技术进展

存储闸门理论：解决存储墙问题的一种新方案

在当今的计算机系统中，处理器 (CPU) 与存储之间存在着巨大的性能落差。这一日益增大的落差被人们称为存储墙 (memory wall) 问题。存储墙是一个困扰业界多年的大问题 [1]。解决存储墙问题对支持大数据的应用、对于研制新一代的 E 级超级计算机至关重要。在 2015 年 11 月 10~12 日召开的中国计算机学会全国高性能计算学术年会 (CCF HPC-China 2015) 上，来自美国伊利诺理工大学 (Illinois Institute of Technology) 的孙贤和 (Xian-He Sun) 教授在主题演讲中报告了其最新研究成果：存储闸门理论 (memory sluice gate theory) [2]，给出了一个解决存储墙问题的新方案。

存储闸门理论将层次存储系统形象地比喻为水渠，将各级存储层次的数据操作类比为水闸，每个存储层次都能支持一部分访存操作。但是它也会将一些访存发送给下一级，并且通过存储的并发性控制流量。就像水闸能够阻挡一部分水流，并在适当的时机通过排洪道加快流量一样。存储闸门理论的核心内容是匹配，即在每级存储层次上都能达到软件需求与硬件性能的匹配，从而降低甚至完全消除访存开销。这种匹配包括最佳存储层次数的制定，每个层次上的软、硬件配置，也包括访存局部性和访存并发性的最佳调配，从而将数据不断地输送给处理器。进而消除存储墙的影响。

难能可贵的是孙贤和教授还提出了一整套具体可行的方法去实现存储闸门理论。他用 C-AMAT (Concurrent-AMAT, 并发式平均存储访问时间) [3]，一个新的存储性能优化模型，做闸门调度器 (gate calculator) 去优化每一个水渠闸门的局部匹配。用一个新提出的层次性能匹配 (Layered Performance Matching, LPM) 方法 [4] 找出存储层次系统的总体最佳系统匹配，并通过实例证明通过存储闸门理论的性能匹配可以将某些程序的存储性能提高 150 倍 [2,4]，彻底消除了存储墙的影响。一个应用程序的运行时间 (total running time) 是其纯粹计算时间加上其存储停顿时间 (memory stall time)。存储停顿时间是处理器等待存储系统把数据输送上来的时间。如果一个应用程序的存储停顿时间占运行时间的 60%，那么纯粹计算时间仅占运行时间的 40%，此时存储停顿时间是纯粹计算时间的 1.5 倍。如果在优化后，此应用程序的存储停顿时间降到仅占纯粹计算时间的 1%，那么它的存储停顿时间就减少了 150 倍，也就是说存储系统的性能提高了 150 倍。在科学计算中，现在的存储停顿时间平均占运行时间的 50%~70%。对大数据的应用程序，存储停顿时间所占比例会高达 90%，或 90% 以上。孙贤和教授认为，通过存储闸门理论的性能匹配，科学计算的存储停顿时间应当可以减少到总运行时间的 20%~40%。现在的存储系统是以优化数据局部性为主导的。存储闸门理论的优化是以数据局部性和数据并发性双主导的，加了一个优化的新维度，还加了一个硬件动态支持新力度。他认为如果硬件动态支持完全到位，存储闸门优化把存储停顿时间减少到计算时间的 10%，甚至 1% 是完全可能的。把存储停顿时间减少到总体运算时间的 20%~40% 是一个非常合理、略带保守的推断。

