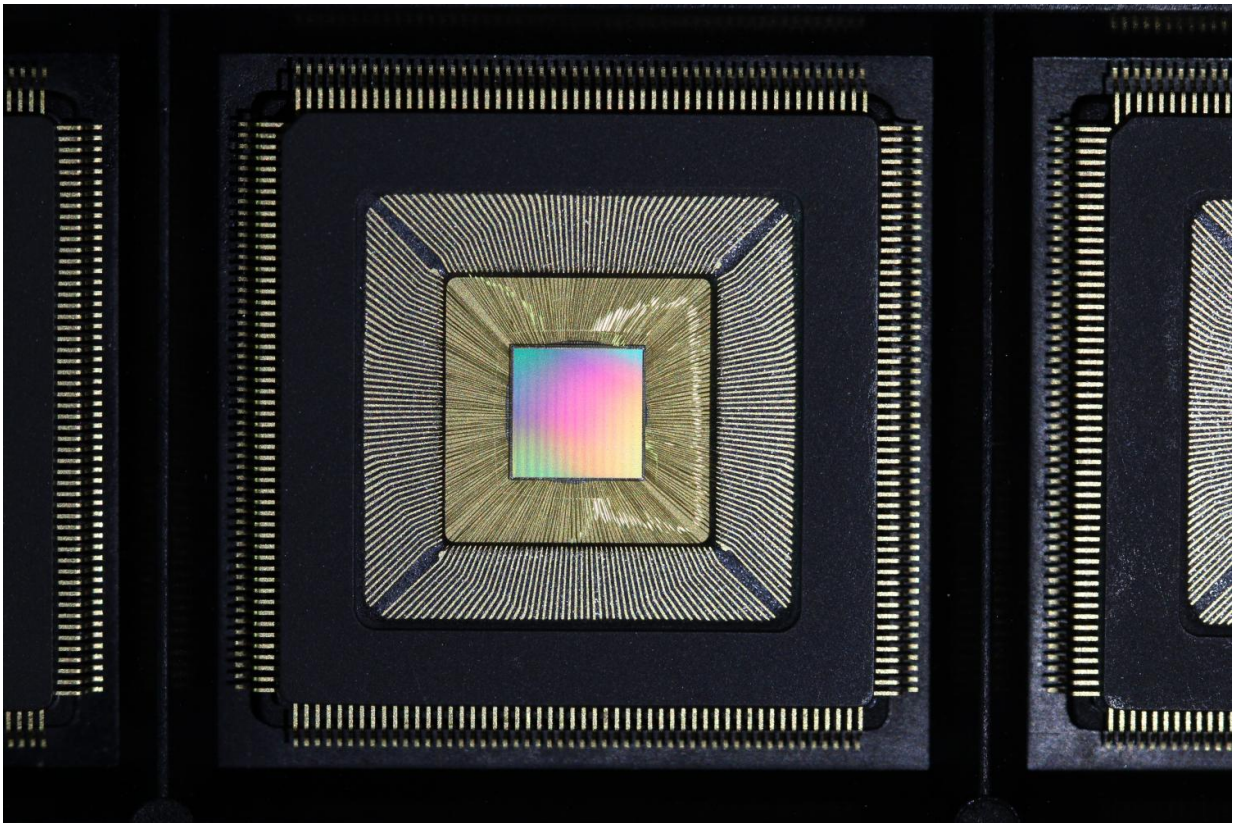


New microchip demonstrates efficiency and scalable design

August 23 2016



Unidded Princeton Piton Processor. Multicolored die is wirebonded inside chip package. Credit: Princeton University

Princeton University researchers have built a new computer chip that promises to boost performance of data centers that lie at the core of

online services from email to social media.

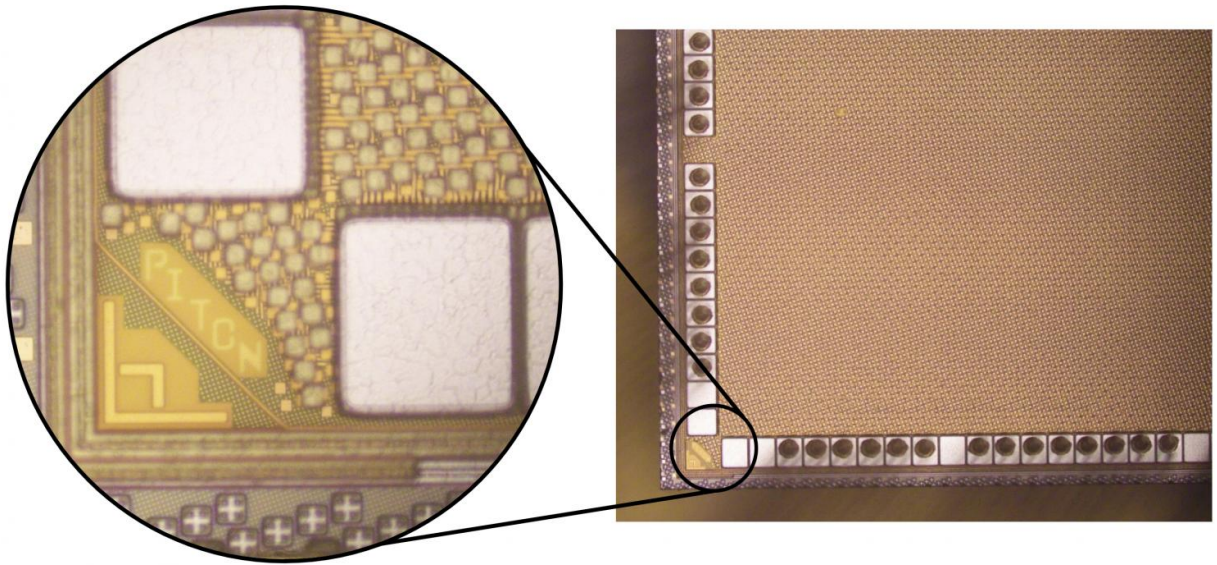
Data centers - essentially giant warehouses packed with computer servers - enable cloud-based services, such as Gmail and Facebook, as well as store the staggeringly voluminous content available via the internet. Surprisingly, the computer chips at the hearts of the biggest servers that route and process information often differ little from the chips in smaller servers or everyday personal computers.

