A. Summary

The goal of this work is to develop computational methods for large-scale correlated electronic structure calculations which provide better performance than the traditional methods and effectively utilize massively parallel computers. The resolution of the identity (RI) and orbital invariant (OI) approaches are both very promising for different types of chemical problems, but both require further development before they can be considered as viable alternatives to traditional correlated calculations in the general case.

This proposal describes a research program that will carry out the necessary developments, using second-order many-body perturbation theory (MP2) as the initial context in which to develop the RI and OI approaches:

- Understand the requirements on, and the behavior of, the RI fitting basis, developing a strategy to derive suitable fitting sets automatically from the user's chosen AO basis.
- Examine the sparsity of electronic interactions in large systems, apropos of both the OI approach and other methods which depend on locality.
- Incorporate sparsity effectively into the Global Array programming model, in accord with the "experimental" findings above.
- Develop and implement analytic gradients for OI- and RI-MP2.
- Apply these methods to problems in the areas of metal binding to 18-crown-6 ethers, and electronic materials.

Since the primary focus of this proposal is the development of new *methodology* for the treatment of large molecules, the impact of this work will extend well beyond this particular project. These tools, which will be available as part of the NWChem computational chemistry package, can and will be applied to a broad range of chemical problems, of course. But the ideas behind them also have application in other, higher level correlated methods. Similarly, the specific knowledge gained about the behavior of electronic interactions in large systems will benefit other methods under development which rely on locality, including attempts to apply "linear scaling" techniques to correlated methods.

In that goal of this research program is to push the boundaries of large-scale calculations, it is also important to take full advantage of the capabilities of high-performance computers, which means parallel implementations. This work will use the NWChem parallel computational chemistry code, and the Global Array programming model as its base. Building on the substantial experience with parallel correlated electronic structure techniques embodied in NWChem, parallelism can be incorporated into this work with almost no additional effort or risk. The benefits of this work will include enhancements to the public domain Global Array Toolkit, which is widely used in the chemistry community and elsewhere, as well as the practical experience, on which others will be able to draw to further the incorporation of parallel computing into the standard "tool box" of our field.

Finally, this work includes demonstrations the effectiveness of these approaches by applying them to problems of environmental and materials interest, such as the design of organic light emiting diodes and the binding of metal cations to crown ethers.

B. Table of Contents

Replace with NSF Form 1359.

C. Project Description

C.1 Introduction

Computational chemistry has, in recent years, been transformed into a virtually indispensable tool used by a large cross-section of the discipline. The ability to model "real world" systems with the necessary sophistication to obtain chemically meaningful results has helped give rise to a remarkable level of synergy between experimental and theoretical treatments of chemical problems. This, in turn, has fueled interest in further expanding the role of computational chemistry to even larger, more sophisticated, and more demanding simulations. In order to meet these challenges, we must insure our ability to effectively utilize cutting edge high-performance computer systems, and simultaneously look for new approaches which reduce the computational cost of these simulations.

As the combination of hardware and software improvements enable ever larger simulations, it has become clear that we must also examine the traditional electronic structure methods used in these calculations with an eye to improving their performance. Somewhat ironically, while the cost of electronic structure calculations increases (often quite rapidly) as the chemical systems increase in size, the opportunities for methodological improvement also increase – as the number of interactions (represented by integrals, excitation amplitudes, etc.) increase, the importance of any individual contribution decreases, and of course the magnitude of these interactions fall off with distance as well.

In density functional theory (DFT) and Hartree-Fock self-consistent field (SCF) methodology, there has been a great deal of excitement recently over so-called "linear scaling" techniques such as those based on the Fast Multipole Method [13–16, 36, 52, 67–70, 82, 86, 94–98], which promise calculations at a cost which depends linearly on the size of the basis set (N) rather than the formal $O(N^3)$ or $O(N^4)$ for these methods. While DFT and SCF methods are quite popular, higher level (correlated) methods are crucial in providing reliable results in many chemical applications. Such methods, including many-body perturbation theory, coupled cluster, and configuration interaction, scale formally as $O(N^5)$, $O(N^6)$, and higher as the sophistication and complexity of the method increase. So far, correlated methods have received relatively little attention aimed at improving their performance, no doubt due, at least in part, to their complexity relative to SCF and DFT. Although multipole methods and other approaches now being developed for SCF and DFT will undoubtedly be applied to correlated methods, their utility in the more complex methods is unclear at present. Moreover, correlated methods offer both challenges and opportunities which the linear scaling techniques do not address. It is therefore important to examine other promising techniques which can provide alternatives to ideas imported from SCF and DFT usage, or better yet work with them in a complementary fashion to provide additional performance improvements. This proposal considers two approaches to correlated electronic structure calculations which are well-suited for large-scale calculations, in a sense, "taking advantage" of certain features of large chemical systems to improve performance: the "resolution of the identity" integral approximation (RI), and the orbital-invariant formulation (OI).

