



作者: 张双虎 来源: 中国科学报 发布时间: 2021/12/21 11:29:23 选择字号: 小中大

阿波罗：芯片测“功”唯快不破

随着集成电路技术的不断发展，芯片规模越来越大、集成度越来越高。与此同时，在设计和使用芯片时，掌握并管理其功耗就变得越来越大。

为在设计或使用过程中管理好芯片功耗，杜克大学电子与计算机工程系教授、杜克大学计算进化智能实验室联合主任陈怡然团队开发出一种新的人工智能模型——阿波罗（APOLLO）。该模型在硬件上可进行每秒数十亿次计算，理论上能用于预测任何类型计算机处理器(CPU)的功耗，且仅需要极低的额外硬件开销。它能提高处理器效率并为人们开发新型微处理器提供帮助，相关性已在高性能微处理器上得到验证。

日前，该相关成果在第54届IEEE/ACM年度微体系结构国际研讨会（MICRO-54）上发布，并获本届唯一“最佳论文奖”。

功耗难题

在现代计算机处理器中，计算频率可达每秒数十亿次。随着晶体管密度和时钟频率显著增加，功耗也急剧增加。

功耗越大，芯片产生热量越快。如果不能及时散热，设备温度就会升高，这将导致设备不能正常工作甚至损坏。同时芯片内部的功耗与电流需求突然波动，会导致内部电磁问题，从而影响处理器的速度。此外，高功耗产生热量也对封装、散热系统等提出了更高要求。因此，功耗管理已经成为贯穿芯片设计和应用全流程的问题。

“CPU运行中涉及非常多的信号，其功耗每时每刻都在发生变化。”该论文第一作者、陈怡然团队成员谢知遥告诉《中国科学报》，“因此，人们设计CPU、或CPU运行时，需要对其功耗有非常充分的了解，这样才能设计出好的CPU，让其发挥最佳性能。”

在设计芯片时，设计工程师们往往依赖行业标准的功耗分析工具。比如，根据单个信号网络的切换以及这些网络驱动的电容性负载进行计算。“人们想了解芯片的功耗，通常的方法是进行模拟。这类方法比较准确，但缺点是计算成本很高。”谢知遥解释说，“模拟的方法非常多，相关的模拟软件技术上也很成熟，但它模拟计算起来很慢，不能满足人们对掌握芯片即时功耗的需求。”

在实际应用中，CPU需要运行各种不同的程序，每个程序对应的功耗各不相同；即使运行同一个程序，在不同时刻，芯片的功耗也会发生瞬时变化。CPU被设计并制造出来后，如果全新场景的实际功耗比预先设计的高很多，就可能出现功耗问题，但要详细掌握芯片在不同工况下的即时功耗并非易事。

“在过去20多年里，功耗模型已被反复研究，但准确、快速、低成本、自动化地进行功耗分析仍然难以实现。”陈怡然告诉《中国科学报》，“由于很多功耗模型是设计师们针对某一款CPU人工调试而成的，这造成了巨大的人力成本。同时，随着CPU设计日趋复杂，想要人工设计准确的功耗模型变得越来越困难。”

因“简”而快

“这种方法的核心是个非常简洁的功耗模型。”谢知遥说，“和以前通过模拟计算的方式不同，我们这个模型并非进行准确的模拟，而是进行快速的估算。”

处理器工作时，芯片内部涉及数百万个关键信号，该模型通过人工智能技术，自动选取极少量（约100个）与功耗最相关的CPU信号作为输入，然后建立一个快速的线性模型，用于对每个周期的功耗进行预测或监测。

“这100个典型信号可以形成一个‘轻量级’的、非常简洁的功耗模型。”谢知遥强调说，“而且，信号选取（不是随机抽取）由人工智能自动完成，它摆脱了对工程师的依赖，从而让模型的计算速度大大提升。”



- 相关新闻 相关论文
- 1 宇宙膨胀速度超预期添新证
 - 2 七个物理量子位组成的逻辑量子位实现
 - 3 本科生导师制探索重在“教研相长”
 - 4 南京大学：为新时代培养拔尖领军人才
 - 5 “中国天眼”牵手贵州大学破解大数据难题
 - 6 多地发现奥密克戎后疫情防控有哪些变化？
 - 7 卓越工程师孟祥飞：矢志造重器 为国增底气
 - 8 研究发现水产养殖限用抗生素可降低生态风险



