CS147 - Lecture 17

Kaushik Patra
(kaushik.patra@sjsu.edu)

1

- Parallel Processing Systems
- Symmetric Multi-Processor System
- Multi-Core System

Reference Books / Source:

1) *Chapter 17, 18 of 'Computer Organization & Architecture' by Stallings*
2) *http://www.oracle.com/technetwork/systems/opensparc/opensparc-t1-page-1444609.html*

Parallel Processing Systems ...

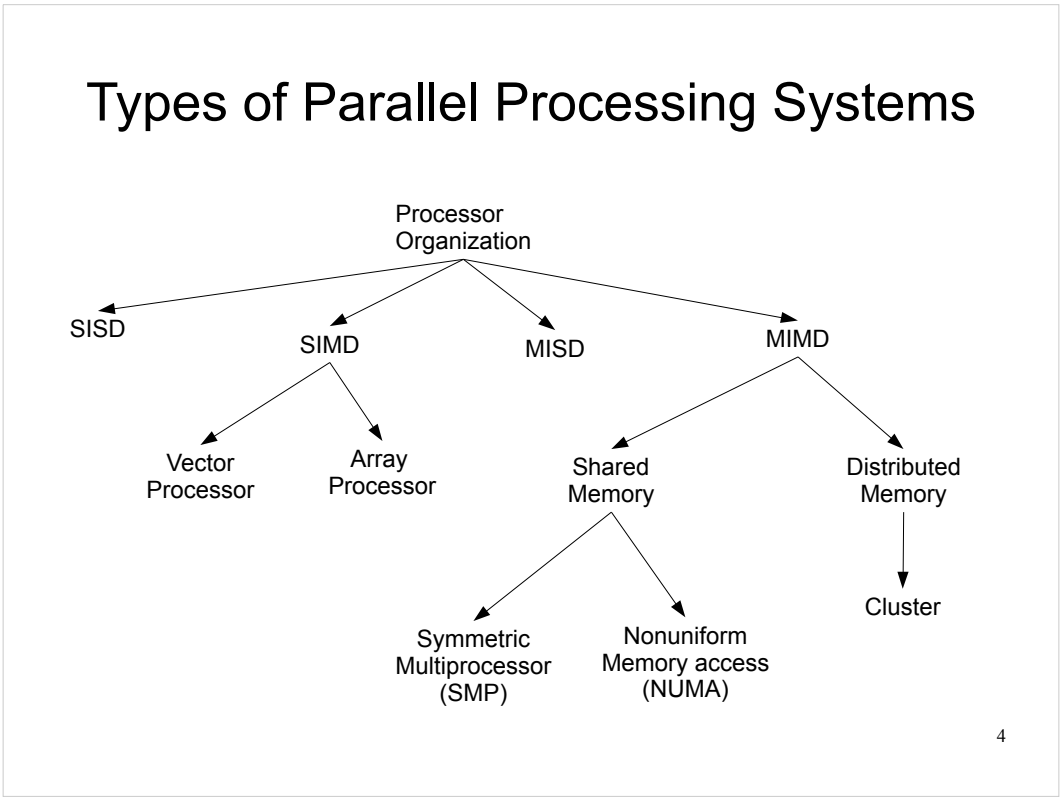
2

Types of Parallel Processing Systems

- **Single Instruction Single Data (SISD) System.**
  - Processor executing single instruction on single data stream.
  - e.g. Uniprocessor Systems
- **Single Instruction Multiple Data (SIMD) System.**
  - Processor executing single instruction on multiple data stream.
  - e.g. Vector or Array processor.
- **Multiple Instruction Single Data (MISD) System.**
  - Processor executing multiple instruction on single data stream.
  - Not commercially available
- **Multiple Instruction Multiple Data (MIMD) System.**
  - Processor executing multiple instruction on multiple data stream.
  - e.g. Symmetric Multiprocessor (SMP), Cluster and NUMA system.

