From Sir Isaac to the Sloan Survey
Calculating the Structure and Chaos Owing to Gravity in the Universe*

George Lake‡† Thomas Quinn‡ Derek C. Richardson‡

Abstract

The orbit of any one planet depends on the combined motion of all the planets, not to mention the actions of all these on each other. To consider simultaneously all these causes of motion and to define these motions by exact laws allowing of convenient calculation exceeds, unless I am mistaken, the forces of the entire human intellect.
—Isaac Newton 1687

Epochal surveys are throwing down the gauntlet for cosmological simulation. We describe three keys to meeting the challenge of N-body simulation: adaptive potential solvers, adaptive integrators and volume renormalization. With these techniques and a dedicated Teraflop facility, simulation can stay even with observation of the Universe.

We also describe some problems in the formation and stability of planetary systems. Here, the challenge is to perform accurate integrations that retain Hamiltonian properties for 10^{13} timesteps.

1 The Scientific Importance of Cosmological N-body Simulation.

Simulations are required to calculate the nonlinear final states of theories of structure formation as well as to design and analyze observational programs. Galaxies have six coordinates of velocity and position, but observations determine just two coordinates of position and the line-of-sight velocity that bundles the expansion of the Universe (the distance via Hubble's Law) together with random velocities created by the mass concentrations (see Figure 1). To determine the underlying structure and masses, we must use simulations. If we want to determine the structure of a cluster of galaxies, how large must the survey volume be? Without using simulations to define observing programs, the scarce resource of observing time on \$2 billion space observatories may be mispent. Finally, to test theories for the formation of structure, we must simulate the nonlinear evolution to the present epoch.

This relationship to observational surveys defines our goal for the next decade. The Sloan Digital Sky Survey (SDSS)[18] will produce fluxes and sky positions for 5×10^7 galaxies with redshifts for the brightest 10^6. Our ambitious observational colleagues have cut

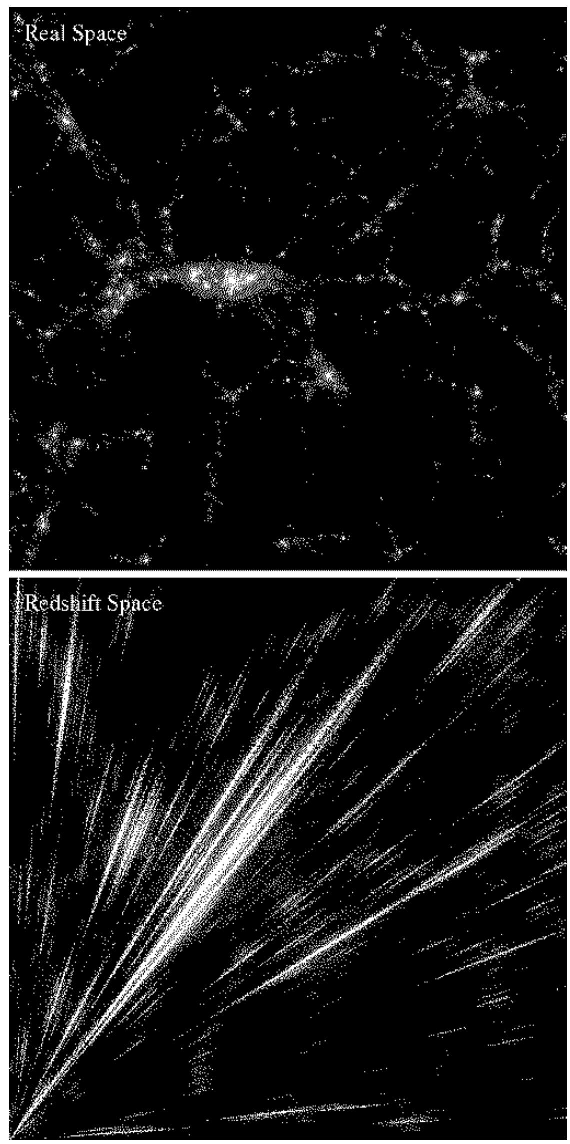

Figure 1: The top panel shows a slice of a cosmological simulation (100 Mpc on a side and 10 Mpc thick). The bottom panel shows the same slice viewed by an observer at the lower left corner in "redshift" space. To refine and test analysis techniques, we need simulated "skies" where the underlying truth is known.

*Support by NASA HPCC/ESS, IRP and ATP
†NASA HPCC/ESS Project Scientist
‡Department of Astronomy, University of Washington

steel and ground glass to survey a "fair volume" that *we must simulate*, but we need $N = 10^{12}$ to do this. Direct summation of the gravitational forces using fixed timesteps would take 10^{10} Teraflop-years.

We will explain why this is a unique time to survey the Universe as well as describing the technical breakthroughs required to create a better survey of the cosmos. We will then present the three keys to a realistic float count: 1) spatially adaptive potential solvers, 2) temporally adaptive integrators and 3) volume renormalizations. Another goal of this paper is to define "high quality simulations" and the niche science that can be done with $N \sim 10^8$.

2 From SDSS to The Ultimate Digital Sky.

You guys aren't trying hard enough.
—-Steve Shectman 1993

2.1 SDSS. The quote above was comparing our simulation plans to upcoming large surveys, SDSS and 2MASS—an all-sky infrared survey at 2 microns. Both of these projects will produce about 10 Terabytes (TB) of data, not a coincidence at all.

The efficiency of a photometric sky survey is:

$$\epsilon_{photo} = q\Omega D^2$$

where q is the detector quantum efficiency, Ω is the solid angle of the field covered by the detectors and D^2 is the collecting area of a telescope of diameter D. The sky surveys done by 48-inch Schmidt survey machines had $\epsilon = 0.30\mathrm{m}^2\mathrm{deg}^2$, assuming an optimistic photographic quantum efficiency of 0.5%. SDSS uses a 2.5m telescope with an array of 30 (2048x2048) CCDs with a pixel scale of 0.4 arcseconds and a quantum efficiency of roughly 50%. The value of ϵ is roughly 4.8, more than an order of magnitude higher than the Schmidt Telescopes[18].

