنند. اما آزمایشات ما این را نشان نمی‌دهد!

البته با کمی تعمق می‌توان فهمید که چرا این‌گونه نشده است. آنها اختلاف قوی‌ترین GPU را با ضعیف‌ترین CPU در نظر می‌گیرند. بنابراین اگر از CPU قوی‌تر و یا از GPU ضعیف‌تر استفاده کنیم، این اختلاف کاهش می‌یابد، هر چند همیشه و در همه حال GPU بهتر از CPU عمل می‌کند. در سیستم تست ما هم گرچه از کارت گرافیک خیلی خوبی استفاده کردیم، اما پردازنده خوبی هم داشتیم که به طبع مانع از ایجاد اختلاف فاحش می‌شد.

ضمناً آنها کدینگی را در نظر می‌گیرند که بیشترین اختلاف را در GPU نسبت به CPU دارد (کدینگ‌های مختلف، دارای نسبت‌های متفاوتی از اختلاف بین GPU و CPU هستند.)

نتیجه

استفاده از GPU به جای CPU علی‌رغم تمام محدودیت‌هایی که دارد، سرعت و راندمان بالایی را به ارمغان می‌آورد. اگر بتوان روزی تمامی پردازش‌های CPU را توسط GPU انجام داد، تحولی شگرف در سرعت پردازش به وجود می‌آید و دیگر چیزی به عنوان CPU وجود خارجی نخواهد داشت. البته مشروط بر اینکه برنامه‌ها قادر به ارسال تعداد زیادی دستورالعمل به GPU باشند و در سطح بسیار بالایی از پردازش موازی پشتیبانی کنند.

اگر قصد خرید کارت گرافیکی دارید، حتماً مدلی که CUDA یا AVIVO را پشتیبانی می‌کند، خریداری کنید. هر چند با خرید این نوع کارت‌ها به تنهایی امکان استفاده از قابلیت فوق را ندارید، اما می‌توانید از برنامه‌هایی که برنامه‌نویسان و تولیدکنندگان براساس این قابلیت نوشته‌اند، استفاده کنید و لذت سرعت بالا را حس کنید. مثلاً برنامه‌هایی که برای تبدیل فرمت‌های مختلف ویدیویی به یکدیگر وجود دارند، با استفاده از CPU این کار را انجام می‌دهند، حال آنکه برنامه‌های دیگری براساس CUDA یا AVIVO نوشته شده‌اند که همان کار را بدون استفاده از CPU چندین برابر سریع‌تر انجام می‌دهند. ضمناً دوستداران بازی هم می‌توانند بازی‌های جدیدتر را که براساس قابلیت فوق عرضه می‌شوند، بسیار طبیعی‌تر و روان‌تر بازی کنند.

به طور خلاصه می‌توان گفت که استفاده از GPU به جای CPU دو مزیت اصلی دارد:

- سرعت بسیار بالاتر در انجام محاسبات
- آزاد شدن CPU و امکان استفاده از آن در پردازش‌های دیگر

کلام پایانی

تا چندی پیش در جوامع علمی مطرح بود که پردازنده‌ای طراحی شود که دارای GPU هم باشد و برای کاربردهای سبک گرافیکی مورد استفاده قرار گیرد، اما حالا قضیه برعکس شده است. GPUها کار پردازنده را انجام می‌دهند.

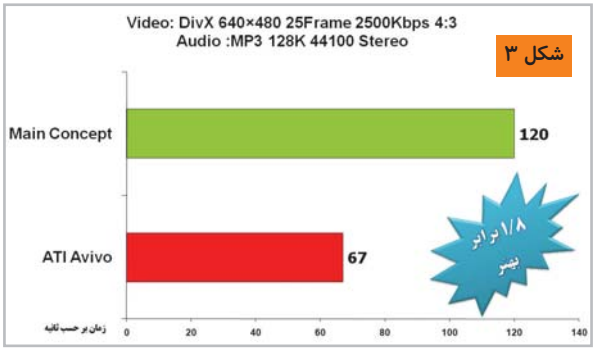
هر چند GPUها فعلاً تمامی پردازش‌های CPU را نمی‌توانند انجام دهند، اما دورنمای این فناوری چیز دیگری را نشان می‌دهد. آری، شمارش معکوس برای به تاریخ پیوستن CPU شروع شده است. البته نگارنده معتقد است که GPU جای CPU را نمی‌گیرد (به دلیل تفاوت معماری آنها و کاربردهای متفاوت آنها) بلکه GPU نقش پردازنده کمکی یا Coprocessor را بازی خواهد کرد، با این تفاوت که این پردازنده کمکی در بسیاری از پردازش‌ها از پردازنده اصلی قوی‌تر عمل می‌کند! ■

تشکر ویژه

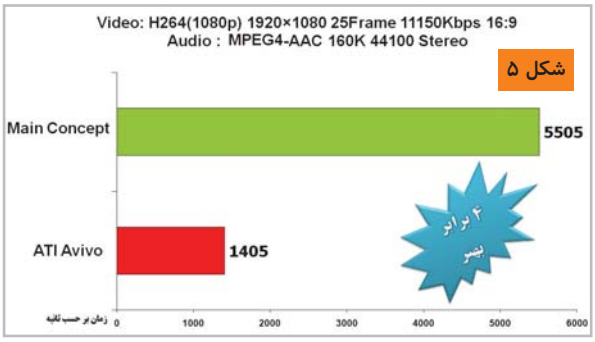
با تشکر از شرکت محترم فونیران که کارت گرافیک مورد استفاده در این مقاله را در اختیار لابراتوار رایانه‌خبر قرار دادند.

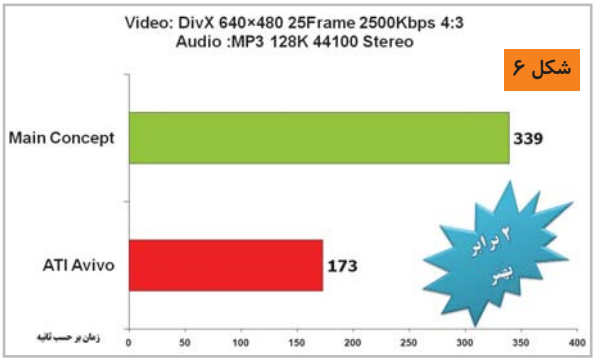
شماره تلفن : ۸۸۷۲۸۷۲۲
www.phoniran.com

تبدیل فایل ۲۰۰ مگابایتی با فرمت MPEG 2 به فرمت‌های مختلف



تبدیل فایل یک گیگابایتی با فرمت VOB به فرمت‌های مختلف





www.ComputerNews.ir