Preliminary Report of the Task Group for an NSF/NASA Computational Effort in Gravitational Wave Science

1 Preamble

In February, 2002, the Physics Division of the NSF and the Astronomy & Physics Division of NASA's Office of Space Science set up a task group to study certain theoretical issues critical to the LIGO and LISA projects. The charge to the group is reproduced in the Appendix. The group was requested to prepare a preliminary report by early April, and a more complete report several months later. A workshop for Source Groups for Interferometric Gravitational Wave Detectors was held in conjunction with the LIGO Science Collaboration meeting at Livingston, LA, March 20–23, 2002. A large fraction of the community involved in source simulation science (in the broadest sense) attended the workshop. A significant amount of the input for this report came from this workshop. In addition, several individuals submitted material to the task group in response to e-mailings to the community at large. This document is the preliminary report that was requested.

2 Introduction

Our understanding of the cosmos changed radically in the 20th century when new electromagnetic "windows" were opened onto the universe, especially the radio-wave and X-ray windows. Even more radical changes are likely in the coming decade, when NSF and NASA open windows onto the universe that utilize a fundamentally new kind of radiation: ripples in the fabric of space and time called "gravitational waves." NSF's Laser Interferometer Gravitational-Wave Observatory (LIGO) will initiate the gravitational-wave analog of optical astronomy, while the joint NASA/ESA Laser Interferometer Space Antenna (LISA) will initiate the gravitational analog of radio astronomy.

LIGO and LISA will unveil phenomena that have never before been seen or have been seen only foggily — e.g.,

- collisions of black holes (the most violent events that occur in the modern universe, but events that humans have never seen, since they produce only gravitational waves, not electromagnetic); and
- the gradual inspiral of a neutron star into a supermassive hole, producing gravitational waves that carry to earth a detailed map of the hole's space-time curvature

The more we know in advance about the cosmos's gravitational waves, the more successful we will be in finding the waves' imprints in the noisy data that LIGO and LISA generate. For example, a prior knowledge of the gravitational waveforms from black-hole collisions, when incorporated into LIGO's data analysis algorithms, should result in a two-fold to ten-fold increase in the rate of detection of the collision waves, perhaps even making the difference between detection and non-detection in LIGO's initial interferometers. Similarly, without predicted waveforms it will be impossible for LISA to detect the map-carrying waves from neutron stars spiraling into supermassive holes; with reliable waveforms, detection is likely.

In most cases, the computations of gravitational waveforms require supercomputer simulations of the waves' sources — e.g., simulations of the collisions of black holes or of the implosion of a rapidly rotating stellar core. Many of the needed simulations entail solving numerically Einstein's general relativistic field equations, i.e. "numerical relativity", but in some cases it should be adequate, at least initially, to carry out simulations using Newton's laws of gravity or other approximations to Einstein's laws.

The need for source simulations will not end when waves are detected; rather, it will increase substantially: To extract the information carried by the waves (e.g., the maps of a black hole's space-time curvature) will require detailed comparisons between the observed waveforms and the waveforms predicted by simulations. Thus, when waves are being detected, source-simulation software will become a crucial component of the LIGO and LISA data analysis.

Our understanding today of the waves' details is quite poor, and the capabilities of source-simulation software are far below LIGO's and LISA's needs. In part this is due to the complexities and difficulties of the simulations (especially those using numerical relativity), and in part it is due to the substantially subcritical size of the community attempting the simulations and the meager resources available to the source-simulation community. For LIGO, the need for good source simulations is a near crisis because the first LIGO science run begins this year. For LISA, the most urgent need focuses on one wave source (inspiral of white dwarfs, neutron stars, and small black holes into supermassive holes), which must be understood in the next two years as a key input into firming up the LISA science requirements.

2.1 The Relation Between Theory and Experiment

Many reports on groundbreaking research in the physical sciences have emphasized that large experiments need an appropriate level of theoretical support to maximize their scientific payoff. We quote from the McKee-Taylor decadal survey committee report, *Astronomy and Astrophysics in the New Millenium*:

> The new initiatives recommended below are motivated in large part by theory, which is also key to interpreting the results. Adequate support for theory, including numerical simulation, is a cost-effective means for maximizing the impact of the nation's capital investment in science facilities.

and

> This report recommends a number of projects with technologically advanced instrumentation that will enable observers to extend the frontiers of knowledge. In many instances, astrophysical theorists provide the ideas that guide the choice of instrumentation, the decisions about what to observe, and the interpretation of data. Adequate support of theory is therefore essential in optimizing the nation's investment in astronomy and astrophysics.
>
> To encourage theorists to contribute to the planning of missions and facilities and to the interpretation and understanding of the results, one or more explicitly funded theory challenges should be integrated with most moderate or major initiatives...
>
> Numerical simulation will play an important role in the theory challenges, both in advancing our understanding and in enabling detailed comparison with observation. The theory challenges should significantly enhance the effectiveness of theoretical research related to specific missions.

