apeNEXT: A MULTI-TFLOPS COMPUTER FOR SIMULATIONS IN LATTICE GAUGE THEORY

apeNEXT-Collaboration: F. Bodin[1], Ph. Boucaud[2], N. Cabibbo[3], F. Di Carlo[3], R. De Pietri[4], F. Di Renzo[4], H. Kaldass[5], A. Lonardo[3], M. Lukyanov[5], S. De Luca[3], J. Micheli[2], V. Morenas[6], O. Pene[2], D. Pleiter[7], N. Paschedag[5], F. Rapuano[3], L. Sartori[8], F. Schifano[8], H. Simma[5], R. Tripiccione[8], P. Vicini[3]

ABSTRACT

We present the APE (Array Processor Experiment) project for the development of dedicated parallel computers for numerical simulations in lattice gauge theories. While APEmille is a production machine in today's physics simulations at various sites in Europe, a new machine, apeNEXT, is currently being developed to provide multi-Tflops computing performance. Like previous APE machines, the new supercomputer is largely custom designed and specifically optimized for simulations of Lattice QCD.

1 Introduction

For many non-perturbative problems in quantum field theory, numerical simulations on the lattice offer the only known way to compute various quantities from first principles. Much progress has been made during recent years [1], e.g. in calculating

[1]IRISA/INRIA, Campus Université de Beaulieu, Rennes, France
[2]LPT, University of Paris Sud, Orsay, France
[3]INFN, Sezione di Roma, Italy
[4]Physics Department, University of Parma and INFN, Gruppo Collegato di Parma, Italy
[5]DESY Zeuthen, Germany
[6]LPC, Université Blaise Pascal and IN2P3, Clermont, France
[7]NIC/DESY Zeuthen, Germany
[8]Physics Department, University of Ferrara, Italy

the light hadron spectrum, the light quark masses, the running coupling constant α_s or observables in heavy quark physics. Further-on, lattice simulations allow the study of phenomena, like chiral symmetry breaking and confinement, or of phase transitions in the early universe.

The computer resources required for such simulations are huge and critically depend on the physical parameters, like quark masses, and the formulation of the theory on the lattice, e.g. with improved chiral properties. To make the necessary resources available, various research groups engage in the development of massively parallel computers which are specifically optimized for this kind of applications. One of these projects is APE [2] which is currently developing its fourth generation of machines, apeNEXT, within the framework of an European collaboration by INFN (Italy), DESY (Germany) and the University of Paris Sud (France).

2 apeNEXT Architecture

apeNEXT is designed as a massively parallel computer following the Single Program Multiple Data (SPMD) programming model. All machine functionalities, including the network and memory interfaces, are integrated into one single custom chip running at a clock frequency of 200 MHz. The nodes run asynchronously and are implicitly synchronized by communications.

The processor core is a 64-bit architecture and all floating point operations use IEEE double precision format. A so-called *normal* operation $a \times b + c$, with a, b, and c complex numbers, can be started at each clock cycle. The peak performance of each node is therefore 1.6 GFlops.

Each node is a fully independent processor with 256–1024 MBytes private memory (standard DDR-SDRAM with ECC) storing both program and data. Since apeNEXT has very long instruction words (VLIW), microcode de-compression and instruction buffers are used to reduce conflicts between data and instruction loading. A large number of 256 $(64 + 64)$-bit registers allows efficient data re-use.

The high-performance communication network has a 3-d torus topology. Each link moves one byte per clock cycle and the startup latency of about 20 clock cycles, i.e. 100 ns, is very short. Communications and arithmetic operations can be carried out in parallel. By loading local or remote data through dedicated pre-fetch queues one can almost perfectly exploit the concurrency between memory accesses, network transfers, and arithmetic operations.

On an apeNEXT processing board 16 nodes are mounted, and 16 boards are interconnected on the backplane within a crate. Larger systems are assembled by connecting together several crates using external cables. Thus, a single rack system with 2 crates has 512 nodes and provides a peak performance of 0.8 TFlops. The footprint is about 1 m^2 only and the estimated power consumption is < 10 kW. Due to this very moderate power dissipation air cooling is still possible.

apeNEXT systems are accessed from a front-end PC via a custom designed host-interface PCI-board. It uses a simple I2C link for bootstrapping and basic operating system requests, while I/O operations are handled via high-speed data links. By connecting one or more processing boards of an apeNEXT machine with host-interface boards the overall I/O bandwidth can be adjusted to the user needs.

Two high-level programming languages, TAO and C (with suitable extensions for parallelisation) are supported on apeNEXT. After further optimisation of the intermediate assembly, the microcode is generated, scheduled and compressed.

3 Status and Outlook

The design of apeNEXT has been finished and prototype processors are to be delivered in autumn 2003. All other hardware components exist and have been tested. The final procurement costs are estimated to be 0.5 €/MFlops peak performance.

The full VHDL design of the machine has been used for benchmarking. Key lattice gauge theory kernels were measured to run at a sustained performance of $O(50\%)$ or more. While the apeNEXT design is optimized for simulating lattice QCD, we expect the architecture to be suitable also for other applications.

European research groups in lattice gauge theory aim for large dedicated installations of massively parallel computers during the next years. In Edinburgh (UK) a 10 TFlops QCDOC machine will be installed by 2004 and in Italy a 10 TFlops apeNEXT installation is planned. In Germany researchers from various groups joined the Lattice Forum (LATFOR) initiative. They have set up a broad research program [3], assessed the required computing resources, and proposed a 15 and 10 TFlops installation of apeNEXT hosted by NIC at DESY (Zeuthen) and GSI (Darmstadt).

References

1. For an overview see “Proceedings of the XXth International Symposium on Lattice Field Theory,” Nucl. Phys. B (Proc. Suppl.) 119 (2003).
2. R. Alfieri *et al.* [APE Collaboration], arXiv:hep-lat/0102011; R. Ammendola *et al.*, arXiv:hep-lat/0211031; F. Bodin *et al.*, arXiv:hep-lat/0306018.
3. R. Alkofer *et al.*, http://www-zeuthen.desy.de/latfor .