孙贤和教授说，水渠 (sluice) 和闸门 (gate) 这两

个术语的选择是具有深刻含义的。Sluice 强调的是渠。 这渠有层次（包括寄存器、多级缓存、主存、Disk 的整个数据移动的通道）；有大小（取决于并行性），有宽窄（取决于位宽），有快慢（取决于传输频率）。Gate 强调的是匹配和调节，是通过门的开与合实现渠的匹配和调节。（存储）渠的主要功能是在数据传输过程中减小处理器与存储之间的性能落差。从设计层面来讲，存储闸门系统结构与 20 世纪 70 年代提出的数据流 (dataflow) 系统结构在概念上完全相反。在数据流体系结构中，计算是跟着数据走的。在存储闸门系统结构中，数据是通过存储渠源源不断输送到计算处理单元的，也就是说数据是在既定的渠里跟着计算走的。从技术层面来讲，现有的存储系统都是通过优化数据局部性来优化系统性能。存储闸门理论证明，在每一个存储层次上优化数据局部性无法达到性能最优，优化存储系统的总体数据局部性也无法达到性能最优，存储系统优化一定要考虑数据并发性。但同时存储闸门理论也证明，一个存储系统很难同时支持数据存储局部性最优和数据存储并发性最优。系统性能的最佳点是数据局部性和数据并发性的平衡点，是最大程度地把数据访存与计算的时间重叠 (overlapping) 起来。存储闸门理论的性能匹配可以找到这一最佳点。这是存储闸门理论的威力所在。

孙贤和教授指出，存储闸门理论进行流量控制的要点是：(1) 充分利用存储系统中已存在的或可支持的并行硬件结构，提高存储并发性，同时认识到增加访存需求并发度，可能会引起数据访问之间在共享通道（片上网络、末级缓存、芯片通信管脚等）上的争用和排队延迟；(2) 充分利用数据访问的局部性，增加数据的复用率，同时认识到局部性会导致纯粹缺失 (pure miss)，会使得处理器的运算因为缺失数据而停顿下来；(3) 按照对性能的影响程度，区别对待每一项数据和每一次数据访问，将一个存储通道分配给不同的进程（线程或程序），获得的性能收益一般是不一样的。孙贤和教授还指出，存储闸门理论的数据并发性包括访存需求并发性，像预取、runahead。但更多的是指存储系统硬件的并发性，像多端口 (multi-port)、多组 (multi-bank)、流水线 (pipelining)、非阻塞 (non-blocking)、多通道 (multi-channel) 等技术。存储硬件的并发性在于提高存储系统的性能。在适当的时刻，适当的地点，适量地增加硬件的并发性可以像海绵一样吸收一部分流量，也可以像在水渠里分出了一些子水渠，子水渠还有子水渠，在层次存储系统里逐层放大，以加快水的流量。而存储访存需求并发性的一大作用在于增加存储与计算的重合。如果这种重合不存在，或无法达到，增加存储访存需求并发性往往会增加存储系统的负担，对提高系统性能可能适得其反。数据局部性好并不一定好，增加访存需求并发性很可能会产生负效果，存储硬件并发性是存储流量动态调节的关键。存储闸门理论对很多存储问题都提出了完全不同的见解和解决方案。存储闸门理论是解决存储墙问题的一个革命性的进步。

多年以来存储墙问题就像一座难以逾越的高山一样挡在计算机系统性能提高的道路上。很多人一直都认为存储墙问题只能依赖存储硬件器件性能的提高来解决。存储闸门理论给出了一个从系统结构出发解决存储墙问题的方案，打开了一个解决存储墙问题的全新方向。

HPC-China 2015 大会主席，中国科学院计算技术研究所所长孙凝晖教授说，存储闸门理论回答了我们以前许多无法回答的问题，对下一代计算机的硬件研制、软件开发具有重大的指导性意义。■ （成集）

参考文献

[1] 赵沁平等. 10000个科学难题信息科学卷. 科学出版社, 2011.

[2] http://hpcchina2015.csp.escience.cn/dct/attach/Y2xiOmNsYjpwZGY6MTA1MjI2

[3] Xian-He Sun and Dawei Wang, Concurrent Average Memory Access Time, in *IEEE Computers*, vol.47, no.5, pp.74~80, May, 2014.

[4] Y.-H. Liu and X.-H. Sun, LPM: Concurrency-driven Layered Performance Matching, in *Proc. of the 44th International Conference on Parallel Processing* (ICPP'15), Beijing, China, Sept. 2015.