



Cluster Versions of CrystalWave and OmniSim Hardware Recommendations

□ Introduction

If you do not yet have a cluster, then this document contains suggestions on how to choose a cluster which will run optimally on our CrystalWave and OmniSim Cluster Editions. Certain high end features available with some clusters may provide no benefit to you and waste your money. On the other hand some relatively cheap additions may substantially enhance your attained performance.

This document relates entirely to the clustering of the FDTD engines in CrystalWave and OmniSim.

□ Factors affecting FDTD cluster speed

Memory Bandwidth

FDTD is a memory intensive application. Therefore it is important to get as much memory bandwidth in the whole cluster as possible. Total memory bandwidth is given simply by:

$$(\text{Memory bandwidth of one node}) \times (\text{number of nodes})$$

You may see terms like PC2100 or PC2700 which means a memory bandwidth of 2100MB/s or 2700MB/s.

Below is a table showing the total memory bandwidth of different memory modules. Note that the actual speed at which the memory is run is also a function of the chipset.

Technology	Type	Alt Name	Nominal frequency	System Memory Bandwidth (max theoretical)		
				Single Channel	Dual Channel	Quad Channel
DDR	DDR200	PC1600	200MHz	1.6 GB/s	3.2 GB/s	6.4 GB/s
DDR	DDR266	PC2100	266MHz	2.1 GB/s	4.2 GB/s	8.4 GB/s
DDR	DDR333	PC2700	333MHz	2.7 GB/s	5.4 GB/s	10.8 GB/s
DDR	DDR400	PC3200	400MHz	3.2 GB/s	6.4 GB/s	12.8 GB/s
DDR	DDR500	PC4000	500MHz	4.0 GB/s	8.0 GB/s	16.0 GB/s
DDR-2	DDR2-400	PC2-3200	400MHz	3.2 GB/s	6.4 GB/s	12.8 GB/s
DDR-2	DDR2-533	PC2-4200	533MHz	4.2 GB/s	8.4GB/s	16.8GB/s
DDR-2	DDR2-667	PC2-5300	667MHz	5.3 GB/s	10.6 GB/s	21.2 GB/s
DDR-2	DDR2-800	PC2-6400	800MHz	6.4 GB/s	12.8 GB/s	25.6 GB/s
RDRAM		PC1066	1066MHz	2.1 GB/s	4.2 GB/s	8.4 GB/s
XDR/XDIMM	XDR 3.2GHz		3.2GHz	6.4 GB/s	12.8GB/s	25.6GB/s

Table 1: Memory Bandwidths by Technology. N.b. the quoted nominal frequencies are not comparable – different technologies define the frequency in different ways.

CPU Speed

Obviously more is better but at some point it wont speed up your simulations if the memory cannot supply data fast enough to the CPU. This is often the case.

Dual Core CPUs and multi-CPU nodes

These architectures increase CPU compute speed but do not increase the total memory bandwidth of the node (but see Opteron below). On many machines our FDTD engine uses most of the machine's memory bandwidth. Therefore running two FDTD engine nodes on a dual CPU (or dual-core) machine you will not get the speed doubling might expect – typically compute speed will increase by from 1.2x to 1.6x. **Opteron** systems work very differently from Intel systems – each Opteron has a separate memory bus. So if you have a dual-Opteron machine (not just a dual-core Opteron!) then you get double the total memory bandwidth and you will get close to the 2x speed-up.

Network Performance

The faster the interconnect between nodes the smaller the simulation you can do efficiently. If you are doing a problem using 100MB or more per node, then your network is unlikely to be slowing your simulations down at all. However if you have a problem that is using only 500kB per node then network latency and bandwidth will likely be significant – the faster the network the better in that case.

From these points we can offer the following guidelines:

- Choose a compute node with a high memory bandwidth. This is a function of a) the memory speed in MHz and b) the chipset design - the chip that sits between the CPU and memory. Refer to Table 1 above.
- As of writing most modern systems use “DDR-2” memory. DDR-2 can run at higher data frequencies than the older DDR – see Table 1 above.
- A “Dual Channel DDR-2” design is faster than a standard DDR-2 design. Basically “dual channel” can in principle *double* your memory bandwidth so worth looking out for! Similarly “Quad-Channel” will double the bandwidth again.
- Look also at the memory frequency. 400MHz is common now and some systems have 533MHz memory frequencies. Again refer to
- Look at the FSB (front side bus) frequency. This controls the speed at which the CPU talks to the chipset. The faster the better. 800MHz is common now and some systems have 1024MHz.
- CPU frequency. Don't pay for the fastest CPUs – they tend to be very expensive. Two nodes of 2.8GHz CPUs each with 1GB of memory will probably run almost twice the speed of one node with a 3.6GHz CPU and 2GB of memory.
- Level-2 cache size. A bigger cache may substantially speed up certain simulations and make little difference to others. This will be determined by the dimensions of your simulations – the number of FDTD grid cells in each direction.
- Do not pay significant extra money for multiple CPUs on each node (SMP) or dual-core CPUs. Currently we do not take advantage of these and even if we do in the future it is unlikely to give you large performance gains.
- Hyperthreading. This is a similar issue to dual-core discussed above. Currently we do not take advantage of hyperthreading.
- Network: we currently support only TCP/IP cluster interconnect and recommend a Gigabit Ethernet fabric. It is not much more expensive than 100MB/s Ethernet and may prove useful in certain circumstances. We do not currently support Infiniband or Myrinet interconnects and in any case our view is that they are unlikely to speed up your simulations at all, except possibly in very unusual cases.
- Ethernet switches: buy a Gigabit Ethernet switch but do not spend money on the fastest low-latency switches – it won't speed up your FDTD simulations except in very special circumstances. The important thing is for the total bandwidth of your switch to allow all nodes to communicate at or close to 1Gb/s at the same time. So if you have a 16 port switch it should have a bandwidth of 8Gb/s to 16Gb/s. Be careful of a switch topology with two layers of switches, such that some nodes are “closer” to each other than others. For example imagine a cluster of 64 nodes where groups of 16 nodes are connected to 4 switches and the 4 switches are then connected to each other. This may potentially create a bottleneck if e.g. 8 nodes connected to switch-1 wanted to simultaneously talk to 8 nodes connected to switch-2 – all 8 nodes would have to share a 1Gb/s link.
Some switch vendors offer stacked switches where the interconnect bandwidth between switches in the stack is much higher than 1Gb/s. Anything above 8Gb/s is likely to be adequate for most applications.
- Hard disk: not important.