Application-to-Core Mapping Policies to Reduce Memory Interference in Multi-Core Systems

Reetuparna Das

Rachata Ausavarungnirun

Onur Mutlu

Future many-core processors are likely to concurrently execute a large number of diverse applications. How these applications are mapped to cores largely determines the interference between these applications in critical shared hardware resources. This paper proposes new application-to-core mapping policies to improve system performance by reducing inter-application interference in the on-chip network and memory controllers. The major new ideas of our policies are to: 1) map network-latency-sensitive applications to separate parts of the network from network-bandwidth-intensive applications such that the former can make fast progress without heavy interference from the latter, 2) map those applications that benefit more from being closer to the memory controllers close to these resources.

Our evaluations show that both ideas significantly improve system throughput, fairness and interconnect power efficiency. Averaged over 128 multiprogrammed workloads of 35 different benchmarks running on a 64-core system, our final application-to-core mapping policy improves system throughput by 16.7% over a state-of-the-art baseline, while also reducing system unfairness by 22.4% and average interconnect power consumption by 52.3%.

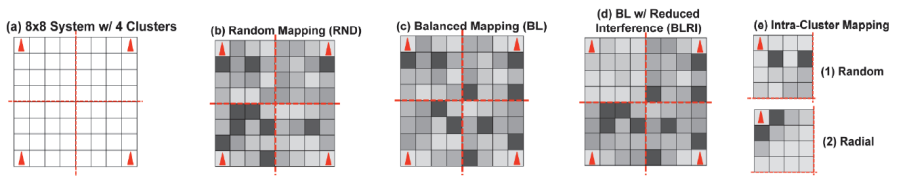

Figure 1: Visual examples of clustering and different inter-cluster and intra-cluster mapping policies

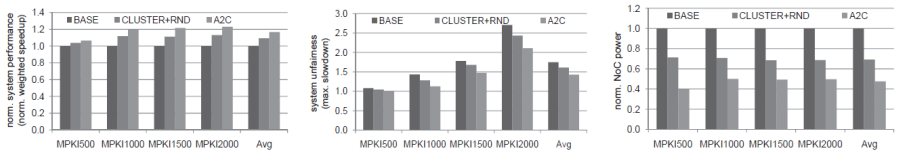

Figure 2: (a) System performance (b) system unfairness and (c) interconnect power of A2C for 128 workloads