The median seeing of a good site is ~ 0.8 arcseconds; optimal sampling sets the pixel scale at 0.4 arcseconds. SDSS will observe $\sim \pi$ steradians of the sky in 5 colors for a total pixel count of 8.2×10^{11} times 5 colors of 2 bytes fixing the archive at 8.2 TB. As the surveyed angle will lie between π and 4π steradians with the pixel size set by the atmosphere, the data volume of a large survey is fixed.

The increase in the spectroscopic efficiency is far more stunning. SDSS expects to identify 10 million galaxies, 50 million stars and a few hundred thousand quasars. Spectra will be taken of a million galaxies and roughly 100,000 quasars and a lesser number of stars. The spectroscopic efficiency is given by:

$$\epsilon_{spectra} = qN_{spectra}D^2(S/N)^{-2}$$

where q and D were defined before, $N_{spectra}$ is the number of spectra that can be taken simultaneously and (S/N) is the signal to noise required to get a redshift. The redshift surveys done 30 years ago used telescopes that were 3–5m, measured spectra one at a time and required a $(S/N) \sim 10$ as they were measured by holding photographic plates up to lamps, hence $\epsilon_{spectra} \sim 8\times10^{-5}m^2$. The SDSS spectrograph will have 640 fibers fed to high q CCDs. Digital spectra require a $(S/N) \sim 1$ for redshifts, so $\epsilon_{spectra} \sim 5 \times 10^3 m^2$. The throughput advance is nearly 8 orders of magnitude.

SDSS will produce some "lighter" data products. A 50 GB catalog will give derived quantities for all "found objects" while there will be an archive of postage stamp images that will be ~ 0.5 TB. The spectroscopic data inherently adds little volume to the total database. In most surveys, 75% of the sources are within one magnitude of the limit. The limiting magnitude is defined by requiring that only 5% of the sources are fraudulent (in a 5 color survey, this means that over one quarter of the catalog is nonexistent). Thus, for most of the sources, the colors are the highest resolution spectra that may be extracted.

2.2 The Nearly Ultimate Digital Sky (NUDS). How large can our Digital Sky become? SDSS will probably cost ~\$50M. We're about to describe a project that couldn't be done for \$50B, but isn't completely outside the realm of current technology. NUDS uses 10m space telescopes, so the pixel size of $\sim 10^{-2}$ arcseconds is determined by the diffraction limit. If we survey 4π steradians using 20 colors, the archive is roughly 0.1 Exabyte with 2×10^{15} pixels (10^4 SDSS archives). The found object catalog will saturate at 200 billion sources, given by the surface density of objects seen on the Hubble Deep Field (HDF). If you observe sufficiently long, the sources no longer pile up at the faint limit of the survey. At the HDF limit, the source counts are turning over as we see the limit of the Universe defined by a combination of its curvature and seeing back in time to an epoch before galaxies were starting to form (we are seeing everything in our visible horizon). The star-galaxy ratio is inverted from that seen with SDSS; there are many more galaxies than stars. Each of them is resolved and must be saved in the "postage stamp archive". The postage stamp archive becomes meaningless, as a casual inspection of the HDF shows that postage stamps of individual objects would cover roughly half of the sky! So the 0.1 petabyte source catalog is NUDS' only "light data product".

2.3 The Ultimate Digital Sky (UDSS). UDSS requires a new kind of detector that registers the direction, energy and time of arrival of each photon. One then archives photon data rather than pixel data. Considering all the photons that hit a 10m telescope in 10 years, our rough estimate of the data volume is $\sim 10^{23}$ bytes which we call an Avogadro-byte. If we had a real observational astronomer here, they'd tell you why they have to have it.

3 Following the Progress of Simulations

3.1 A Brief History of N. Over the last 20 years, the N of our simulations has increased as: $log_{10}N = 0.3 * (Year - 1973)$. To simulate 10^{12} particles, can we just wait until 2013? Computers in 1974 had speeds of 10^6 flops, now they are $\gtrsim 10^{10}$ flops—a 10^4 speed improvement. However, our current algorithms equal this speed-up by requiring 10^4 fewer floats per timestep to advance 10^9 particles than the algorithms used in 1974 code. To reach $N = 10^{12}$ with computer speed and algorithms contributing equally, we would have to achieve the impossible and make the cost of advancing a particle equal to few dozen pairwise force evaluations!

We note that both the spectroscopic survey efficiency and our simulation throughput have increased by 10^8 in the last decades. However, the spectroscopic efficiency has saturated; none of D, q, $N_{spectra}$ or (S/N) can change by large factors. In the future, simulation will play an increasingly greater role.

3.2 Declaration of N-dependence. There are a variety of problems where $N \sim 10^6$ represents a minimum ante. For example, clusters of galaxies are extremely important for determining cosmological parameters such as the density of the Universe. Within a cluster, the galaxies are 1-10% of the mass, and there are roughly 10^3 of them. If the galaxies have fewer than 10^3 particles, they dissolve before the present epoch owing to two-body relaxation in the tidal field of the cluster. To prevent this, we need $N > 10^7$ per cluster. Scaling to the Sloan Volume yields $N \sim 10^{12}$.

There are $\sim 10^{20}$ solar masses within the SDSS volume, so even 10^{12} is a paltry number as each particle would represent 10^8 solar masses. We need ten-fold more to represent the internal structure of galaxies. N will always be far smaller than the true number of particles in the Universe and will compromise the physics of the system at some level. We can only make sure that: 1) the physics being examined has not been compromised by discreteness effects owing to N-deprivation and 2) gravitational softening, discrete timesteps, force accuracy and simulation volume don't make matters even worse. N is not the figure of merit in most reported simulations—it should be! *The N–body Constitution* in the Appendix provides a set of *necessary but not sufficient* guidelines for N-body simulation.

