$Covf - 9010439 - 1$ 

UCRL-JC--104063

DE92 003016

# **OBTAINING GIGAFLOP PERFORMANCE FROM** PARTICLE SIMULATION OF PLASMAS

# David V. Anderson, Bruce C. Curtie, Dana E. Shumaker University of California **Lawrence Livermore National Laboratory** Livermore, CA

This paper was prepared for the<br>Proceedings of the Fifth International Symposium on Science & Engineering on Cray Research Computers<br>October 22-24, 1990 London, England



**June 1990** 

#### **DISCLAIMER**

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

# OBTAINING GIGAFLOP PERFORMANCE FROM PARTICLE SIMULATION OF PLASMAS

David V. Anderson, Bruce C. Curtis**,** Dana E. Shumaker

National Energy Research Supercomputer Center Lawrence Livermore National Laboratory Livermore, California 94550 USA

Eric J. Horowitz

!

• **C**o**m**pu**t**e**r S**c**ien**ces **C**o**r**po**rat**io**n,** Bal**t**im**o**re Ma**r**ylan**d** U**SA**

# A**B**S**TRACT**

**I**n **t**h**e** nume**ri**cal s**im**ulat**i**o**n o**f **p**las**m**a phen**om**ena **t**he**r**e are two fu**nd**a**m**ental approaches that are **g**enerally followed. In the continuum approach one models the **e**volution of the fluid moment equations derived from the appropriate Boltzmann equation of the plasma. Alternatively, in the particle a**p**p**ro**a**ch** a la**r**g**e** g**roup** o**f** s**im**ul**a**te**d ch**arg**ed** pa**r**t**i**cles are m**o**ved acc**ord**in**g** t**o** the se**lf**-c**o**nsistent e**l**ect**rom**a**g**netic fie**ld**s which pa**r**t**l**y **d**epen**d o**n the cha**rg**e an**d** cu**rr**ent **d**ensities **of** these sa**m**e partic**l**es. Alth**o**u**g**h the pa**r**tic**l**e si**m**ulati**o**n meth**od** has **b**een t**r**a**d**it**io**nally the **mor**e expensi**v**e **of** the tw**o**, it is **m**uch m**or**e capa**b**le **of g**i**v**in**g** a**d**equate acc**o**unt **o**f **m**any i**m**p**or**tant kinetic phen**om**ena. **W**ith the a**dv**ent **o**f **v**ect**or** mu**l**tip**ro**cess**or** supe**r**c**om**pute**r**s, such as the C**r**ay**-**2 **or** C**r**ay Y-**M**P, we ha**v**e **l**ea**r**ne**d** t**o** a**d**apt pa**r**ticle simu**l**ati**o**n c**od**es t**o** expl**o**it the pa**r**al**l**el featu**r**es **of** these **m**achines. Yet, in spite **of** such **d**e**v**e**lo**pments, the pa**r**tic**l**e simu**l**ati**o**n c**od**es ha**v**e **r**emaine**d** much sl**o**we**r** than the maximum machine spee**d**s. **W**e ha**v**e in**v**esti**g**ate**d** new techniques that **f**u**r**the**r o**pti**m**ize these **m**eth**od**s t**o br**in**g** the speeds **of** these partic**l**e si**m**ulati**o**ns int**o** the **g**i**g**afi**o**p **r**an**g**e. **R**ecent p**rogr**ess in this a**r**ea su**gg**ests that the use **of** pa**r**t**i**c**l**e simulati**o**n **m**eth**od**s wi**l**l **b**ec**o**me c**o**mpetiti**v**e w**i**th the a**l**te**r**nati**v**e flui**d** m**od**e**l**s especial**l**y when it is **r**ea**l**ize**d** that **g**i**g**afl**o**p pe**rfo**rmance ma**k**es the**m** much m**or**e aff**ord**able.

### I*N***TR**O**DU**C**T**IO*N*

Pe**r**haps the simp**l**est an**d** m**o**st **g**ene**r**a**l** meth**odo**l**og**y **f**o**r** simu**l**atin**g** the **b**eha**v**i**or of** plasmas is to **fo**llo**w** the cha**rg**e**d** pa**r**ticles in thei**r** self-c**o**nsistent e**l**ectroma**g**netic fiel**d**s.[1,2,3] **To do** this, **o**ne s**olv**es the c**o**up**l**e**d** Ne**w**t**o**n-**M**ax**w**el**l** equati**o**ns **for** the m**od**ele**d** p**rob**lem. **T**hou**g**h st**r**ai**g**htf**or**wa**rd**, this app**ro**ach is **v**e**r**y intensi**v**e c**o**mputati**o**nally b**o**th in **r**e**g**a**rd** to st**or**a**g**e an**d** time **r**equi**r**e**m**ents. **T**hus, hist**o**rica**ll**y, m**o**st mo**d**e**l**s o**f** plasmas ha**v**e emp**lo**ye**d** a flui**d r**ep**r**esentati**o**n ino**rd**er t**o** make the ca**l**cu**l**ati**o**ns aff**ord**a**bl**e. **T**hese pa**r**ticle simu**l**ati**o**n m**od**e**l**s a**r**e emp**lo**ye**d** in the stu**d**ies **of** ast**ro**physics, particle accele**r**at**or**s, s**ol**i**d** state physics, m**ol**ecu**l**a**r d**ynamics, as

well as **pl**asm**a p**h**y**sics. In t**h**is **p**a**p**e**r** we t**ry** to s**h**ow how thes**e p**a**r**ti**cl**e *co*des *c*an be made to run more emcientl*y* on some *c*on**t**emporary shared memory supercomputers*,* to make them competitive wi**t**h fluid codes.

