. 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?

As technology advances, computing becomes more accessible. Transistors continue to shrink, allowing them to be embedded in almost any device. Many everyday items—such as homes, cars, glasses, and headphones—are now equipped with computing capabilities due to the miniaturization of chips. This makes integrating computing into daily life much simpler. Additionally, as the cost of transistors decreases, various products can incorporate chips at affordable prices, enabling them to be networked and transformed into smart devices.

* 1. 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?

There are two main reasons why increasing processor clock frequency is becoming more challenging: power density and dissipation. The limits of air cooling and slower voltage regulation have become bottlenecks. To manage power consumption, we lowered processor frequencies and increased the number of cores per chip. However, adding more cores raises power density, making it harder to fit additional transistors. Furthermore, higher clock frequencies lead to greater power dissipation, and this limitation has become the primary factor restricting server CPU performance.

* 1. Why is Moore’s Law slowing down? Why did Dennard Scaling end?

Moore's Law, proposed by Intel co-founder Gordon Moore in 1965, observes that the number of transistors on a microchip doubles roughly every two years, resulting in increased computing power. While this trend held true for a period, transistors cannot be miniaturized indefinitely. Once they reach a certain size, further reduction becomes physically impossible, eventually slowing down Moore's Law. Similarly, Dennard scaling, which suggests that as transistors shrink, both voltage and power usage should decrease proportionally, has faced challenges. The issue lies in the fact that power density didn't scale as expected due to two reasons: supply voltage wasn't reduced at the necessary constant rate, and clock frequencies were increased faster than the scaling allowed.

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

The majority of memory energy consumption arises from DRAM access and the cache hierarchy, with a significant portion attributed to energy used by the Last Level Cache (LLC). DRAM access is particularly energy-intensive, consuming around 1 to 2 joules per operation, which is substantially more than the energy required for internal processor operations or internal memory access. Additionally, DRAM is relatively slow and less efficient in terms of processing. The large LLC plays a crucial role in reducing the need for frequent DRAM access in systems with extensive caches.

* 1. What solutions to Computing’s Energy Problem does Professor Mark Horowitz’s envision?

Professor Horowitz emphasized the importance of enabling more application experts to take part in designing the efficient hardware and software systems they require. He advocates for experts to define the specific hardware that best fits their needs and to be directly involved in its design. Additionally, he advised against relying solely on general-purpose processors, suggesting that systems should be optimized for particular applications to achieve more efficient collaboration between hardware and software, ultimately reducing energy consumption. His approach encourages hardware and software engineers to create chip architectures tailored to their needs, even using software programming to generate hardware circuits. I believe this idea is groundbreaking and holds significant long-term potential.

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.

During the debate, Dave supported proprietary ISAs, arguing that they provide stability and a strong ecosystem, allowing the ISA to better meet customer needs while driving product improvement through competition with other proprietary ISAs.

David, however, favored Open-Source ISAs, noting that proprietary designs limit functionality and pricing. He believes that open ISAs can foster innovation, lower costs, reduce errors, and invite broader participation, which would also support academic research.

I also support Open-Source ISAs. While Dave highlights the strong ecosystem of proprietary ISAs, open-source ones can be regulated with industry standards and still build their own ecosystem. They enable people with different needs to design tailored architectures, which can be shared online for debugging and cost reduction. Dave’s concerns about control can be addressed through clear industry guidelines and safety standards.

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

This article explores the necessity and benefits of open-source ISAs. It begins by highlighting that while Linux revolutionized the computer industry, the key interface, ISA, remains proprietary. In Part 2, the author outlines several issues with proprietary ISAs, including patent restrictions that hinder innovation, the untapped potential of skilled designers, and the lack of strong optimization. Additionally, proprietary ISAs may vanish if the company behind them goes bankrupt.

The author argues that ISAs should be as open and free as the Linux system, proposing RISC-V as the ideal open-source option. RISC-V offers several advantages: it features a modular design with a small core instruction set that can be customized with standard extensions. Its instructions are more compact, saving space, and it also supports a 128-bit address space, making it suitable for future large-scale data storage needs, particularly in cloud computing and other fields.