These recommendations have already had an impact on several NASA missions. Both Hubble and Chandra have significant theory components. A similar strategy should be adopted for LISA (NASA) and LIGO (NSF).

Recommendation # 1: Commensurate with the experimental costs of LIGO and LISA, a vigorous theory program should be instituted immediately to address the science needs of these projects.

2.2 NSF/NASA Coordination

Quoting again from the decadal survey:

> The enormous scale of many astronomical problems requires a coordinated national approach. In many cases, investigations that span different wavelength bands and disciplinary boundaries are needed in order to achieve a fundamental understanding of the phenomena under study. Interagency coordination and cooperation are often essential for such a multidisciplinary approach. Both ground- and space-based facilities can be used to address the scientific themes identified by the committee as ripe for progress throughout this decade. Such facilities are traditionally supported by NSF and NASA, respectively.

The computational science needs of LISA and LIGO, discussed more fully in section 3 below, contain many elements of overlap. For example, a code to calculate the gravitational waveform from the merger of black holes will work equally well for solar mass black holes, a LIGO source, and supermassive black holes, a LISA source. Moreover, many of the skills for addressing one kind of problem can be used in another — people trained in one area of source simulation will likely also work in other similar areas. The Task Group believes there is a compelling case, not just for a computational effort, but that it should be coordinated between NSF and NASA.

Recommendation #2: A program of computational projects addressing the science needs of LIGO and LISA should be instituted and coordinated between NSF and NASA.

2.3 Scale of the Effort

In section 3 we enumerate and describe the computationally intensive research projects that are needed by LIGO and LISA. These projects all focus on *source-simulation research*, defined broadly (so it includes not only simulating the gravitational-wave sources themselves, but also simulating the astrophysical environments of the sources and developing data analysis algorithms, most of which have tight interfaces to the source simulations).

The total funding currently available in the US for such source simulations is about $1 million dollars per year (almost all of it from the approximately $4.5 million theory component of NSF's gravitational physics program, with a small contribution from NASA's $7.5 million Astrophysics Theory Program). To meet LIGO's and LISA's needs, we estimate that the funding for simulations must be ramped up by a factor of at least five, to about $5 million per year, with the ramp up occurring as rapidly as the necessary manpower can be assembled from adjacent fields and by training of new graduate students. (This estimate is justified in section 4.) This increase to a $5 million source-simulation budget is comparable to the size of the NSF and NASA programs in which the source simulations now reside, and could decimate those programs if they were forced to absorb the costs. For comparison, the price tag of $5 million per year over, say, ten years is roughly 5 per cent of the cost of LIGO and LISA together.

In estimating the US needs for source simulations, we presume that our international partners will put a comparable level of funding into this field. Already they are contributing substantially more than the US: Germany alone is putting as much funding into source simulation work at a single institute (the Albert Einstein Institute) as the entire US expenditure.

Our estimate of 5 per cent of the cost of LIGO and LISA together may seem expensive, especially in light of the figure in the McKee-Taylor decadal report of 3 per cent of the cost of astronomical projects for related theoretical support. This is a valid concern that will have to be addressed more fully than we have had time to do yet. For the time being, we note the following:

(i) The charge to the committee included estimating the cost of the computationally intensive research associated with LIGO and LISA. This is our best estimate of what it will take to get the job done.

(ii) In the short term, there are greater needs for LIGO than for LISA, as discussed in the next section.

(iii) There are other theoretical needs for LISA than those discussed in this report. However, they are not computationally intensive and thus may not require the scale of new effort envisaged here.

(iv) Many of the science needs described below involve simulations in numerical relativity. These simulations will have a valuable byproduct: They will produce computational infrastructure that will greatly accelerate explorations of the nonlinear dynamics of spacetime curvature in domains not accessible to gravitational-wave or other observation. Examples from the past and present are: the discovery of critical behavior in gravitational collapse, the discovery of toroidal black holes, explorations (relevant to string theory) of the stability of black strings in five dimensions, and explorations of the dynamics of spacetime near generic singularities.

The hardware needs of this effort are discussed in section 5. The annualized cost is estimated at about $1.5 million for dedicated hardware plus system administrators. By grouping the hardware in regional centers, one can minimize the cost of maintenance and system personnel. This is accounted for in the $1.5 million figure. In addition, the science needs will require access to the very largest national facilities for production runs. Any costs associated with this access have not been accounted for here.

The needs of LISA and especially LIGO are so urgent, that the buildup to our recommended level of effort and funding should be carried out at a rate limited only by the time required to attract the required human talent into the field. We estimate for this a buildup timescale no longer than about three years.