By designing their [chip](#) specifically for massive computing systems, the Princeton researchers say they can substantially increase processing speed while slashing energy needs. The [chip architecture](#) is scalable; designs can be built that go from a dozen processing units (called cores) to several thousand. Also, the architecture enables thousands of chips to be connected together into a single system containing millions of cores. Called Piton, after the metal spikes driven by rock climbers into mountainsides to aid in their ascent, it is designed to scale.

"With Piton, we really sat down and rethought computer architecture in order to build a chip specifically for data centers and the cloud," said David Wentzlaff, an assistant professor of electrical engineering and associated faculty in the Department of Computer Science at Princeton University. "The chip we've made is among the largest chips ever built in academia and it shows how servers could run far more efficiently and cheaply."

Wentzlaff's graduate student, Michael McKeown, will give a presentation about the Piton project Tuesday, Aug. 23, at [Hot Chips](#), a symposium on high performance chips in Cupertino, California. The unveiling of the chip is a culmination of years of effort by Wentzlaff and his students. Mohammad Shahradd, a graduate student in Wentzlaff's Princeton Parallel Group said that creating "a physical piece of hardware in an academic setting is a rare and very special opportunity for

computer architects."



A microscope photo of the corner of the Princeton Piton Processor with zoom in on Piton logo (Inset). Credit: Princeton University

Other Princeton researchers involved in the project since its 2013 inception are Yaosheng Fu, Tri Nguyen, Yanqi Zhou, Jonathan Balkind, Alexey Lavrov, Matthew Matl, Xiaohua Liang, and Samuel Payne, who is now at NVIDIA. The Princeton team designed the Piton chip, which was manufactured for the research team by IBM. Primary funding for the project has come from the National Science Foundation, the Defense Advanced Research Projects Agency, and the Air Force Office of Scientific Research.

The current version of the Piton chip measures six by six millimeters. The chip has over 460 million transistors, each of which are as small as 32 nanometers - too small to be seen by anything but an electron

microscope. The bulk of these transistors are contained in 25 cores, the independent processors that carry out the instructions in a computer program. Most personal computer chips have four or eight cores. In general, more cores mean faster processing times, so long as software ably exploits the hardware's available cores to run operations in parallel. Therefore, computer manufacturers have turned to multi-core chips to squeeze further gains out of conventional approaches to computer hardware.

In recent years companies and academic institutions have produced chips with many dozens of cores; but Wentzlaff said the readily scalable architecture of Piton can enable thousands of cores on a single chip with half a billion cores in the data center.

"What we have with Piton is really a prototype for future commercial server systems that could take advantage of a tremendous number of cores to speed up processing," said Wentzlaff.

The Piton chip's design focuses on exploiting commonality among programs running simultaneously on the same chip. One method to do this is called execution drafting. It works very much like the drafting in bicycle racing, when cyclists conserve energy behind a lead rider who cuts through the air, creating a slipstream.



An annotated CAD tool layout of the Princeton Piton Processor showing 25 cores. Credit: Princeton University

At a data center, multiple users often run programs that rely on similar operations at the processor level. The Piton chip's cores can recognize these instances and execute identical instructions consecutively, so that

they flow one after another, like a line of drafting cyclists. Doing so can increase energy efficiency by about 20 percent compared to a standard core, the researchers said.

A second innovation incorporated into the Piton chip parcels out when competing programs access computer memory that exists off of the chip. Called a memory traffic shaper, this function acts like a traffic cop at a busy intersection, considering each programs' needs and adjusting memory requests and waving them through appropriately so they do not clog the system. This approach can yield an 18 percent performance jump compared to conventional allocation.

The Piton chip also gains efficiency by its management of memory stored on the chip itself. This memory, known as the cache memory, is the fastest in the computer and used for frequently accessed information. In most designs, cache memory is shared across all of the chip's cores. But that strategy can backfire when multiple cores access and modify the cache memory. Piton sidesteps this problem by assigning areas of the cache and specific cores to dedicated applications. The researchers say the system can increase efficiency by 29 percent when applied to a 1,024-[core](#) architecture. They estimate that this savings would multiply as the system is deployed across millions of cores in a [data center](#).

The researchers said these improvements could be implemented while keeping costs in line with current manufacturing standards. To hasten further developments leveraging and extending the Piton architecture, the Princeton researchers have made its design open source and thus available to the public and fellow researchers at the OpenPiton website: <http://www.openpiton.org>

"We're very pleased with all that we've achieved with Piton in an academic setting, where there are far fewer resources than at large, commercial chipmakers," said Wentzlaff. "We're also happy to give out

our design to the world as open source, which has long been commonplace for software, but is almost never done for hardware."

More information: parallel.princeton.edu/piton/

Provided by Princeton University

Citation: New microchip demonstrates efficiency and scalable design (2016, August 23)
retrieved 3 May 2024 from

<https://techxplore.com/news/2016-08-microchip-efficiency-scalable.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--