- 一周新闻排行
- 1 我国风湿免疫学先驱和开拓者蒋明逝世
 - 2 北京大学教授童坦君意外摔倒受伤 医治无效逝世
 - 3 《自然》预测2023年科学大事
 - 4 陈嘉庚科学奖和陈嘉庚青年科学奖获奖名单公布
 - 5 共和国一日痛别两院士
 - 6 2022年中科院期刊分区表发布
 - 7 魏龙：精细与有恒的核技术追光者
 - 8 研究发现一种与严重儿童肥胖有关的新的遗传机制
 - 9 穿越疾风暴雨，国产大飞机C919腾飞背后
 - 10 一种神经影像特征可用于预测药物和食物渴求强度

- 编辑推荐博文
- IJERPH: 担心被自动化取代，工人们焦虑得夜不能寐
 - 2022年，一个充满大事件的一年
 - 美国加州一名音响工程师的哲思（139）
 - 2022年最新版《狂犬病》专著选介（19）
 - 科学家根据需要顺时针或逆时针转动单个分子
 - 太阳系的74Se来源与超新星爆发过程
- 更多>>

CPU在运行的时候，人们会对功耗进行控制，如果功耗过高，就必须想办法将它降下来。事实上，工程师们也积累了不少办法，可以让功耗降下来，但所有这些需要一个前提，就是人们能够快速且准确地掌握芯片功耗的即时变化。

遗憾的是，在计算功耗时，工程师们常面临一个无法跨越的时间“窗口”。比如，之前快速功耗模型的计算速度可能是每毫秒进行一次——人们只能估算一个1毫秒的功耗平均值。

除信号选取和模型训练是自动完成之外，阿波罗的训练数据也通过算法自动生成，这让其因“简”而快成为可能。

论文表明，阿波罗可以在几分钟内获得几千万时钟周期的功耗，而传统基于emulator的工业界方法需要长达两个星期时间。在准确率方面，阿波罗也可以可以达到90%至95%。另外，阿波罗的功耗分析可以精确到每个时钟周期，之前任何方法在这种速度下都无法获得这样高分辨率的功耗分析。

“每毫秒计算一次的时间分辨率远远不够，因为在1毫秒之内，有可能其功耗前半程很低，后半程很高，虽然我们计算出了功耗数值，但它并不能反映处理器的真实工作状态。”谢知遥说补充说，“我们的模型可以控制在0.3纳秒（0.000001 毫秒）。就是每0.3纳秒，就可以计算一次实时功耗值。”

面向应用

因为以前无法了解处理器的即时功耗，工程师们在进行芯片设计时，为避免应用中功耗波动带来的影响，往往会为功耗设计一定的余量。

现在，阿波罗可以在极短时间内，对芯片内部的即时功耗变化进行反馈。即使芯片内部由于复杂的交互出现了电压的快速波动，工程师们也能及时采取措施，对CPU功耗和温度进行管理，从而发挥CPU的最佳性能，并保障其运行稳定。

“因为阿波罗非常简洁，它的成本和面积都很小（约占用CPU芯片0.2%的面积），所以可以把它做成硬件，整合进芯片内部。”谢知遥说，“这样，在使用过程中，一旦监测到处理器功耗的异常变化，我们就可以让它采取一些自我调节措施，把功耗降下来。”

如果在设计芯片时，加入这样的功耗管理模块，不但性能可以得到提升，其运行的稳定性也能得到保证。

在该项研究中，ARM公司提供了很多帮助，目前这一模型已经在ARM的多款高性能商业处理器上进行过验证，显示了较领先的芯片性能。

“这不是一项纯理论研究，而是具备非常强的工程性研究特点，或者说是一项几乎就能应用在产品里的研究。”陈怡然说，“在商业化应用之前，该模型可能还需要在更多的平台上进行测试和综合评估。至于它能不能最终落地，什么时间能成为产品，会受到多种因素影响，还需要时间来证明。”

相关论文信息：

<https://doi.org/10.1145/3466752.3480064>

版权声明：凡本网注明“来源：中国科学报、科学网、科学新闻杂志”的所有作品，网站转载，请在正文上方注明来源和作者，且不得对内容作实质性改动；微信公众号、头条号等新媒体平台，转载请联系授权。邮箱：shouquan@stimes.cn。

打印 发E-mail给:

关于我们 | 网站声明 | 服务条款 | 联系方式 | 中国科学报社 京ICP备07017567号-12 京公网安备 11010802032783

Copyright © 2007-2022 中国科学报社 All Rights Reserved

地址：北京市海淀区中关村南一条乙三号

电话：010-62580783