LA-UR -90-4225

CONF-9009319--1

Received by OSTI

JAN 0 8 1991

LA-UR--90-4225

DE91 005879

TITLE: QCD ON THE CONNECTION MACHINE

AUTHOR(S): RAJAN GUPTA

SUBMITTED TO: PROCEEDINGS OF CONFERENCE ON LARGE SCALE COMPUTING IN THE 21ST CENTURY HELD AT CAPE COD, BOSTON, MA 9/28-10/03/90.

Los Alamos

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

QCD on the Connection Machine

Rajan Gupta †
T-8, MS-B285, Los Alamos National Laboratory, Los Alamos, NM 87545

ABSTRACT

In this talk I give a brief introduction to the standard model of particle interactions and illustrate why analytical methods fail to solve QCD. I then give some details of our implementation of the high performance QCD code on the CM2 and highlight the important lessons learned. The sustained speed of the code at the time of this conference is 5.2 Gigaflops (scaled to a full 64K machine). Since this is a conference dedicated to computing in the 21^{st} century, I will tailor my expectations (somewhat idiosyncratic) of the physics objectives to reflect what we will be able to do in 10 years time, extrapolating from where we stand today. This work is being done under a joint LANL-TMC collaboration consisting of C. Baillie, R. Brickner, D. Daniel, G. Kilcup, L. Johnsson, A. Patel, S. Sharpe and myself.

† Talk presented at *Large Scale Computing in the* 21^{st} *century*, Cape Cod, OCT 1990

1. Introduction

In the last two decades elementary particle physics has undergone a certain synthesis. Phenomena at energy scales up to 100 GeV (these high energies are achieved in experiments done at accelerator laboratories around the world) are described by four fundamental forces. Excluding gravity, they form what is called the Standard Model (SM) of elementary particle interactions. The fundamental fermions come as three families each of quarks and leptons (the top quark has yet to be detected) which interact through the exchange of vector bosons: gluons (strong force), photons (electromagnetic force) and the $W^{\pm}$, Z (weak force). A brief description of these forces is as follows:

(1) Gravity: This, the weakest of forces, is characterized by Newton's constant G_N and despite the smallness of its strength, $G_N M_p^2 \sim 6.7 \times 10^{-39}$, it is responsible for the large scale structure of the Universe. The long range gravitational force couples to mass and its attractive nature gives rise to the stability of planets, stars, solar systems and Galaxies. Our understanding of this force is still at a classical level in terms of Einstein Gravity. Even at the classical level the theory possesses rich structure like pulsars, quasars, super-nova, neutron stars, black holes, *etc.*. Its incorporation into a unified quantum model is still an open problem in spite of the numerous attempts that have been carried out in the last 60 years. At present there are a number of very interesting numerical simulations being done to understand the formation of stars and galaxies and I refer you to the talk by Larry Smarr for a glimpse into this field.

(2) Strong Interactions: The fundamental particles carrying the strong force are quarks and gluons. The quarks come in three colors (red, blue and green) and are arranged as 3 families of doublets. They interact by exchanging gluons. The gluons themselves carry the color charge which at the hadronic scale (characterized by the mass of a proton, ~ 1 GeV) has strength of $\mathcal{O}(1)$. The theory is highly non-linear and standard analytic methods (which typically rely on expansion in a small parameter like the coupling) fail because the coupling is large. Thus we have not been able to carry out the routine procedure used to solve a quantum system: first solve for the ground state and then calculate the excitation spectrum in terms of binding between the constituents. Numerical simulations of the discretized version of QCD provide a first principle method of attack and are the subject of this talk.

(3) Electromagnetic Interactions: QED is the best understood and verified field theory. The strength of the electric charge is given by the fine structure constant $\alpha_{em} = 1/137$. The electron is the prototype of the elementary particle carrying unit electric charge, while the quarks carry 2/3 or 1/3 units of electric charge. This force is responsible for the formation of atoms, molecules and solids with electrons, protons and neutrons as the basic building blocks. While we understand QED at the microscopic level, the fields of Quantum Chemistry (many electron systems), macro-molecules like proteins (many atom systems), and condensed matter (properties of bulk matter) pose tremendous challenges for supercomputing. The complexity arises because of the vast number of degrees of freedom even though they are governed mainly by the non-relativistic Schrodinger equation. If I were pressed for a definition of a general purpose computer, it would be the one which can solve these three problems efficiently. This is because of the wide class of algorithms and data structures that arise in these calculations.