3

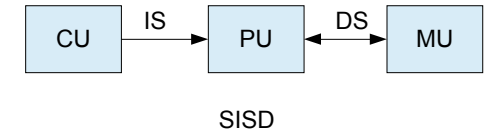
- The SISD system was very common in personal computing space couple of decades ago. It was the single processor environment with simple and limited computation ability. The single processor used to process a single stream of data through operation. The parallelism (pseudo) is achieved using timesharing mechanism among multiple programs orchestrated by operating system.
- The SIMD is now a days very common in graphics processing where array (single or two dimension) of numbers are needed to be processed. A single instruction is operated on multiple data stream at the same time.
- MISD is more in theoretical / laboratory / experimental domain.
- MIMD is very common now a days, even in personal computing environment. Multicore system is very common and cheap available on market. Multiple instructions are executed in parallel processing different data stream at time.



- This is a family tree classification chart of parallel processing system. The SIMD can further be classified into vector processing and array processing. Vectors are the single dimension array of numbers. A vector processor can perform single function on one or two vector variable in single execution. Similarly array processor can perform single function on one or two 2-dimensional array variable in single execution.

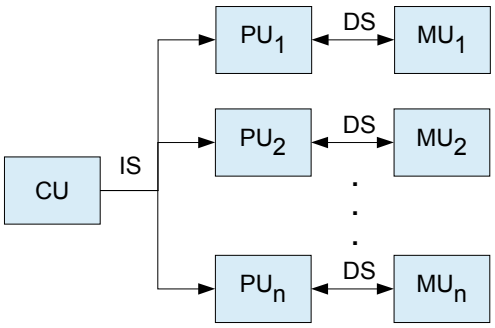
- The MIMD can further be classified into shared memory type MIMD and distributed memory type MIMD. For a shared memory MIMD system different processor shares common memory at some hierarchy level (i.e. they may have their own local memory, but local memory is ultimately connected to a shared common memory at some point in the hierarchy). Shared memory system can be further classified into the Symmetric Multiprocessor (SMP) system and Non-uniform memory access (NUMA) system.The system used in PC nowadays are mostly SMP. At compute ranch / firm (where hundreds of machines are connected together in network) of industry, NUMA is common nowadays. On the other hand distributed memory MIMD only has local memory associated with processors and the processors are connected through network to exchange information between them.

Parallel Systems Organization



SISD

```
CU = Control Unit
IS = Instruction Stream
PU = Processing Unit
DS = Data Stream
MU = Memory Unit
```



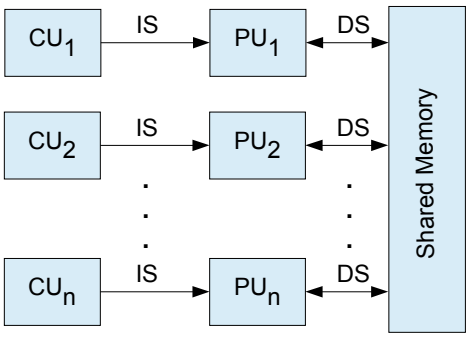
SIMD

5

- The above diagram shows top level organization for SISD and SIMD machines. SISD has single control unit (CU), single processing unit (PU) and single memory unit (PU). Control unit execute single instruction on scalar data type. The PU process the scalar data from MU and store the result back into MU.

- SIMD has single CU, but multiple PU and MU. Each PU is associated with its own memory unit. CU executes single instruction on vector / array data type using multiple PU. The vector / array data is loaded from the multiple MU at once. The vector / array result stored into these multiple MU at once.

Parallel Systems Organization



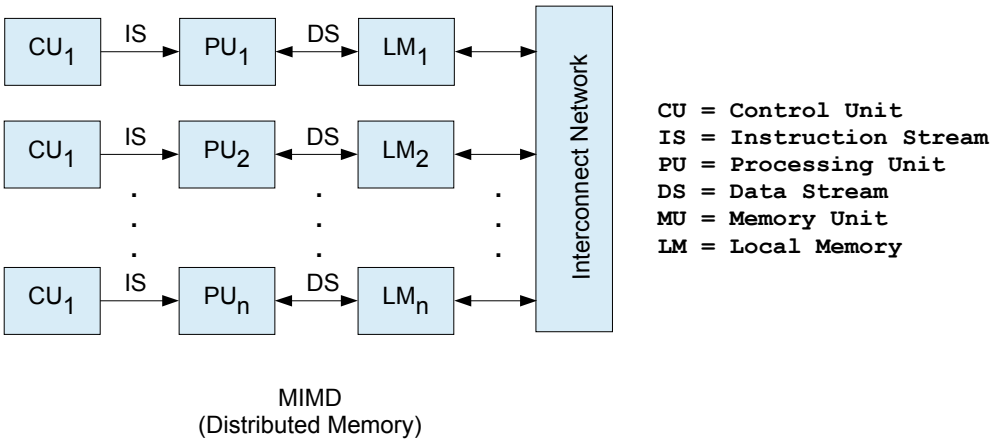
MIMD
(Shared Memory)

