Quantum Chromodynamics
Simulating Quarks and Gluons

The goals of high energy and nuclear physicists are to identify the fundamental building blocks of matter, and to determine the interactions among them that lead to the physical world we observe. Major progress has been made towards these goals through the development of the Standard Model of High Energy and Nuclear Physics. The Standard Model consists of two quantum field theories. One provides a unified theory of the weak and electromagnetic interactions, and the other, Quantum Chromodynamics (QCD), provides a theory of the strong, or nuclear, interactions. The Standard Model identifies quarks as the fundamental building blocks of strongly interacting matter, and gluons as the carriers of the strong forces. QCD explains how quarks and gluons interact to form the particles we directly observe, such as protons, neutrons and the host of short lived particles produced in accelerator collisions. It also describes how they interact to form atomic nuclei.

The Standard Model has been enormously successful. It correctly describes a vast amount of data produced in high energy and nuclear physics accelerator experiments, and in cosmic ray experiments. It has passed every experimental test to which it has been put. However, our knowledge of the Standard Model is incomplete because it has been difficult to extract many of the most interesting predictions of QCD. The only existing method for doing so from first principles and with controlled systematic errors is through large scale numerical simulations within the framework of lattice gauge theory.

QCD simulations are needed to determine a number of the basic parameters of the Standard Model, to make precision tests of it, and to obtain a quantitative understanding of the physical phenomena controlled by the strong interactions. Despite the many successes of the Standard Model, it is expected that by probing to sufficiently short distances or high energies, one will find a more encompassing theory. It is not that the Standard Model is expected to be proven wrong, it is simply expected to have a limited range of applicability, just as classical mechanics does. A central objective of the experimental programs in high energy and nuclear physics is to search for a breakdown in the Standard Model and new physics beyond it. However, one cannot determine whether a theory breaks down without knowing what it predicts, so QCD simulations are an integral part of this effort. Thus, numerical studies of QCD play an important role in efforts to obtain a deeper understanding of high energy and nuclear physics.

Impact on Science:

High energy and nuclear physicists seek to understand matter at the smallest distance scales or largest energy scales. However, these fields impact science at all scales, including the largest probed by astrophysics. Because of the inherent interest and scientific importance of high energy and nuclear physics, the United States, the European Community and Japan all support very large experimental programs in these fields. Major goals of the experimental programs are to: 1) verify the Standard Model of High Energy Physics, or discover its limits, 2) determine the properties of strongly interacting matter under extreme conditions, and 3) understand the internal structure of nucleons and other strongly interacting particles. Lattice QCD simulations are essential to research in all of these areas.

Most strongly interacting particles decay via the weak interactions. In many cases an experimental measurement of the decay properties, coupled with a lattice QCD calculation, will provide a direct measurement of one of the fundamental parameters of the Standard Model. By determining the same parameter from different experiments and lattice QCD calculations, one can check the consistency of the Standard Model. A major, international effort is in progress to perform experiments required to determine Standard Model parameters associated with heavy quarks (charm, bottom and top). These are among the least well known parameters in the Standard Model, and their precise determination would provide a very important test of the theory. As Figure 1 illustrates, lattice QCD calculations are a vital part of this effort. Each solid colored band is the allowed region of the Standard Model parameters ρ and η from a particle experiment and corresponding lattice calculation. For the Standard Model to be correct the solid bands must overlap, and ρ and η must lie in the region of overlap. the figure on the left shows the constraints as they exist today. The figure on the right shows the constraints as they would exist if the errors in the lattice QCD calculations were reduced to 3% with no improvement in the experiments. Thus, in this case, as in many other tests of the Standard Model, it is essential that improvements in lattice QCD calculations keep pace with those in experiments in order to reap the full return on the very large investments being made in the experiments.