(4) Weak Interactions: Both quarks and leptons carry the weak charge and interact through the exchange of $W^{\pm}$ and Z bosons. These intermediate vector bosons, unlike the photons and gluons, are very heavy so the force is extremely short ranged. The force is characterized by the Fermi coupling constant $G_F m_p^2 \approx 10^{-5}$ which is extremely small and perturbation expansion in terms of Feynman diagrams should in principle suffice. This is indeed true except in cases where the initial and final states in the quantum process include quarks. Then, in addition to the exchange of a vector boson (which is essential for the process to occur), the quarks can exchange any number of gluons. Thus strong interactions sneak in, and these QCD corrections cannot be computed by analytical methods in any reliable way. Lattice techniques again provide a first principle method for calculating the QCD corrections to weak interaction matrix elements. The results of these calculations will provide the most stringent tests of the SM. In the last five years we, and a number of other collaborations, have made significant progress in establishing the technology necessary to relate numerical results to experimental data.

Despite the many successes of the Standard Model, there are a number of deep fundamental questions whose answers lie beyond the SM. Examples of these are (a) the origin of masses for the quarks and leptons, (b) why did nature need three families of quarks and leptons, (c) the origin of the Higgs boson and details of the electroweak symmetry breaking. These are very much topics of current research, nevertheless, progress on these fronts is complementary to our ability to solve QCD. So the three broad goals of QCD calculations are (1) to verify that QCD is the fundamental theory of strong interactions, (2) to make quantitative predictions for where and how the SM fails so that one can make progress in our search for physics beyond the standard model, and (3) to develop first principle techniques for extracting predictions which can be tested against experiments for a strongly interacting theory.

Let me regurgitate why QCD calculations are important and interesting. QCD is a well defined mathematical theory of one of the fundamental forces of nature. It has very few free parameters and has an extremely rich structure that can be compared with very precise measurements.

The limitations of present numerical calculations are computational. To simulate the theory we have to discretize space time on a 4-dimensional grid and carry out a statistical Monte Carlo integration. This introduces systematic errors which are being steadily understood and brought under control by the advent of better algorithms and more powerful computers. The goal is to develop the technology to answer questions to some level of precision along with an estimate of what further resources are required to decrease the errors by a factor of 10 say. Included in this program, of course, is the possibility of reformulating the problem by which one can achieve much larger breakthroughs than possible with brute computational power.

2. Implementing QCD on the CM2

QCD simulations are ideally suited to any reasonable parallel machine that can do floating point arithmetic efficiently. The reasons for this are (1) the lattice grid is regular and the physics is homogeneous, (2) the algorithm is simple (the basic operation is complex arithmetic built up from a 3×3 matrix times 3×1 vector primitive), (3) it requires by and large only nearest neighbor communications, (4) the ratio of computations to communications is large (approximately 10:1), (5) the same basic operation is done over and over again on data that evolves by a Markov process and (6) there exist algorithms that are robust against round-off errors. Thus, towards the second half of 1988 our curiosity got translated into real interest with the advertisement that the CM2 had a peak speed of 29 Gigaflops and that there was a major resource of computational power available if one was willing to be a guinea pig.

In late 1988 Ralph Brickner started converting the pure gauge QCD code to *Lisp for the CM2 with help from TMC. His efforts provided examples of the basic constructs, communication calls, and also how to interact with the system in general. He has been the lead person in the computational part of the project. We used *Lisp since at that stage it had the most developed compiler and it also allowed system diagnostics. We started work in earnest to implement the full QCD code (SU(3) with 2 flavors of dynamical Wilson fermions) once LANL announced the purchase of a full 64K machine at the end of 1988. This is computationally the hardest problem and we would require a dedicated Cray YMP 8/32 to update a 16^4 lattice. This project is somewhat ambitious even for the CM2 and we have carried out what will only be the first generation calculations on a $16^3 \times 32$ lattice with quark masses down to the strange quark.

