Arm十年最大更新：V9架构正式发布

https://mp.weixin.qq.com/s/WGJoQJfmsRS6nDk6nx0omA

None Tue Mar, 30 23:59

来源：内容由半导体行业观察（ID:icbank）编译自「anandtech」，谢谢。

自Arm在2011年10月首次发布Armv8架构以来，已经过去了近十年的时间。这对Arm来说是一个相当可观的十年，因为在这段时间内，他们的指令集架构受到移动市场和服务器市场的高度关注，并铆足劲在包括笔记本电脑和台式机设备市场发力。过去多年里，Arm对ISA进行了改进，也对体系结构进行了各种更新和扩展。当中一些可能很重要，有些可能也是一瞥而过。

今天，作为Arm的Vision Day活动的一部分，该公司正式发布了该公司的新一代Armv9架构的首个细节，为Arm未来十年内成为下一个3000亿芯片的计算平台奠定了基础。

读者可能会问的一个大问题是，Armv9与Armv8究竟有何不同，能让架构获得如此大的提升。

确实，从纯粹的ISA角度来看，v9可能不像v8相比v7那样实现根本性的跳跃，后者引入了AArch64，一个完全不同的执行模式和指令集，该指令集与AArch32相比具有更大的微体系结构分支，例如扩展寄存器，64位虚拟地址空间和更多改进。

Armv9继续使用AArch64作为基准指令集，但是在其功能上增加了一些非常重要的扩展，以保证architecture numbering的增加，并且允许Arm不仅可以获得对AArch64进行某种软件重新基准化v9的新功能，还能保持我们多年来在v8上获得的扩展。

Arm认为新架构Armv9有三个主要支柱，即安全性、AI以及改进的矢量和DSP功能。对于v9，安全性是一个非常重要的主题，我们将深入探讨新扩展和功能的新细节，但是首先谈到的DSP和AI功能应该很简单。

新的Armv9兼容CPU所承诺的最大的新功能可能是开发人员和用户可以立即看到的——SVE2作为NEON的后继产品。

可伸缩矢量扩展（SVE）的于2016年首次亮相，并首次在富士通的A64FX CPU内核中实现，该芯片已为日本排名第一的超级计算机Fukagu提供支持。SVE的问题在于，新的可变矢量长度SIMD指令集的第一次迭代的范围相当有限，并且更多地针对HPC工作负载，缺少了许多仍由NEON涵盖的更通用的指令。

SVE2于2019年4月发布，旨在通过用所需指令补充新的可扩展SIMD指令集来解决此问题，以服务于类似DSP等目前仍在使用NEON的工作负载。

除了增加的各种现代SIMD功能外，SVE和SVE2的优势还在于其可变的向量大小，范围覆盖了128b到2048b，让其无论在什么硬件运行，都允许向量的可变粒度为128b。如果纯粹从向量处理和编程的角度来看，这意味着软件开发人员将只需要编译一次其代码，并且如果将来某个CPU带有本地的512b SIMD execution pipelines，该代码将能够充分利用单元的整个宽度。同样，相同的代码将能够在具有较低硬件执行宽度能力的保守设计上运行，这对于Arm设计从物联网、移动到数据中心的CPU而言至关重要。在保留Arm体系结构的32b编码空间的同时，它还可以完成所有这些工作。然而类似X86这样的架构则需要根据矢量尺寸增加新的指令和扩展。

机器学习也被视为Armv9的重要组成部分，因为Arm认为在未来几年中，越来越多的ML工作负载将变得司空见惯，当中包括了对性能或电源效率有至关重要要求的场景中。那就让在专用加速器上运行ML工作负载变成长久的需要，与此同时，我们还会继续在CPU上运行较小范围的ML工作负载。

矩阵乘法指令（Matrix multiplication instructions）是此处的关键，它将代表生态系统中将更大范围采用v9 CPU作为基本功能所迈出的重要一步。

通常，我认为SVE2可能是保证升级到v9的最重要因素，因为它是更确定的ISA功能，可以在日常使用中与v8 CPU区别开来，并且可以保证软件生态系统能够正常运行，这与现有的v8堆栈有所不同。对于服务器领域的Arm来说，这实际上已经成为一个相当大的问题，因为软件生态系统仍在基于v8.0的软件包基础上，不幸的是，该软件包缺少了最重要的v8.1大型系统扩展。

使整个软件生态系统向前发展，并假设新的v9硬件具有新的体系结构扩展功能，这将有助于推动事情发展，并可能解决某些当前情况。

但是，v9不仅涉及SVE2和新指令，它还非常注重安全性，在安全性方面我们将看到一些更根本的变化。

介绍机密的计算架构

在过去的几年中，安全性和硬件安全性漏洞已成为芯片行业的头等大事，Spectre，Meltdown等漏洞的出现及其所有同级边信道攻击都表明，重新思考如何保证安全成为了一个基本需求。

Arm希望用来解决这一总体问题的方法是通过引入Arm机密计算体系结构（Arm Confidential Compute Architecture：CAA）来重新设计安全应用程序的工作方式。

在继续之前，我想提箱一下，今天的披露仅仅是对新CCA运作方式的高层次解释，Arm说，有关新安全机制的确切工作原理的更多细节将在今年夏天的晚些时候公布。

CCA的目标是从当前的软件堆栈情况中获得更大的收益，在当前的软件堆栈情况下，在设备上运行的应用程序必须固有地信任它们所运行的操作系统和虚拟机管理程序。传统的安全模型是基于以下事实建立的：更高特权的软件层被允许查看较低层的执行，然而当操作系统或系统管理程序被以任何方式损害时，这就可能成为了一个问题。

