

**暗硅 Dark Silicon:** 由于功耗的限制，一个很高端的多核处理器同时只能有很少一部分电路可以同时工作，其余处于非工作状态的门电路叫做“暗硅”；

**Dennard Scaling:** 单位面积功耗保持不变的条件下，半导体工艺制成每前进一代，频率能提高40%，即CPU越来越快，该定律忽视了leakage current和threshold voltage两个因素。之后CPU走向多核时代。

自2005年以来，处理器设计者增加了核心数量来达到摩尔定律预测的性能增长，而不是专注于单核性能的提升。处理器架构转向多核也是对Dennard定律的失败一种响应，然而多核的性能快也会像单核一样收到限制。本文通过结合设备缩放、单核缩放和多核缩放来建模多核缩放的限制，以此衡量未来五代技术一系列并行工作负载的加速潜能。对于设备缩放，我们同时使用ITRS和一组更保守的设备缩放参数。为了模拟单核缩放，我们结合了来自超过150个处理器的测量值，以推导出面积/性能和功率/性能的帕累托最优边界。最后，为了建立多核尺度模型，我们建立了一个详细的上限性能和下限核心功率的性能模型。我们研究的多核设计包括单线程cpu和具有对称、不对称、动态和组成拓扑的大规模线程类gpu多核芯片组织。研究表明，无论芯片组织和拓扑结构如何，多核缩放的能力限制程度没有得到计算界的广泛重视。即使是在22 nm时（仅一年后），21%的固定尺寸芯片也必须断电，而在8 nm时，这个数字也会增长到50%以上。到2024年，在常用的并行工作负载中，只有7.9×的平均加速，与每代性能加倍的目标有近24倍的差距。

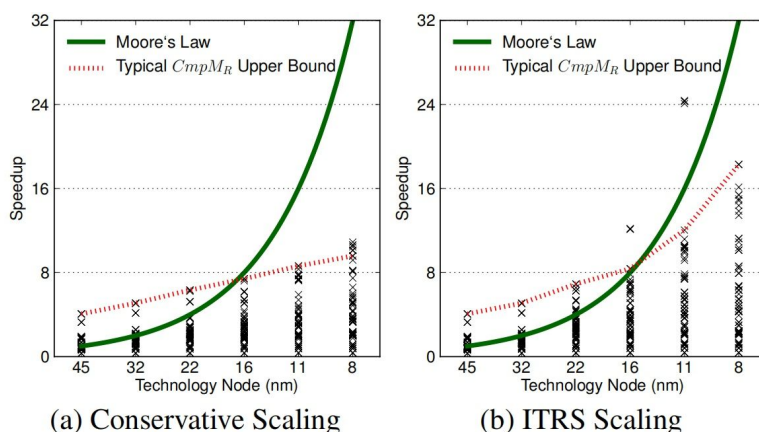


Figure 9: Speedup across process technology nodes over all organizations and topologies with PARSEC benchmarks

图9总结了单个散点图中的所有加速投影。对于每个技术节点上的每个基准测试，我们绘制了8种可能的配置，(CPU、GPU) × (对称、非对称、动态、组成)。实曲线表示性能摩尔定律或每个技术节点的加倍性能。如前所述，由于功率和并行性的限制，在可实现的目标和摩尔定律的期望目标之间存在着显著的差距。ITRS缩放的结果略好。通过保守尺度估计，与摩尔定律相比，在8nm技术节点上存在至少22×的加速间隙。假设ITRS缩放，在8 nm处的间隙至少为13×。

几十年来，Dennard 缩放允许在每个新工艺节点上使用更多晶体管、更快晶体管和更节能的晶体管，这证明了开发每个新工艺节点所需的巨额成本是合理的。Dennard 扩展的失败导致业界竞相走多核路径，这在一段时间内允许并行和多任务工作负载的性能扩展，从而处理器扩展经济效益得以保持。但随着多核微缩的好处开始减弱，必须找到晶体管效用的新驱动力，否则工艺微缩的经济性将被打破，摩尔定律将在我们达到最终制造极限之前结束。一个基本问题是在不久的将来可以从多核路径中提取多少性能。