A typical particle code is executed in four **p**hases:

t

- 1. Inter**p**olate **t**he ele**c**trom**a**gnetic fields from **t**he grid **t**o the **p**arti**c**le positions (*fiel*d *int*e*rpolation*)
- **2**. Using these fields*,* **p**ush **pa**rti**c**les by adv**a**n**c**ing positions and velo*c*i**t**ies a**cc**ording to Newton*'*s l**a**w (*push*)
- 3. Interpolate from parti**c**les to deposit **c**harge **a**nd **c**urrent densities **o**n the grid (de*position int*e*rpolation*)
- 4. **S**olve the Maxwell field equati*o*ns **o**n the grid (*solve*)

In many **p**a**r**ticle codes the majority *o*f the time is spent in the first three **p**hases. **E**ven when the solution of the field equations is most time *c*onsuming*,* we will assume that this **p**hase **c**an be s**t**r**a**ighforwardly o**pt**imized by standard te**c**hniques. In this **p**a**p**er we wish t**o** em**p**hasize the methods for **t**he o**p**timization of the first th**r**ee **p**hases*,* because we believe it holds the key to obtaining **p**erformance i**n** the gigaflo**p** range.

### **Vector Para**l**lel Algorith**m**s in QNJD**

**Prior to 1985 most particle** c**odes we**r**e** o**nly pa**r**tia**l**ly vecto**r**ize**d**, pa**r**tly becau**s**e the m**ac**hin**es **a**v**a**il**a**bl**e had li**mi**ted vecto**r **ha**r**dw**ar**e** a**nd partly bec**a**u**s**e the** alg**o**ri**th**ms **w**er**e not well dev**e**loped. A notab**l**e exception to this w**as **the wo**r**k of** B**une**man **et** al[**4**] **who w**r**ote highly optimized p**ar**ticle cod**es **in C**al as**semb**l**y l**ang**uage fo**r **the C**ra**y-1 computer. Th**is s**ituation began to ch**a**n**ge **with the int**r**oduction of v**ec**to**r m**ultip**r**oc**e**s**s**o**r **m**ac**hin**es s**uch** as **the C**ray X**-**MP**, Cr**ay**-**2**, Cr**ay Y**-**MP**,** as **well** as s**o**me **o**f Ja**p**anese manufactur**e.** In the past few years, nearly ali of the restrictions on vectori**z**ation and multitaski**ng** have been removed. The hybrid particle code QNJD, with fluid electrons and **p**article ions, **p**ioneered sever**al** of thes**e** dev**e**lopments[5] and was fully paralleli**z**ed. "Parallelized" denotes the simultaneous use of both vectori**z**ation and multitaskin**g**.

The first phase of the time ste**p**, as indicated in the list above**,** *:*requires the interpolat**i**on of the fields, known on the mesh**,** to the particle positions. Since the particles, stored in so-called particle tables, tend to become randomly distributed in space, it follows that the fields used to move the particles must be accessed randomly from the computer memory. **I**n terms of the written al**g**orithm this is express**e**d in terms of an indirect index structure, such as *EX(I(N))*. This structure, which frustr**a**ted vectorization on the earlier vector computers**,** may now be indirectly vectorized on many of the newer vector computers by operations known as the "gather" or the "scatter." The gather mechanism is **u**sed to load memory items into the CPU while the related scatter operation is used to store into the memory from the CPU. QNJD employed the **ga**ther procedure to allow vector**i**zation

and used multitasking to carry out the *field interpolation* phase. It is crucial to note that this form of indirect vector**i**zation is considerably slower, on the Cray-2, than normal di**r**ect vectorization of cons**t**ant stride ar**r**ays; at best the gather operation can run at one fou**r**th the normal vector speed. Its speed is usually worse than that because random access of memory tends to be slowed by memory bank and quadrant conflicts. But, in most cases, it is still faster than scala**r** coding.

Du**ri**ng th**e** next phas**e**, the particle push, the orbits are ali ind**e**p**e**ndent and thus trivially parallelized.

Th**e** third phas**e** is a m**o**r**e c**omplicat**e**d in**t**erpolation th\_,a used in **t**h**e** first phas**e** and requi**re**s bo*t*h a gather and a scat**t**e**r** ope**r**a**t**ion to pe**r**form th**e** deposition of charge and cur**r**ent d**e**nsity cont**r**ibutions f**r**om the vari**ou**s **p**ar**t**icl**e**s. **T**h**ere** is a **possib**l**e** re**cu**rsi**on i**n **t**hi**s ste**p **beca**use **d**iff**ere**n**t** partic**l**es m**ay try to** inc**r**e**m**en**t t**he s**am**e mesh **lo**ca**t**i**o**n s**i**mul**t**ane**o**us**ly;** c**o**nsequen**t**ly this aspec**t** p**r**e**v**ente**d v**ec**tor**iza**t**i**o**n an**d** mu**ltit**askin**g**. One **r**eme**dy** t**o t**his p**rob**lem, emp**loy**e**d** in **QN3D**[**6**], was **to br**eak **t**he **d**ep**o**siti**o**n ph**a**se in**to** se**v**e**r**a**l** su**b**-phases each **of** which **a**ccu**m**u**l**a**t**e**d t**he **d**ensi**t**ies **fro**m **a** c**orr**esp**o**n**d**in**g** su**bgro**up **of** p**ar**ticles. **By** ch**oo**sin**g** e**a**ch su**bgro**up such **t**ha**t** n**o t**w**o** pa**rt**ic**l**es **o**ccupie**d t**he s**a**me mesh ce**l**l i**t t**hen **b**eca**m**e i**m**p**o**ssib**l**e f**or** c**o**nflic**t**s **to o**ccu**r**. **QN3D** use**d bot**h **v**ec**tor**iz**at**i**o**n an**d** mul**t**i**ta**s**k**ing wi**t**hin e**a**ch su**b**-phase **to** p**ara**lleli**z**e **t**he *deposition interpolation*.

