NYU Tandon School of Engineering

Fall 2024, ECE 6913

Homework Assignment 2

Instructor: Azeez Bhavnagarwala, email: ajb20@nyu.edu

Course Assistant Office Hour Schedule

On Zoom: 9:30AM – 11AM Monday, Tuesday, Wednesday, Thursday, Friday Join

Zoom: https://nyu.zoom.us/j/99424200816

Aishwarya Kandam, ak11049@nyu.edu
Aditya Krishna, ak11137@nyu.edu
Srinatha Sivareddy, ss18364@nyu.edu
Manoj Srinivasan, ms14845@nyu.edu
Aditya Ojha, ao2612@nyu.edu

Homework Assignment 2 [released Sunday Sept 8th 2024] [due Friday Sept 27th 11:59PM]

You *are allowed* to discuss HW assignments with anyone. You are *not allowed* to share your solutions with other colleagues in the class. Please feel free to reach out to the Course Assistants or the Instructor during office hours or by appointment if you need any help with the HW.

Please enter your responses in this Word document after you download it from NYU Classes. *Please use the Brightspace portal to upload your completed HW.*

1. **Please read the ISSCC 2014 Keynote Publication by Professor Mark Horowitz**

   "Computing's Energy Problem (and what we can do about it)" [1]

1)How does Technology Scaling decrease the cost of Computing? How do reductions in the cost of manufacturing a transistor enable widespread use of computing devices?

Ans -1.1) Technology scaling, specifically through Moore's Law, has decreased the cost of computing by increasing the number of transistors on a chip without increasing the overall chip size. This exponential increase in transistor density enables higher performance at lower power per transistor. As transistors become cheaper to manufacture, the cost per operation reduces significantly. This decrease in cost allows computing devices to be integrated into a wide range of applications and products, such as smartphones, home appliances, and vehicles, making computing ubiquitous in daily life.

2) Why did scaling processor clock frequency become more difficult in the last 20 years? How did Power Density/Dissipation become the primary constraint on server CPU performance?

Ans - 1.2) In the 1990s, clock frequency increased faster than voltage could scale down, leading to an exponential increase in power dissipation, as power is proportional to both voltage squared (V^2) and frequency (f). The inability to lower the threshold voltage further due to leakage currents halted voltage scaling. Around the early 2000s, processors hit the "power wall" of approximately 100W, the limit for air cooling, which caused power density to become the limiting factor for performance scaling. As a result, further increases in clock frequency became infeasible without exceeding acceptable power and thermal limits.

3) Why is Moore's Law slowing down? Why did Dennard Scaling end?

Ans -1.3) Moore's Law is slowing down as feature size approaches physical and economic limits. Manufacturing smaller transistors becomes more difficult and costly. Dennard scaling, which predicted that as transistors shrink, their power density remains constant, broke down because voltage could no longer scale down at the same rate as feature size, leading to increased power density and heat dissipation issues. Leakage currents, which rise with smaller feature sizes, also contribute to the end of voltage scaling.

4) What component of energy consumption by Memory (in General Purpose Computing processors) is substantial ?

Ans - 1.4) The energy consumed by memory, especially off-die DRAM, is substantial in general-purpose computing processors. DRAM accesses consume significantly more energy compared to on-chip cache accesses or simple processor operations. For example, a DRAM access may cost 1-2nJ, while a cache access only costs 10pJ. This large energy disparity makes memory access one of the largest contributors to energy consumption in modern processors.

5) What solutions to Computing's Energy Problem does Professor Mark Horowitz's envision?

Ans -1.5) Professor Mark Horowitz suggests following solutions to computing's energy problem,

Specialization: Using specialized hardware accelerators for specific tasks that are orders of magnitude more efficient than general-purpose processors. This approach can lead to significant energy savings, particularly in convolution-like operations.

Energy-efficient memory systems: Reducing memory energy consumption by optimizing cache design to minimize access energy and introducing new technologies that lower the power cost of memory I/O.

Parallelism: While parallelism helps, it only provides limited energy savings. The key lies in lowering the energy per operation rather than just increasing the number of operations.

Application-hardware co-optimization: Creating better synergy between applications and the hardware that runs them, ensuring that hardware is well-matched to the task to reduce unnecessary energy consumption.

2) This assignment requires you to review 2 references on RISCV beginning with a summary transcript [2] of the Debate on Proprietary Vs Open-Source Instruction Sets at the 4th Workshop on Computer Architecture Research Directions, June 2015 sponsored by the ACM.

This Debate between Professor David Patterson (author of the textbook you are using) and Dave Christie of AMD highlights all of the key technical and business arguments for and against an Open-Source ISA such as RISC V as of 2015 (the same year the RISC V Foundation was established). A Technical Report from EECS UC Berkeley highlights the technical reasons for Open ISAs [3] providing a more detailed discussion on the advantages offered by open-source ISAs

(1) Articulate *your* views on the topics debated in [2]. Justify your views.