Recommendation #3: The proposed program of computational projects should be funded at an annual rate of $5 million for personnel and $1.5 million for hardware, with a buildup to this level as fast as the required human talent can be found and trained.

3 Science Needs

The advent of ground-based gravitational wave detectors such as the ground-based LIGO, and the space-based LISA, is causing a transformation of large areas of classical general relativity. The difficulty of the gravitational wave measurements brings significant challenges in detection, signal extraction, and source identification. With this, problems such as the inspiral and merger of binary black holes are no longer subjects of purely academic interest. Rather, detailed understanding of these sources is essential to meet the challenges raised by the detectors, and to usher in a vibrant era of discovery in gravitational wave astrophysics.

A large number of these urgent theoretical problems surrounding LIGO and LISA are computationally intensive, requiring significant innovative numerical techniques as well as resources in hardware, software, and personnel. These problems can be categorized into several types, as follows.

(i) Numerical simulations of sources of gravitational radiation aimed at providing information (e.g. theoretical waveforms) to be used in LIGO's gravitational-wave search algorithms, and in developing and scoping out search algorithms for LISA.

(ii) Accurate source-simulation software for use in extracting the astrophysical information carried by the waves, after the waves have been detected. The simulations produced with this software must be much more accurate than those used in the wave searches; while one can do fairly well in searches using somewhat crude theoretical waveforms, these crude waveforms will seriously distort the extracted information.

(iii) Simulations of the environments in which gravitational-wave sources form and evolve. These simulations are targeted to improve our a priori estimates of the sources' parameters for use in the wave searches, and to improve our estimates of the numbers of sources, and thence the distance we must look to see them and their wave strengths. These estimates are used in planning LIGO's and LISA's wave searches and in setting LISA's science requirements and setting the target sensitivity of advanced LIGO detectors.

(iv) Development of and scoping out of computationally intensive data analysis algorithms; in most cases, these computationally intensive algorithms are expected to have tight interfaces to the source simulations [items (i)—(iii)]. The actual implementation of data analysis algorithms is carried out within the LIGO and LISA Projects and is not a part of this effort.

For simplicity of prose, we shall use the phrase *source simulations* to refer collectively to all four categories of computational research, even though category (iv) is not strictly speaking the simulation of sources but instead is the scoping of algorithms that largely spring from source simulations.

A list of specific problems that are computationally intensive and urgent for LIGO and LISA is as follows:

3.1 Late, Fully Relativistic Inspiral of BH/BH and NS/BH Binaries

Gravitational waves from the inspiral of black hole (BH) and neutron star (NS) binaries carry detailed information about the binary's masses, spins, and orbit; this information will be crucial for interpreting the more interesting, final merger waves. Gravitational waves from the late inspiral phase will also carry detailed information about the nature of gravity (spacetime curvature) inside the binary, information that can be used for high-precision tests of general relativity and for observing relativistic effects never before seen (e.g., the influence of wave "tails" on radiation reaction).

Analytic, post-Newtonian computations can be used to compute the waveforms from inspiraling compact binaries, until the orbital speed reaches about 1/4 the speed of light; this orbital speed occurs when there are about $30(M/4\mu)$ cycles of inspiral waves left, where M is the total mass and μ the reduced mass. At this point, the spacetime curvature between the two bodies is so strong that the post-Newtonian expansion fails or at least becomes suspect. In addition, spin-induced orbital precession is likely to be of considerable importance (ignoring it in wave searches may reduce the event rate by a factor of two or more); this precession will severely complicate the simulations. Thus, the only way to compute reliably the gravitational waves during the late inspiral phase as a function of the binary's parameters (masses and spins) is via numerical relativity. The waveforms obtained from numerical simulations will be needed in both the searches [(i) above] and in accurately extracting the binary's parameters from observed waves [(ii) above]

Numerical simulations will be especially important in LIGO for binaries with total mass between ~ 10 solar masses and ~ 100 solar masses, since the late inspiral waves contribute a large fraction of the integrated gravitational-wave signal. Indeed, these simulations might be crucial to LIGO's first wave detections. For LISA, the waves from supermassive BH/BH binaries will be very strong (with signal to noise ratios of hundreds to tens of thousands); correspondingly, to take full advantage of the data when extracting their information, the simulations must be highly accurate. This will allow the dynamics of the late inspiral to be studied in detail. In addition, these simulations could be used to remove the dominant supermassive BH/BH signal to permit the discovery and study of much weaker sources; this can be achieved only with the aid of highly accurate simulations.

3.2 BH/BH Mergers

The final collision and merger of two black holes in a binary is our best observational window into the nonlinear dynamics of spacetime curvature. Because of the great strength and nonlinearity of the curvature and its large-amplitude dynamics, no approximation techniques can give much insight into this source; instead, we must rely almost entirely on 3+1 dimensional numerical simulations.