In **t**he las**t** ph**a**se of QN**3**D we sol**v**e an **a**pproxim**at**e form of M**a**xwell's equ**at**ions. Since we **a**re **t**rying **t**o limi**t t**he scope of **t**hi**s** p**a**per **t**o **t**he issues of speeding up **t**he o**t**her **t**hree ph**a**ses of **t**he calc**u**l**at**ion, we sh**a**ll no**t** discuss **t**he field sol**v**er here.

O**t**her workers h**av**e **a**lso been exploring simil**a**r ide**a**s for op**t**imi**z**ing **t**he speeds of p**a**r**t**icle codes. **F**or ex**a**mple, Heron **a**nd Ad**a**m[7] h**av**e developed **a** deposi**t**ion **a**lgori**t**hm which is org**a**ni**z**ed such **t**h**at** mos**t** of **t**he loops employ direc**t v**ectori**zat**ion, **t**hereby reducing some of **t**he o**v**erhe**a**d of g**at**her-sc**att**er forms of **v**ectori**zat**ion.

### Op**timiz**i**n**g **the** I**nter**p**olation Phases in ST**O**R**M

**QN3D, though fu**l**ly ve**c**torized** a**nd** m**ultit**as**ked w**as a**bout 40 tim**es s**lower t**ha**n th**e **pe**ak s**p**eeds o**f** a **Cr**a**y-**2. T**h**e *deposition interpolation* phase was **t**he slowes**t** of **t**he four phases. Thus, in 1989, a new **v**ersion of QN3D was developed **t**ha**t** employed a differen**t** deposi**t**ion **t**echnique in hopes of op**t**imizing **t**he *deposition inte*rp*olation* phase. I**t** also had a new field solver and added a hybrid fluid componen**t** for one species of ions in the model. The code was renamed STORM, as it was qui**t**e differen**t** from its predecessor. We have built a special **v**ersion of i**t** which we ha**v**e dubbed SQA*L* for the purpose of conduc**t**ing s**t**udies into op**t**imization in **t**he **C**ray-2 environment.

#### Or**g**an**iz**a**tio**n **of t**his **P**aper

We begin by developing alterna**t**ive **t**echniques **t**ha**t** allow a vecto**r**-pa**r**allel **t**rea**t**ment and a**t t**he same time avoid the slower gather- scat**t**er operations entirely. Some indirect indexing is still there, but it is in the ou**t**er l**o**ops.

Then we discuss a particle ordering scheme and show how it might allow us to ge**t** higher perform**a**nce in our present computing environment **a**nd how i**t** might extend into p**a**rallel computing environments.

The structure of this paper is as follows: We begin in Section 2 with a dis*c*ussion of strategies for high performance and then in Section 3 we give concrete examples of implementing some of these ideas in the SQAL code. **W**e propose other techniques to be tested in Section 4. In Section 5 we *c*on*c*lude wi**t**h a discussion about how we expect to achieve gigaflop performance in our present computing environment.

### **STR**A**T**E**GIES F**O**R** H**IGH PERF**O**R**M**ANCE**

**A**s s**t**a**t**ed **abov**e**, th**e **m**aj**o**r **ob**s**t**acle **to** g**oo**d **p**erf**orman**ce in **a part**i**c**l**e co**de is **t**he tendency for the **pa**r**t**icles **t**o be randomly lo**c**ated in space which leads to **t**he use of indirec**t** inde**x**ing **a**nd whi*c*h c**a**n le**a**d **t***o* slowly exe*c*u**t**ing code. Tha**t** is, the **C**PU*'*s idle mos**t** of the **t**ime while **t**hey wai**t** for the memory accesses **t**o complete. This is a **c**ri**t**ical issue for **t**he **C**ray-2 and less important for **t**he **C**ra*y* Y-M**P**. And it may be im**p***o*r**t**an**t** for **t**he **C**r**ay**-3. If **t**he speed of **ac**cessin**g** memory can be in**c**reased*,* there is **t**he **p**ros**p**ect of keeping **t**he **C**PU*'*s busy. Un**t**il we do **t**hat*,* i**t** m**a**kes little sense to o**pt**imize **t**he **a**rithme**t**i**c**al as**p**ects.

Our first objec**t**ive in making these codes more e**f**ficient is to elimina**t**e **t**he use of ga**t**her-sca**t**ter vectoriz**at**ion by replacing i**t** wi**t**h direct vectoriza**t**ion (in which **t**he s**t**ride is uniform.) **W**e **t**hen develop efficien**t** meth*o*ds for a*c*cessing memory. La**t**er*,* we shah consider a scheme **t**hat orders **t**he par**t**icles by posi**t**ion.

#### **SQAL EV**O**LUTI**O**N**

**A test code, d**u**b**be**d** SQ**AL, w**as **bu**i**lt from the** S**T**O**RM cod**e **to** a**llow us to test th**e **v**a**r**i**o**u**s opt**i**m**iza**t**i**o**n **ide**a**s. The first ver**si**on of S**Q**AL stored the field** a**rr**a**ys** a**nd d**ep**os**i**tion** a**rr**a**ys such th**a**t e**a**ch** fi**eld component w**a**s stored** i**n** a **s**e**p**a**r**a**te 3D stor**ag**e block. They were** s**tored** w**ith wh**a**t** i**s known as n**a**t**u**r**a**l ord**e**r**ing **such that** i**ncrementin**g **the** i**ndex** i**n the** x **d**i**r**e**ct**i**on corresponds to** u**n**i**t str**i**d**e a**ccess of the memory.**

### **Field In**t**erpolations with Interleavin**g

**For** linear in**t**e**r**p**olat**ions **to** a pa**rt**ic**l**e p**os**i**t**i**on,** in **t**he *field interpolation* phase, we must access **t**he fields a**t t**he 8 surrounding grid vertices. These can be loaded from memor*y* as **4** short vectors each of leng**t**h 2. Very li**tt**le improvement in speed can be obtained from the use of such short vec**t**ors.

