

به نام خدا

L^AT_EX

فاطمه علی ملکی
امیررضا جهانگیری
محمدحسین چهکندی
مهدی حق وردی
خدیجه نظری



دانشگاه اصفهان

مقدمه

معماری کامپیوتر - دیروز تا امروز

اجزا

معماری‌های مختلف

معماری کامپیوتر در آینده

هوش مصنوعی و معماری کامپیوتر

مقدمه

- در این ارائه به بررسی معماری کامپیوتر می‌پردازیم
- ابتدا سرگذشت و روند تکاملی معماری را بررسی می‌کنیم،
- سپس به معرفی اجزای اصلی یک کامپیوتر می‌پردازیم،
- پس از آن به داخل CPU می‌رویم و معماری‌های متفاوت آن را می‌بینیم،
- سپس در مورد آینده‌ی معماری کامپیوتر صحبت می‌کنیم
- و در آخر، تاثیر هوش مصنوعی به روی معماری کامپیوتر را بررسی می‌کنیم.

معماری کامپیوتر - دیروز تا امروز

تکامل معماری کامپیوتر

- در دنیای امروزی کامپیوترها برای اهداف زیاد و توسط افراد زیادی استفاده می‌شوند،
- کارها و اتفاقاتی که زمانی غیر قابل تصور بود، برای جامعه‌ی ما بسیار بدیهی و مرسوم است،
- تکنولوژی معماری کامپیوتر در طول سالیان متمادی، عمدتاً به دلیل پیشرفت‌های تکنولوژی ساخت قطعات الکترونیکی، پیشرفت علوم کامپیوتر و نیازهای افراد پیشرفت کرده است.

نسل اول کامپیوترها

- در سال ۱۹۳۷، اولین کامپیوتر با استفاده از لامپ‌های خلاء توسط پروفسور ایکن اختراع شد.
- در سال ۱۹۴۷، دانشگاه پنسیلوانیا کامپیوتری به نام ENIAC را طراحی کرد که از مبنای دودویی برای نمایش اطلاعات استفاده می‌کرد.
- معماری کامپیوترهای این دوره (و تمام دوره‌ها)، بر اساس مدل Von Neumann بود (و هست)، که شامل
 ۱. واحد حافظه،
 ۲. واحد پردازش،
 ۳. واحد کنترل و
 ۴. واحد ورودی/خروجیمی‌شود.

نسل دوم کامپیوترها

- در دهه ی ۱۹۵۰، ترانزیستورها به جای لامپ های خلاء در کامپیوترها استفاده شدند،
- این باعث کاهش حجم و افزایش سرعت کامپیوترها شد.
- در این دوره کامپیوترهای دیجیتال و مینی کامپیوترها شروع به ظهور کردند

- در دهه‌ی ۱۹۶۰، مدارهای مجتمع (IC) جایگزین ترانزیستورها شدند.
- استفاده از ICها باعث افزایش قابلیت پیچیدگی و کارای کامپیوترها شد.
- این به این معنی‌ست که تعداد بیشتری ترانزیستورها را در یک تراشه کوچک‌تر قرار دادند و این امر به کامپیوتر امکان انجام محاسبات پیچیده‌تر و سریع‌تر را می‌داد.
- کامپیوترهای این دوره (و دوره‌های بعدی) از معماری مجموعه دستورات (Instruction Set Architecture) استفاده می‌کردند.
- معماری کامپیوتر IBM 360 از معماری‌های مشهور این دوره است.

نسل چهارم کامپیوترها

- در دهه ۱۹۷۰، ریزپردازنده‌ها به جای ICها استفاده شدند.
- این باعث افزایش قابلیت انعطاف پذیری و کاهش هزینه‌ی ساخت کامپیوترها شد.
- در این دوره معماری کامپیوترها شخصی و کامپیوترهای قابل حمل توسعه یافت.

