

به نام خدا

L<sup>A</sup>T<sub>E</sub>X

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.

- در دهه‌ی ۱۹۵۰، ترانزیستورها به جای لامپ‌های 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

- در سطح بالا CPU از دو قسمت اصلی تشکیل شده که خود به قسمت‌های دیگری تقسیم می‌شوند:

۱. Data Path

این قسمت عملیات‌های ریاضی و محاسبات را انجام می‌دهد.

Data Path از قسمت‌هایی همچون:

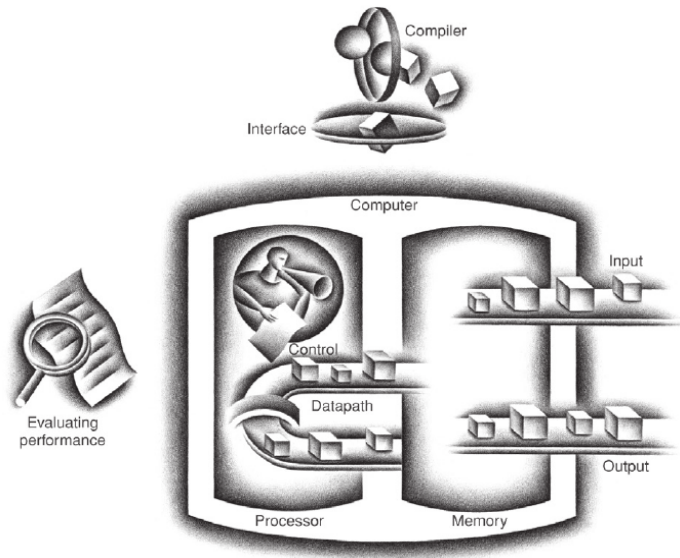
۱.۱ Register File

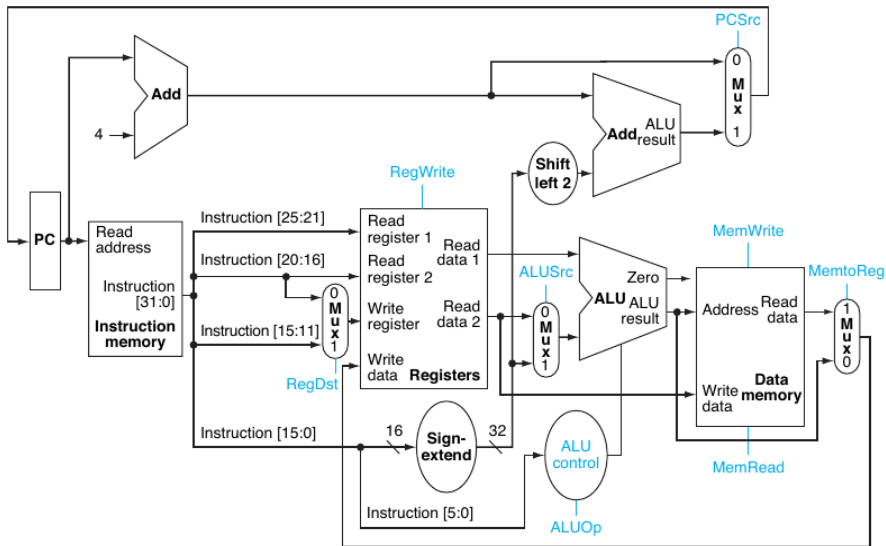
۲.۱ ALU

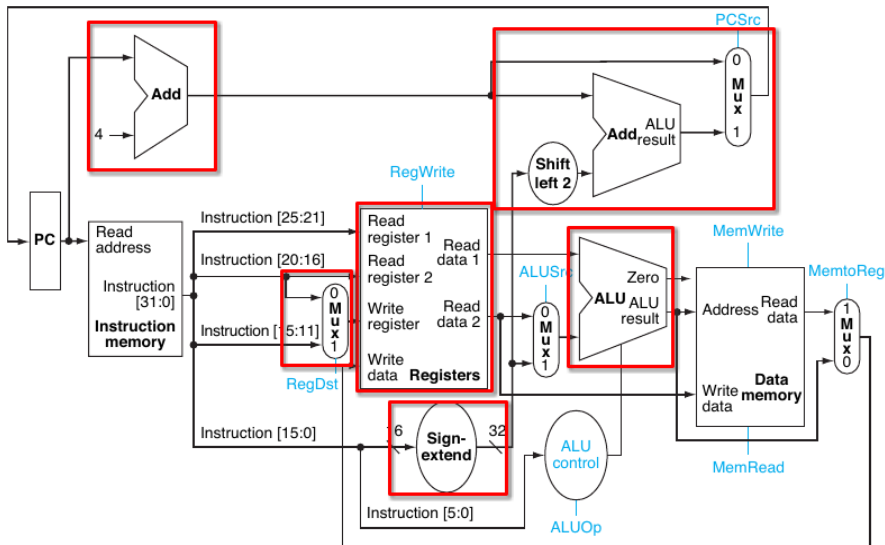
۳.۱ چندین Multiplexer و واحدهای جمع و extend تشکیل شده است.