**W**e *c*an improve **t**he situation *c*onsiderabl*y* b*y* processing **t**he par**t**icles **64** a**t** a time and b*y* using an interleaved storage sche**m**e. What we do is store the fields such that their ordering in memory is

EX (N),EY(N)**,**EZ (N),BX(**N**)**,**BY(**N**)**,**B**Z**(**N**),**E**X(**N** + 1)**,**..

such that twelve grid quantities are stored contiguously as we move from one grid vertex to the next one in the *x* coordinate. Here, N is taken to represent the grid offset index, which essentially labels the grid points using "normal" ordering (the x index turning fastest.) When we load this vector, of length 12, from memory we can potentially achieve good performance in direct vectori*z*ation mode which should outperform the gather-scatter modes used earlier. These field variables are then stored in a temporary field array FAVL(LCT,IVX,NL) where NL ranges over 64 particles per pass, IVX ranges over the 8 vertices of the enclosed grid cell, and LCT ranges over the 6 different field components.

Even with the various optimizations done in Fortran to get good performance, we found that the pe**rf**ormance of the *field in*t*erpola*t*ion* was only about 23% faster than the original STORM This relates to the fact that our vectors are relatively short making the loop overhead rather expensive. Table 1. shows the various measured access **r**ates compared to the theoretical **m**axim**u**m rat**e**s**.** Th**e co**l**u**mn labell**e**d SQAL**0**6 **gi**v**es** t**he s**pee**ds** a**chie**ved using Fortran coding for the loading of the field quantities; later on we shall discuss the version SQAL08 that uses assembly coding and local memory to perform the loading. In Fortran our actual access rate of roughly 45 megawords per second is about 21 percent of the theoretical maximum rate of almost one access every clock period. The speeds of these versions of SQAL are shown in Table 2. If we assume that the access speed labelled IDEAL could be achieved, then the *Field Inte*r*polation* speed of 59 MFLOPS would increase to about 106 MFLOPS. Even then, we do not have enough speed on a single processor to infer gigaflop speeds on an 8 processor multitasking machine. We must work harder to further optimize the code.



Table 1: Measured acce**s**s rates for STORM and two versions of SQAL are compa**re**d to **t**he th**e**o**reti**cal (**i**deal**) r**ates fo**r** a **C**ra**y**-2. **In** th**e** *field inte*r*polation* Loop 12 we read the fields on the **g**rid into temporary arrays which are subsequently used in the interpolation of field quantities to the particle positions. SQAL06 u**s**es four-way unrolled Fortran for acce**s**sin**g** memory. SQAL08 im*p*rove**s** on this by employin**g** local memory for the **g**rid indice**s** and Cal-2 assembler to enhance **the** load**ing** o**per**at**ions.** Fo**r** t**he** *Deposition In*\_*erpolatiott*, Loop 40 does bot**h** load**s** and stores as it accumulate**s** the char**g**e a**n**d current den**s**ities.

#### D**e**position Int**erp**o**l**ations with Interleaving

The STORM code used a vectorized method for depositing charge and cur-

rent densities to the grid from the particle quant**i**ties. Si**n**ce each part**i**cle contributes to eight grid vertices and since there are four fields  $(\rho, J_x, J_y, J_z)$ , there are 32 quantities to be generated. This was accomplished by cons**t**ructing 32 contiguous arrays, each spanning **t**he en**t**ire grid. A gather-scatter loop was used to **f**ill these a**rr**ays with the grid contributions. Since only one particle was processed a**t** a time, **t**here was no possibility for a con**f**lict. But the**r**e was also no easy way to generalize this method for pa**r**allel computation. A**f**ter all the 32 grid a**rr**ays we**r**e **f**illed, **t**he actual g**r**id ar**r**ays o**f**  $(\rho, \mathbf{J}_x, \mathbf{J}_y, \mathbf{J}_z)$  were constructed by summing these partial arrays.

**Somewhat** i**n analo**g**y to the t**r**ea**m**ent of the** *field interpolation***, we** m**odiif**e**d** t**he** c**ode to p**r**o**c**ess the** *p***artic**l**es in** gr**oups of** 6**4. That made it possible to gen**e**rate al**i **of t**h**e interpolation wei**g**hts** b**y d**i**re**c**t vector mod**e**. • The 32 temporary ar**r**ays wer**e **inte**r**leaved into a lar**g**er array in which th**e **t**h**e de**p**osition cou**l**d** *p***roceed into** e**i**g**ht conti**g**uous elements at a time. These 8** are the four quantities  $(\rho, \mathbf{J}_x, \mathbf{J}_y, \mathbf{J}_z)$  at some grid point followed by the **same four quantities at the adjacent grid poin**\_ **in the** x **di**r**e**c**tion. These** "**interleaved**" **temporary arrays we**r**e then fil**l**ed in d**i**re**c**t vector mode. As** b**efor**e**, the four** *p*r**imary arrays were constru**c**ted** b**y summin**g **ove**r **these temporary ar**r**ays.**



Table 2: Spe**e**d mea**s**urement**s** ar**e g**iven for th**e** three **ess**ential re**g**ion**s** of the **p**article code**s** STORM, SQAL06, and SQAL08. The fourth re**g**ion, the s*olve*, i**s** not treated here. W**e g**ive the performanc**e** in term**s** of the popular yard**s**tick of micro-**s**econd**s** per particle per time**s**tep a**s** wel**l** a**s** in term**s** of **th**e mea**s**u**r**ed **s**peed in me**g**a-flop**s**. The boundary treatment i**s** mo**s**tly loads, **s**tores and lo**g**ic and with relatively little floating point arithmetic it i**s** difficu**l**t to achieve a hi**g**h megaflop rate.

