Available online at www.sciencedirect.com
SCIENCE DIRECT®
Nuclear Physics B (Proc. Suppl.) 140 (2005) 176–182
NUCLEAR PHYSICS B PROCEEDINGS SUPPLEMENTS
www.elsevierphysics.com

The apeNEXT project*

The APE Collaboration: F. Bodin[a], Ph. Boucaud[b], N. Cabibbo[c], F. Di Carlo[c], R. De Pietri[d], F. Di Renzo[d], W. Errico[e], H. Kaldass[f], A. Lonardo[c], S. de Luca[c], J. Micheli[b], V. Morenas[g] O. Pene[b], D. Pleiter[h], N. Paschedag[f], F. Rapuano[i], D. Rossetti[c], L. Sartori[e], F. Schifano[e], H. Simma[f], R. Tripiccione[j], P. Vicini[c]

[a]IRISA/INRIA, Campus Université de Beaulieu, Rennes, France

[b]LPT, University of Paris Sud, Orsay, France

[c]INFN, Sezione di Roma, Italy

[d]Physics Department, University of Parma and INFN, Gruppo Collegato di Parma, Italy

[e]INFN, Sezione di Pisa, Italy

[f]DESY Zeuthen, Germany

[g]LPC, Université Blaise Pascal and IN2P3, Clermont, France

[h]NIC/DESY Zeuthen, Germany

[i]Physics Department, University of Milano-Bicocca, Italy

[j]Physics Department, University of Ferrara, Italy

In this talk I report on the status of the apeNEXT project. apeNEXT is the last of a family of parallel computers designed, in a research environment, to provide multi-teraflops computing power to scientists involved in heavy numerical simulations. The architecture and the custom chip are optimized for Lattice QCD (LQCD) calculations but the favourable price performance ratio and the good efficiency for other kind of calculations make it a quite interesting tool for a large class of scientific problems.

1. The APE project

The APE project started in 1984 with the design and manufacture of a 1 GFlops parallel computer for LQCD. The original group had only theoretical and a few experimental physicists involved [1]. During the years students in physics and computer science have joined the collaboration to build the large community that today designs and uses APE computers. A further enlargement of the collaboration came by the end of the APEmille project when the DESY and the Orsay group joined in. Table 1 summarizes the main features of all computers of the APE family.

As one can see in table 1, all APE machines before apeNEXT are based on a SIMD architecture. Since APEmille, however, the possibility of local addressing, to increase the ease of programming in certain classes of problems, has been introduced. The term "flexible" in the topology row refers to the possibility of performing hardware-controlled next-to-nearest-neighbour communication. The number of registers, the clock speed and the word size have increased following the evolution of technology. An important step, taken during the APE100 initial phase, was the custom design of the VLSI chips. At that time, in fact, a certain number of software packages for the schematic capture, the simulation and the VLSI synthesis became available to ordinary users outside the design centers of large companies. This allowed the collaboration to develop

*Talk given by F. Rapuano at the Lattice Conference 2004, Fermilab (IL), USA.

0920-5632/$ – see front matter
doi:10.1016/j.nuclphysbps.2004.11.359

Table 1
The family of APE processors. The year in parenthesis is the time when the project was concluded. Physics runs in general have started quite earlier on prototypes or small scale machines.

	APE(1988)[1]	APE100(1993)[2]	APEmille(1999)[3]	apeNEXT(2004)[4]
Architecture	SISAMD	SISAMD	SIMAMD	SPMD
Number of nodes	16	2048	2048	4096
Topology	flexible 1D	rigid 3D	flexible 3D	flexible 3D
Memory	256 MB	8 GB	64 GB	1 TB
Registers (Word Size)	64(32)	128(32)	512(32)	512(64)
Clock speed	8 MHz	25 MHz	66 MHz	200 MHz
Peak speed	1 GFlops	100 GFlops	1 TFlops	7 TFlops

the main building blocks of APE machines with all and only those devices relevant for the functionalities needed for our purposes. In particular one of the main characteristics of all APE floating point processors is to perform, at each clock cycle a "normal" operation $a \times b + c$, where a,b and c are complex numbers. This operation is the most performed operation in a LQCD program so it is crucial that the processor perform this operation as fast as possible to obtain high efficiencies in the application programs. Furthermore a custom design allows to keep power consumption as low as possible which is a relevant point when one plans to have thousands of such components in a parallel machine. A final comment can be made regarding clock speeds. One may notice that in all projects the clock frequency is quite low while gaining on floating point speed from parallelism of operations. A low clock frequency allows to avoid sophisticated technical solution for boards manufacturing and reduces the possibility of errors due to tight timings in control and data signals thus improving operating reliability. Without going into the details of each machine one can summarize the main rules that have always been followed in the design of APE computers:

- The computer should be very efficient for LQCD calculations (and in fact efficiencies up to 65% have been reached in the Dirac operator kernel calculation) but reasonably efficient for other fields.
- A large number of registers for efficient code optimization with no need for difficult to manage cache memories.
- A microcoded architecture with a very long instruction word (VLIW) to have all devices under program control at each clock cycle.
- Reliable and safe hardware solutions.
- A large effort in the system software design to give the user high quality programming and optimization tools.

2. apeNEXT architecture

apeNEXT [4] is the last of the processors designed by the APE group. The goal of the project is to reach multi-TFlops performances needed for state-of-the-art LQCD simulations with fermion loops. The project started in the year 2000, the general ideas were presented at the Bangalore Lattice Conference. It has been concluded by the first quarter of this year with the successful test of all components in a 16 then enlarged to 256 nodes prototype. A large mass production will start by October this year. apeNEXT has an important technical innovations with respect to the previous generations of APE machines: one custom VLSI chip integrates all main functionalities, including network devices. Fig. 1 shows the layout of this chip called J&T.

This $(1.5cm)^2$ chip performs, with 64 bits accuracy, all operations that were performed at 32 bit in the 30x50 cm^2 board shown in fig 2. Furthermore J&T contains 7 bidirectional 200