۲. Control Unit

واحد کنترل پردازنده، به data path، مموری و دستگاه‌های I/O دستورات لازم برای اینکه چه کاری را باید انجام دهند، می‌دهد.



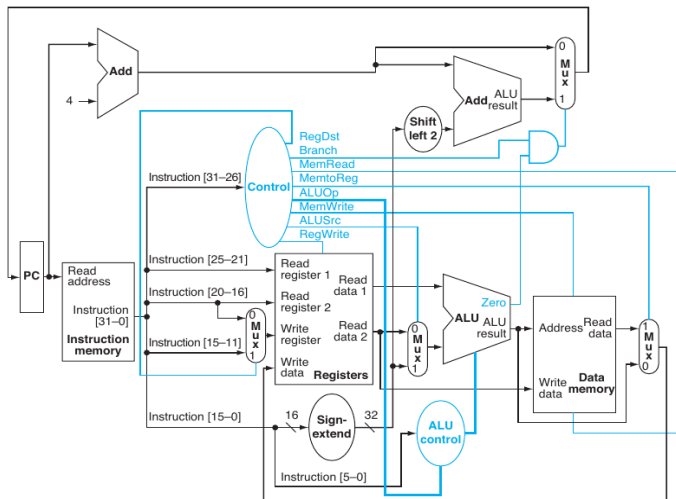




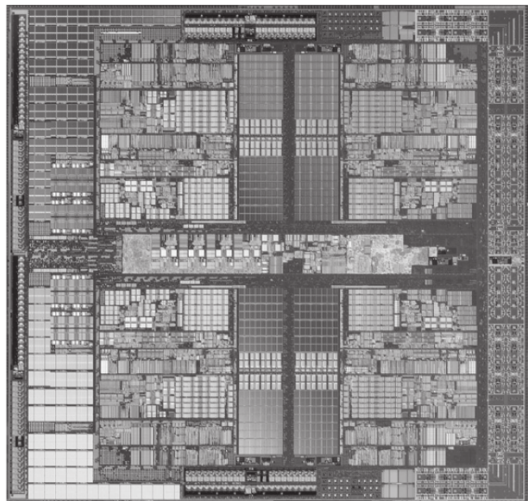


- وقتی برای یک معماری، Instruction Set نوشته می‌شود، به این معنی است که هر یک از instruction ها یک معنی می‌دهد، یک سری رجیستر خاص را نیاز دارد و باید از مسیر متفاوتی از داخل data path رد بشود،
- کدگشایی و کنترل کردن مسیر گذر یک instruction و داده‌هایش به عهده‌ی Control Unit است.

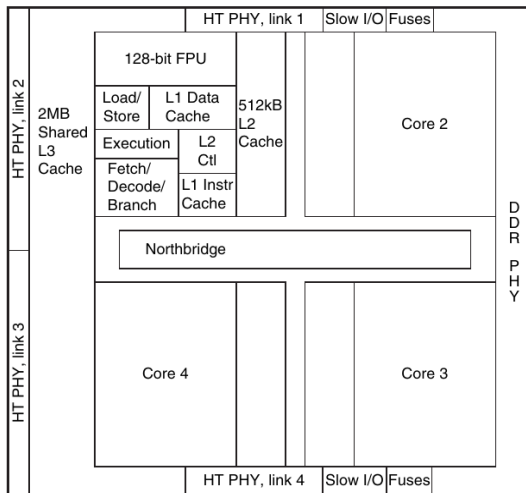
# Data Path and Control Unit



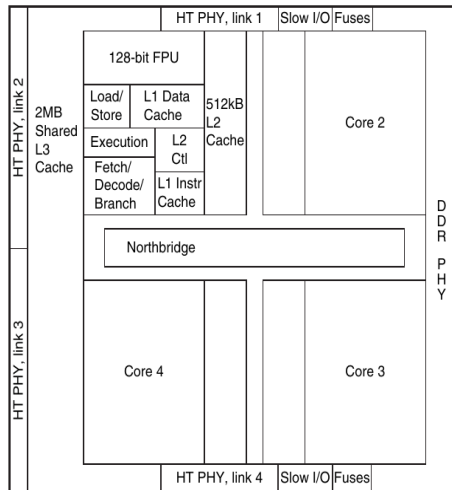
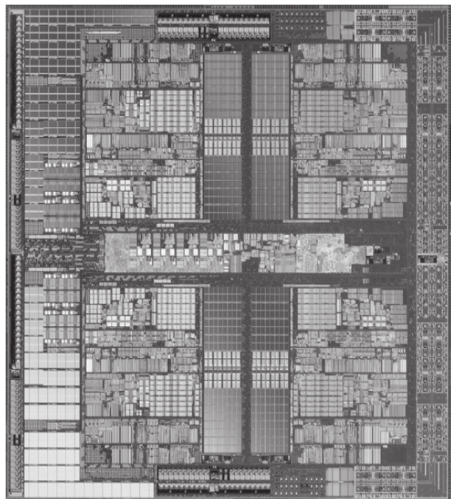
# AMD Barceloca Microprocessor