In Table 2., in the row labelled *Deposition In*t*erpolation*, we see that timings for this phase improved significantly, up 86% from the results of STORM. The improved speed of 26 MFLOPS is still far too slow to give the code performance anywhere near a gigaflop, even if we were able to multitask it at level 8. We are a factor of at least 4 too slow. Again, as in the revised *field in*t*e*r*pola*t*ion* phase, we seem to be limited by the access rate to the common memory. Referring to Table 1. **w**e see that much of the improved performanc**e** of the *Deposition Interpolation* may be attributed to the significantly improved memory access. Even so, it falls far short of the theoretical a**s**ymptotic ra**t**e, at abou**t** the 13 percen**t** level• If we were to a**ss**ume that the memo**ry** acce**ss** ra**t**e was that given under "**I**DEAL" in Table *2*, then the **s**peed of thi**s** pha**s**e would inc**r**ea**s**e to 10*7* MFLO**P**S.

**I**t i**s** clea**r** tha**t** fu**r**the**r** modifications mu**st** be **s**ought to inc**r**ease the**s**e **s**peeds**.** On the *C***r**a**y**-*2* we can t**ry** to u**s**e local memo**ry**, we can op**t**imize with as**s**embl**y** code, o**r** we can seek longe**r** vec**t**o**rs**. One can al**s**o con**s**ide**r** changing the enti**r**e code and da**t**a s**tr**uc**t**u**r**e to minimize acce**ss**es to common memo**ry.** La**s**tl**y**, **t**he**r**e is the option to u**s**e a faste**r** compu**t**er, **s**uch as a C**r**a**y** Y-MP**.**

#### U**s**ing As**s**embler Code and Local Memo**r**y

**I**n m**e**asuring **the** tim**e** sp**e**nt in **the** diffe**r**ent s**e**c**t**ions of t**he** *field interpolat*i*on* rou**t**ine (in SQAL0**5**) we found **t**h**a**t roughly **4**0% of **t**he **t**ime is spen**t** in the • **p**erformance *o*f the "ari**t**hme**t**ical" loo**p**s and **6**0% of the **t**ime is used in reading **t**he fields from common memory. In terms *o*f microseconds per par**t**icle per time step we were using 2.2**6** to lo**a**d **t**he fields **a**s compared to .98 for **t**he in**t**erpolation and .**66** for **t**he push. It became app**a**ren**t***,* tha**t** to make further progress we mus**t** optimize **t**he loading of the fields.

Our first a**pp**roach was to use For**t**ran o**pt**imization **t**echniques. **W**e unroll**e**d **t**he loo**p** accessing **t**he memory such **t**ha**t** i**t p**rocessed four p**a**r**t**icles **p**er **pa**ss. This versi*o*n, SQAL0**6**, gave a modes**t** s**p**eed-up of **t**his loop of **a**pproxima**t**ely **9**%. To be**tt**er unders**t**and **t**hese slow speeds, we **t**urned **t**o **t**he assembly code. **W**e wan**t**ed to analyse the assembly code generated by the compiler to learn if i**t** wa*s* op**t**imal and if no**t** recode **t**he loo**p** ei**t**her wi**t**h better For**t**ran or in **C**al-2 assembly code.

The code generated by **t**he compiler for **t**he inner loop was quite good*,* bu**t** we discovered a memory bot**t**leneck in **t**he ou**t**er loop. A subs**t**antial **a**moun**t** of time was spen**t** lo**a**ding **t**he loca**t**ions of the grid cells in scalar mode. To remove the bot**t**leneck we loaded **t**he grid cell locations into local memory in vec**t**or mode in a separa**t**e loop. This ne**tt**ed an addi**t**ional 41% speed-up. Although this was done in For**t**ran, i**t** was based on our re**a**ding of the assembly code generated by **t**he compiler.

In the inner loop, the limi**t**ing fac**t**or was **t**he shor**t** vec**t**or leng**t**h. Since the **C**ray-2 supports a memory bandwidth of [vec**t**or length] words in [vec**t**or leng**t**h+8] clock period**s**, and **t**he field interpola**t**ion loop used a **v**ec**t**or leng**t**h of 12, we were achieving only sligh**t**l*y* be**tt**er **t**han one word every two clocks. **W**e improved on **t**ha**t** ra**t**e by loading four of the length 12 vectors into consecutive local memory locations, so the subsequen**t** store in**t**o common memor*y* could be done with a vec**t**or length four times greater. This modifica**t**ion, coded in assembl**y** language, produced a maximum rate of nearl*y* **t**wo words moved ever**y t**hree clocks and resulted in ano**t**her 20% speed-up beyond our bes**t** Fortran version. Thus over**a**ll **w**e have g**a**ined a speedup of **69**% in the memor*y* **a**ccess rate. This can be seen in Table 1 by comparing the versions SQAL0**fi** with SQAL08 in the row labeled *F*i*eld Int*er*polation Loop I*\_. And compared to STORM, we have more than d**o**ubled the speed of this code segment.

We were working on a simi**l**ar modification for the memory access of the *deposition in*t*erpola*t*ion*, as this paper was going to press. Thus the nota**t**ion "NA" in Table 1.

# **T**H**E**O**R**ET**I**CA**L** EST**I**MA**T**ES

**I**n the fo**re**going w**e pre**s**e**nted **r**esu**l**ts from our test **c**ode SQAL that showed that the use of interleaving, local memory*,* direct vectorization**,** and even assembler coded memory accesses substantially improved the speed of the interpolation phases. Yet this was insufficient to attain the speed on a single Cray-2 processor above 100 megaflops. In what follows we make estimates of speedups that might be obtained with further code modifications or by using a Cray Y-MP.