The main physical effect of discreteness is the energy exchange that results from two body collisions. Gravity has a negative specific heat owing to the negative total energy (sum of gravitional binding and kinetic energy) of a bound ensemble, like a star cluster. As a star cluster evolves, stars are scattered out by collisions and leave with positive energy. The remaining stars have greater negative energies, the cluster shrinks, the gravitational binding energy increases and the stars move faster. In galaxies and clusters of galaxies, the timescale for this to occur is 10^3 to 10^6 times the age of the Universe. In many simulations, the combination of discreteness in mass, time and force evaluation can make the timescale much shorter leading to grossly unphysical results. So, we must use N sufficient that physical heating mechanisms dominate over numerical heating or the numerical heating timescale is much longer than the time we simulate. We inventoried all the physical heating mechanisms experienced by galaxies in clusters and discovered a unique new phenomena we call "galaxy harassment" [29, 30].

4 All the N that fits.

There are two constraints on our choice of N. The cost of computing a full cosmological simulation is $\sim 10^{5.7}N^{4/3}$ floats (the scaling with $N^{4/3}$ arises from the increased time resolution needed as interparticle separation decreases). The memory needed to run a simulation is $\sim 10^2 N$ bytes. If we fix N by filling memory, the time to run a simulation is 10 days $\times$ (bytes/flop rate$(N/30\text{Million})^{1/3}$). Current machines are well balanced for our Grand Challenge simulations. With Gigaflops and Gigabytes, we can perform simulations with $N \sim 10^{7.5}$. With Teraflops and Terabytes, we can simulate 10^{10} particles. Simulations with $N \sim 10^{12}$ lie in the nether world of Petaflops and Petabytes.

5 Parallel Adaptive N-body Solvers.

Performance gains of the recent past and near future rely on parallel computers that reduce CPU–years to wall-clock–days. The challenge lies in dividing work amongst the processors while minimizing the latency of communication.

The dynamic range in densities demands that spatially adaptive methods be used. Our group has forsaken adaptive mesh codes to concentrate on tree-codes[2, 6] that can be made fully spatially and temporally adaptive. The tree-codes use multipole expansions to approximate the gravitational acceleration on each particle. A tree is built with each node storing its multipole moments. Each node is recursively divided into smaller subvolumes until the final leaf nodes are reached. Starting from the root node and moving level by level toward the leaves of the tree, we obtain a progressively more detailed representation of the underlying mass distribution. In calculating the force on a particle, we can tolerate a cruder representation of the more distant particles leading to an $O(N \log N)$ method. We use a rigorous error criterion to insure accurate forces. Since we only need a crude representation for distant mass, the concept of "computational locality" translates directly to spatial locality and leads to a natural domain decomposition.

The force of a distant cell can be calculated using higher moments other than monopole. Our experiments have settled on hexadecapoles as the most efficient order. Since a hexadecapole requires 319 floats to evaluate while a monopole is only 38, the greater serial efficiency becomes a bigger win in parallel as fewer off-processor cells are fetched for a fixed error criterion. In order to use hexadecapoles in cosmological simulations, we had to generalize periodic boundary conditions to arbitrary order[44].

Experience has shown that the domain decomposition and the data structure for force calculation must be the same for efficient calculation. One approach is to build an oct-tree for forces[6] and hash it to produce the domain decomposition[5].

We use a balanced k–D tree for both the domain decomposition[43] and the data structure. The tree is constructed by recursively bisecting the particle distribution along the longest axis. The lowest level nodes of this tree contain several particles (usually 8 to 32) whose force calculations are collectively optimized. Each 2^m domains is a single rectangular region of the upper m levels of the tree. Pointers are unnecessary, each node in the tree can be indexed so that finding children, parents and siblings in the tree are simple bit shift operations[5]. The same shift operations provide a natural ordering to move through the particle list and calculate forces. By using such a natural order, there is a large correlation in the off-processor data from one particle to the next. An efficient caching strategy can exploit this.

This simple tree structure provides a fast and effective means to create portable parallel codes for N-body simulations, nearest–neighbor searching and group finding. The same code runs under PVM, MPI, the Cray shmem library, KSR's pthreads and INTEL's NX system. Strategies vary a bit from machine to machine. The Cray T3D-512 version has a peak rate of nearly 20 Gigaflops with a sustained rate of over 10 Gigaflops. The IBM SP2 version is ~ 1.7 times faster.

6 Hierarchical Timestepping.

The great advance in the calculation of forces has come from hierarchical methods that are spatially adaptive. As the number of particles in a cosmological simulation grows, so do the density contrasts and the range of dynamical times ($\propto 1/\sqrt{\text{density}}$). If we take the final state of a simulation and weight the work done on particles inversely with their natural timesteps, we find a potential gain of of ~ 50. Temporal adaptivity is one of the last algorithmic areas where we can squeeze an order of magnitude improvement.

The most commonly used time integration scheme for N-body simulations is leapfrog:

$$\begin{aligned}
Drift,\ D(\tau/2); \quad \underline{r}_{n+1/2} &= \underline{r}_n + \frac{1}{2}\tau\underline{v}_n, \\
Kick,\ K(\tau); \quad \underline{v}_{n+1} &= \underline{v}_n + \tau\underline{a}(\underline{r}_{n+1/2}), \\
Drift,\ D(\tau/2); \quad \underline{r}_{n+1} &= \underline{r}_{n+1/2} + \frac{1}{2}\tau\underline{v}_{n+1}
\end{aligned}$$

where $\underline{r}$ is the position vector, $\underline{v}$ is the velocity, $\underline{a}$ is the acceleration, and τ is the timestep. The operator $D(\tau/2)K(\tau)D(\tau/2)$ evolves the system under the Hamiltonian

$$H_N = H_D + H_K + H_{err} = \frac{1}{2}\underline{v}^2 + V(\underline{r}) + H_{err},$$

where H_{err} is of order τ^2[40].

The existence of this surrogate Hamiltonian insures that the leapfrog is symplectic—it is the exact solution of an approximate Hamiltonian. Errors explore the ensemble of systems close to the initial system rather than an ensemble of non-Hamiltonian time evolution operators near the desired one.