We got the basic code written in three weeks and were able to start production runs from the day the CM2 arrived at LANL. The debugging was done by cross comparison with the original code running on a Cray XMP. During Feburary and March 1989 we received tremendous help from J. P. Masser and C. Lasser at TMC to optimize the code, to set up check-pointing and conversion to unix readable format. The physics analysis has been carried out on the Cray's. The sustained speed of our initial code was $0.9 Gigaflops$ when scaled up to the full machine.

It was apparent to all that there were many software improvements that could be made with relatively small changes that would yield a big pay-off in the overall performance of the CM2. To make this effort comprehensive, TMC took up QCD as one of the test problems in its effort to develop an optimized compiler and to guide its future software and hardware. I will mention the two major enhancements that increased the code performance by a factor of 6 in the present CM2. I am going to have to use a fair amount of jargon associated with the CM2, however, I assume that anyone who wants to penetrate the rest of this section is already familiar with the CM2.

The Slicewise compiler: For floating point operations, the fieldwise data storage of the CM2 is not optimum. The data has to go through the transposer before a floating point word can be pipelined through the Weitek co-processor. Also, the pipe was over the number of physical sites resident in the memory of the bit serial CM processor (number of VP's) and not the sites per weitek chip (basic node for floating point operations). For QCD, and for many other problems, optimum performance is obtained with a larger pipe consisting of the sites and the loop over internal indices per site. This could be achieved if (1) the data is stored as 32 bit words (called Slicewise), (2) the memory access bypasses the transposer and (3) the loops are unrolled by hand to get the largest pipe for a given problem. To get a head-start before TMC developed a compiler to do this, Ralph Brickner in collaboration with L. Johnnston's group at TMC did this by coding in CMIS directly. The overall structure is: the data is transposed before the computationally intensive inner loop from fieldwise to slicewise, and the pipeline is set up by hand. The matrix arithmetic is done by macros written in CMIS and by using the Weitek registers with maximum efficiency.

The pipe is unrolled using a code generator, which given a lattice geometry and the conditionals on the data produces a list of addresses and operations which are loaded into the micro-sequencer at the start of the job. We suggest that any pioneering endevour of this kind by individuals interested in writing in microcode use a code generator to unroll the pipe (*i.e.* develop their own mini compiler) rather than hand coding indices. The optimizing compiler developed by TMC for the general user that will work on slicewise data is called the Slicewise compiler, and should be available by the end of 1990.

Multiwire News: Having optimized computations, the bottleneck became communications. TMC engineers realized that physical problems in 1 to 4 dimensions do not use all of the 12 communications channels available in a 12-dimensional hypercube topology (the communication node is also defined as per weitek chip). Mark Bromley and gang at TMC were able to visualize the possibility of simultaneous bidirectional communications in all four dimensions (this feature is called multiwire news). The catch was having data sitting in the right place. One needed independent communications buffers for departing and arriving data. Here, even the harshest critics of the original CM2 hardware have to accept the serendipitous presence of the now defunct transposer. These three 32 word registers are used as the buffers for multiwire news communication.

What did this effort get us? A speed up from 0.9 to 6.0 Gigaflops. And we did not sacrifice anything to get this performance -- we are using our best algorithm (which is also optimum on the cray) to do the physics we want to. The only drawback is that this compute power will allow us to do a good job on a box with $16^3 \times 32$ grid points while we need 128^4 for hard numbers.

If, by now, you have the feeling that I am a TMC groupie let me assure you that our relationship with TMC has been far less acquiescent. Our excitement stems from having worked in collaboration with a company that has produced a truely massively parallel computer and has provided a working programming paradigm (data parallelism) that mimics real physical problems. TMC's drive to be the best and to explore the frontiers of supercomputing in close cooperation with users sets them apart from all other manufacturers that I am familiar with. There is a lesson here for the manufacturers who are busy fighting windmills and worrying about the Japanese threat. To those of us whose interest is in solving problems, the subtleties of SIMD versus MIMD, special purpose versus general purpose, Amdahl's law versus ???'s bottleneck translates into just one measure – throughput. Right now, for QCD and problems in it's class, the CM's are the machines to compare against.

