doi:10.1631/FITEE.1800501

题目：高通量计算的兴起

概要：近年来，云计算、人工智能和物联网等新兴计算应用的出现，对计算机系统设计提出 3 个共同要求：高利用率、高吞吐量和低延迟。在这里，这些被称为“高通量计算”的要求。我们进一步提出一种新指标，称为“系统熵”，用于测量计算机系统的混乱程度和不确定性。我们认为，与追求高性能和低功耗的传统计算系统的设计不同，高通量计算应致力于实现低并发性。然而，从计算机体系结构角度看，高通量计算面临两大挑战：（1）如何充分利用应用程序数据的并行和并发执行来实现高吞吐量；（2）如何获得低延迟，即便在具有高利用率的数据路径中发生严重争用的情况下。为应对这两个挑战，引入两种技术：片上数据流体系结构和标签化冯诺依曼体系结构。构建了两个可以实现高吞吐量和低延迟的原型，显著降低了系统熵。

关键词：高通量计算；系统熵；信息高铁