Leapfrog is a second-order symplectic integrator requiring only one costly force evaluation per timestep and only one copy of the physical state of the system. These properties are so desirable that we have concentrated on making an adaptive leapfrog. Unfortunately, simply choosing a new timestep for each leapfrog step evolves $(\underline{r}, \underline{v}, \tau)$ in a manner that may not be Hamiltonian, hence it is neither symplectic nor time-reversible. The results can be awful[10]. Time reversibility can be restored[20] if the timestep is determined implicitly from the state of the system at both the beginning and the end of the step. This requires backing up timesteps, throwing away expensive force calculations and using auxiliary storage. However, we can define an operator that "adjusts" the timestep, A, yet retains time reversibility and only calculates a force if it is used to complete the timestep[34]. This is done by choosing A such that it commutes with K, so that $DAKD$ is equivalent to $DKAD$. Since K only changes the velocities, an A operator that depends entirely on positions satisfies the commutation requirement. The "natural definition" of timestep, $\propto 1/\sqrt{\text{density}}$, is ideal but it is difficult to define when only a few particles are within the region of interest. Synchronization is maintained by choosing timesteps that are a power-of-two subdivision of the largest timestep, τ_s. That is, $\tau_i = \frac{\tau_s}{2^{n_i}}$, where τ_i is the timestep of a given particle. We are currently experimenting with this approach and encourage others to look at variants.

7 Volume Renormalizations.

The power of this technique has recently been shown in the simulation of rare quasar formation sites[22]. A large scale simulation is first done at modest resolution (particle mass of $10^{10.5}\mathrm{M}_\odot$)– galaxies "weigh" just $\sim$ 100 particles. Current simulations with this resolution cover volumes of (100 Mpc)^3 with $\gtrsim$ 10^7 particles. Within these volumes, regions of interest are identified. These can be sites of galaxy/QSO formation occurring at high redshift, a large cluster of galaxies or a structure matching our local group.

Next, initial conditions are reconstructed using the same low-frequency waves present in the low resolution simulation but adding the higher spatial frequencies. To reduce the number of particles and make the nonlinear simulation possible using the same cosmological context, we construct another set of initial conditions with particles whose mass, and therefore mean separation, increase with the distance from the center of our volume. Note that because tides are important in the formation of the filaments it is not sufficient to extract just the central region. Finally, the higher resolution simulations can be done adding gas dynamics using TREESPH[19].

Using this approach, we can simulate structures like the local Virgo cluster of galaxies using 10^7 particles. A larger cluster like Coma requires 10^8 particles. We can get the gross dynamics of the local group, capturing the larger satellites with 10^7 particles.

8 Simulating the Sloan Volume.

Our proposed program to simulate the Sloan Volume before the new millenium is as follows:

- Simulate the entire volume (800 Mpc)^3 with $N = 10^{10}$, each with a mass of $10^{10.5}\mathrm{M}_\odot$.
- "Renormalize" dozens of groups, clusters, etc. and simulate with 10^8–10^9 particles.

The total cost for the first simulation is roughly a Teraflop-year and requires a machine with a Terabyte of memory. The second sequence of simulations should be designed to have roughly equal computational cost, but will require less memory.

9 The Fate of the Solar System.

Advances in hardware and numerical methods finally enable us to integrate the Solar System for its lifetime. Such an integration is a 1,000 fold advance on the longest accurate integration ever performed[27] and can address numerous questions:

Is the Solar System stable? Do all the planets remain approximately in their current orbits over the lifetime of the Solar System, or are there drastic changes, or perhaps even an ejection of a planet?

What is the affect of orbital changes on the planetary climates? According to the Milankovich hypothesis, climate variations on the Earth are caused by insolation changes arising from slow oscillations in the Earth's orbital elements and the direction of the Earth's spin[9]. Remarkably, the geophysical data (primarily the volume of water locked up in ice as determined by the $^{18}O/^{16}O$ ratio in seabed cores) covers a longer time than any accurate ephemeris.

How does weak chaos alter the evolution of the Solar System? An empirical relationship was found between the Lyapunov timescale and time for dramatic change such as planet crossings and ejections:

$$\frac{T_{eject}}{T_{dynamical}} \propto \frac{T_{Lyap}}{T_{dynamical}}^{1.8}$$

but the relation was only investigated for $T_{Lyap} \sim 10 - 100 T_{dynamical}$ and there is large scatter[28].

What is the stability of other planetary system? How are the giant planets related to terrestrial planets on stable orbits? Without such a cleansing of planetesimals from the Solar System by giant planets[14], the bombardment of the Earth by asteroids would be steady and frequent throughout the main sequence lifetime of the Sun[47]. The chaos produced by Jupiter and Saturn may have played a role in insuring that planetesimals collided to form the terrestrial planets[1], but too much chaos will eject planets in the habitable zone. While a search for giant planets is the only technically feasible one today, it may be the ideal way to screen systems before searching for terrestrial planets.

10 Methods for Evolving the Solar System.

10.1 The Legacy of Laplace. When Laplace expanded the mutual perturbations of the planets to first order in their masses, inclinations and eccentricities, he found that the orbits could be expressed as a sum of periodic terms—implying stability. Poincaré[33] showed that these expansions don't converge owing to resonances. Using the KAM theorem, Arnold[4] derived

[1] In Ancient Greek, chaos was "the great abyss out of which Gaia flowed".