# AMD Barceloca Microprocessor Sketch



# AMD Barceloaca Microprocessor



## معماری‌های مختلف

– اگر شما الان بخواهید یک کامپیوتر بخرید، از بین معماری‌های مختلف دو معماری اصلی پیش روی شما هستند:

۱. x86

۲. ARM

- این معماری بر پایه‌ی معماری Inter 8008 که در سال ۱۹۷۲ معرفی شد، است.
- در واقع کدهایی که برای این معماری نوشته شده‌اند را می‌توان برای آخرین CPU های Intel یا AMD اسمبل و اجرا کرد.
- پس از Intel 8008 معماری‌های Intel 8088، 8086 16-bit و سپس 80186، 80286 و ... معرفی شدند و در کل x86 نام گرفتند.
- پردازنده‌های شرکت‌های Intel و AMD همگی بر پایه این معماری هستند.



- نام این معماری برگرفته از **Advanced RISC Machines** که قبل تر از **Acorn RISC Machine** گرفته شده بود، است.
- پردازنده‌های آرم، بخاطر
  - قیمت ارزان،
  - مصرف انرژی کم و
  - تولید گرمای کم
- برای دستگاه‌های سبک و دارای باتری، مثل تلفن‌های هوشمند و لپ‌تاپ‌ها بسیار مناسب هستند.
- بین سال‌های ۲۰۲۰ تا ۲۰۲۲ سریع‌ترین سوپر کامپیوتر دنیا (Fugako) هم از پردازنده‌های معماری آرم استفاده می‌کرد.
- چیپ‌های سری M شرکت اپل هم از معماری آرم استفاده می‌کنند.

# 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.

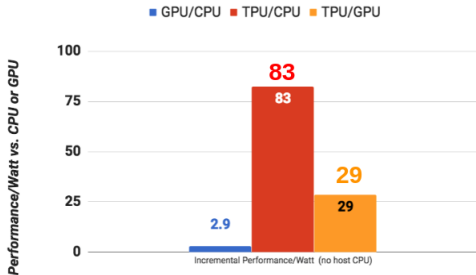
- با رخ دادن انقلابی در هوش مصنوعی به نام **یادگیری ماشین** که به ضرب ماتریسی متکی بود، نیاز به پردازنده‌های مخصوص ضرب تانسورها برای اجرا سریع‌تر و دقیق‌تر الگوریتم‌ها یادگیری ماشین احساس می‌شد،
- واحد پردازش تانسور (Tensor Processing Unit (TPU)) شتاب‌دهنده‌ی یادگیری ماشین است که توسط گوگل طراحی شده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

- 
- Figure 1: Dataflow of the proposed architecture. The diagram illustrates the data flow between various components. On the left, a 'PCIe Gen3 x16 Interface' (green) connects to a 'Host Interface' (blue) at 14 GiB/s. The 'Host Interface' connects to a 'Unified Buffer (Local Activation Storage)' (blue) at 10 GiB/s. The 'Unified Buffer' connects to a 'Systolic Data Setup' (blue) at 167 GiB/s. The 'Systolic Data Setup' connects to a 'Matrix Multiply Unit (64K per cycle)' (yellow) at 167 GiB/s. The 'Matrix Multiply Unit' connects to 'Accumulators' (blue), 'Activation' (yellow), and 'Normalize / Pool' (yellow). The 'Matrix Multiply Unit' also connects to a 'Weight FIFO (Weight Fetcher)' (blue) at 30 GiB/s. The 'Weight FIFO' connects to 'DDR3 DRAM Chips' (green) at 30 GiB/s. The 'DDR3 DRAM Chips' connect to 'DDR3 Interfaces' (green) at 30 GiB/s. The 'DDR3 Interfaces' connect to the 'Host Interface' at 14 GiB/s. The 'Host Interface' also connects to 'Control' (red) blocks. The 'Control' blocks are connected to the 'Unified Buffer', 'Systolic Data Setup', 'Matrix Multiply Unit', 'Accumulators', 'Activation', 'Normalize / Pool', and 'Weight FIFO'. A legend indicates: Green = Off-Chip I/O, Blue = Data Buffer, Yellow = Computation, Red = Control. A note at the bottom states 'Not to Scale'.

## 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
- 1<sup>st</sup> Results November 2018