When considering methods appropriate to large-scale correlated calculations, it is also appropriate to consider how such calculations will be performed. Pushing the boundaries of problem size usually entails the use of high-performance computers. Despite the seemingly constant introductions of faster and faster CPUs, the performance of "real" applications not keeping pace – performance is in practice often limited by the speed of memory, bus, or I/O rather than by the CPU. This is especially true of electronic structure calculations, which make intensive use of *all* resources, not just CPU. Consequently, virtually all modern high-performance computers employ multiple processors to provide improved performance, and this trend will continue. In fact, parallelism is not limited to high-end machines. In the product lines of most computer vendors, one can find multi-processor systems suitable for research groups, or even as individual desktop machines. There is enough accumulated experience to show that automatic parallelization by compilers, and related approaches are not sufficient to produce computational chemistry software capable of effectively exploiting these architectures, but at the same time, with appropriate algorithms, it is possible to drive them quite efficiently. Explicit consideration of parallel computing is, therefore, necessary in order to achieve the largest calculations and beneficial even for problems to be tackled with the local computational resources of a research group or individual.

The resolution of the identity approximation and the orbital-invariant formulation addressed in this proposal can be applied to a wide range of correlated methods, and some preliminary work has been done with both approaches in methods such as second- and fourth-order perturbation theory, coupled cluster, and configuration interaction. However neither approach is currently considered a generally applicable alternative to the traditional formulation of these methods for a variety of reasons. A key goal of my research program is to to advance the state of the art in large-scale correlated calculations, through the development of novel approaches such as RI and OI, as well as by furthering the practical use of parallel computing. As a vigorous first step towards that goal, this proposal considers a deeper investigation of the resolution of the identity approximation and the orbital invariant formulation in the context of second-order many-body perturbation theory (abbreviated MBPT(2) or MP2). This builds upon and extends my previous work with RI-MP2 and OI-MP2, and will provide valuable insights which will greatly facilitate extending the proposed work to higher-level correlated methods.

C.2 Second-Order Many-Body Perturbation Theory

MP2 is the simplest and computationally least expensive *ab initio* electronic structure method which includes the effects of dynamic electron correlation. In many cases the MP2 method provides acceptable results for properties for which correlation is important, including reaction energetics, transition states, and electronic and vibrational spectra. Where MP2 itself is not sufficient, it provides an inexpensive starting point which can be improved through the use of higher levels of perturbation theory, coupled cluster, or configuration interaction approaches. As a consequence of its computational and scientific efficiency, MP2 is the most commonly used *ab initio* method after Hartree-Fock. In the context of this proposal, MP2 provides the most readily accessible framework in which to refine the RI and OI approaches, while sharing many characteristics with higher-level methods so that the proposed work has ramifications well beyond just MP2.

The traditional form of the MP2 energy expression (in spin-orbital form, following the usual notation[1]) can be given as,

$$E^{(2)} = \frac{1}{2} \sum_{i,j,a,b} \frac{(ia|jb)[(ia|jb) - (ib|ja)]}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}. \tag{C.1}$$

The energy evaluation itself has a formal $O(N^4)$ scaling, but the four-index transformation required to produce the molecular orbital (MO) integrals from the atomic orbital (AO) integrals scales as $O(N^5)$.

For large systems which are spatially rather extended and use modest atomic basis sets, one intuitively expects that distant parts of the molecule do not interact strongly, leading to sparsity in the two-electron integrals. Empirical observation [1, 19] and more rigorous theoretical analysis [25] suggest that the number of non-negligible two-electron integrals over Gaussian-like atomic orbitals (AOs) has an asymptotic dependence of $O(N^2(\ln N)^2)$ compared to the formal $O(N^4)$. For such systems, an approach which can take advantage of the locality of interactions, such as "orbital-invariant" formulations (OI-MP2 when applied to the MP2 method) has the potential to reduce the scaling significantly. At the other end of the spectrum are spatially compact systems with extensive atomic basis sets. Both factors mitigate against any useful sparsity in such cases. But an approach such as the "resolution of the identity" integral approximation which takes advantage of the near-redundancy in the basis set to simplify integral evaluation, offers significant performance improvements.