#### **C**r**itical P**a**th** A**na**ly**sis f**or **Cra**y**-2 and C**ray **Y-**MP

B**y knowing** s**omething** a**bout the perfo**rma**nce of v**ar**ious in**s**tructio**ns **on** the computer, we can make a best case estimate of how fast different code s**eg**men**t**s mi**g**ht run**.** W**e h**av**e m**a**de** so**me** s**imp**lifyin**g** ass**ump**t**io**n**s** s**uc**h as**:**

- 1. P**e**rf**e**ct ov**e**rlap of adds and multipli**e**s
- 2. On**e** Dir**e**ct V**e**ctor M**e**mory Acc**e**ss p**e**r Clock P**e**riod
- 3. One Gather Acc**e**ss Ev**e**ry 4 Clock P**e**riods (Cray-2)
- 4. One Path to M**e**mory (Cray-2)
- 5. On**e** Gath**e**r Acc**e**ss Ev**e**ry Clock P**e**riod (Cray Y-MP)
- 6. Thr**ee** Paths to M**e**mory (Cray Y-MP)
- *7*. On**e** Scatt**e**r Acc**e**ss p**e**r Clock P**e**riod
- 8. I**g**nore Loop Ov**e**rh**e**ad
- 9. I**g**nor**e** Startup and M**e**mory Bank Conflicts

W**e** hav**e** don**e** su**ch** an analysis for th**e** *field inte*rp*olation***,** the *push***,** and • the *deposition in*t*e*rp*ola*t*ion*. Also, since the memory performance seems to be a critical issue for this particle code, we have also made estimates for the Cray Y-MP which has a slower clock cycle but a much faster memory in comparison to a Cray-2. Table 3. shows the limiting performance that could be obtained from STORM or SQAL under these assumptions.

#### Ordered Parti**c**le S**c**heme

It should be evident from the foregoing that the methods suggested for optimi**z**ing SQAL may not be adequate to bring it to **a** performance level near a gigaflop. Even if we assume that we could multitask the code at the highest level, neither the Cray-2 (with 4 CPU's) nor the Cray Y-MP (with 8 CPU's) will reach a gigaflop. However, the Y-MP will operate **a**t a larger fraction of a giga**fl**op than **t**he Cray-2. For this reason, we will now restrict our discussion to the Cray Y-MP.

Theoretical Asymptotic Best Case Speeds (MFLOPS)				
Code	<b>STORM</b>	<b>SQAL</b>	<b>STORM</b>	SQAL
Segment	$(Cray -2)$	$(Cray-2)$	(Cray Y-MP)	(Cray Y-MP)
<b>Field Interpolation</b>	180	215	286	271
and Push	(1.16)	(1.06)	(74)	(.53)
Deposition	133	191	259	207
Interpolation	(1.03)	(.83)	(.53)	(.76)
Combined	157	204	277	241
<b>Phases</b>	(2.19)	(1.89)	(1.27)	(1.60)

**T**a**bl**e **3: S**p**eed** e**stim**a**tes** ar**e giv**e**n for the Cr**a**y-2** a**nd for th**e **Cr**a**y** Y**-MP und**e**r e**x**tr**e**mely optimistic** as**su**m**ptions. The numbers in p**a**r**e**nth**e**sis give th**e execution times in micro-seconds per particle. It is evident that the Cray Y-MP **• outperforms th**e **Cr**a**y-2 inspite of its slow**e**r clock cycle. When combin**e**d with th**e **multit**as**kin**g **potentia**l **of th**e**s**e m**achin**e**s, t**he Y**-MP would** a**p**p**e**a**r to be th**e **m**a**chine to use to obt**a**in perform**a**nc**e **in the gig**a**tio**p **r**a**ng**e**.**

Also th**ere is** a s**e**riou**s pr**o**b**l**e**m wh**e**n **we c**ons**i**d**e**r **ho**w a c**o**d**e such** as SQAL **c**oul**d** be multi**t**ask**e**d**.** T**h**e *field interpolation*, **th**e *push***,** and th**e** *solve* are trivially multita**s**ked becaus**e** there i**s** no dat\_ conflict. Unfortu**n**ately, there is a conflict in the *deposition inte*rp*olation* that mu**s**t be re**s**olved inorder to multitask that phas**e** of the calculation. The problem is that two different processors may try to update the same grid points at the sam**e** time.

One approach to this problem recogni**z**es that th**e** errors generated by forcing the *deposition interpolation* **t**o multitask are relatively small and therefore acceptable. **S**uch an "asynchronous" ver**s**ion would generate lrreproducible results which is less than satisfactory even if the errors are **s**mall.

Th**e** other approach us**e**s a particle sorting schem**e** to eliminate an**y** possible conflicts of multitasking. The idea her**e** is to order the particles **s**uch they are grouped together according to the grid cell they occupy. When grouped this way, great economie**s** result because the number of memory accesses required goes down dramatically. It also becomes advantageous to **s**tore portions of fields in local memory when they are reused many time**s**. In this scheme they would be accessed as many times as there are particles in the cell being processed. We shall describe an algorithm that rebuilds the particle tables after each push in such a way that less than 4*n* operations are required to ensure that the *n* particles are properly sorted into their grid cells.

SQAL and STORM both carry the particle variables *ip*, *jp*, and *kp* in the particle tables; these indices specify the coordinates of the grid cell containing the particle. It is convenient to also carry the offset index

$$
ic(n) = ip(n) + im * (jp(n) - 1) + im * jm * (kp(n) - 1)
$$

where *im* and *jm* are the grid dimesnions in the x and *y* coordinates. The

quantities *lp*, *jp*, *kp* an**d** *ic* are compute**d** as the last pa**r**t of the *push* phase. By keepin**g** the **o**l**d v**a**l**ue **of** *ic* as *ic*o*lcl* we can test (*i*c-*icol***d**) t**o** i**mm**ea**d**iately **d**eter**m**ine which pa**r**ticles ha**v**e **b**een pushe**d o**ut **of** their f**o**rmer cel**l**s an**d** which **o**nes ha**v**e **b**een **r**etaine**d** in them. Just a**f**te**r** the push we **r**ec**o**nst**r**uct the pa**r**tic**l**e ta**bl**es t**o** keep them **ord**e**r**e**d** with respect t**o** the **gr**i**d** ce**ll**s. **T**h**i**s can **b**e **do**ne as **fo**ll**o**ws:

- 1. C**o**nst**r**uct a t**r**ansit ta**b**le **of** pa**r**ticles **d**epa**r**tin**g** cells
- 2. S**or**t the t**r**ansit ta**b**le **ord**e**r**in**g** pa**r**tic**l**es **b**y cells
- 3. C**o**nst**r**uct t**r**ansit p**o**inte**r**s t**o o**l**d** pa**r**ticle table
- 4. C**o**nstruct **r**etaine**d** p**o**inte**r**s t**o old** pa**r**tic**l**e ta**bl**e
- 5. All**o**cate a new pa**r**ticle ta**b**le
- 6. P**ro**cess the **gr**i**d** cells in ascen**d**in**g ord**e**r** an**d for** each cell:
	- **Mov**e the pa**r**ticle **d**ata **of r**etaine**d** particles to new table
	- **M**o**v**e the pa**r**tic**l**e **d**ata **of d**epa**r**tin**g** particles t**o** new ta**bl**e

As we envisa**g**e this **m**eth**od**, we wi**ll** n**o**t **mov**e the pa**r**tic**l**e **d**ata until we ha**v**e **d**ete**r**mine**d** the **m**appin**g of** the **old** "se**r**ial" num**b**e**r**s int**o** the new **o**nes. F**or** *nm* pa**r**tic**l**es, c**o**nst**r**uctin**g** the t**r**ansit ta**bl**e will **r**equi**r**e *nrn* inte**g**e**r** a**dd**s as we**ll** as a few times *nm* **log**ical **o**pe**r**ati**o**ns. O**f** the *rtrn* particles **o**n**l**y a **fr**acti**o**n (pe**r**haps 10%) will **b**e **d**epartin**g** pa**r**ticles. Say the**r**e a**r**e *I* **of** these. **T**he s**or**t wi**l**l **r**equi**r**e **o**n the **ord**e**r of** *l* In *I* **o**pe**r**ati**o**ns **of** the inte**g**e**r** a**r**ithmetic an**d** l**og**ica**l v**arieties. **T**he**r**e wil**l b**e many **f**ewe**r** t**r**ansit an**d r**etaine**d** pointe**r**s than pa**r**ticles, pe**r**haps **o**n the **ord**e**r of** 20% **of** the num**b**e**r of** pa**r**tic**l**es. Once the pointers are set, moving the particle data requires  $2 * nq * nm$  memory accesses. He**r**e, *nq* is the num**b**er **of** att**r**i**b**utes pe**r** pa**r**tic**l**e**-** en**v**isa**g**e**d** as 12 at **l**east.

• A **gr**eat **d**eal is **g**aine**d b**y usin**g** these s**or**te**d** pa**r**ticles. Instea**d** o**f** accessin**g** each fie**ld gr**i**d** quantity 8 times f**or** e**v**ery pa**r**tic**l**e, this wo**r**k can **b**e **• r**e**d**uce**d** t**o** 4 times **for** e**v**e**r**y cell. If the**r**e are 10 pa**r**ticles pe**r** cell, this implies a 20 **fo**l**d r**e**d**ucti**o**n in mem**or**y accesses in the *field inter*p*olation* phase. A simi**l**a**r r**e**d**ucti**o**n occu**r**s in the *deposition interpolation* **b**ecause the accumu**l**atin**g gr**i**d** quantities a**r**e **k**ept in **v**ect**or r**e**g**iste**r**s unti**l** ali pa**r**ticles in a ce**l**l ha**v**e c**o**nt**r**ibute**d**. N**o**t **o**nly **do**es the time spent **do**in**g** mem**or**y accesses **d**ec**r**ease, **b**ut s**o**me o**f** the a**r**ithmetic can be **f**u**r**the**r o**pt**i**mize**d b**y keepin**g** f**r**equently **r**euse**d** fiel**d** quantities in the l**o**cal memo**r**y **or** re**g**iste**r**s.

It is the c**o**st o**f** the s**or**tin**g** an**d** the mappin**g from** the **old** pa**r**tic**l**e ta**bl**e t**o** the new one that pa**r**t**l**y **o**ffsets these **g**ains. **T**hese c**o**sts a**r**e minimize**d b**y so**r**tin**g** on**l**y the **d**epartin**g** pa**r**tic**l**es an**d** by the **f**act that the pa**r**tic**l**e ta**bl**e **d**ata is n**o**t m**ov**e**d** unti**l** p**o**inte**r**s to the new pa**r**ticle ta**bl**e are **d**ete**r**mine**d**.