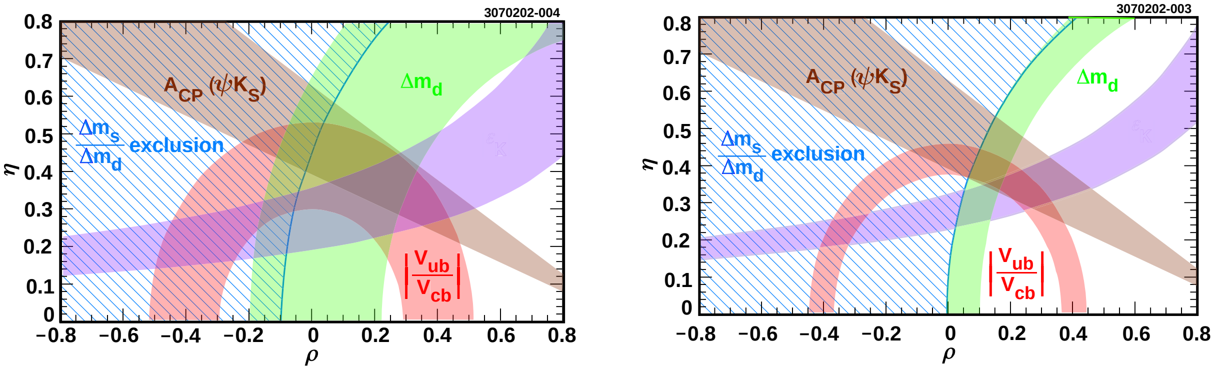

Figure 1: *Constraints on the Standard Model parameters ρ and η (one sigma confidence level). For the Standard Model to be correct, they must be restricted to the region of overlap of the solidly colored bands. The figure on the left shows the constraints as they exist today. The figure on the right shows the constraints as they would exist with no improvement in the experimental errors, but with lattice QCD uncertainties reduced to 3%.*

At ordinary temperatures and densities quarks and gluons are confined in elementary particles. However, at more than 100 million Kelvin, as occurred in the first moments in the expansion of the Universe, or at nuclear matter densities exceeding one billion grams per cc, as may occur in the cores of compact stars, quarks and gluons are liberated and form a quark–gluon plasma, an entirely novel form of matter. Creating and studying the early Universe in microcosm is one of the principal and exciting goals of experiments at the Relativistic Heavy Ion Collider at the Brookhaven National Laboratory. Since the plasma would exist only for a fleeting moment before decaying into ordinary particles, it cannot be directly observed. Confirming its existence depends on model calculations that rest on results from first principles numerical simulations. Results are urgently needed to support the experimental effort.

The following observables, which can all be computed on the lattice, are required: 1) the temperature and order of the ordinary-matter/plasma phase transition; 2) the equation of state (energy density and pressure *vs.* temperature); 3) the strange quark excess and strangeness fluctuations; 4) the weakening of quark binding as the temperature and/or density increases; 5) the plasma modes of oscillation; and 6) phase diagram and equation of state at non–zero baryon density.

Knowing the temperature of the phase transition and equation of state tells us whether the energy of a heavy ion collision is sufficient to produce the plasma. If the transition is strongly first order, super–cooling and phase separation could occur with dramatic consequences for the subsequent evolution of the final state. Excess strangeness production and strangeness fluctuations have been proposed as a signature for plasma production, but a first principles quantification is needed. The weakening of binding in heavy quark–antiquark systems is another signature. Plasma modes may also have a measurable effect in the production of muon–antimuon pairs. Finally, the phase diagram at nonzero density is central to a characterization and possible indirect detection of quark-matter cores inside dense stars.

Protons, neutrons, and other strongly interacting particles, which comprise most of the known mass of the universe, are dramatically different from any other quantum systems that have ever been explored. The quark and gluon constituents of a proton are absolutely confined within it at ordinary temperatures and densities, and cannot be removed, unlike electrons that can be removed from an atom, or nucleons that can be removed from a nucleus. Unlike the photons, which carry the electromagnetic forces in atoms, the gluons, which carry the strong force, are essential constituents of the proton, accounting for half of its momentum and undergoing significant excitations. Whereas most of the mass in atoms and nuclei comes from the masses of the constituents, most of the mass of the proton arises from the gluon interactions. Indeed if the quarks became massless, the mass of the proton would hardly decrease. Because quarks and gluons interact so strongly and non–linearly, the only known way to calculate the remarkable properties of the strongly interacting particles they form is by numerical studies of QCD.