It has been estimated that using a large, modern parallel computer (512 node IBM SP) and an efficient parallel code, it will be possible to perform exact MP2 gradient calculations with up to 1200 basis functions (C_1 symmetry) before multi-passing through the integrals is necessary [39] (slightly larger for MP2 energies only). By targeting both ends of the spectrum of chemical problems, the novel methods discussed in this proposal will enable calculations on a broad variety of systems which would be considered beyond the reach of traditional methods. In the spatially compact/large basis regime are "benchmark" calculations, where large basis sets are applied to obtain high-quality results with a given method, including extrapolations to the complete basis set limit. Computational thermodynamics and other applications which require accurate reaction energies would fall into this category. On the spatially extended/modest basis end of the spectrum are, for example, a wide range of systems of interest from the materials standpoint–for molecular electronics, optical properties, nanomaterials, and numerous other applications. Precisely where along this spectrum the RI and OI approaches are applicable is not understood at present, and this is one of the questions the proposed work addresses. However there is reason to believe that the range of applicability of both approaches is large, and perhaps partially overlapping.

In earlier research, Robert Harrison and I demonstrated parallel implementations of both of these methods for energy evaluation [8, 9]. This proposal builds on that earlier work to develop RI-MP2 and OI-MP2 to the point at which they can be put to routine use as serious alternatives to traditional MP2 calculations for large-scale calculations, especially those requiring high-performance (parallel) computers. After introducing the methods in

[1] $i, j, \ldots m$ represent occupied molecular orbitals, $a, b, \ldots e$ virtuals, and the ϵ_p are the canonical SCF eigenvalues.

more detail, I will describe the necessary extensions to that previous work which constitute the heart of this proposal.

C.3 The Resolution of the Identity Approximation to MP2

The so-called "resolution of the identity" approach [32, 51, 90] has been applied to SCF [34, 90], MCSCF [87–89], DFT [27–29], MP2 [9, 31, 32, 93], and coupled cluster methods [76]. It is very closely related to the fitting of the density commonly used in density functional theory [22, 60] and to the multiplicative integral approximation of Van Alsenoy [3]. The essence of the RI approach is the replacement of pairwise products of the atomic orbital (AO) basis functions, $\{\chi_p\chi_q\}$, by linear combinations of a "fitting" basis, $\{\alpha_t\}$. If $\{\alpha_t\}$ is the Gaussian Product Theorem (GPT) result of the AO basis $\{\chi_p\}$ with itself, the substitution is exact, while smaller fitting bases result in approximations. The success of this approach depends on the observation that the AO product space is usually nearly linearly dependent and can be spanned (or nearly so) by a smaller set of functions [5, 66]. Such an approximation should work particularly well for calculations involving large atomic basis sets, which provide more redundancy in the AO product space and at the same time would provide little sparsity that might be exploited in other approaches to reduce the computational effort.

When applied to the familiar the four-center two-electron repulsion integrals (ERIs),

$$(pq|rs) = \int \chi_p(r_1)\chi_q(r_1)\frac{1}{|r_1 - r_2|}\chi_r(r_2)\chi_s(r_2)d^3r_1d^3r_2, \qquad \text{(C.2)}$$

the RI approximation results in combinations of two- and three-center one- and two-electron integrals. Several criteria can be chosen for optimizing the fitting basis representation of the AO product space, resulting in different expansions of the four-center ERIs [90]. The "V approximation"

$$(pq|rs) = \sum_{t,u}(pq|t)V_{tu}^{-1}(u|rs), \qquad \text{(C.3)}$$

where V_{tu} is the ERI $(t|u)$, is representative and currently seems to be the fit of choice based on results of Vahtras and coworkers, and similarity to the variational density fitting procedure widely used in DFT. This choice can be thought of as a least squares fit of the electrostatic potential of the product space density as a linear combination of the electrostatic potential arising from the fitting density.