For LIGO, BH/BH inspirals and mergers are likely to be the strongest sources, and for total masses above about $25\,M_\odot$ ($M_\odot$ = solar mass), the merger waves are likely to constitute a large portion of the integrated signal strength. The initial LIGO's searches for these waves are seriously hampered by not having significant waveform information from the simulations; such information, when available, will likely improve the event rate for BH/BH discovery by a factor ~ 2 and could make the difference between discovery or non-discovery of waves by initial LIGO interferometers. Once these BH/BH waves are discovered, comparison with the results of simulations will play an essential role in extracting the information the waves carry.

For LISA, the supermassive BH/BH merger situation is the same as for BH/BH inspiral: the merger waves should be so strong that they will easily be found. Given the high signal to noise expected, these mergers present outstanding opportunities to study spacetime dynamics in these most extreme sources. As with the inspiral phase, simulations of the gravitational radiation emitted will be crucial to extracting specific information about the binary system and to removing these waves from the data so as to see weaker sources; to accomplish these tasks, the simulations must have very high accuracy, far higher than in the case of LIGO.

3.3 Tidal Disruption of Neutron Stars by Black Holes

LIGO has the possibility to probe the structures of neutron stars and the equation of state of neutron-star matter (at densities up to 10 times nuclear) by observing the tidal disruption (tearing apart) of neutron stars by companion black holes. This tidal disruption should occur at the end-point of inspiral in many NS/BH binaries, though in some the NS will be swallowed by the BH whole. The onset of tidal disruption is expected to occur at gravitational wave frequencies ~ 400 to ~ 1000 Hz, where LIGO's sensitivity is moderately worse than its peak sensitivity, around 100 – 200 Hz. Correspondingly, each NS/BH binary will likely be discovered around 100 – 200 Hz by its inspiral waves; the challenge, then, will be to dig the higher-frequency tidal-disruption waves out of the noise, and measure their details. This digging and measuring will rely crucially on numerical simulations of the tidal disruption. A first cut at these simulations can be carried out using Newtonian or post-Newtonian descriptions of the neutron star, and black-hole tidal gravity deduced from general relativity. However, for fully reliable signal extraction and measuring, fully relativistic (numerical relativity) simulations will be needed. By contrast with the inspiral and merger waves from BH/BH binaries where simulations are greatly needed in the initial LIGO time frame, the tidal disruption waves are likely to be sufficiently weaker as to be a target only of the advanced LIGO detectors.

3.4 NS/NS Transition from Inspiral to Plunge/Merger

NS/NS binaries, with their relatively small total masses, should emit waves at somewhat higher frequencies than NS/BH and BH/BH binaries. As a result, the NS/NS inspiral waves, in the LIGO frequency band, are emitted when the binary is not yet highly relativistic and thus can be computed reliably using analytic, post-Newtonian techniques without the use of computer-based simulations. Since the waves from the physical collision of the stars are likely to be at relatively high frequencies (above 1500 Hz), advanced LIGO interferometers have little chance to see them. However, gravitational waves emitted during the transition from inspiral to plunge/collision are estimated to lie in the band ~ 500 to ~ 1000 Hz, where advanced LIGO detectors can see them. These transition waves will carry detailed information about the neutron stars' structures (especially their radii) and about the nuclear equation of state. This information can only be extracted with the help of numerical simulations of the end of inspiral, the plunge, and the beginning of the physical collision. Some insight has already come from Newtonian and post-Newtonian simulations, but for accurate information extraction, fully relativistic simulations will be necessary. Computer-based simulations of NS/NS collisions have also the additional value of serving as a developmental tool for numerical relativity techniques that can be applied elsewhere, and for understanding scenarios for the production of gamma-ray bursts (which are an important astrophysical source to be used in triggered searches for gravitational waves).

3.5 Stellar Collapse

LIGO has a good possibility of watching two types of stellar collapse, and LISA, one. Specifically, LIGO may see the collapses of stellar cores to form neutron stars, or proto neutron stars or black holes, including some collapses that trigger supernovae and perhaps others that do not; it may also see the accretion-induced collapse of some white dwarf stars in binaries. LISA may see the collapses of supermassive stars.

The physics of these collapses may be quite complex, including, e.g., large-scale convection, significant influences of neutrino transport and magnetic fields, and dynamical or secular bar formation due to rapid rotation coupled with viscous forces and gravitational radiation reaction. With this, it is clear that three-dimensional simulations will play a key role in providing insights about the emitted waves, for use in LIGO's and LISA's wave searches and physics extraction.