In the three decades since the discovery of quarks in the nucleon, a tremendous experimental effort has been devoted to exploring the quark and gluon structure of hadrons at Bates, Brookhaven National Laboratory (BNL), Cornell, Fermi National Accelerator Laboratory (FNAL), Jefferson National Accelerator Facility (JLab), the Stanford National Accelerator Laboratory (SLAC), and international accelerators. The distributions of quarks and gluons in nucleons have now been measured in exquisite detail. Surprising puzzles have arisen from these experiments, such as the discovery that hardly any of the spin of the proton comes from the spin of valence quarks, in striking contrast to atoms and nuclei, whose spin arises primarily from the spin of valence fermions. Other experiments measure the strange quark content of the nucleon and seek evidence for hadrons with exotic quantum numbers that would expose novel QCD phenomena. Lattice calculations are essential to quantitatively understand these experiments from first principles, and to obtain physical insight into how QCD works.

Scientific Opportunities:

There is a wide range of scientific opportunities in lattice QCD. As in most areas of computational science, the problems that can be addressed depend critically on the computational resources available. Figure 2 provides an indication of what could be achieved over the next decade.

Figure 2: *Figure is in accompanying power point and postscript files. Figure Caption: Resources required to reach major research goals.*

The U.S. lattice QCD community currently sustains just under 1.0 Tflop/s. That is, it has continuous use of computers which together sustain this rate. No one machine realizes this performance. These resources come from allocations at DOE and NSF supercomputer centers, special purpose computers located at Columbia University and BNL, and optimized clusters at FNAL and JLab. They have allowed the determination of a limited number of key quantities to an accuracy of a few percent. Among these quantities are the strong coupling constant, the masses of the c and b quarks, and the decay constants of the π and K mesons. These resources have also enabled the development and testing of new formulations of lattice QCD that will significantly improve the accuracy of future calculations. As might be expected, these more sophisticated formulations require significantly more computer power than the simpler ones they replace.

In order to provide timely support to the experimental programs in high energy and nuclear physics, the U.S. lattice QCD community will need to sustain 100 to 200 Tflop/s within the next few years, and multiple Pflop/s within the next ten years. Resources at this level would enable a wealth of exciting calculations with existing algorithms and codes. Here we mention a few examples.

The study of the decays of mesons with one light and one heavy quark via the weak interactions is a key component in the program to make precision tests of the Standard Model. Two examples are the B and D mesons, which contain a heavy bottom and charm quark, respectively. Their decays to final states that do not contain strongly interacting particles are characterized by decay constants. Calculations currently in progress are expected to determine these decay constants to an accuracy of approximately 5%. However, an accuracy of 1% is needed for Standard Model tests. With current algorithms this improvement will require a hundred fold increase in computing resources. The Cleo-C experimental program in progress at Cornell will measure the D meson decay constant to high accuracy, thereby providing a validation of the lattice calculation. However, the B meson decay constant will be extremely difficult to measure, so the lattice determination of it will be critical. With computing resources of this magnitude, there are a host of other processes that can also be calculated with accuracy sufficient to have an important impact on our understanding of the Standard Model. These include the decays of B and D mesons to final states containing strongly interacting particles, and the mixing of neutral K and B mesons with their anti–particles. Computational resources at the Pflop/s level would, for example, enable a high accuracy evaluation of ε'/ε. This quantity characterizes the violation of CP symmetry in the K meson system. (C stands for charge conjugation, the interchange of particles and anti-particles, and P for parity, the inversion of space). There has recently been a very accurate experimental measurement of this quantity, and an equally accurate lattice determination would provide an important test of the Standard Model. More generally Pflop/s scale resources would allow use of the most sophisticated of the recently developed lattice formulations with realistic quark masses. Such simulations would significantly improve the accuracy of all lattice QCD calculations.

Turning to the study of high temperature QCD, current calculations predict the temperature of the phase transition at zero baryon density with an accuracy of 15%, but have not yet established whether it is first order, or merely a crossover, at realistic quark masses. The equation of state and strangeness fluctuations have been determined only on coarse lattices where discretization artifacts are not completely under control. Simulations have confirmed the weakening of quark binding between very heavy quarks, but for light quarks the methodology has not been adequately developed. It is possible to follow the phase boundary slightly away from zero baryon number density, but unsolved sign fluctuation problems prevent ready simulation at high baryon number density.