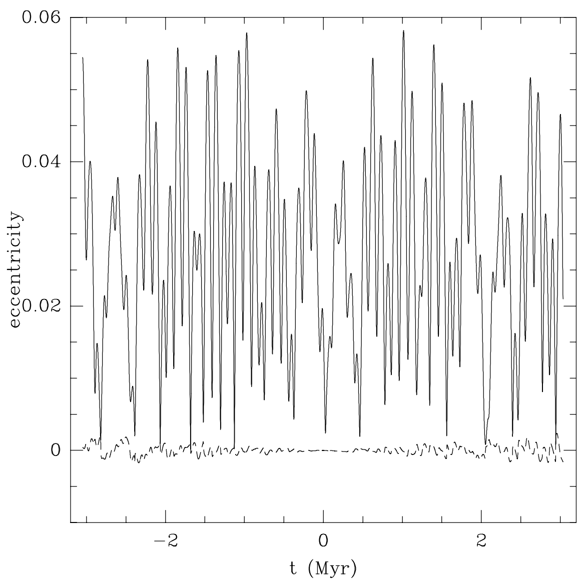

Figure 2: A comparison of the Earth's eccentricity as calculated by evolving the secular system and explicitly integrating the planet's orbits. The solid line shows the Earth's eccentricity in the direct integration, and the dashed line shows the difference in Earth's eccentricity between the two methods.

constraints on planet masses, eccentricities, and inclinations sufficient to insure stability. The Solar System does not meet his stringent conditions, but this does not imply that it is unstable.

Laskar[23] tested the quasi-periodic hypothesis by numerically integrating the perturbations calculated to second order in mass and fifth order in eccentricities and inclinations, ~150,000 polynomial terms. Fourier analysis of his 200 million year integration reveals that the solution is not a sum of periodic terms and implies an instability that is surprisingly short, just 5 Myr.

10.2 The Legacy of Newton The second method for attacking the stability problem is to explicitly integrate the planets' orbits (Table 1). As early as 1965, Pluto's behavior was suspicious[11]. In the last ten years, projects using general purpose computers such as LONGSTOP[31] have battled special purpose machines like the Digital Orrery[3, 45] for top honors in computing the orbits of the outer planets. The Digital Orrery integrations show a Lyapunov exponent of (1/20 Myr^{-1}) for Pluto's orbit, while LONGSTOP finds that Pluto is locked in a complicated system of resonances so that the Lyapunov exponent could be extremely sensitive to fine details of the initial conditions.

10.3 New Integration Methods. LONGSTOP integrations were not limited by CPU time but by roundoff error[35]. New fourth order "mixed variable symplectic" (MVS) integrators reduce roundoff error such that a 10 billion year integration of all nine planets is possible with 128 bit precision. These new integrators were used for an accurate integration of all nine planets and the Earth's spin axis for 3.05 Myr into the past, and future[36, 27]—roughly the limit of 64 bit precision. General relativity was included in an extremely elegant way[40]. A comparison with the heroic secular perturbation calculation shows remarkable agreement over their common range including the existence of the secular resonance claimed to be responsible for the chaos[27], but all planetary orbits appear to be regular over the 6 Myr interval. (see Figure 2)

MVS integrators[39] separate the Hamiltonian into the Kepler part and the mutual planetary kicks:

$$H_{Solar\ System} = H_{Kepler} + H_{planetary\ kicks} + H_{err}$$

and use fourth order leapfrog integrators[49] composed of steps: $D(\tau/2)K(\tau)D(\tau/2)$ as in §6, but the *Drift* operator moves the particle on a Kepler ellipse and the *Kick* operator calculates the mutual interactions of the planets. H_{err} is of order τ^4. The accuracy of the integrator at a given timestep is increased by the ratio of Jupiter's mass (the biggest kicker) to the sun's mass.

Sussman and Wisdom[46] performed a 100 Myr integration on a somewhat special purpose computer that used the potential approximation to General Relativity[32], but was otherwise similar to the 6 Myr integration[36]. They found an initial divergence timescale of 12 Myr, but a 4 Myr divergence dominates after 60 Myr. The 4 Myr divergence occurs much later in the outer planets than in the inner planets hinting at two distinct mechanisms (or three since Pluto has its own distinct chaotic behavior). They were cautious about identifying the underlying dynamical mechanism for the chaos, noting that an angle can alternate between libration and circulation owing to the projection of a high-dimensional trajectory onto a plane as well as from a separatrix. Understanding the source of Solar System chaos awaits an analytical demonstration that the resonances involved are sufficiently strong and close for resonance overlap.

Nonetheless, the Solar System is almost certainly chaotic. Laskar[24] looked at the fate of Mercury and estimates the chance of ejection in the next few billion years approaches 50%. Our belief in the regularity of the Solar System would be dashed if the ejection of Mercury were in the historical record. There could have been a dozen or more planets just a few billion years ago. At the very least, the chaotic motion leads to a horizon of predictability for the detailed motions of the planets. With a divergence timescale of 4-5 Myr time[23, 46], an error as small as 10^{-10} in the initial conditions will lead to a 100% discrepancy in 100 Myr. Every time that NASA launches a rocket, it can turn winter to spring in a mere 10 Myr.[2] (Don't let this go beyond this room, environmental impact statements are already tough enough.)

Table 1: Solar System Integration History

Year	Ref	Length (Myr)	# Planets	GR?	Earth's Moon?
1951	[15]	0.00035	5	no	no
1965	[11]	0.12	5	no	no
1973	[12]	1.	5	no	no
1986	[3]	217.	5	no	no
		3.	8	no	no
1986	[32]	100.	5	yes	no
1988	[45]	845.	5	no	no
1989	[38]	2.	9	no	no
1991	[36]	6.	9	yes	yes
1992	[46]	100.	9	yes	yes
1999	us	10,000.	9	yes	yes

We have started a 9 Gyr integration—4.5 Gyr into the past when the Solar System was formed and 4.5 Gyr into the future when the Sun becomes a red giant. One basic requirement is a computer with fast quad precision to overcome roundoff problems. Table 2 shows that the IBM 3CT is the current machine of choice. To understand any chaos, we will need to see it by an independent means and devise methods to determine its underlying source.

