The Future of Microprocessors

--读书报告

微处理器这个概念是在近几年才开始兴起，但在十五年之前就有学者对它进行了详尽的研究并且对它的未来发展趋势做了预测，现在来研读由Kunle Olukotun和Lance Hammond在ACM通讯上所做的这篇文献，我从中收获良多。为什么要研读两位大师的文献，是因为这两位大师对微处理器产业的发展起到了至关重要的作用，其中Kunle Olukotun还是设计了第一款主流的多核芯片，被视为多核之父的计算机界泰斗。

在这篇论文中开篇作者就对现代计算机的性能增长做了一个直观的描述，之后引出了处理器设计师对于并行性效率的探索的话题。之后在改进处理器以提升性能这里，有两种设计，第一个设计是增加在每一个时钟周期内能够调度的指令的数量，第二个设计是增加处理器的时钟频率，这两个技术在现在的处理器中被比较广泛的运用，因为这种技术已经十分的成熟，而且确实能够加快指令的执行速度。但是这两种技术加持下的指令都是按照顺序执行的，这种技术初期的时候发展是十分迅速的，但是随着技术和硬件的不断成熟和进步，再想利用这两项技术来对处理器的性能进行改进就变得十分艰难，这一点可以从原文中的图2看出。

从图2还可以看出，由于指令集并行技术的有效利用，这样所以导致了曲线的急剧上升，但是之后由于这种技术的潜力被完全开发，曲线又渐渐趋于水平。从这里就可以看出处理器的性能提升由此是进入了瓶颈的。其中的原因一方面是因为制造每个周期内执行多个指令的超标量处理器的制造工艺十分的昂贵，另一方面是由于每一个流水等级的变短使得流水线的设计变得十复杂。

之后论文中提到了关于处理器的功耗问题，我们可以从图3中看出随着处理器的性能不断更新换代，其功耗也在不断地呈幂级数增长，虽然性能得到了提升，但是散热技术没法得到相应的增长。这也给处理器的性能增长造成了阻碍。以上的因素结合在一起，就使得厂商和处理器设计师们转移了自己的研发重点，而不再去专注于提升处理器单核的时钟频率。在现今数据量和服务请求越来越大的大背景下，用多核处理器去代替单核处理器就成为了一种趋势。

但是光是简单的将两个核心组合在一起显然是不行的，从文中图4也可以看出，设计一个多核CPU显然是有不同的选择的，按我的理解来看，产品是依据市场推出的，所以在单核CPU性能提升被限制的情况下，实现多核CPU的目的就是为了提高吞吐率，让用户能有更好的体验。这里作者提出了一种思想，就是所谓的”latencty is unimportant”，运用这样的思想，一个典型的服务将在一瞬间处理数以百计的请求，这样就可以让处理器在保持繁忙的同时还保持高效的工作。”latency is unimportant”这个思想可以进一步运用，运用每个核的算力，牺牲更多的时延以换取更高的吞吐率。因为用户关心的是他们的计算机是不是响应他们的指令尽可能地快，而不是关心处理器同时处理指令的能力。工程师通过这种巧妙的trade-off，实现了对处理器性能的提升。

通过阅读这篇文献，我了解了处理器设计师是怎样在初期提升处理器的性能，以及之后在遇到瓶颈之后转而去研发多核处理器的过程。我也明白多核处理器不仅仅是简单的多个核心进行拼装，设计它的时候显然是考虑到功耗，用户体验，实现难度等一系列因素的。研读这篇15年前的论文，我受益良多。我明白一件事就是困难总是会出现的，发现问题并且解决问题的过程是十分重要的，这是人成长过程中最宝贵的财富。