3.6 The Inspiral of WD's, NS's and Small BH's into Supermassive BH's

The gravitational waves from the inspiral of a compact, stellar-mass object (white dwarf, NS, or small BH) into a supermassive BH should carry a detailed map of the spacetime curvature of the quiescent supermassive hole. These maps will be one of the most interesting outputs of LISA, and correspondingly this source is of high priority for LISA. The event rate for these inspirals and the challenge of extracting their maps are the principal input for defining the floor (minimum) of the LISA noise curve. Scoping out the event rate and map extraction for this source entails computationally intensive research on data analysis algorithms [category (iv) above] and on the environments in which gravitational-wave sources form and evolve [N-body simulations; category (iii) above]. These N-body simulations entail studying the formation and evolution of star clusters around supermassive black holes, including computations of the rates of capture of compact objects, and the statistical distribution of their orbital parameters.

Direct N-body simulations, even with the Japanese special purpose computers (GRAPE), are still limited today to unrealistically low N ($\sim 10^4$). Monte Carlo simulations can be used with up to $N \sim 10^6$–10^7 stars, but are still extremely computationally intensive, especially when the effect of binaries is included. Very few groups are currently working on these methods.

After the floor of the LISA noise curve has been specified, there will remain major computational work to enable LISA to actually detect these waves and extract their maps: Highly accurate waveforms must be computed using highly accurate orbits and the Teukolsky formalism for first-order perturbations of black holes plus nonlinear self-force computations. These waveforms must be incorporated into search and information-extraction algorithms, which themselves must be developed and optimized. The searches for these waves and extracting their information will likely be the most computationally intensive aspects of LISA data analysis, and correspondingly the development and optimization of the necessary algorithms will be very computationally intensive.

3.7 Formation and Evolution of BH/BH Binaries in Globular Clusters and other Star Clusters

For LIGO, there is a high-priority N-body computational challenge [category (iii) computational project]. The most likely source for initial LIGO is the inspiral and merger of stellar-mass BH/BH binaries. The range of total masses detectable by initial LIGO interferometers is a few to $1000\,M_\odot$. Main sequence binaries in the field (in the bulk of a galaxy) are thought to produce BH/BH binaries with total masses no larger than roughly 40 or $50\,M_\odot$. However, heavier binaries may form in globular clusters and other star clusters, by stellar-dynamical processes: individual holes form by evolution of individual stars; the holes sink to the center of the cluster via dynamical friction, find each other, and form a loosely bound binary; the binary tightens its binding due to interactions with other holes and stars and other binaries, until gravitational radiation reaction takes over and drives the binary to merge; the merged, more massive hole captures a new black-hole companion, and the process repeats. There are many pitfalls to this scenario, such as possible ejection of the BH/BH binary from the cluster due to a 3-body interaction before it has become tight enough to merge by radiation reaction. However, recent 3-body, 4-body, and N-body simulations of stellar and black-hole interactions in star clusters suggest that the above scenario may play out sufficiently successfully to provide an interesting BH/BH rate for LIGO, perhaps even for the initial LIGO interferometers.

It is of considerable importance for initial LIGO to firm up these numerical simulations and thereby provide estimates of the black-hole masses, orbital eccentricities, and event rates that LIGO might see. These estimates are needed in planning LIGO's observations, and most especially in developing search algorithms: Until now all the algorithm development has focused on BH/BH systems with total masses below about $50\,M_\odot$, with roughly equal masses, and with circular orbits. The recent simulations suggest we may need to deal with eccentric orbits, with fairly extreme mass ratios, and with masses that extend up to $1000\,M_\odot$. The comments about the difficulties of N-body simulations for galactic nuclei with central supermassive black holes in section 3.6 apply equally to these simulations of globular clusters.

3.8 Finding Signals in Confusion-Limited Noise

At low frequencies, the gravitational wave signal detected by LISA will be dominated by a superposition of signals from galactic binary stars. At higher frequencies, there are likely to be strong signals (e.g., from BH/BH mergers) that will mask interesting weaker signals, and possibly also a masking background due to the inspiral of compact bodies into supermassive BHs; see Sec. 3.6. Accordingly, there is a great need to develop methods for separating thousands of simultaneous wavetrains of diverse sorts from a single time series [category (iv) research]. Such methods will need to use both abstract statistical theory and concrete algorithms designed around particular families of waveforms of likely interest in the LISA band. The scoping out and development of these methods is likely to be computationally intensive and will interface with simulations of the masking sources and the sources of the sought waves. An example was discussed in Sec. 3.6. This research must also include developing techniques for estimating the spectra of residual confusion backgrounds of unfitted sources, in the context of the specific LISA mission architecture.