نسل پنجم کامپیوترها

- در دوره‌ی نسل پنجم کامپیوترها که از دهه‌ی ۱۹۸۰ شروع شد، تحولات مهمی در معماری کامپیوتر رخ داد.
- در این دوره کامپیوترهای موازی که قدرت پردازش با استفاده از چندین واحد پردازشگر به صورت همزمان را داشتند، طراحی و ساخته شدند.
- کامپیوترهای برداری، یکی دیگر از پیشرفت‌ها این دوره بود. این کامپیوترها مجهز به پردازنده‌هایی بودند که مخصوص انجام عملیات به روی بردارها و ماتریس‌ها بودند و برای برنامه‌های علمی و مهندسی که با این داده‌ها سروکار داشتند بسیار مناسب بودند.
- در این دوره استفاده از ICهای فوق بزرگ (VLSI) و ICهای فوق بزرگ (ULSI) نیز رایج شد.
- در این دوره شاهد ظهور کامپیوترهای شخصی (Personal Computer) و سیستم‌های توزیع شده (Distributed Systems) هستیم.

اجزا

معماری‌های مختلف

معماری کامپیوتر در آینده

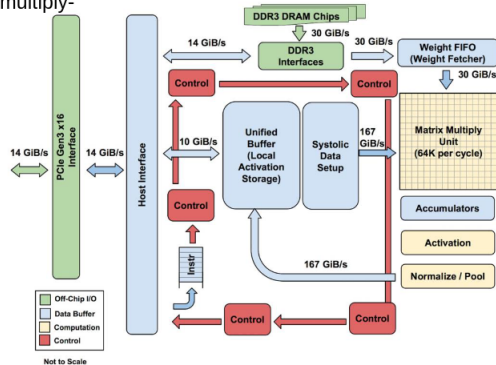
هوش مصنوعی و معماری کامپیوتر

- در دهه‌های ۸۰ و ۹۰ میلادی، کامپیوترها هر ۱۸ تا ۲۴ ماه (قانون مور) سریع‌تر می‌شدند.
- این یعنی اگر شما امسال یک کامپیوتر می‌خریدید و دوستان شما یک سال بعد از شما کامپیوتر جدیدی می‌خریدند، کامپیوتر آنها بسیار سریع‌تر می‌بود
- اما امروزه، تنها راه پیشرفت در معماری کامپیوتر ساخت سخت‌افزار برای یک کاربرد خاص است.
- برای مثال پردازنده‌ها گرافیکی (GPU) برای انجام محاسبات گرافیکی بسیار کارآمد هستند. آنها می‌توانند میلیون‌ها ضرب ماتریس در یک هر ثانیه انجام بدهند.

- با رخ دادن انقلابی در هوش مصنوعی به نام **یادگیری ماشین** که به ضرب ماتریسی متکی بود، نیاز به پردازنده‌های مخصوص ضرب تانسورها برای اجرا سریع‌تر و دقیق‌تر الگوریتم‌ها یادگیری ماشین احساس می‌شد،
- واحد پردازش تانسور (Tensor Processing Unit (TPU)) شتاب‌دهنده‌ی یادگیری ماشین است که توسط گوگل طراحی شده است. TPU ها برای ضرب ماتریس‌ها بسیار کارآمد هستند که برای آموزش شبکه‌های عصبی ANN ضروریست.

TPU: High-level Chip Architecture

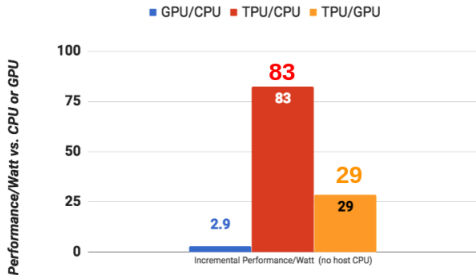
- The Matrix Unit: 65,536 (256x256) 8-bit multiply-accumulate units
- 700 MHz clock rate
- Peak: 92T operations/second
 - $65,536 * 2 * 700M$
- >25X as many MACs vs GPU
- >100X as many MACs vs CPU
- 4 MiB of on-chip Accumulator memory
+ 24 MiB of on-chip Unified Buffer (activation memory)
- 3.5X as much on-chip memory vs GPU
- 8 GiB of off-chip weight DRAM memory



Perf/Watt TPU vs CPU & GPU

Using production applications vs contemporary CPU and GPU

Measure performance of Machine Learning?



See MLPerf.org (“SPEC for ML”)

- Benchmark suite being developed by 23 companies and 7 universities
- 1st Results November 2018