With recently developed improved lattice formulations, and resources sustaining 100–200 Tflops/s it will be possible to determine the temperature and order of the phase transition, the equation of state of the quark–gluon plasma, the strange quark excess and binding, and observe the weakening of quark binding, all at a resolution sufficient for modeling heavy ion collisions. Pflop/s resources will be required to study the properties of strongly interacting matter at high densities. There have been fascinating proposals regarding the possibility of superconducting phases at very high densities, which can be studied from first principles only with lattice methods. These studies have potential impact on our understanding of the early development of the universe and on the cores of neutron stars.

Resources at the 100-200 Tflop's level will lead to major advances in our understanding of the internal structure of the nucleon. For example, they will enable a precision calculation of electromagnetic form factors characterizing the distribution of charge and current in the nucleon. This calculation is necessary to understand recent precise measurements of proton and neutron form factors at JLab and measurements at Bates and JLab of the contribution of strange quarks to these form factors. They will also make possible calculation of the quark structure of the nucleon at a level that will provide fundamental understanding of experiments at JLab using 6–12 GeV electron beams and HERMES and RHIC–spin high-energy experiments. In addition, they will enable calculation of transitions to excited nucleon states revealing the degree of deformation present in these states and allow calculation of the potential between selected hadrons to explore the QCD origin of nuclear forces. Pflop/s resources would enable study of the gluon structure of the nucleon, in addition to its quark structure. They would also allow precision calculation of the spectroscopy of strongly interacting particles with unconventional quantum numbers, guiding experimental searches for states with novel quark and gluon structure.

The calculations listed above would significantly deepen our understanding of the Standard Model, and therefore of the basic laws of physics.

Research Issues in Lattice QCD:

QCD is formulated in the four–dimensional space–time continuum. In order to carry out numerical studies it is necessary to reformulate the theory on four–dimensional space–time lattices. Physical observables are expressed in terms of Feynman path integrals, which can be evaluated numerically in the lattice formulation. In current simulations these integrals involve hundreds of millions of variables. The fact that the quarks are fermions severely complicates the evaluation of these multi–dimensional integrals by giving rise to non–local couplings among the integrands. Nevertheless, well tested algorithms exist to accurately evaluate these integrals. These algorithms are hybrids of molecular dynamics and Monte Carlo methods.

Once simulations have been performed on the lattice, two major challenges remain in order to obtain results for the continuum theory. One is to perform extrapolations to the limit of zero lattice spacing. In recent years improved formulations of QCD on the lattice have been developed that sharply reduce finite lattice spacing artifacts, thereby significantly reducing errors arising from the extrapolation to the continuum limit. The advantage of being able to work at increased lattice spacings can be seen from the fact that the computational resources required for a simulation grow approximately as the seventh power of the inverse of the lattice spacing. A second challenge arises from the fact that the up and down quarks have masses that are much smaller than those of the other quarks, and of the typical energy scale in QCD. Because the required computational resources grow approximately as the third power of the inverse of the quark mass, it has not been possible to perform simulations at the physical masses of the up and down quarks. Instead one performs simulations at a range of unphysically large values of these masses, and performs extrapolations to the physical ones. These extrapolations are guided by theory based on a property of QCD known as chiral symmetry. The improved lattice formulations also reduce chiral symmetry violating lattice artifacts, bringing these extrapolations under significantly better control. Thus, the improved lattice formulations of QCD are having a dramatic impact on our ability to perform accurate calculations. Indeed, in the last several years, the advances arising from improved lattice formulations have far outweighed those from increases in computing power. Continued work in this area will be a very important component of research in QCD in the coming years. However, it should be emphasized that in order to implement these more sophisticated formulations of lattice QCD much more computing power is required than to implement the older ones. Indeed, it seems likely that it will not be possible to use the most promising of them with realistic quark masses until one has computing resources capable of sustaining multiple Pflop/s.

Between 70% and 90% of the computer resources in lattice QCD calculations are spent on repeated inversions of the Dirac operator for the quarks in a background field produced by the gluons. On the lattice, the Dirac operator is a sparse matrix, and iterative techniques, such as the conjugate gradient algorithm, are typically used to perform this inversion. To date the rather random nature of the non–zero elements of the Dirac operator have hampered the application of standard multi–scale techniques to accelerate the inversion. This is a problem on which a successful collaboration between application physicists and applied mathematicians could reap large benefits.