Table 2: Digital Orrery Speeds

machine	Years of Evolution per cpu year
IBM 3CT	$\sim 10^9$
SGI R8000	$\sim 10^8$
HP/735	$\sim 10^7$
Sparc-10	$\sim 10^6$
Dec alpha	no quad precision

10.4 Parallel Methods for Calculating the Solar System. Because there are only nine planets, distributing the planets among different processors does not promise great speed gains through parallel computation. We employ a different form of parallelism—the "time–slice concurrency method" (*TSCM*)[41]. In this method, each processor takes a different time–slice; processor 2's initial conditions are processor 1's final conditions and so on. The trick is to start processor 2 with a good prediction for what processor 1 will eventually output, and iterate to convergence. This is analogous to the waveform relaxation technique used to solve some partial differential equations[16]. However, Kepler ellipses are a good guess to the orbits for a timescale that is proportional to the ratio of the Sun's mass to Jupiter's. Tests show that it is extremely efficient to iterate to convergence in double precision (typically 14 iterations each costing 10-15% of a quad iteration), then perform just two iterations to get convergence in quad. In this way, the total overhead of the full 16 iterations can be less than a factor of 4. There are still many algorithmic issues to be addressed.

For long-term integrations, TSCM has been formulated in a way that preserves the Hamiltonian structure and exploits the nearness to an exactly soluble system; otherwise errors grow quadratically with time. *TSCM* will enable us to integrate $\sim$ 0.5 Gyr per day on a 512 node SP2—a speed–up over real–time of 10^{11}. This will make it feasible to study the stability of other solar systems. Detailed development and implementation will be much more challenging than for previous methods, and our high quality serial integration will be required for comparison and validation.

Finally, we will use a new technique to gauge the origin of instabilities (the "tangent equation method")[42]. In the past, it was common to integrate orbits from many slightly different initial conditions. While that works, it is more rigorous and also more economical to integrate the linearized or tangent equations—the equations for *differences* from nearby orbits. We will integrate the tangent equations along with the main orbit equations.

11 Cosmology meets Cosmogony: Simulating the Formation of Planetary Systems

Theories of Solar System formation are traditionally divided into four stages[25]: collapse of the local cloud into a protostellar core and a flattened rotating disk (Nebular Hypothesis); sedimentation of grains from the cooling nebular disk to form condensation sites for planetesimals; growth of planetesimals through binary collision and mutual gravitational interaction to form protoplanets (planetesimal hypothesis); and the final assembly to planets with the remaining disk cleansed by ejections from chaotic zones.

The cosmology code described in §5 & §6 is ideal for the third stage of Solar System formation, particularly in the inner regions where gas was not a primary component and gravitational interactions dominate the evo-

[2] Are the integrations meaningful given this sensitivity to the initial conditions? We investigate Hamiltonian systems that are as close to the Solar System as possible. KAM theory tells us that the qualitative behavior of nearby Hamiltonians should be similar. While the exact phasing of winter and spring is uncertain after millions of years, the severity of winter or spring owing to changes in the Earth–Sun distance and the obliquity are predictable.

lution. The first stage entails magnetohydrodynamics, the complicated small-particle physics and gas dynamics of the second stage is still not well understood, and the fourth is the purview of long-term stability codes.

All that is required for a detailed simulation of the third stage is a model of the collisional physics and a code capable of dealing with a large number of particles. Previous simulations of the planetesimal stage have been restricted to of order 10^2 particles[7], or examined a local patch of a disk with Kepler shear[1] with one or two "external perturbers" to mimic the action of giant planets[47]. Our cosmology code has the potential to treat as many as 10^8 particles simultaneously, a million-fold improvement that makes us enthusiastic! Only statistical methods[48] employing prescriptions for the outcomes of encounters have been used to peek at this regime.

We reach an important threshold at $N \sim 10^8$ in our ability to follow planetesimal evolution. At early times, the relative velocities between planetesimals are small and inelastic physical collisions lead to "runaway" growth of planetary embryos[7]. Eventually gravitational scattering increases the planetesimal eccentricities to such an extent that collisions result in fragmentation, not growth. The embryos will continue to grow owing to their large mass, but at a slower rate as their "feeding zones" are depleted[21]. The total mass of our planetary system is $448M_\oplus$ or $3.6 \times 10^4 M_{lunar}$, while the inner planetesimal disk amenable to simulation had a mass $\sim 10^3 M_{lunar}$. To capture both growth and fragmentation[48] requires a minimum particle mass of $10^{-5} M_{lunar}$ leading to our target $N \sim 10^8$.

A detailed direct simulation of planet formation can address a variety of important questions, including: Was there runaway growth of a few embryos or a continuously evolving homogeneous mass distribution? How does the primordial surface density alter the evolution? What is the dominant physical mechanism that drives the late stages of growth—perturbations by the giant gas planets or intrinsic gravitational instabilities? What fixes the spin orientation and period of the planets—uniform spin-up from planetesimal accretion[26] or a stochastic process dominated by the very last giant collisions[13]? Is it feasible that the Earth suffered a giant impact late in its growth that led to the formation of the Moon[8]? How much radial mixing was there and can it explain observed compositional gradients in the asteroid belt[17]?

12 Summary

.

An old joke is that when theorists give talks, only they believe the result; while when observers speak, they are the only ones that do not. Simulation frees theory from artificial simplifications and determines its nonlinear consequences after applying the projections and biases of observational (not experimental) data. However, all too often it becomes the "worst of both worlds". The results may rest on the choice of "sub-grid physics" and the output is freely rotated and projected to create a picture that is compared to real observations. In our approach, we strive to:

- Carefully choose problems where the most important physics is above the grid; problems dominated by gravity are best.
- Implement any subgrid physics faithfully.
- Adhere to clear standards for high quality simulations.
- Create data products with sufficient integrity to be put into the critical path of observational programs.
- Collaborate with observational astronomers to insure the utility of the simulation data.

In this approach, we embrace future computing technology and ask, "What simulations will solve these problems once and for all"? In many cases, problems must be deferred, and the next year is spent on an appropriate niche problem where this goal can be achieved. However, with the approach to Teraflop and Petaflop computing, we hope that the niches grow to fill the space of problems that we wish to attack.

References

[1] S. J. Aarseth, D. N. C. Lin and P. L. Palmer, *Evolution of planetesimals. II. Numerical simulations*, Astrophys. J., 403, (1993), pp. 351–76.

