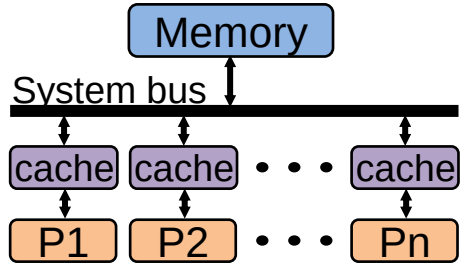
Computer Structure

Multi-Processors

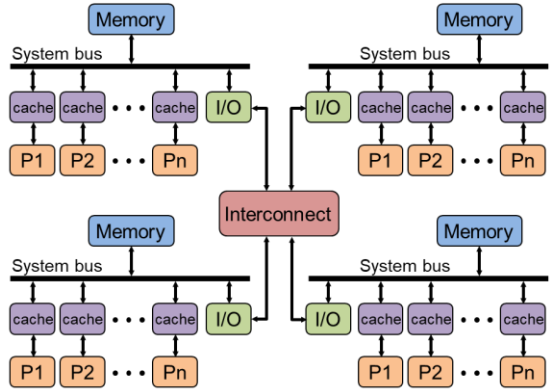
Lihu Rappoport and Adi Yoaz

Multi-Processor Memory

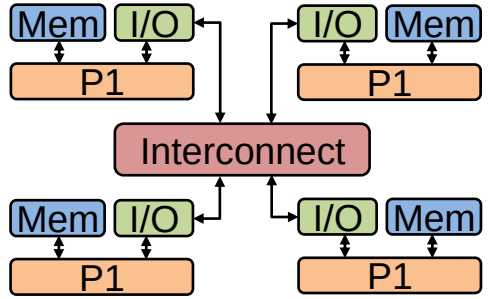
- **Centralized shared memory (SMP – Symmetric Multiprocessor)**
  - Uniform Memory Access (UMA)
    - All processors access all memory locations at a similar latency
    - Data sharing through memory reads/writes
  - Memory bandwidth becomes bottleneck: not scalable



- **Distributed Shared-Memory (DSM)**
  - Also called NUMA (non-uniform memory access)
  - All processors can address all memory locations, like in SMP
  - Variable latency: local access faster than remote access



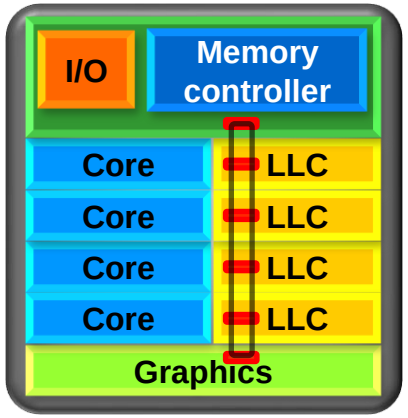
- **Message-Passing (multi-computers / clusters)**
  - A computer cluster is a set of connected computers that work together
  - A processor can directly address only its local memory
  - Communicates with other processors by sending/receiving messages
    - Puts extra burden on SW for handing messages
  - Scales to many processors