Ans- Articulating Views on Proprietary vs. Open-Source ISAs (Summary of Debate)

The debate between **David Patterson** and **Dave Christie** at the 4th Workshop on Computer Architecture Research Directions in 2015 centered on the merits of proprietary ISAs (like x86 and ARM) versus open-source ISAs (like RISC-V). Here are the key points and my perspective.

a) Proprietary vs. Open-Source ISAs

Christie's Argument:
Christie emphasized the importance of established ecosystems around proprietary ISAs. He argued that success comes not from the ISA itself but from the ecosystem of tools, compilers, and hardware support that grows around it. ARM and x86, for example, dominate due to this ecosystem rather than inherent technical superiority. He questioned whether an open-source ISA could develop the same level of commercial support.

Patterson's Argument:
Patterson advocated for open ISAs, pointing out that the barriers to entry for proprietary ISAs—high licensing fees and closed ecosystems—stifle innovation. Open ISAs like RISC-V democratize hardware design, enabling more developers, especially startups and academia, to contribute to processor design. He argued that RISC-V's modular, extensible architecture provides flexibility that proprietary ISAs lack.

My View:
While proprietary ISAs have thrived due to well-established ecosystems, open ISAs represent the future of innovation. RISC-V's flexibility and accessibility are critical in today's environment, where domain-specific accelerators and IoT devices demand customizability. Open ISAs empower a wider base of developers to experiment and innovate without legal or financial constraints.

b) Ecosystem and Innovation

Christie's Perspective:
Christie argued that proprietary ISAs build stronger ecosystems because companies have a vested interest in maintaining them. He suggested that the barriers to developing an open ISA ecosystem—especially reaching critical mass in terms of software and tool support—are too high for open-source alternatives like RISC-V to compete effectively.

Patterson's Perspective:

Patterson countered that open-source software ecosystems, such as Linux, have succeeded without initial commercial backing. He pointed to RISC-V's growing momentum, supported by companies and academic institutions, as proof that open ISAs can foster innovation and create robust ecosystems.

My View:
Open ISAs are still in their early stages, but the collaborative nature of open-source development offers a compelling advantage. The success of Linux and other open-source platforms shows that community-driven ecosystems can rival or surpass proprietary ones. RISC-V is already gaining traction, and its flexibility in supporting domain-specific hardware will only accelerate its adoption in diverse industries.

c) ISA Lock-in and Specialization

Christie's Argument:Christie downplayed concerns about ISA lock-in, noting that modern software has become more ISA-agnostic, making it easier to switch between architectures. He questioned whether we need another ISA like RISC-V when ARM and x86 already serve most market needs.

Patterson's Argument:
Patterson highlighted the limitations of proprietary ISAs, which are often controlled by a few companies, limiting who can design with them. In contrast, open-source ISAs like RISC-V provide long-term stability and avoid the risk of being tied to the fortunes of a single company. RISC-V's modularity makes it well-suited for domain-specific computing, an increasingly critical aspect in IoT and AI.

My View:
As the computing landscape shifts toward specialization, open ISAs provide an essential framework for customization. Proprietary ISAs often include legacy features, limiting their adaptability to specific applications. RISC-V's open, modular architecture allows designers to tailor the ISA to their exact needs, making it the ideal solution for the growing trend toward heterogeneous computing.

(2) Review and summarize technical reasons for Open-Source ISAs in [3].

Ans- The UC Berkeley Technical Report *"Instruction Sets Should Be Free: The Case For RISC-V"* outlines below technical reasons for adopting open-source ISAs like RISC-V.

a) Barriers of Proprietary ISAs
Proprietary ISAs like ARM and x86 are closed ecosystems that impose significant licensing costs and legal restrictions. These barriers limit innovation by restricting ISA design to a small number of companies. Additionally, licensing negotiations can be slow, further hindering time-to-market for new designs. The report argues that open ISAs remove these barriers, fostering a more competitive and innovative environment.

b) Innovation through Free ISAs
Open-source ISAs allow more developers, from startups to academia, to experiment and innovate. Shared open-core designs lower development costs, reduce time-to-market, and improve security by making hardware designs transparent. RISC-V, for instance, is already being used in various projects and has a growing ecosystem of tools and software.

c) Flexibility and Future-Proofing
RISC-V's modular architecture supports a base-plus-extension model, enabling flexibility for specific applications while maintaining a stable core. This flexibility makes RISC-V ideal for a wide range of devices, from IoT to data centers. Additionally, RISC-V is designed to accommodate future needs, such as 128-bit addressing and quadruple-precision floating point, ensuring its relevance for decades to come.

d) Scalability and Cost Efficiency
RISC-V is particularly well-suited for IoT devices, where cost and power efficiency are critical. By offering an open, scalable ISA, RISC-V makes it feasible to design processors for devices that need to cost as little as $1. This cost efficiency is essential as the number of connected IoT devices continues to grow.

References

[1] M Horowitz, *"Computing's Energy Problem (and what we can do about it)"* Plenary Session 1.1, 2014 ISSCC Digest of Tech. papers, Feb 2014 [*PDF attached*]

[2] M Hill et al, "Proprietary Versus Open Instruction Sets" 4th Workshop on Computer Architecture Research Directions, June 2015, ACM [*PDF attached*]

[3] K Asanovic et al, "Instruction Sets Should Be Free: The Case For RISC-V", EECS, University of California Berkeley, Technical Report No. UCB/EECS-2014-146, Aug 6 2014 [*PDF attached*]