CU = Control Unit
IS = Instruction Stream
PU = Processing Unit
DS = Data Stream
MU = Memory Unit

6

- The shared memory MIMD system has multiple CU and corresponding PU, but data is loaded and stored from same shared memory. Multiple CU can execute multiple instructions in parallel using their own dedicated PU. The PU access data from shared memory. Often PU has associated local memory which the PU can exclusively accessed. The local memory content is synced with the shared memory time to time.

Parallel Systems Organization



MIMD
(Distributed Memory)

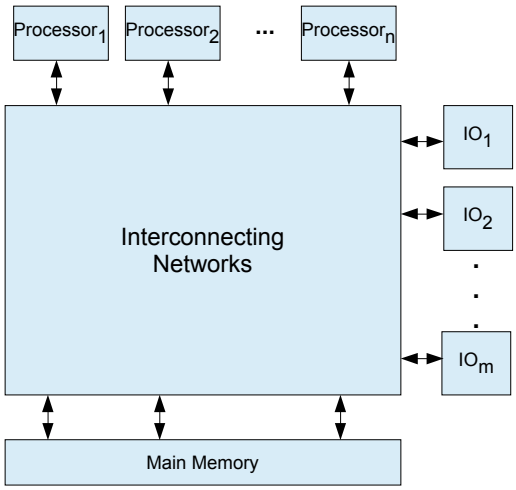
7

- The distributed memory MIMD is very similar to shared memory MIMD. However, each PU is associated with its own local memory (LM). Each of the (CU, PU, LM) unit is connected with each other through network. They can exchange information to each other through this network.

Symmetric Multi-Processor (SMP) System ...

8

SMP Organization



- Two or more compatible processors.
- Shares same main memory
- Shares access to IO devices.
- Processors can perform same functions.
- Controlled by integrated OS

9

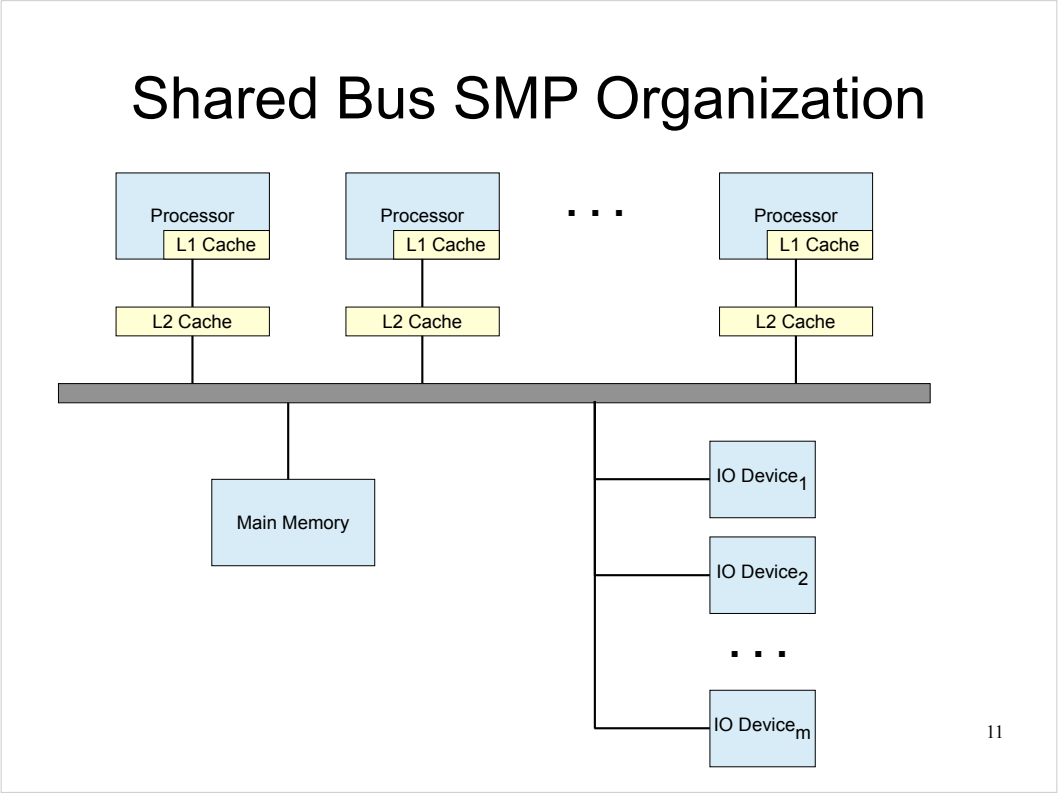
- The shared memory symmetric multiprocessor (SMP) contains two or more compatible processor. By compatibility we mean that all the processor in the group can executed same instruction and produce same result. All these processors may have their own local memory. However, all these processor shares same main memory and also shares same group of IO devices. Memory and IO devices are shared through interconnecting network. The operation is this type of system is controlled by integrated operating system which controls scheduling of different programs on the processors in multi-programming environment.