CCA引入了动态创建“realms”的新概念，可以将其视为对OS或虚拟机管理程序完全不透明的安全容器化执行环境。系统管理程序将仍然存在，但仅负责调度和资源分配。而“realm”将由称为“ealm manager”的新实体管理，其被认为是一段新的代码，大致大小约为hypervisor的1/10。realm内的应用程序将能够“证明”领域管理器以确定其是否可信任，这对于传统的虚拟机管理程序而言是不可能的。

Arm并没有深入探讨究竟是什么造成了realm与操作系统和虚拟机管理程序的非安全世界之间的这种隔离，但听起来确实像硬件支持的地址空间，但它们无法相互交互。

使用realms的优势在于，它极大地减少了设备上运行的给定应用程序的信任链，并且OS对安全性问题变得越来越透明。与当今需要企业或企业使用带有授权软件堆栈的专用设备的情况相反，需要监督控制的关键任务应用程序将能够在任何设备上运行。

MTE（memory tagging extensions）并不是v9的新功能，而是随v8.5一起引入的，MTE或内存标记扩展旨在帮助解决世界软件中两个最持久的安全问题。缓冲区溢出（Buffers overflows）和无

用后使用（use-after-free）是持续的软件设计问题，在过去的50年中，这些问题一直是软件设计的一部分，并且可能需要花费数年的时间才能对其进行识别或解决。MTE旨在通过在分配时标记指针并在使用时进行检查来帮助识别此类问题。

未来的Arm CPU路线图

这与v9没有直接关系，但是与即将到来的v9设计的技术路线图紧密相关，Arm还谈到了有关他们在未来2年中对v9设计的预期性能的一些观点。

Arm谈到了移动市场在今年如何将带有X1的设备性能提升了2.4倍（此处我们仅指ISO流程设计的IPC），该性能是几年前推出的Cortex-A73的两倍。

有趣的是，Arm还谈到了Neoverse V1设计及其如何达到A72类似设计性能的2.4倍，并透露他们期待着他今年晚些时候发布的首批V1设备。

对于代号为“Matterhorn”和“Makalu”的下一代移动IP内核，该公司公开了这两代产品的合计预期IPC增益为30％，其中不包括SoC设计人员可以获得的频率或任何其他其他性能增益。这实际上代表着这两种新设计的世代增加了14％，并且如幻灯片中的性能曲线所示，这表明相对于自A76以来Arm在过去几年所管理的工作而言，改进的步伐正在放缓。不过，该公司指出，进步速度仍然远远超过行业平均水平。但潭门也坦言，这被一些行业参与者拖累了。

Arm还提供了一张很有意思的幻灯片，该幻灯片旨在关注系统侧对性能的影响，而不仅仅是CPU IP性能。从这里提供的一些数据可以看到，例如每5ns的内存延迟中有1％的性能，这是我们现在已经广泛讨论了几代的数字，但是Arm在这里还指出，排除了是否通过改善内存路径，增加缓存或优化频率功能来改善实现的其他各个方面，他们可以使用整整一代的CPU性能提升，我认为这是对SoC供应商当前保守方法的一种评价，这些方法没有充分利用X1内核的预期性能余量，并且随后也未达到新内核的预期性能预测。

Arm继续将CPU视为未来最通用的计算模块。尽管专用的加速器或GPU将会占有一席之地，但它们很难解决一些重要问题，例如可编程性，保护性，普遍性（本质上是在任何设备上运行它们的能力）以及经过验证的正常工作的能力。当前，计算生态系统在运行方式上极为分散，不仅设备类型不同，而且设备供应商和操作系统也不同。

SVE2和Matrix乘法可以极大地简化软件生态系统，并允许计算工作负载以更统一的方法向前迈进，该方法将来将可以在任何设备上运行。

最后，Arm还分享了有关Mali GPU未来的新信息，并透露该公司正在开发VRS等新技术，尤其是Ray Tracing。这一点令人非常令人惊讶，也表明AMD和Nvidia引入RT推动的台式机和控制台生态系统也有望将移动GPU生态系统推向RT。

Armv9设计即将在2022年初面世

今天的公告以一种非常高级的形式出现，我们希望Arm在接下来的几个月中，在公司通常的年度技术披露中，更多地谈论Armv9的各种细节和新功能，例如CCA。

总的来说，Armv9似乎是更基本的ISA转变（可以看作SVE2）与软件生态系统的总体重新基准的结合，以汇总v8扩展的最后十年，并为下一个十年奠定基础Arm体系结构。

Arm于去年下半年已经谈论过Neoverse V1和N2，我确实希望N2至少最终是基于v9而设计发布的。Arm进一步透露，更多基于Armv9的 CPU设计（可能是移动端Cortex-A78和X1的后续产品）将于今年推出，而新的CPU可能已经被通常的SoC供应商所采用，并且有望成为在2022年初在商用设备中出现。

★ 点击文末【阅读原文】，可查看本文原文链接！

*免责声明：本文由作者原创。文章内容系作者个人观点，半导体行业观察转载仅为了传达一种不同的观点，不代表半导体行业观察对该观点赞同或支持，如果有任何异议，欢迎联系半导体行业观察。

今天是《半导体行业观察》为您分享的第2631内容，欢迎关注。

推荐阅读

★谁在“威胁”台积电？

★先进制程晶圆厂淘汰赛：20年锐减九成

★半导体设备开启“Turbo”模式

半导体行业观察

『**半导体第一垂直媒体**』

实时 专业 原创 深度

识别二维码，回复下方关键词，阅读更多

晶圆 | 集成电路 | 设备 | 模拟芯片 | 射频 | 传感器 | 美国 | 光刻

回复 **投稿**，看《如何成为“半导体行业观察”的一员 》

回复 **搜索**，还能轻松找到其他你感兴趣的文章！

点击阅读原文，可查看本文
原文链接！