3.9 Other Source-Simulation Research

There are a number of other important source-analysis challenges, especially for LISA, that entail a mixture of non-computational and computational research. The computationally intensive components of these challenges should be included in the proposed NASA/NSF effort. We describe briefly three of these source-analysis challenges:

3.9.1 Fate of Merging Supermassive BH's in Galactic Mergers

What are the formation and merger histories of galaxies and the supermassive black holes in their nuclei? What happens (mechanisms of energy loss, changes in orbital parameters) between the radius where the black hole gravity strongly dominates ($r = 0.1GM/\sigma^2$, say, where σ^2 is the galaxy's central velocity dispersion) and the radius where gravitational radiation takes over? What are the statistical predictions (and prediction uncertainties) for the fates and the rates of successive mergers of binaries of various masses at various redshifts?

3.9.2 Tides and Mass Transfer in Short-period WD's

Many white dwarf binaries in the LISA band will have non-gravitational-wave period changes because of effects like tides and mass transfer. Astrophysical studies of these effects are needed to give a better statistical characterization of the binary source populations in the LISA band, based on astronomical data.

3.9.3 Stochastic Primordial Background Spectra

Theoretical prediction is needed of stochastic primordial background spectra due to inflation, phase transitions, brane worlds, and other sources involving new physics.

4 Personnel Needs

Gravitational-wave source simulations, like other simulation-science efforts, are carried out most efficiently by intermediate-scale research groups or collaborations — groups of, say, roughly 15 people including several professorial-level scientists, a half dozen postdoctoral-level scientists, a half dozen graduate students, and one or more personnel who create and maintain computational infrastructure. (Note: we are not implying that such a collaboration must necessarily be housed in a single location.) The budget to support such a collaboration is *a minimum* of $1 million per year in salary, fringes, and overhead.

The scale of the science needs outlined in the previous section suggests that personnel equivalent to about five such groups are needed. (Not every need identified above requires a separate research collaboration at the level of 15 people, but may be incorporated with other needs. Conversely, the BH/BH problem may require more resources because of its technical difficulty and time urgency.) Thus the personnel costs of the effort are at least $5 million per year.

5 Computational Infrastructure for Source Simulation

5.1 Requirements

Simulations of astrophysical processes relevant to gravitational-wave detectors generally require large scale computational engines. The codes to be run on these systems use significant interprocessor communication. Traditionally, computational source analysts have used large supercomputer facilities around the US to achieve the necessary compute power. While this remains a necessary method to access resources for production runs and large scale testing, there is a need for improved access (by way of better turn-around time for jobs) to these facilities. Moreover, rapid prototyping and testing would be significantly enabled by access to compute clusters with priority given to numerical relativity studies. NASA apparently does not have provisions in its ATP grants for hardware for large scale computation. Whatever mechanism is introduced to handle this effort will have to address this need. NSF has large infrastructure in supercomputing that could be used for the joint purposes of the LIGO + LISA programs, especially under a mandate for NSF/NASA cooperation.

5.2 University Based Compute Clusters

Several numerical relativity groups are acquiring and deploying small-scale computing clusters at their home institutions for use during the prototyping and early testing phase of numerical codes. Access to these machines will have immediate impact on the rate of code development. However, it is clear that the current resources are not sufficient and provision of intermediate scale clusters at various institutions around the US is desirable. Since clusters in the 32–64 node range appear optimally suited for the prototyping needs, the need for support personnel during deployment and for long-term maintenance of a cluster must be balanced against the need for resources at any single institution. One solution which reflects this point is that consortia of institutions be provided with compute clusters to be housed at a single institution. In this way, technical support staff would be located at that institution reducing the personnel overhead associated with many small computing clusters distributed around many different sites. A usage model for university based clusters supported by NASA and NSF for numerical source simulations should be developed to fairly support both small and large groups alike. This usage model could be based on technology of the computational Grid currently under development — it would require significant support.

5.3 Supercomputer Access and Grid Computing

The ability to rapidly prototype code via downscaled runs appears well-matched to the concept of the computational grid. For numerical relativity to exploit this national resource as it emerges, it is imperative that the numerical relativity community become engaged in the process of requirements definition and design. At first, this will be viewed as a distraction from the immediate tasks at hand; however the investment will provide dividends in both the intermediate and longer term. The need implicit in participating in the tiered hierarchy of the grid will be in one of several forms, depending on the type of participation. Groups making use of grid facilities and not contributing in kind to the grid would likely need to be able to pay for the use. Groups with computational resources they are willing to register and share on the grid could do so with the expectation of receiving proportionate access to much greater resources than they alone control. Given the need to mobilize the entire numerical source simulation community, this again suggests that multi-institution consortia should be assembled to develop the necessary software tools to exploit the grid for numerical relativity.

We have to emphasize, however, that much of the work in source simulation involves solution of partial differential equations (e.g., Einstein's equations or the equations of hydrodynamics). Solving these equations on many machines in parallel may produce debilitating bottlenecks because of slow communication between nodes. It is too early to tell whether grid computing will avoid this problem, or whether very large machines with fast communication will continue to be needed for production runs. It is likely that for the next 3–5 years at least, numerical relativity will still require access to the largest supercomputers in the nation.