When the RI approximation is applied to MP2, the working equations are obtained by the straightforward insertion of Eq. C.3 into the MP2 energy expression (Eq. C.1), resulting in the equation

$$E^{(2)} = \frac{1}{2}\sum_{i,j,a,b,t'}\frac{(ia|t')(t'|jb)[(ia|t')(t'|jb) - (ib|t')(t'|ja)]}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}, \qquad \text{(C.4)}$$

where for convenience, the V^{-1} term has been subsumed symmetrically into the three-center integrals, following Rendell and Lee [76].

At the heart of RI's advantage is the is the reduced volume of integrals to be calculated and manipulated–down from $O(N^4)$ to $O(N^3)$, taking N_{fit} to be $O(N)$. The transformation of integrals from the AO to MO basis is thus $O(N^4)$, while a new $O(N^5)$ step is introduced, forming approximate four-center integrals from products of three-center integrals using Eq. C.3, which is a large matrix multiplication. When the RI approximation is applied to more complex methods, intermediates are typically produced with fewer indices than in the exact equations (i.e. three instead of four indices), and later contracted or combined to produce approximations of exact quantities, in analogy with the RI-MP2 case, but with the additional advantage that smaller intermediates are carried through substantial parts of the method.

Even with relatively small chemical systems (4 atoms, 140 basis functions, 396 fitting functions) RI-MP2 calculations have shown as much as a 12.4× performance improvement over exact MP2 on a single processor workstation [10]. When the parallel implementation was applied to the K^+:12-crown-4 complex (461 AO and 1227 fitting basis functions), the RI-MP2 calculation was approximately 17 times faster than the exact MP2 [9], while accurately reproducing the exact MP2 binding energy. Although a sizable system for an exact MP2 calculation, the parallel speedup curves shown in Fig. C.1 suggest that as an RI-MP2 calculation, it is in fact somewhat small. The distinctive features of the RI-MP2 method are well suited to the strengths of modern parallel processors. Distributed matrix multiplication [17, 47, 91] performs much more efficiently than most parallel four-index transformation algorithms which have been developed [4, 18, 20, 24, 38, 46, 54, 84, 92, 99–102], and the reduced number of integrals means disk-based implementations can be used much more readily than for the exact MP2.

Figure C.1: Parallel speedup for the K^+:12-crown-4 RI-MP2 energy on three different MPP platforms. Error bars resulting from the uncertainty associated with the load on the remainder of the machine were obtained from analysis of repeated runs of the same job. Error bars for the Cray-T3D data are too small to be distinguished at this scale.

C.4 The Orbital-Invariant Formulation of MP2

While the resolution of the identity approximation is well-suited to problems in which sizable AO basis sets are applied to molecules which are spatially compact, problems at the other end of the spectrum are equally interesting: those involving modest AO basis sets on large, spatially extended systems. As noted above, in such cases one expects substantial sparsity in the interactions between distant parts of the molecule. It is well known that screening AO integral blocks prior to evaluation will provide some of the benefits of this sparsity, but the basis in which the interactions are represented can play an important role in determining how much of this reduction will be present in practical calculations. Canonical Hartree-Fock MOs, normally used to formulate correlated methods, tend to be delocalized. This means that even if there is significant sparsity in the AO basis representation of the integrals, it will be "filled in" by the delocalization of the MOs, and the number of non-negligible integrals will not approach the asymptotic limit of $O(N^2(\ln N)^2)$.

An orbital-invariant formulation is more general because it allows the use of localized MOs (from Edmiston-Ruedenberg [26], Foster-Boys [12], or Pipek-Mezey [71] procedures, for example) as occupied orbitals (or even AOs projected into the occupied space). Following the framework of the Method of Self-Consistent Electron Pairs (SCEP) [59, 74], the virtual orbitals may in fact be represented in any (non-orthogonal) basis that spans the virtual space. Thus instead of HF molecular orbitals, which do not localize very well, it is possible to use the AO basis itself, greatly improving the spatial locality of the individual functions spanning the virtual space and thereby improving sparsity. Orbital invariant formulations have been reported for a number of perturbation theory, coupled cluster, and configuration interaction-based energies [37, 59, 74, 80], but the approach cannot not be described as being widely used.

