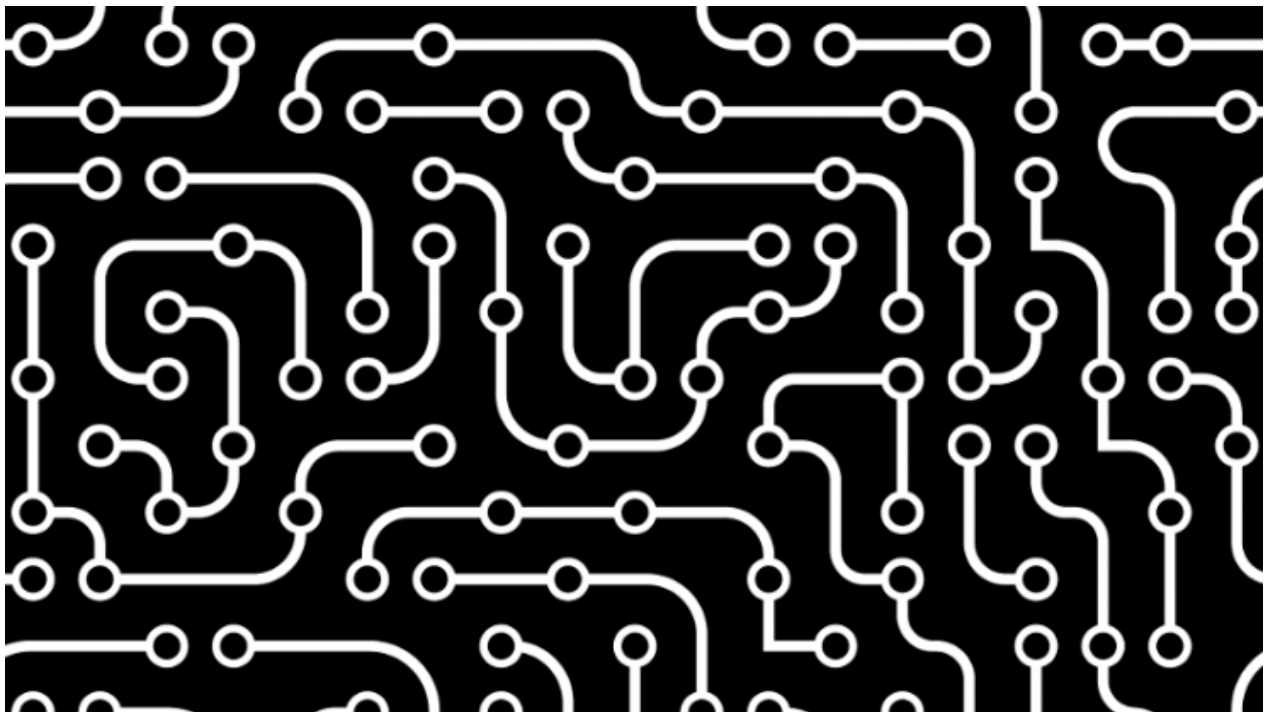# AI Models Microprocessor Performance in Real-Time

DECEMBER 10, 2021 | BY KEN KINGERY

New algorithm predicts processor power consumption trillions of times per second while requiring little power or circuitry of its own

Computer engineers at Duke University have developed a new AI method for accurately predicting the power consumption of any type of computer processor more than a trillion times per second while barely using any computational power itself. Dubbed APOLLO, the technique has been validated on real-world, high-performance microprocessors and could help improve the efficiency and inform the development of new microprocessors.

The approach is detailed in a paper published at MICRO-54: 54th Annual IEEE/ACM International Symposium on Microarchitecture, one of the top-tier conferences in computer architecture, where it was selected the conference's best publication.

microprocessor in the background, which opens many new opportunities. I think that's why people are excited about it."

In modern computer processors, cycles of computations are made on the order of 3 trillion times per second. Keeping track of the power consumed by such intensely fast transitions is important to maintain the entire chip's performance and efficiency. If a processor draws too much power, it can overheat and cause damage. Sudden swings in power demand can cause internal electromagnetic complications that can slow the entire processor down.

By implementing software that can predict and stop these undesirable extremes from happening, computer engineers can protect their hardware and increase its performance. But such schemes come at a cost. Keeping pace with modern microprocessors typically requires precious extra hardware and computational power.

"APOLLO approaches an ideal power estimation algorithm that is both accurate and fast and can easily be built into a processing core at a low power cost. And because it can be used in any type of processing unit, it could become a common component in future chip design."

ZHIYAO XIE

"APOLLO approaches an ideal power estimation algorithm that is both accurate and fast and can easily be built into a processing core at a low power cost," Xie said. "And because it can be used in any type of processing unit, it could become a common component in future chip design."

The secret to APOLLO's power comes from artificial intelligence. The algorithm developed by Xie and Chen uses AI to identify and select just 100 of a processor's millions of signals that correlate most closely with its power consumption. It then builds a power consumption model off of those 100 signals and monitors them to predict the entire chip's performance in real-time.

Because this learning process is autonomous and data driven, it can be implemented on most any computer processor architecture — even those that have yet to be invented. And while it doesn't require any human designer expertise to do its job, the algorithm could help human designers do theirs.

The work is part of a collaboration with Arm Research, a computer engineering research organization that aims to analyze the disruptions impacting industry and create advanced solutions, many years ahead of deployment. With the help of Arm Research, APOLLO has already been validated on some of today's highest performing processors. But according to the researchers, the algorithm still needs testing and comprehensive evaluations on many more platforms before it would be adopted by commercial computer manufacturers.

"Arm Research works with and receives funding from some of the biggest names in the industry, like Intel and IBM, and predicting power consumption is one of their major priorities," Chen added. "Projects like this offer our students an opportunity to work with these industry leaders, and these are the types of results that make them want to continue working with and hiring Duke graduates."

CITATION: "APOLLO: An Automated Power Modeling Framework for Runtime Power Introspection in High-Volume Commercial Microprocessors." Zhiyao Xie, Xiaoqing Xu, Matt Walker, Joshua Knebel, Kumaraguru Palaniswamy, Nicolas Hebert, Jiang Hu, Huanrui Yang, Yiran Chen, Shidhartha Das. MICRO-54, October, 2021. DOI: 10.1145/3466752.3480064

Online: https://dl.acm.org/doi/abs/10.1145/3466752.3480064 (https://dl.acm.org/doi/abs/10.1145/3466752.3480064)