5.4 Need for Technical Support Personnel

Alongside the obvious need for technical personnel to deploy and maintain cluster hardware, there is a further need for technical personnel whose function is porting, optimizing, and maintaining codes to best exploit the available computational resources. In all cases, the current level of support for these activities is essentially nonexistent. When the fundamental problems with scientific analysis codes have been overcome — this is already the case for many codes used to model neutron stars within certain approximation schemes — it becomes essential to optimize the codes for the architectures to be used for large-scale testing and running. This optimization process requires dedicated personnel with technical knowledge of the target platforms and a strong numerical background. A limited amount of support for these activities is already available through the supercomputing centers, but this would be inappropriate for the resource allocation model described above. Moreover, insufficient expertise is dedicated to these problems in numerical relativity.

5.5 Need for Computational Software Infrastructure Development

The software infrastructure for large simulations is generally quite similar. This means that development of a small number of tools which solve data management, parallelization, and other common problems is essential. The Cactus infrastructure initially developed at the Albert Einstein Institute is a good example of a useful tool. The development and documentation of this product has required many years of effort and has several dedicated support personnel. While it is a useful tool, it does not solve every problem. There remains a need to further develop the infrastructure provided by the Cactus team; in some cases this may involve learning from its design and implementation to develop similar, improved tools. Development, documentation, deployment, and maintenance of software libraries to provide the parallelization, data management, adaptive-mesh-refinement, etc., will be essential to the long-term improvement of the tools needed for numerical relativity. The personnel for these efforts must include significant computer science representation alongside the physicists and astrophysicists interested in source modeling.

5.6 LIGO Data Analysis Activities and the Computational grid

The technical manpower shortage in support of source analysis activities related to LIGO and LISA science mirrors a similar shortage for LIGO data analysis activities. The LIGO Scientific Collaboration (LSC; including the LIGO Laboratory) has adopted the multi-tiered model of the computational grid as it is being promoted by several U. S. government agencies, including the NSF. While this model is an ideal way to distribute resources across the community, there is an attendant need for technical support personnel to fully exploit the potential of the computational grid.

In a traditional national laboratory or computing center, there are a large number of professional and scientific staff personnel assigned to the facility in order to enable it to operate efficiently and to provide the user support that is required. Unfortunately, in the currently emerging multi-tiered grid, the hardware resources are easily provided by the NSF, but the critical personnel needed in order to enable these centers to operate have not been provided. The LSC has situations in which many \$100K's of hardware are able to be operated only by making use of undergraduate students, graduate students, and postdoctoral scholars. While this provides immediate hands-on exposure to future generations of scientists, the level of support is usually inadequate to support large numbers of users from outside the institution which houses the hardware. Moreover, the Committee is also concerned that too great a reliance on physics postdoctoral scholars and graduate students for this work could prove prejudicial to their long term careers.

This problem for the data analysis community will also arise in the source-simulation community if it follows the same grid-oriented model and is not able to acquire the adequate mix of personnel needed to successfully adopt this computing paradigm.

5.7 Hardware/Support Costs

A 64-node cluster of cutting-edge processors and the fastest commodity inter-machine communication technology costs about \$250K, or \$300K including maintenance for three years. (Each cluster should be thought of as having a three-year life-span before a major upgrade is necessary.) This gives an annualized cost of \$100K. A regional center with two such clusters and a system administrator would have an annual cost of \$300K. With one center associated with each of the five "groups" of researchers described in section 4, this gives an annual cost for hardware and direct support of \$1.5 million. Initially the centers could be phased in over two years at this rate of funding.

The cost of personnel for computational software infrastructure development was accounted for in section 4.

6 Demographics of Researchers

The following considerations underlie our recommendations:

- As mentioned earlier, gravitational-wave source simulations are carried out most efficiently by groups of roughly 15 people. There currently is only one source-simulation group in the world with this critical size: the group at the Albert Einstein Institute in Golm, Germany; and it is being stripped subcritical this year by raiding from other institutions. All the American groups and collaborations are substantially subcritical.

- In the mid 1990s NSF funded a “Binary Black Hole Grand Challenge Project” which made substantial progress in developing the numerical-relativity tools needed for source simulations. Many of the outstanding young numerical relativists who were trained in this Project have left the field, mostly going into industry. This was triggered in part by NSF’s cutback on funding of the field at the end of the Grand Challenge and in part by a paucity of tenure track opportunities. As a result, until about one year ago the gravitational-wave source simulation community in the US was shrinking. This shriveling has led today to an extreme dearth of source-simulation graduate students, especially in the US.