Rather than repeat the detailed derivation of the orbital invariant MP2 (OI-MP2) equations, which have already been published [8, 73], I will briefly sketch the relationship between OI and traditional formulations. The MP2 energy of Eq. C.1 can also be written with the excitation amplitudes, t_{ij}^{ab} made explicit, as Note that the SCF eigenvalues ϵ_p of Eq. C.1 have been replaced with the equivalent notation f_{pp}, denoting the diagonal elements of the MO basis Fock matrix.

To accommodate localized orbitals, which do not diagonalize the occupied-occupied (o-o) and virtual-virtual (v-v) blocks of the Fock matrix, the amplitude equation above must include all elements of the o-o and v-v blocks of the Fock matrix instead of just the diagonals (the energy equation is unaffected),

$$0 = r_{ij}^{ab} = (ia|jb) - (ib|ja) + \sum_e (t_{ij}^{ae} f_{be} + f_{ae} t_{ij}^{eb}) - \sum_m (t_{im}^{ab} f_{mj} + f_{im} t_{mj}^{ab}), \forall i, j, a, b. \quad \text{(C.5)}$$

The residual r_{ij}^{ab} is introduced at this stage as a reminder that in this more general form, the equation must be solved iteratively for t, with the residual measuring "how far" the current t is from the final solution.

The equations are now recast in the form of coupled matrix equations. The amplitudes are collected into a set of matrices, $\{\mathbf{t}_{ij}\}$ which are labeled by virtual orbitals a and b in their row and column dimensions. Similarly, the integrals are collected as a set of exchange operators, $\mathbf{K}_{ij}^{ab} = (ia|jb)$. The MP2 energy is now represented as

$$E^{(2)} = \frac{1}{4} \sum_{i,j} (\mathbf{t}_{ij} \cdot (\mathbf{K}_{ij} - \mathbf{K}_{ij}^{\dagger})), \quad \text{(C.6)}$$

and the amplitude equation becomes

$$0 = \mathbf{r}_{ij} = \mathbf{K}_{ij} - \mathbf{K}_{ij}^{\dagger} + \mathbf{F}\mathbf{t}_{ij} + \mathbf{t}_{ij}\mathbf{F} - \sum_m f_{im} \mathbf{t}_{mj} - \sum_m \mathbf{t}_{im} f_{mj}, \forall i, j. \quad \text{(C.7)}$$

Any transformation $\mathbf{C}$ of the MOs which satisfies the orthonormality constraint $\mathbf{C}^{\dagger}\mathbf{S}\mathbf{C} = \mathbf{1}$ for some metric $\mathbf{S}$ may be used to transform the excitation space. Orthogonality of the occupied and excitation spaces can be guaranteed by appropriate use of projection operators during solution of the amplitude equations [59]. This results in the most general form of the MBPT(2) equations, In this form, the occupied orbitals can be localized (noncanonical) and the virtual space can be any appropriate set of orbitals. The AO basis itself provides a very localized representation which can be used rather than localized virtual orbitals, which are often hard to obtain.

The orbital invariant formulation makes it possible to retain a great deal more sparsity than is possible using the traditional formulation. This lays the groundwork for approaches that can take advantage of the localization and sparsity. One such idea is the "local correlation" approximation, originated by Pulay and Saebø, which extends the localization concept to the excitation amplitudes by neglecting excitations which go beyond some local neighborhood of the originating orbital. Although relatively little work has been done based on the orbital-invariant approach, it should yield an MP2 procedure, for example, with an $O(N^3)$ scaling compared to the traditional approach's $O(N^5)$. Saebø and coworkers have reported good performance for the several applications of this approach [78, 81, 85] using their sequential implementation. Very recently, Maslen and Head-Gordon [58] have given another derivation of the orbital-invariant MP2 which leads naturally to a local correlation ansatz which is slightly different from that used by Pulay and Saebø, though the concept is very much the same.

C.5 Relationship to Other Methods

Methods specifically targeted at large-scale correlated calculations have not yet received the kind of attention that has been given to reducing the scaling of SCF and DFT calculations in recent years. The orbital invariant and resolution of the identity approaches represent rather different ideas from those used in linear scaling methods, and I anticipate that as the "linear scaling revolution" is extended to correlated methods, these various approaches will be combined in synergistic fashion to provide even greater benefit.

When considering the spatially compact/large basis set regime, where sparsity is negligible, there do not seem to be any alternatives to the RI-MP2 approach at this time. When sparsity does come into play, there are several approaches have been put forward.