[2] A. W. Appel, *An efficient Program for Many-Body Simulation*, SIAM J. Sci. Stat. Comput., 6, (1985), pp. 85–93.

[3] J. H. Applegate, M. R. Douglas, Y. Gursel, G. J. Sussman and J. Wisdom, *The outer Solar System for 210 million years*, Astron. J., 92, (1986), pp. 176–94.

[4] V. I. Arnold, *Small Denominators and the Problem of Stability in Classical and Celestial Mechanics*, in Report to the IVth All-Union Mathematical Congress, Leningrad, (1961).

[5] J. Barnes, *An efficient N-body algorithm for a fine-grain parallel computer*, in The Use of Supercomputers in Stellar Dynamics, P. Hut and S. McMillan, eds., Springer Verlag, New York, (1986), pp. 175–80.

[6] J. Barnes and P. Hut, *A Hierarchical O(NlogN) Force-Calculation Algorithm*, Nature, 324, (1986), pp. 446–50.

[7] C. Beaugé and S. J. Aarseth, *N-body simulations of planetary formation*, Mon. Not. R. astr. Soc., 245, (1990), pp. 30–39.

[8] W. Benz, W. L. Slattery and A. G. W. Cameron, *The origin of the Moon and the single-impact hypothesis. I*, Icarus, 66, (1986), pp. 515–35.

[9] A. Berger, J. Imbrie, J. Hayes, G. Kukla and B. Saltzman, eds., *Milankovitch and climate: understanding the response to astronomical forcing*, Reidel, Dordrecht, (1984).

[10] M. P. Calvo and J. M. Sanz-Serna, *The development of variable-step symplectic integrators with application to the two-body problems*, SIAM J. Sci. Comput., 14, (1993), pp. 936–52.

[11] C. J. Cohen and E. C. Hubbard, *Libration of the Close Approaches of Pluto to Neptune*, Astron. J., 70, (1965), pp. 10.

[12] C. J. Cohen, E. C. Hubbard and C. Oesterwinter, *Elements of the outer planets for a million years*, Astron. Pap. Amer. Ephem., 22, (1973), pp. I.

[13] L. Dones and S. Tremaine, *On the origin of planetary spins*, Icarus, 103, (1993), pp. 67–92.

[14] M. Duncan and T. Quinn, *The Long-Term Dynamical Evolution of the Solar System*, Ann. Rev. Astron. Astrophys., 31, (1993), pp. 265–89.

[15] W. J. Eckert, D. Brouwer and G. Clemence, *Coordinates of the five outer planets 1653–2060*, Astron. Pap. Amer. Ephem., 12, (1951), pp. 1–47.

[16] C. W. Gear, *Waveform methods for space and time parallelism*, J. Comp. and Appl. Math., 38, (1991), pp. 137–47.

[17] J. C. Gradie, C. R. Chapman and E. F. Tedesco, *Distribution of taxonomic classes and the compositional structure of the asteroid belt*, in Asteroids II, R. P. Binzel, T. Gehrels and M. S. Matthews, eds., Univ. Arizona Press, Tucson, (1989), pp. 316-35.

[18] J. E. Gunn and G. R. Knapp, *The Sloan Digital Sky Survey*, in Sky Surveys, Protostars to Protogalaxies, T. Soifer, ed., Ast. Soc. of Pacific conference series, #43, (1992), pp. 267–79.

[19] L. Hernquist and N. Katz, *TREESPH, a unification of SPH with the hierarchical tree method*, Astrophys. J. Suppl., 70, (1989), pp. 419–46.

[20] P. Hut, J. Makino and S. McMillan, *Building a better leapfrog*, Astrophys. J. Lett., 443, (1994), pp. L93–6.

[21] S. Ida and J. Makino, *Scattering of Planetesimals by a Protoplanet: Slowing Down of Runaway Growth*, Icarus, 106, (1993), pp. 210–27.

[22] N. Katz, T. Quinn, E. Bertschinger and J. M. Gelb, *Formation of quasars at high redshift*, Mon. Not. Roy. astr. Soc., in press, (1994).

[23] J. Laskar, *A numerical experiment on the chaotic behaviour of the Solar System*, Nature, 338, (1989), pp. 237–8.

[24] J. Laskar, *Large-scale chaos in the Solar System*, Astron. Astrophys., 287, (1994), pp. L9–12

[25] J. J. Lissauer, *Planet formation*, Ann. Rev. Astron. Astrophys., 31, (1993), pp. 129–74.

[26] J. J. Lissauer and V. S. Safronov, *The random component of planetary rotation*, Icarus, 93, (1991), pp. 288–97.

[27] J. Laskar, T. Quinn, and S. Tremaine, *Confirmation of resonant structure in the Solar System*, Icarus, 95, (1992) pp. 148–52.

[28] M. Lecar, F. Franklin and M. Murison, *On predicting long-term orbital instability: a relation between the lyapunov time and sudden orbital transitions*, Astron. J., 104, (1992), pp. 1230–6.

[29] B. Moore, N. Katz and G. Lake, *On the destruction and overmerging of dark halos in dissipationless N-body simulations*, Astrophys. J., 457, (1996), pp. 455-9.

[30] B. Moore, N. Katz and G. Lake, *Galaxy harassment and the evolution of clusters of galaxies*, Nature, 379, (1996), pp. 613–16.

[31] A. M. Nobili, A. Milani and M. Carpino, *Fundamental frequencies and small divisors in the orbits of the outer planets*, Astron. Astrophys., 210, (1989), pp. 313–36.

[32] A. M. Nobili, 1986, *LONGSTOP and the masses of Uranus and Neptune* in Solid Bodies of the Outer Solar System, ESTEC, Noordwijk, (1986), pp. 9-13.

[33] H. Poincaré, *Les methodes nouvelles de la mechanique celeste.* Gauthiers-Villars, Paris, (1892).

[34] T. Quinn, N. Katz, J. Stadel and G. Lake, *Time stepping N-body simulations*, in preparation.

[35] T. Quinn and S. Tremaine, *Roundoff error in long-term planetary orbit integrations*, Astron. J., 99, (1990), pp. 1016–23.