3. Successes of Lattice QCD

In this section I will summarize what are the important questions that we can address using Monte Carlo simulations of the Euclidean version of QCD discretized on a regular 4-dimensional grid. The list is by no means complete. An excellent set of references that discuss the status in more detail are the conference proceedings from the yearly Lattice meetings. I suggest the last four (refs. [1], [2], [3] [4]) for further reading. In addition to the topics that I have mentioned above, members of the lattice community are also working on problems in statistical mechanics, condensed matter, spin glasses and more recently protein folding (a spin off from the study of 2-dimensional gravity, random surfaces and matrix models). Such cross fertilization has been a source of new ideas and gives a broader justification for the numerical approach.

(1) To a certain extent algorithm development for lattice QCD has, so far, been a clever adaptation of methods known in fields with longer history (mostly statistical mechanics). On the other hand the history of supercomputing coincides with that of QCD. We have seen an increase by $\approx 10^5$ in computing power in the last 12 years. This increase was necessary to even test existing algorithms and to see if they are suited to QCD. As a result we now have the tools to simulate QCD with 2 flavors of quarks with no additional approximations other than that of lattice discretization. So we can now evaluate systematic errors and estimate how much more compute power will be required to calculate hard numbers.

(2) Numerical simulations provide the most convincing demonstration that confinement and asymptotic freedom coexist in pure gauge theory. There exists a linear piece in the heavy $q\bar{q}$ potential that dominates at long distances ($> 10^{-13}$ cm *i.e.* > 1 fermi). For the theory with quarks the linear rise is screened by the formation of mesons. The length scale at which screening manifests itself depends on the quark mass and present calculations are just about entering this regime. Future calculations need to map out the potential over length scale of 0.01 fermi to about 1 fermi. This factor of 100 can be covered by patching together several calculations which span roughly a factor of 10. This potential can then be compared with phenomenological potentials used by experimentalists and we can match the predicted spectrum against the heavy onium states.

(3) Simulations show (a) spontaneous breaking of chiral symmetry at zero temperature and (b) the presence of the chiral symmetry restoration transition at small quark masses at $T \approx 150 MeV$. The details of the transition like the transition temperature, latent heat (if it is a first order transition) depend on the number of flavors and the dynamical quark mass. These still have large errors but the technology to calculate T_c and the equation of state in the vicinity of the transition is in place. These results are sought after by the relativistic heavy ion program, and the relevant experiments (RHIC at Brookhaven) will hopefully start producing data towards the end of this century.

(4) Reproducing the hadron spectrum will be a demonstration that QCD is the correct theory of strong interactions. There are, roughly speaking, two scales that control systematic errors in this calculation: the correlation lengths of the pion and the nucleon. Since $\xi_\pi = 7\xi_N$ and emperical studies show that the lattice dimension L should be $> 4\xi_\pi$ we need at least a 128^4 lattice to overcome systematic errors associated with finite volume effects or with the lattice granularity. Calculations on a quenched $32^3 \times 64$ lattice are being done with one gigaflop year of computing. We therefore expect definite quenched results with a tera-flop computer having a terabyte of mem-

ory. On the other hand full QCD calculations on similar size lattices will require a Peta-flop machine.

(5) Matrix elements of weak interaction operators between hadronic states: As stated in the introduction, these matrix elements provide stringent tests of the standard model and the possibility of distinguishing between the many enhancements to the SM that have been proposed to explain the cracks. For example the best determined quantity at the moment is the B parameter in kaon decays that is related to the CP violation parameter ϵ. These calculations have required significant analytical work to relate lattice results to continuum physics. This work has been done over the last seven years with numerical simulations providing major guidance and confirmation of the analytical results. The computational requirements in this area are similar to the spectrum calculations except that for certain observables we will need much larger statistics.

(6) In addition to mesons and baryons, QCD predicts the existence of glueballs. These are color singlet states that can be described as bound states of just gluons. There are two kinds, those whose quantum numbers allow mixing with mesons (these are the normal glueballs) and those which cannot be described as a bound state of a quark and an antiquark (called exotics). Glueballs have not been identified in experiments though a few favorite candidates exist. If lattice calculations can pin point the energy range to search in, and experiments confirm the prediction, then we have confirmation that QCD is the correct theory. This would be especially significant in the case of the exotics. In present calculations the signal is reasonable in the normal states but the relevance of these results is negated by the large energy shift expected due to mixing with mesons, while there is almost no signal in the exotic channels.