Going beyond the standard approaches, several algorithmic obstacles hamper the application of lattice methods to a variety of tantalizing new problems. For non–zero baryon densities the measure of the Feynman path integrals is not positive definite, making importance sampling techniques extremely expensive, if applicable at all. Recent developments appear to enable simulations at low baryon density, in the region relevant to the study of heavy–ion collisions. However, new ideas are needed to carry out studies of QCD at large densities. A related issue is the inability of current algorithms to deal with time dependent problems, such as the calculation of scattering cross–sections.

A more fundamental issue arises for lattice theories of the Standard Model when parity violation is included. No known lattice formulation of left handed neutrinos has been found that does not introduce mirror right handed neutrinos, which do not exist in nature. This is troubling as the lattice is the most precise way to define a quantum field theory non-perturbatively. Nevertheless this is not a practical problem for experimental predictions since weak interactions in most cases can be handled perturbatively.

Technology Barriers:

Lattice QCD is a mature field with an extensive code base. Three major codes, CPS, MILC and

SZIN are publicly available. Each of these packages contains tools to carry out simulations with a variety of lattice formulations, and to study a broad range of physics problems. Each contains over 100,000 lines of code. The MILC and SZIN codes run on virtually all commercial machines, as well as on clusters, while the CPS code runs on special purpose computers designed at Columbia University, the most recent of which is the QCD on a Chip (QCDOC).

A major effort is in progress under the DOE SciDAC Program to develop a QCD Applications Program Interface (QCD API), which will enable the U.S. lattice community to use special purpose computers, clusters and commercial machines with high efficiency, while preserving the large investment in existing codes. This project includes the creation of standards for communication interfaces, optimized low level algebraic kernels, optimized high level operators, and run-time functions. Existing application codes can easily be ported to the QCD API by linking to libraries that are being created for each of the targeted architectures.

A number of major benefits are expected from the SciDAC effort. It has already been demonstrated that the QCD API will enable existing codes to be ported to, and obtain excellent performance on, the QCDOC with a minimum of effort. The high level structure of the QCD API should allow new applications to be developed easily, enabling the physicists to focus on QCD, rather than the code. In addition, with the existence of a true community code young scientists entering the field will be able to move to new collaborations, as they progress from graduate student to postdoc to junior faculty, without having to learn new software packages.

This SciDAC Software Project involves nearly all the senior lattice gauge theorists in the U.S., as well as a number of computer scientists and computer engineers located at universities and national laboratories. The effort currently supports approximately 10 FTE, and has the participation of many more on a voluntary basis. As the community moves to computing platforms sustaining hundreds of Tflop/s and then Pflop/s, this software effort will need to grow by at least 50% in order to maintain and enhance the code that is being currently developed.

Resources Required:

There are numerous exciting scientific opportunities in lattice QCD, and it is essential that the U.S. lattice community take advantage of them expeditiously. Important experiments that depend on lattice QCD calculations for their interpretation are in progress or are planned for the near future, so we must move quickly to support the experimental program in a timely fashion. Recent advances in algorithms, particularly new formulations of QCD on the lattice, now enable calculations of unprecedented accuracy, provided the required computational resources are available. Furthermore, we must act now to recover U.S. computational leadership in lattice QCD. Lattice gauge theory was invented in the United States, and U.S. physicists have traditionally been intellectual leaders in the field. However, for the last five years greater investments in computational resources have been made in Europe and Japan. Within the next year European lattice theorists will have dedicated resources sustaining over ten Tflop/s. If U.S. physicists are to regain leadership of the field, and be able to attract outstanding young scientists, we must act now to obtain comparable resources. Within the next five years the U.S. lattice community needs to sustain hundreds of Tflop/s, and by the end of the decade multiple Pflop/s. Simplifying features of QCD simulations provide a pathway for doing so through the construction of special purpose hardware.

The underlying homogeneous and local character of Quantum Chromodynamics implies that the corresponding numerical methods are very well suited to parallel computing and can be efficiently mounted on a massively parallel computer with a mesh communications network. This offers important opportunities for highly cost effective and very large scale computing since mesh machines can easily scale to tens of thousands of processors.