Advantages of SMP System

- **Performance**
  - Through parallel programing.
- **Availability**
  - Implements redundancy in system.
- **Incremental Growth**
  - Performance enhancement by additional processor.
- **Scaling**
  - Ranges of products performance vs. price.

10

• Since having multiple processor executing programs in the same time, performance / throughput improvement compare to uniprocessor system is evident. Single program performance can also be improved by implementing the program using parallel programing paradigm. Secondly, Having multiple processor of same type also gives the advantage of redundancy in the system. This means, even if one processor is down (due to hardware failure) the whole system can continue to run. The only affect of processor failure is that the throughput of the system to come down. Thirdly, we can keep attaching additional processors to the system (If the overall system allows it) to increase performance on demand. Last but not the least, SMP gives system vendors to preconfigure several ranges of products varying performance (PC vs. Enterprise system) and prices. This serves the business need to a computer system vendor quite well.



- The most common implementation for SMP system is using BUS interconnect. All the components (processors, main memory, IO devices (they may be through local bus)) are connected with each other through a common bus. Only one device can gain control on the bus at a time to send data to other devices as needed. For example, one processor system can get control of the bus an send request for a data transfer from main memory. Once request is sent, the processor release control of the bus and wait till the main memory sends the data in. The main memory receives the request and gain control of the bus to transfer data to requesting processor.

Features of Shared Bus System

- **Addressing**
  - Modules must be distinguishable.
- **Arbitration**
  - To arbitrate competing request for bus control.
- **Time sharing**
  - Exclusive access to bus to single component (master).

12

- To correctly transfer / exchange information between devices attached to a bus, an addressing method should be developed where each device can be addressed uniquely. In that way, data can be targeted to a specific device.
- To give control to the bus to a specific device on a multiple bus control request from several devices, arbitration must be done. A bus arbitrator is introduced for this. Different bus sharing protocol implement arbitration policy differently.
- There must be exclusive control to the bus to a single component. This is very important to avoid any data corruption. If multiple component tries to transfer data through bus at the same time, data will be corrupted and no devices can operate successfully in successive stages.

Advantages of Shared Bus System

- **Simplicity**
  - Similar logic for connectivity as in single processor system.
- **Flexibility**
  - Easy to expand system.
- **Reliability**
  - Passive system of connection, thus system remains operational in attached device failure scenario.

13

- Compare to other interconnect system (e.g. cross-bar) bus system can be implemented in hardware in relatively simple way – because bus system concept is relatively simple to start with.
- Bus system supports hierarchy of buses. This means one bus system can talk with other bus system as if they are talking to yet another device. For this, each bus system in the hierarchy should be addressable uniquely similar to a device. For example, USB (Universal Synchronous Bus) can be expanded for multiple devices with no extra effort by just attaching USB expander. It can further be expanded by attaching another expander to the other expander.
- Bus is a passive system – only active part is the arbitrator. Therefor, if any component attached to a bus is failed, rest of the system can continue to work.

Multicore System ...

