
Kurzfassung

Moderne Prozessoren, speziell auch im Serversegment, werden mit zunehmender Anzahl an Kernen ausgestattet. Dies allein führt zu steigender Komplexität der Prozessoren, speziell wenn es um die Verwendung von Speicher geht. Eine ungünstige Verteilung von Daten kann die Performance von Anwendungen erheblich beeinflussen. Gleichzeitig sind energetische Aspekte beim Design und der Verwendung von Prozessoren zu berücksichtigen.

In dieser Arbeit untersuche ich die Speicherhierarchien von AMD EPYC Rome und Intel Xeon Cascade Lake SP Serverprozessoren. Dafür werden etablierte Benchmarks genutzt. Aufgrund der unterschiedlichen Designs unterscheiden sich die Zugriffslatenzen auf Speicher, speziell zu entfernten Caches, wobei beim AMD Prozessor komplexere NUMA Eigenschaften auffallen als beim Intel Prozessor.

Die Intel Prozessorkerne bietet AVX-512 Unterstützung, wodurch der theoretische Durchsatz an Daten doppelt so hoch sein kann wie bei AMD, die AVX Instruktionen verarbeiten können. Bei der Messung der Bandbreiten der Caches kann der Intel Xeon Cascade Lake SP Prozessor den theoretischen Vorteil von AVX-512 nicht vollständig ausspielen, speziell durch geringe Taktfrequenzen, geringen tatsächlichen Durchsatz und niedrige Kernzahl. Bei beiden Prozessoren zeigt sich bei der Arbeitsspeicherbandbreite, dass diese ausgereizt werden kann, ohne dass sämtliche Kerne verwendet werden. Sobald die Bandbreite ausgereizt wird, kann diese bei beiden Prozessoren auch mit niedrigerer Taktfrequenz der Prozessorkerne erreicht werden. Aufgrund der höheren Anzahl an Speicherkanälen können AMD EPYC Rome Prozessoren höhere Speicherbandbreiten erreichen. Bei beiden Prozessoren untersuche ich die Einflüsse von Energiesparmaßnahmen auf Speicherbandbreiten und -latenzen. Für den AMD Prozessor werden Systemkonfigurationen aufgezeigt, welche für energetische sowie Performance-Optimierungen genutzt werden können.

Abstract

Modern processors, in particular within the server segment, have higher core counts with each generation. This alone increases the complexity of these processors, which is particularly true for the usage of memory. A disadvantageous distribution of data can significantly impact the performance of applications. At the same time, energetic aspects have to be considered when designing and using a processor.

In this thesis, I investigate the memory hierarchy of AMD EPYC Rome and Intel Xeon Cascade Lake SP server processors. Established benchmarks are used for this purpose. Due to the different designs, significant differences in latencies for memory accesses, in particular for remote cache accesses, are observed. The AMD processor exposes more complex NUMA properties than the Intel processor.

The Intel processor cores supports AVX-512, therefore its theoretical data throughput can be twice that of an AMD core, which support AVX. However, my cache bandwidth measurements show that the Intel Xeon Cascade Lake SP processor cannot utilize that benefit, in particular due to low core frequencies, a low effective throughput and low core count.

Both processors can saturate their main memory bandwidth without using all cores. Once a saturation has been reached, lower core frequencies can be used without reducing the bandwidth. The AMD processor has a higher bandwidth thanks to its higher memory channel count. I investigate the influence of power saving mechanisms on memory latency and bandwidth for both processors. System configurations are demonstrated for the AMD processor that can be used for energetic and performance optimizations.