- In the past 18 months there has been a radical change in the status of source-simulation research. A number of institutions have begun to recognize the need for and promise of source simulations. As one consequence, tenured or tenure-track appointments have been made to source-simulation scientists by the following institutions in the past 18 months: Bowdoin College, Goddard Spaceflight Center, MIT, U. Minnesota, Penn State, LSU (2 appointments), U. Texas Brownsville, UNAM (Mexico), and U. Tokyo (Japan). In addition, offers have been made but declined or faculty searches are underway at several other institutions, including Caltech and U. Washington. By contrast, the rate of new faculty appointments in this field during the preceding two decades was roughly one every three to five years.

- There is a dearth of talented source-simulation postdocs to take advantage of these sudden faculty opportunities and associated postdoctoral opportunities. As a result, 4 of the 10 recent faculty appointments have been made by raiding professorial faculty at other institutions.

From these considerations we conclude that: (i) Universities in the US and abroad are providing the faculty positions that should underly our proposed expansion of source-simulation research. (ii) However the needs outlined here are met, the effort should be configured in such a way as to foster the training of new people in this field — both graduate students and more senior researchers who come in from adjacent fields. (iii) The effort should also foster the development of critical sized groups and collaborations, which include people who create and maintain computational science infrastructure, as well as numerical-relativity, numerical astrophysics, and applied mathematics researchers.

6.1 Fate of Researchers Who Leave the Field

Young scientists trained in gravitational-wave source simulation are much sought after and highly prized in the financial industry (e.g., JPMorgan/Chase), in the computer, software, and other technical industries, in the national laboratories, and in academia. This is because of the power of the research tools they learn as physicists and simulation scientists: particular ways of identifying and scoping out problems and organizing efforts to solve them, as well as the specific mathematical and simulation techniques that are the every-day tools of a simulation scientist. The committee has anecdotal evidence on what has happened to about ten young researchers who have left the field over the last five or so years. The majority left academia entirely, going to work in the financial industry or at various software companies; a few left to work in other scientific areas of academia. Some left eagerly, attracted by intellectual challenges in new fields and/or by high salaries; others left reluctantly, due to the paucity of career opportunities in source simulation (a situation now changed; see section 6). In *all* cases, their current employment makes significant use of the skills they learned as source-simulation scientists.

It may be a struggle, in the first few years of the proposed effort, to attract enough talented young scientists into this field and hold them there in the face of other opportunities, so as to make the effort successful. However, we are optimistic that the the intellectual excitement of LIGO and LISA, and the intellectual challenge of source-simulation research will be sufficiently powerful attractants. Once the effort is in full swing (after about 3 years, we estimate), the lure of outside opportunities will be a positive thing, permitting a healthy but not excessive flow-through of new young talent, from a strong pool of graduate students. If, in the decade of the 2010s, the needs of LIGO, LISA, and their partners and successors dictate a scaling back of source-simulation research, there will almost certainly be ample opportunities for the scientists involved to move into other areas of research, as they did in the 1990s; and as in the 1990s, they will likely become highly prized contributors to our nation's and the world's technical and economic infrastructure.

Committee Members

Saul Teukolsky, Chair
Patrick Brady
Joan Centrella
Pablo Laguna
Albert Lazzarini
Fred Rasio
Kip Thorne
Barry Barish, ex officio
Tom Prince, ex officio

April 8, 2002

Appendix:
Charge to the Gravitational Wave Computation Task Group

Introduction:

The Physics Division of the NSF and the Astronomy & Physics Division of NASA's Office of Space Science are interested in determining the resources (both human and computational) needed to resolve those theoretical[1] issues critical to LIGO and LISA.

LIGO is about to go on line, and LISA, through the LISA International Science Team (LIST), is currently determining its baseline science requirements. Both programs require reliable templates of gravitational waveforms, analysis techniques to extract individual source signals from a large background, and theoretical estimates of GW event rates.

Short-term Task:

By early April, we would like to receive a preliminary report that answers the following questions:

- What are the most urgent needs for LIGO and LISA in the area of theory involving large-scale computing? In what time frame?
- How does the current state-of-the-art in these theoretical areas compare to the needs?
- What are the major science products that the LIGO/LISA GW community needs?
- What changes in the demographics of the numerical relativity, data analysis, and source astrophysics groups will be needed to meet LIGO and LISA needs over the next 5–10 years?
- What, if any, additional resources will be required to support this (growing) community?
- What additional computational infrastructure will be needed? How should this infrastructure be configured (Beowulf clusters, single mainframes, GRID protocols, etc)?
- Not all PHD students will remain in academia. What is the general value of training in this area? How does the production of significant numbers of PhDs with computational expertise benefit the nation?
- Other issues the Task Group believes to be important to this report.

Long-term Task:

To provide a more detailed report on the above issues.

---

[1] "Theoretical" here should be interpreted broadly to include, e.g., algorithms for signal extraction.