Multicore Organization

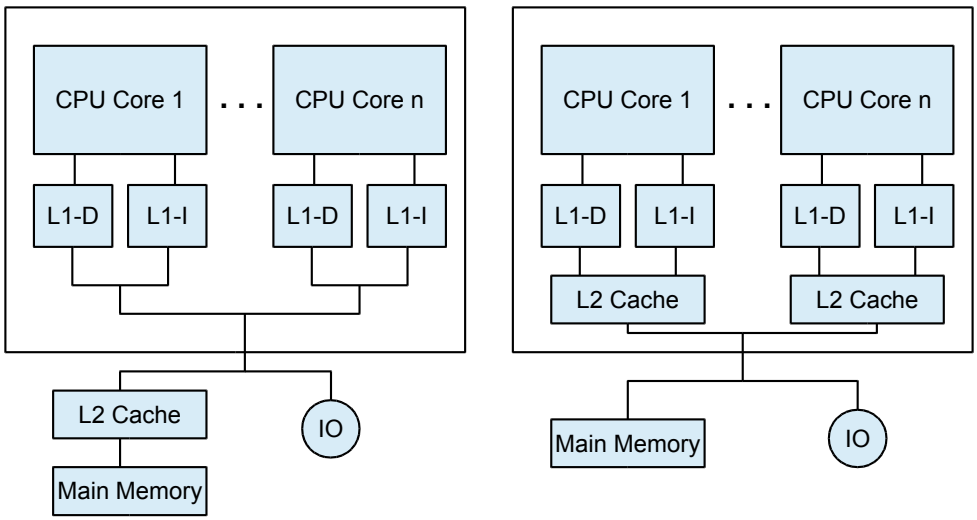
- Multicore system, or chip multiprocessor is a system combining two or more processor (called core) on a single piece of silicon.
- Overall multicore organization can be characterized with the following parameters.
  - Number of cores on the chip
  - Number of levels of cache memory
  - Amount of shared cache memory

15

• The most common implementation for a shared memory MIMD system is the chip multicore systems. It is very common nowadays to have a PC with dual or quad core processor. This system combines multiple processor (called core) on a single silicon.

• A multicore organization is characterized by number of cores (dual or quad is most common in PC system – at industry scale 16 core machine is very common), level for intermediate memory (cache) and sharing between them.