There are four major sources of systematic errors in the calculations: (a) finite volume, (b) lattice granularity, (c) extrapolation in the quark mass and (d) including u, d, s flavors of dynamical quarks with realistic masses. We have used the quenched calculations to understand these, while the focus in simulations with dynamical quarks has been to develop and test algorithms that will work at small quark masses. So within this existing framework of lattice QCD, progress will be incremental as we reduce the errors mentioned above. A large effort is also going into finding new quantities that can be calculated on the lattice.

In a recent review Ken Wilson points out that QCD calculations have the shortest history (12 years) of all fields where supercomputing has played a major role [5]. He implores practitioners to explore alternate algorithms and formulations in addition to demanding more powerful computers. He argues that in these other fields major breakthroughs have come with new formulations rather than brute force computer power. This is a lesson we should always keep in mind. I also agree with his criteria of success: hard numbers that can be compared with experiments or the ability to provide experimentalists with new numbers that influence future experiments. We are far short of that goal, however, I feel that we have reached an important juncture in terms of the tool kit of algorithms, our understanding of systematic errors, the development of the methodology necessary to relate lattice results to the continuum, and at least one well defined way to get answers to a number of important questions.

Let me end this section with a dream that I feel is achievable: a scalable PETAFLOP machine by the year 2000. Hopefully we will have a few new algorithms and possibly a different formulation by then.

4. Computing in the 21^{st} century

I cannot assess the impact that a new formulation may have or what shape a new idea will take. I have already given you a feel for what results enhanced computer power can provide. So let me speculate on how numerical simulations can complement analytical calculations and experiments in a field of basic research. The need for numerical simulations arises when one is dealing with a highly non-linear strongly interacting theory like QCD and when analytical methods fail and model calculations are unreliable since there is no systematic control over the approximations. However, intuition is usually based on these simple models. Thus an important role of numerical simulations is to allow one to quickly sift through ideas and sharpen one's intuition. This can lead to a reformulation of the problem which paves the way for the next generation of simulations with much improved accuracy, or suggest and guide powerful analytic techniques that eventually lead to a complete solution. The demand, therefore, for increased computer power (hardware and software) in a given field will persist until computers are powerful enough to allow scientists to test ideas on a scale of weeks or months. I believe that computing in the 21^{st} century will reach that level for many fields in applied and basic science. It will then be up to us to use it effectively to solve some of the "Grand Challenge" problems.

QCD talks invariably generate the following argument: why should large amounts of computing resources be allocated to a field which is making incremental progress and is clearly short by many order of magnitude in terms of the compute power required? Well my talk certainly did. Considering that three directors of NSF supercomputer centers and some of the program managers were present in the audience, I should have polished my arguments while preparing for the talk. Well in the months following the talk I still have not come up with a good argument to give to a house-person as to why QCD calculations will improve their family's material wealth. On the other hand if it is still important for future generations to ask the questions WHY and HOW, and to be fascinated by science, then fledgling efforts like ours need to be nurtured.

Acknowledgements

It is a pleasure to thank John Mucci for having invited me to participate in this conference. From the talks presented here it has become clear that parallel supercomputing has come of age and is the way of the future for a large class of problems. I hope that TMC's motto "to produce computers that allow scientists to play and dream en route to solving important problems" becomes the slogan of the industry.

References

[1] Field Theory on a Lattice, Seillac, France, *Nuclear Physics B (proc. Suppl.)* 4 (1988).
[2] LATTICE88, Fermilab, Illinois, *Nuclear Physics B (proc. Suppl.)* 9 (1989).
[3] LATTICE89, Capri, Italy, *Nuclear Physics B (proc. Suppl.)* 17 (1990).
[4] LATTICE89, Tallahassee, USA, 10/90, to appear in *Nuclear Physics B (proc. Suppl.)*.
[5] K. G. Wilson, in ref (3) page 82.

END

DATE FILMED

01 / 29 / 91