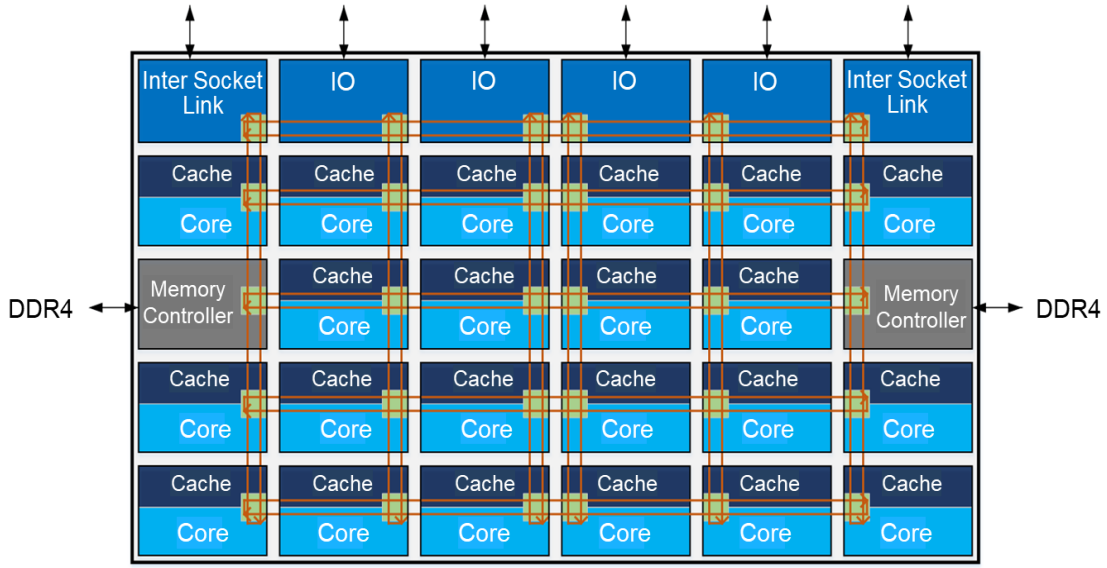
CMP – Chip Multi Core

- **Simple SMP on the same chip**
  - Resources can be shared between CPUs, e.g., shared L3 caches
- **Cheaper than multi-socket SMP**
  - Interface logic integrated on-chip
  - Fewer total chips, single CPU socket
  - Single interface to main memory
- **Less power than multi-socket SMP**
  - On-die communication is more power-efficient than chip-to-chip communication
- **Performance**
  - On-chip communication is faster
  - Efficiency
  - Potentially better use of hardware resources than improving performance of a single-threaded CPU