Multicore Organization



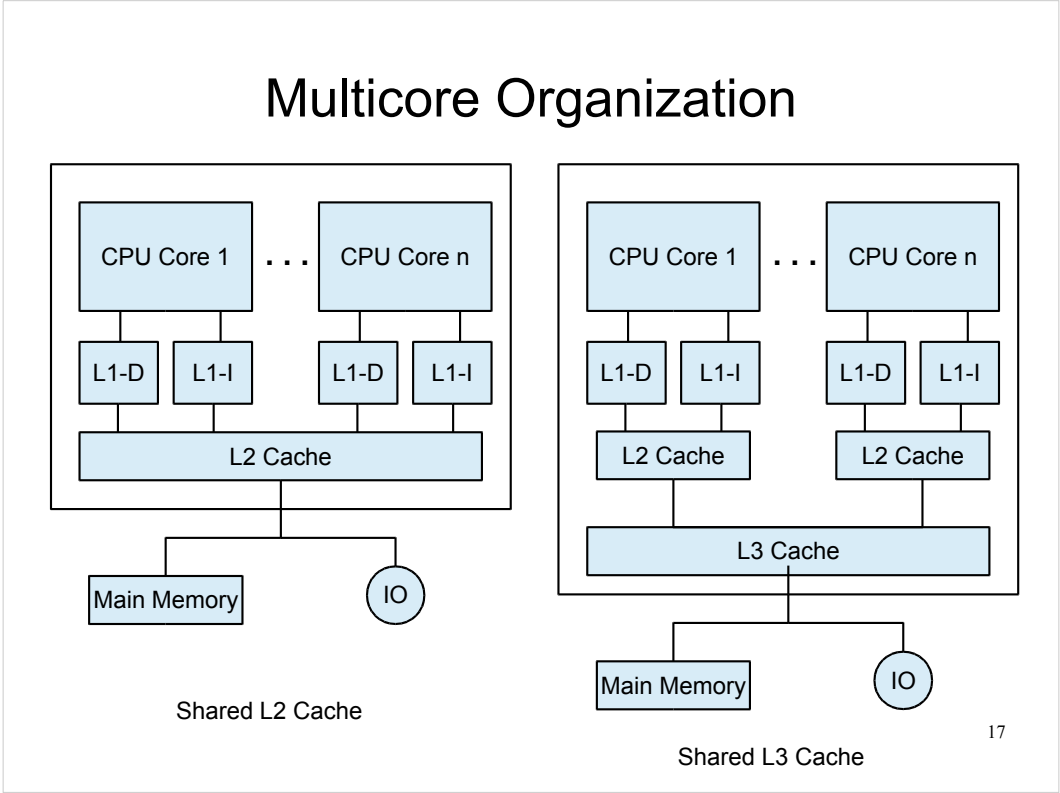
Dedicated L1 Cache

Dedicated L2 Cache

16

- For a dedicated L1 cache multicore organization, the cores and the L1 caches are implemented on single silicon (or single chip). The L2 and main memory / IO reside outside the silicon (or on board).

- For a dedicated L2 cache multicore organization, the cores, L1 caches and L2 caches are implemented on single silicon (or single chip). The L2 cache is dedicated per processor. The main memory / IO reside outside the silicon (or on board).



- For a shared L2 cache multicore organization, the cores, L1 caches and L2 caches are implemented on single silicon (or single chip). The L2 cache is shared among processor. The main memory / IO reside outside the silicon (or on board).

- For a shared L3 cache multicore organization, the cores, L1 caches, L2 caches and L3 are implemented on single silicon (or single chip). The L2 cache is dedicated per processor, where the L3 cache is shared among the processors. The main memory / IO reside outside the silicon (or on board).

Software Performance

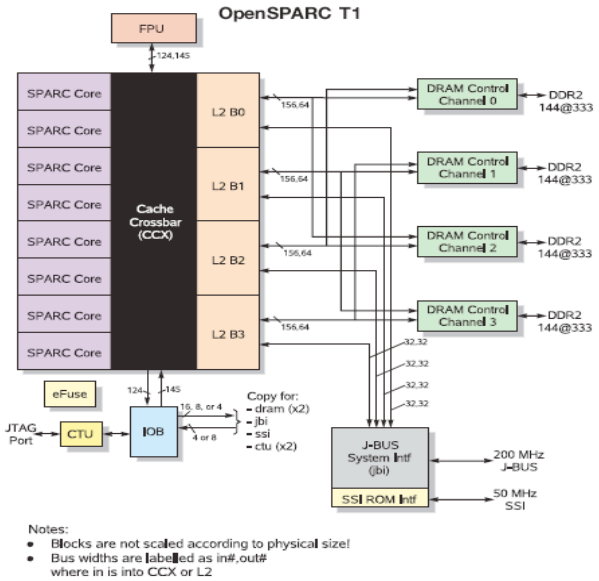
- The potential performance benefit of a multi-core organization depend on the ability to effectively exploit parallel resource.
- A single program speedup depends on how much of it can be parallelized. The speedup follows the Amdahl's law.

$$S = \frac{1}{(1-f) + \frac{f}{N}}$$

18

• Using a multicore machine, throughput of programs in multi-program environment improves. A single program can also take advantage of multicore processor to improve performance by implementing the program in parallel paradigm. However, it is to be noted that there are parts of a program which cannot be parallelized, hence the speed can not be scaled by the amount of available cores. Amdahl's law can be used to analyze the speed up or performance gain as the above formula, where f is the amount of part which can be parallelized and N is the number of cores available in the system.

SPARC-T1 Organization



19

- The SPARC-T1 is commercially available open source processor which follows a shared memory MIMD parallel processing multicore organization. It has 8 SPARC core (which implements hardware threading in the pipeline and has its own L1 cache), 4 L2 cache. Any processor can access any L2 cache through a cache-cross-bar connection (not a bus organization). A crossbar connection is faster than bus organization, but much more complex to be implemented. It works like a digital switch box or telephone exchange centre through which any component can talk to any other component exclusively in parallel with others. L2 caches are connected to external DRAM through DRAM channel controller. This organization also includes a common FPU (floating point processor unit).

CS147 - Lecture 17

Kaushik Patra
(kaushik.patra@sjsu.edu)

20

- Parallel Processing Systems
- Symmetric Multi-Processor System
- Multi-Core System

Reference Books / Source:

1) *Chapter 17, 18 of 'Computer Organization & Architecture' by Stallings*
2) *http://www.oracle.com/technetwork/systems/opensparc/opensparc-t1-page-1444609.html*