[36] T. R. Quinn, S. Tremaine and M. J. Duncan, *A three million year integration of the Earth's orbit*, Astron. J., 101, (1991), pp. 2287–305.

[37] D. L. Richardson and C. F. Walker, *Multivalue integration of the planetary equations over the last one-million years.* Astrodynamics 1987, Soldner et al, eds. Univelt, San Diege, (1987) pp. 1473–56.

[38] D. L. Richardson and C. F. Walker, *Numerical simulation of the nine-body planetary system spanning two million years*, J. Astronautical Sci., 37, (1989), pp. 159–82.

[39] P. Saha and S. Tremaine, *Symplectic integrators for Solar System dynamics*, Astron. J., 104, (1992), pp. 1633-40.

[40] P. Saha and S. Tremaine, *Long-term planetary integration with individual time steps*, Astron. J., 108, (1994), pp. 1962–9.

[41] P. Saha, J. Stadel and S. Tremaine, *A Parallel Integration Method for Solar System Dynamics*, Astron. J., submitted, (1992).

[42] P. Saha, *The Use of Tangent Equations to Detect Chaos in Solar System Dynamics*, in preparation (1997).

[43] J. K. Salmon, *Parallel Hierarchical N-body Methods*, Ph.D. Thesis, California Institute of Technology, 1990.

[44] J. Stadel and T. R. Quinn, *A generalization of Ewald summation to arbitrary multipole order*, in preparation (1996).

[45] G. J. Sussman and J. Wisdom, *Numerical evidence that the motion of Pluto is chaotic*, Science, 241, (1988), pp. 433–7.

[46] G. J. Sussman and J. Wisdom, *Chaotic Evolution of the Solar System*, Science, 257, (1992), pp. 56–62.

[47] G. W. Wetherill, *Possible consequences of absence of "Jupiters" in planetary systems*, Ap. Sp. Sci., 212, (1994), pp. 23–32.

[48] G. W. Wetherill and G. R. Stewart, *Accumulation of a swarm of small planetesimals*, Icarus, 77, (1989), pp. 330–57.

[49] H. Yoshida, *Construction of higher order symplectic integrators*, Phys. Lett. A., 150, (1990), pp. 262–8.

Appendix: The N-body Constitution

Preamble

We, the people of the HPCC consortium, in order to form more perfect N–body simulations, establish justice, insure domestic tranquility, provide for the common defense, promote the general welfare, and secure the blessings of scientific accuracy to ourselves and our posterity do ordain and establish this Constitution for the conduct of numerical N–body simulations.

Article I: On the Gravitational Softening Length

Section 1—The gravitational softening length should be large enough to minimize the effects of two body relaxation.

There should be at least 8 particles in each softening volume in objects of interest.

Article II: On the Size of Time steps

Section 1—Time steps should be chosen to be small enough to eliminate the effects of two body scattering introduced through integration errors.

The timestep must satisfy $\Delta t < \frac{1}{10}\left(\frac{1}{G\rho}\right)^{1/2}$. Near the center of an isothermal sphere this is equivalent to $\Delta t < .38\frac{\epsilon_{grav}}{v_{max}}$ for a standard Plummer softening. For a typical cosmological simulation $N_{steps} = \frac{T_{HUBBLE}}{\Delta t} \sim 4500\frac{10kpc}{\epsilon_{grav}}$. Spline kernel softening requires 33% more timesteps. Hierarchical timestepping is permitted. Some particles must be on the smallest timesteps and any criterion for longer timesteps must be rigorously tested using both isolated models and hierarchical clustering.

Article III: On the Accuracy of Forces

Section 1—Forces should be calculated with a maximum absolute and relative error.

The error should always be less than 0.5% of the force or the rms force, whichever is less.

Article IV: On the Accuracy of the Integrator

Section 1—The integrator must be second order and symplectic (or time reversible) and avoid any correlated higher order error terms.

Retaining second order accuracy is particularly important when the length of the time step in changed during the calculation. However, one must be careful not to introduce correlated third order errors when correcting to second order accuracy. If the integrator is not symplectic then artificial dissipation may result.

Article V: On the Size of the Simulation Volume

Section 1—The simulation volume must be large enough to model all non–linear effects.

Mode couplings between large and small scales can increase power on small scales. In particular, filaments greatly affect the gravitational evolution at small scales. The diameter of the simulation for a CDM spectrum with $\sigma_8 = 0.7$ must be at least $40h^{-1}$ Mpc. Models with more large scale power will require larger volumes.

Section 2—No one object should dominate the evolution of the simulated volume.

To prevent objects from tidally influencing themselves via aliasing and to keep objects from artificially lowering the mean background density, no virialized object may contain more than 1/10 the total simulated mass.

Section 3—The simulation must have a large enough volume that the amplitude in the fundamental mode $\delta_1 < 0.01$ when the simulation is concluded.

Accurate evolution of the largest scale modes requires periodic boundary conditions. Further, without periodic boundary conditions, the accelerations at the edge of the volume are large and error criteria become hard to apply.

Article VI: On the Starting Redshift of the Simulation

Section 1—The simulation must start at a redshift high enough to insure that all represented mass scales are still in the linear regime.

In particular, if the initial conditions are generated using the Zel'dovich approximation, the starting redshift must be sufficiently high to insure that the absolute maximum $|\delta| \lesssim 1$. For example, a CDM simulation with $\sigma_8 = 0.7$ in a cubic volume $40h^{-1}$ Mpc on a side simulated with 128^3 particles would require a starting redshift of at least $z = 50$.

Article VII: What Ratification Shall Establish Constitution

Section 1—This constitution must be ratified by the representatives of at least two different HPCC groups at a constitutional convention by late 1995.

The effects of violating the above articles are not known. Some violations will introduce unphysical dissipative effects while others will introduce errors that act like an artificial heat source.

Status of the N-body Constitution

Several groups have acknowledged the importance of these criteria in their own research papers. However, few have shown a willingness to follow through and perform simulations to the quality specified here.