Ring

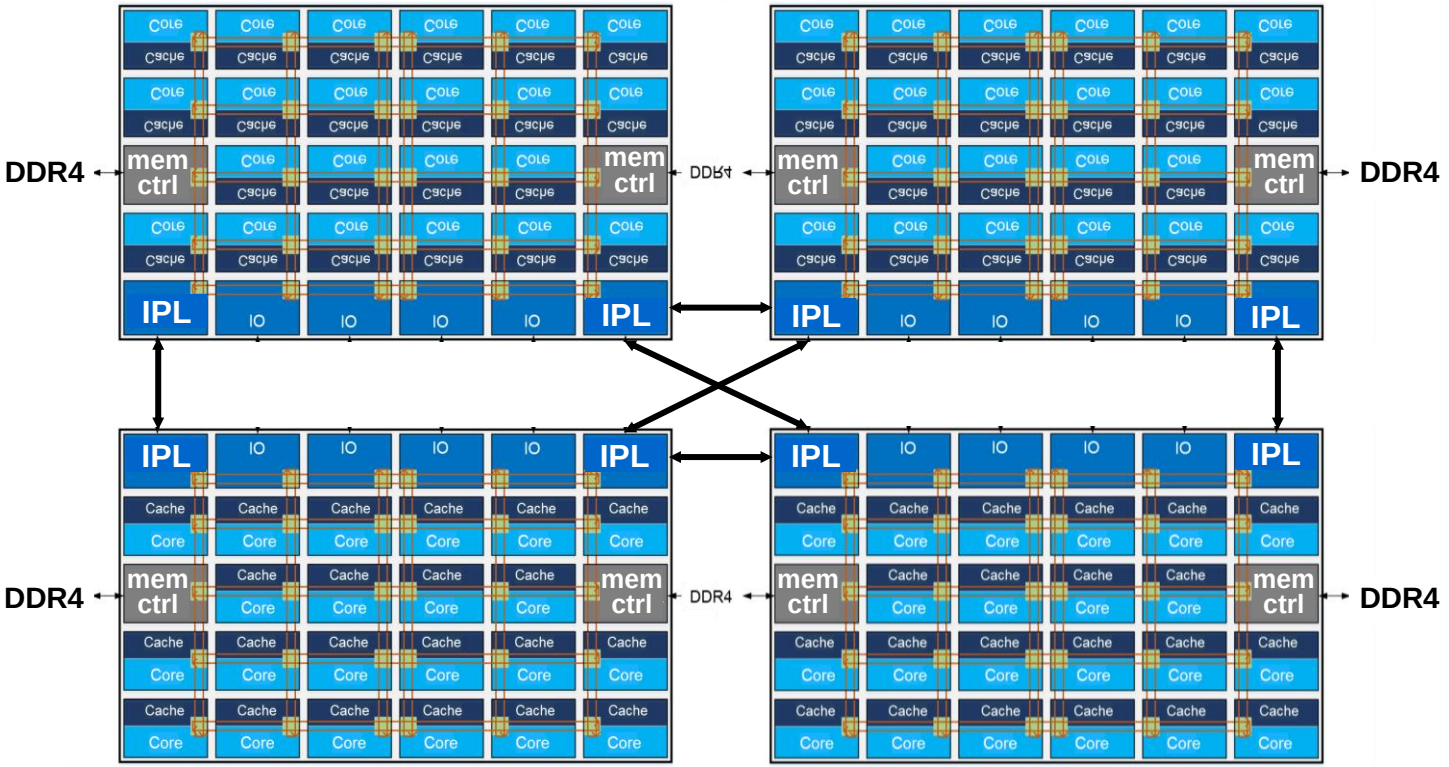


Mesh



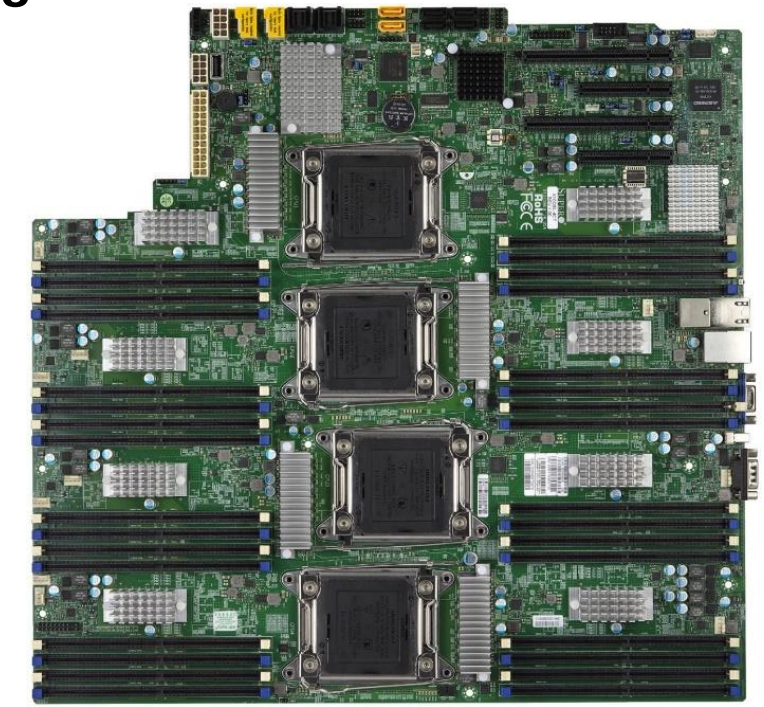
Multi-Socket System

- **Each socket (chip) includes multiple core (CMP)**
- **Each socket has Inter-Processor Link controllers**
  - Sockets are fully connected using a high-speed point-to-point interconnect
  - Coherent interconnect, single shared memory address space
    - Directory-based home snoop coherency protocol



Server Systems

- **Server systems typically support SMP, CMP and SMT together**
  - Multiple sockets per motherboard (SMP)
  - Multiple cores per die (CMP)
  - Multiple thread per core (SMT)
  - E.g., Intel® Xeon® E7-8890 V4
    - SMP: Up to 8 processors per board
    - CMP: 24 cores per processor
    - SMT: 2 threads per core



- **Multiple blades per rack**
  - Blades connect using a network, e.g., InfiniBand

Supercomputers

- **Supercomputers or HPC (High Performance Computers) clusters**
  - A group of servers connected with a dedicated high-speed network
  - Top500 ranks the 500 most powerful non-distributed computer systems
    - Based on HPL – high-performance LINPACK benchmark (a software library for performing numerical linear algebra)

- **Trinity system (Cray XC40) at Los Alamos National Laboratory**

Cores:	979,968
Processor:	Intel Xeon Phi 7250 68C 1.4GHz
Linpack Performance	14,137 TFlop/s
Theoretical Peak	43,902 TFlop/s
Power:	3,843kW