A major challenge in the study of lattice QCD is to bring very large computing resources to bear on lattices of moderate size. For a fixed space-time volume the computational effort required for lattice QCD calculations grows with a high power of the number of lattice points. As a result, substantial benefits can be obtained from improved algorithms that permit increaingly accurate answers to be obtained without signifanctly increasing the number of lattice points (decreasing the lattice spacing), even though these algorithms increase the number of floating point operations per lattice point. Furthermore, it is vitally important to decrease the masses of the up and down quarks towards their physical values. This requires even more significant increases in computation per lattice site. Thus, computers with increasing numbers of processors are likely to be applied to lattices of limited size. These aspects of lattice QCD calculations permit a relatively small amount of memory to be used on each processor, presenting an opportunity to increase the cost effectiveness of custmized machines.

An important consideration in designing customized computers for QCD is the relationship between data movement and floating point operations. The basic operation in QCD simulations is the multiplication of a three component vector, whose elements are complex numbers, by a 3×3 matrix, whose elements are also complex numbers. In performing this operation an average of 1.45 bytes of input data is required for each floating point operation, and an average of 0.36 bytes of output data is produced. In current commercial processors, data movement cannot keep up with floating point calculations unless the problem fits entirely into cache. However, restricting the amount of data on a processor to that which will fit into cache will be counter–productive unless the the inter–processor communications system system is sufficiently robust . One typically decomposes the four–dimensional lattice into domains of L^4 lattice sites, each domain being assigned to one processor. Since the underlying interactions are short range, in order to update variables at a lattice site, one needs information from only a few neighboring ones. Thus, one can update those sites needing only information stored on a processor, while the data needed to update sites on the edges of the domain is collected from neighboring processors. The sustained inter–processor bandwidth (in Megabytes/s) needed to completely overlap communication with floating point calculations interior to an L^4 domain is approximately $0.364MF/L$, where MF is the sustained floating point performance in Megaflops/s. Thus, careful attention must be paid to the balance among memory bandwidth and latency, cache size, floating point performance and inter–processor bandwidth and latency, in order to obtain strong performance.

These specific characteristics of lattice QCD can be effectively exploited using two strategies. With current technology, highly effective machines capable of sustaining hundreds of Gigaflops can be constructed from workstation clusters, a speed that will grow to 1 Teraflops in a year to two. By careful selection of systems meeting the high memory and inter–processor bandwidth requirements of QCD and the use of mesh communication networks, efforts at both FNAL and JLab have resulted in very successful, dedicated machines constructed from commodity components. This strategy requires little design time and can respond quickly to advances in widely available and highly cost effective commercial technology.

A second strategy, exploits the characteristics of lattice QCD to design and construct specifically targeted machines. Here considerably more design effort and time are required. However, highly scalable and very powerful machines result. The present QCDOC project, part of the U.S. SciDAC effort targeting machines sustaining tens of Teraflops, is a significant example.

Metrics of Success:

There is a wealth of data from high energy and nuclear physics experiments that can be used to validate lattice QCD calculations. This includes knowledge of the masses of strongly interacting particles, their decay properties and their internal structure. True success will be the generation of new results with accuracies sufficient to advance our understanding of fundamental physics. A successful program in lattice QCD would lead to calculations of the decay properties of strongly interacting particles to an accuracy needed to make precise tests of the Standard Model. It would enable the determination of the phase diagram and equation of state of the quark–gluon plasma at high temperatures and at baryon densities relevant to heavy–ion collision experiments, and it would provide a quantitative understanding of the internal structure of strongly interacting particles. Finally, a successful lattice QCD program would develop the tools needed to study strong coupling field theories in general, so, when a more encompassing theory than the Standard Model is developed, high energy and nuclear physicists will be able to extract its predictions. Success in all of these areas is possible in the next decade provided the computational resources outlined above become available.