### A**utot**a**s**k**ing Imple**m**e**n**tatio**n **I**s**s**u**e**s

W**e** m**e**nt**io**n**e**d **abo**v**e** th**a**t t**he** SQA**L c**od**e cou**ld n**o**t b**e** mu**l**t**i**t**a**sk**e**d u**n**l**e**s**s** we w**e**re willing to use th**e** so-c**a**ll**e**d asynchronous mod**e**. A cod**e e**mployin**g** th**e** just describ**e**d sort procedur**e**, how**e**v**e**r, can b**e e**asily multit**a**sk**e**d by usin**g** autot**a**skin**g**. As in the **ea**rlier discus**s**ion**,** th**e** *deposition inte*r*pola*ti*o*n is th**e** only phase that is not trivial to multit**a**sk. H**e**re we partition the physical dom**a**in into as m**a**ny **s**ubdomains as we have processors. E**a**ch sub-domain will have its own **s**ub-tabl**e** of particles. Th**e** only difficulty is that som**e** p**a**rticles cross th**e s**ubdomain bound**a**ries in the push. When this occurs**,** th**e**se particles and th**e**ir **a**ttribut**es** must be mov**e**d into other • sub-tabl**e**s in such a way that the ord**e**rin**g** is preserv**e**d. Thi**s** can b**e** don**e** by isolatin**g** from th**e** tabl**e** of d**e**partin**g** p**a**rticl**es** thos**e** that hav**e** crossed , into other subdom**a**ins. Th**e** point**e**rs of **t**h**e**se i**s**ol**a**t**e**d particl**e**s are th**e**n pass**e**d to th**e** tasks rel**e**v**a**nt to th**e**ir *y*ew sub-dom**a**in**s**. Th**e** numb**e**r of such particl**e**s is quit**e** small, down by **a** "su**rf**ace to volum**e**" r**a**tio as compared to th**e** already mode**s**t numb**e**r of dep**a**rtin**g** particles. Implicit in what w**e** have **sa**id h**e**re i**s** th**e** fact that th**e** sorting procedur**e** it**se**lf is multit**a**sk**a**ble since both th**e** particl**e** t**a**bl**e**s a**n**d th**e** grid subdom**a**ins ar**e** partition**e**d among th**e se**ver**a**l processors.

### **RESULTS AN**D **CONCLUSIONS**

We ha**v**e o**bt**a**i**ne**d** a s**i**gn**if**can**t i**mp**r**ove**m**e**nt i**n t.he pe**rform**a**n**ce **of t**he S**T**OR**M** c**od**e by ma**k**in**g** se**v**e**r**al m**od**ifica**t**i**o**ns as e**v**i**d**ence**d** in the sequence **of** S**Q**AL c**od**es. **T**he c**od**e was **r**est**r**uctu**r**e**d b**y **i**nte**r**lea**v**in**g bo**th the fie**ld** a**rr**ays **o**n the **gr**i**d** an**d** the pa**r**tic**l**e tab**l**es**;** the inte**r**lea**v**in**g** al**lo**we**d** us t**o** use **d**i**r**ect **v**e**r**t**or**ization in the piace **of g**a**t**he**r-**scatte**r** c**o**nst**r**ucts use**d** in p**or**ti**o**ns **of** STOR**M**. Ca**r**e**f**ul attenti**o**n t**o** the assem**b**ly c**od**e a**llo**we**d** us to impr**ov**e the F**or**t**r**an **v**e**r**si**o**ns **of** S**Q**A**L** an**d** late**r** t**o** use s**o**me assem**bl**y c**od**e in the sl**o**wes**t** c**od**e **r**e**g**i**o**ns. A**l**th**o**u**g**h we succee**d**e**d** in **do**u**b**lin**g** the spee**d of** the • th**r**ee c**od**e re**g**i**o**ns we a**ddr**esse**d**, we a**r**e still sh**or**t **o**f **d**em**o**nst**r**a**t**in**g g**i**g**afl**o**p spee**d**s in **t**he C**r**ay**-**2 en**v**i**ro**nment.

**T**he pictu**r**e is **mor**e **o**ptimistic if we c**o**nsi**d**e**r** the C**r**ay Y**-**M**P**. F**ro**m **o**ur c**r**i**t**i\_a**l** path ana**l**ysis a t**o**p spee**d of** 241 **M**flops pe**r** p**r**ocess**or** is the the**or**etica**l** machine spee**d l**imit. I**f** we c**o**ul**d** mu**l**titas**k** this at le**v**e**l** 8 we c**o**u**ld** ha**v**e a t**o**p spee**d o**f 1.9**3 g**i**g**afl**o**ps. **T**his is a **big** "i**f**" **b**ecause we ha**v**e sho**w**n that the *deposition in*t*er*p*ola*t*ion* cannot be multitasked unless **o**ne is willin**g** to **a**ccept the slightly wron**g** answ**e**rs th**a**t come from an asynchronous method.

W**e** b**e**li**e**ve th**a**t **a** furth**e**r restructuring of th**e** code to use ord**e**red particles will lead not only to bett**e**r single processor p**e**rformance**,** but to a fully multitaskable code as well. This should give us **a**bout one gig**a**flo\_*,* p**e**rformance on the Cray Y-MP. Even bett**e**r news is th**a**t this v**e**rsion with ordered particles is naturally extendable to more massively p**a**rallel MIMD machines, wheth**e**r built by Cray Research o**r e**thers- particularly if the processors are vector processing units.

#### **Acknowledgements**

 $\sim 10^{-10}$ 

We thank Bruce Cohen, Alex Friedman, Dennis Hewett, and Alan Mankofsky for valuable advice on the many questions that arose in these studies. This work was supported by the U.S. Department of Energy for the Lawrence Livermore National Laboratory under contract W-7405-ENG-48. Cray-1 is a registered trademark and Cray-2, Cray X-MP, and Cray Y-MP are trademarks of Cray Research, Inc.

# **References**

- [1] A. Bruce Langdon and D. C. Barnes, "Direct Implicit Plasma Simulation," in "Multiple Time Scales," Eds. Brackbill and Cohen, Acadmeic Press (1985), 337
- [2] R. J. Mason, J. Comput. Phys. 41 (1981), 233
- [3] A. Friedman, A. B. Langdon and B. I. Cohen, Comments Plasma Phys. Controlled Fusion 6 (1981), 225
- [4] O. Buneman, C.W. Barnes, J.C. Green, D.E. Nielsen, J. Comput. Phys. 38 (1980), 1-44
- [5] Eric J. Horowitz, Dan E. Shumaker and David V. Anderson, J. Comput. Phys. 84 (1989), 279-310
- [6] E. J. Horowitz, J. Comput. Phys. 68 (1987), 56
- [7] A. Heron and J. C. Adam, J. Comput. Phys. 85 (1989), 284,301







DATE FILMED 12-127-191