NEC SX9 Overview

PRACE Code Porting Workshop Cracow 2009
Harald Klimach <klimach@hlrs.de>
The NEC SX Series

• Started in 1983 with single CPU systems
• Continuous development of vector technology up to SX-9 in 2008
• New generation roughly every 4 years
• Software portable across architecture generations
• Shared memory parallel systems starting with SX-3
• Combination of shared memory nodes in a distributed memory system from SX-4 onwards
Vector is Dead

• For almost a decade now it is claimed in HPC, that vector is dead
• But vectorization by now is creeping back into nearly all HPC systems more and more
  – Pipelining essentially part of any modern processor, benefits from vectorization techniques
  – SSE registers steadily growing in size and capability
  – GPUs require threads with little more flexibility than vectors and therefore profit from vectorizable code
• It looks like vectorization is going to be a crucial ingredient in future HPC systems
• Traditional large scale rigid vector systems might soon be dead
• **BUT NOT vectorization techniques themselves**
SX-9 is a Vector System

• The SX-9 architecture is the latest incarnation of the NEC vector processors and the only such system carrying on a continuous development
• Thus it can rely on a long lived software stack for vectorization
• No other vector system survived the HPC trend towards cluster systems
• Cray tried to reemploy a vector system with the X2, but it looks like they have given up the experiment again already
• Applications intended to be run on a vector system need to take care to use its abilities in order to gain performance
• Scalar unit on SX-9 is very poor and is operated only with half the frequency of the vector part
**SX-9 Specifications**

- 256 GB/s memory bandwidth per CPU
- 102.4 GFLOP/s per CPU
  - 2.5 B/FLOP
- 16 CPUs per node
- 512 GB memory per node (up to 1 TB)
- Internode Crossbar Switch
- 32 GB/s (bi-directional) interconnect bandwidth per node (up to 256 nodes)
- Vector part operated at 3.2 Ghz
- 8 times replicated vector pipelines with 2 multipliers + 2 adders
- Hardware Division / Squareroot Unit
- Vector length of 256 words
SX-9 Single Node Features
With its large shared memory and high computing capabilities, even a single node of SX-9 is able to tackle large computing problems without the need for distributed memory parallelism.

Vectorizing applications usually gain more than 25 GFLOPs of sustained performance per processor.

The unique feature of the system is its high memory bandwidth accessing a large amount of memory, as computations are normally acting on data, this is an important feature for many applications.

Dimensions:
1.8x1.1x1.8m

Power consumption:
< 30 kW
SX-9 Memory Hierarchy

- Even the fast main memory interconnect of 256 GB/s yields only a maximum of 2.5 Bytes per FLOP
  - Too small for many applications
  - Memory Gap driving ever more applications to be memory bound
- Fastest memory chunks on SX-9 processors accessible by the programmer: Vector Data Registers (VDR)
  - Controlled by compiler directives
  - Near to immediate access
  - Very useful for temporary, reused data
- Newly introduced with SX-9 was the Adressable Data Buffer (ADB) providing 256 KB memory with 4 Bytes per FLOP
  - Controlled by compiler directives; Similar to Cache
Mind the Gap

• Applications need to take advantage of memory close to the computing units
• This storage is very limited in size, requiring blocking mechanisms to leverage computing chunks of suitable size
• Simple loops can be automatically vectorized by the compiler, use scalars for temporary values in this case
• On the SX-9 the block size is naturally given by the vector length of 256 to take advantage of the VDR
• Though there is the need for blocks, these should be as large as possible in order to hide latencies
• The memory gap is unlikely to get smaller any time soon, thus paying attention to the available memory hierarchy is important for HPC applications
SX-9 at HLRS

• SX-9 builds the vector part of our hybrid system Baku
• There are 12 Nodes
  – 1 Node dedicated for interactive and testing usage
  – Each node with 512 GB main memory
  – Connected with 32 GB/s IXS
• Two Linux Frontends with QC Intel Xeons and 128 GB main memory for pre/post-processing and compiling
• Queue policy
  – Maximum elapsing time 12 h
  – Maximum number of 8 nodes for a single job
• Shared filesystem (GFS) yields around 600 MB/s throughput from a single node (higher transfer rates may be gained with more nodes)