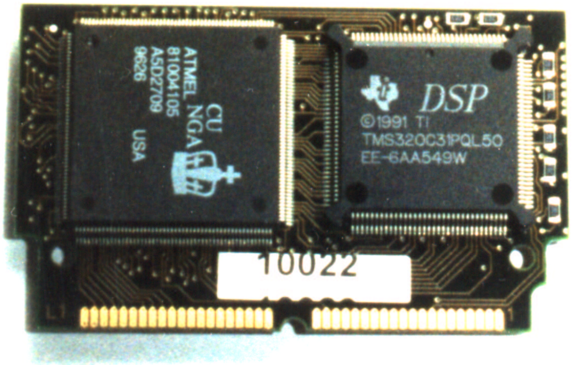

Figure 3: A single node from the QCDSP computer. Approximately 1.8” × 2.6”, this card has a sustained speed of 15 Mflops, a memory of 2 Mbytes and cost $\sim$ \$80 to construct in 1998.

Sidebar

Over the past twenty years a substantial portion of lattice QCD research has been performed on specially tailored computers built either by the research groups themselves or in collaboration with industry. A relatively recent example is the QCDSP computer designed and built by a collaboration centered at Columbia University and completed in 1998. This machine is based on a Texas Instruments DSP (digital signal processor) and has all the components required for a single processing node (DSP, memory and communications) mounted on a miniature PC board, a little larger than a credit card. A picture of such a node (which cost $\sim$ \$80 in 1998) is shown in Figure 3.

An 8K processor machine is located at Columbia and a second 12K processor machine at the RIKEN BNL Research Center (RBRC) at the Brookhaven National Laboratory. Together these machines represent a sustained performance of 0.3 Tflops. With a price performance of about \$10/sustained Mflops the RBRC machine won the Gordon Bell Prize for price/performance at SC98.

For the past three years a collaboration of the group at Columbia, the RBRC, the UKQCD collaboration and scientists at the T. J. Watson Research Laboratory of IBM has been designing a new machine, named QCDOC for Quantum Chromodynamics on a Chip. This computer is based on a system-on-a-chip, applications specific integrated circuit (ASIC) which contains a PowerPC processor, a 1 Gflops double-precision floating point unit, 4 Mbytes of on-chip memory and communications channels with a total bandwidth of 15 Gbytes/sec able to support a six-dimensional mesh network. An added DIMM socket allows off-chip memory up to 2 Gbytes. The first prototypes of this new system are now being tested and a two-node “daughter card” is shown in Figure 4.

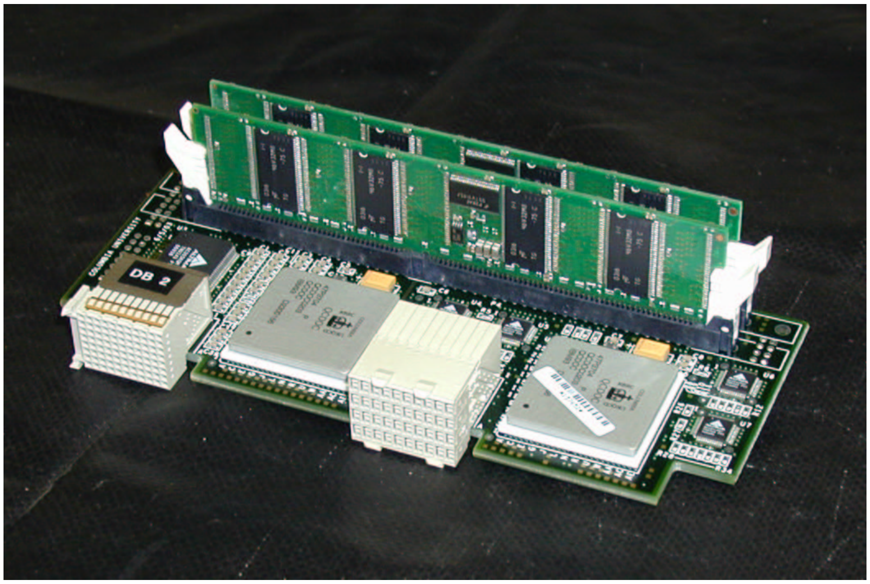

Figure 4: A daughter card holding two QCDOC nodes. This 3” × 6.6” card is expected to cost $\sim$ \$500 in quantity and to sustain $\sim$ 1 Gflops for QCDOC, even when part of a single 30K node partition. The 1.25” × 1.25” silver components are the two QCDOC ASICS. Each of the commodity memory modules mounted vertically on the board contains 256 Mbtyes of 266 MHz ECC DDR SDRAM.