The Laplace Transformation (LT) approach, proposed by Almlöf and Häser [2, 40, 41, 75] allows the use of localized orbitals in a fashion similar to the OI-MP2 equations. Although I believe the approach to be very promising for higher-level correlated methods (where it allows high-complexity terms to be separated into less expensive components, thus reducing scaling), it suffers from several deficiencies at the MP2 level when compared with an orbital invariant formulation.

- The LT approach is approximate in the sense that it relies on a numerical quadrature, while the OI-MP2 is exact. (The quadrature can be evaluated to matching accuracy, but this can be quite expensive.)

- As published, LT-MP2 lacks the flexibility in representing the virtual space present in OI-MP2. The limitation to localized HF virtual orbitals would significantly reduce the sparsity which can be obtained.

- The LT-MP2's numerical quadrature requires multiple integral evaluation and transformation steps, usually the least efficient steps in a parallel code suitable for such large-scale calculations. By contrast, the OI-MP2 only requires a partial transformation, which can be much more efficient, and can utilize memory and disk to reduce or eliminate the need for integral recomputation.

As a result, I believe that the OI-MP2 approach is better suited for application to spatially extended/modest basis calculations.

Pseudospectral approximations have been applied to a number of methods ranging from SCF to full-CI including MP2 and MP3 [53, 55, 56, 77], and might be considered as yet another approach to large-scale calculations. Pseudospectral approximations combine the usual atomic orbital basis with a numerical grid, expanding the Coulomb potential due to AO basis function products onto the grid. Use of a physical grid simplifies taking advantage of the locality of electronic interactions. However it must be pointed out that pseudospectral MP2 relies in essentially the same formulation as presented here to provide localized orbitals which gives the pseudospectral approach its performance advantage. Essentially, pseudospectral MP2 is an integral approximation applied to OI-MP2, and the effort described in this proposal will apply equally with or without the additional approximation.

C.6 The Role of High-Performance Computing

These methods are designed push back the size limitations of correlated calculations by "taking advantage" of certain features of large chemical systems. This naturally suggests that high-performance computers should be taken into account during their development. As noted earlier, the trend in computing is such that "high-performance" must include parallel computers, and increasingly, parallel computing extends all the way down to the desktop.

The NWChem parallel computational chemistry package [42], developed by the Environmental Molecular Sciences Laboratory (EMSL) at the Pacific Northwest National Laboratory provides an excellent framework in which to perform the work proposed here. It is one of the few packages in existence which was designed with the specific intention of providing fully-distributed highly-scalable computational chemistry tools. The Global Array programming model [63–65], which underlies NWChem, provides much of the ease-of-programming common to shared-memory-only programming models, while retaining the generality of the message-passing paradigm, thus greatly simplifying parallel software development. In addition, NWChem embodies a great deal of experience in developing parallel MP2 codes in particular using this model. It presently includes independent fully-direct [102] and semi-direct [21] implementations of exact MP2, the latter including gradients, as well as the parallel RI-MP2 energy [9]; the first parallel implementation of the OI-MP2 method [8] served as a prototype for the design of NWChem and is being reimplemented within the package.

Consequently, the proposed work can be carried out incorporating parallelism with almost no additional effort or risk. The benefits of such an approach, however, extend well beyond this proposal, in that they will lead to further improvements to the Global Array toolkit, and provide additional examples and experience to help the computational chemistry community assimilate parallel computing, now available from desktop to high-end machines, into the standard "tool box".

As a continuing member of the NWChem development team, I have the full support and cooperation of the EMSL for the proposed work, including the use of NWChem (see supporting letter in Section I). Distribution of NWChem is managed by the EMSL, and the intent is to make the package widely available for use on parallel supercomputers. In addition, certain components of NWChem, such as the Global Array Toolkit [63–65] and PeIGS parallel eigensolver [30] have been placed in the public domain and are freely distributed as stand-alone packages to facilitate their use in other applications.

C.7 Proposed Work

Both OI- and RI-MP2 energies have been implemented already and their promise has been demonstrated, both in terms of parallel scalability and improved performance compared to traditional MP2 calculations. The goal of this project is to develop both methods to the point at which they are sufficiently well understood, and provide the capabilities necessary, to offer a useful (more efficient) alternative to traditional MP2 calculations on large systems. In order to accomplish this goal, a number of different issues must be addressed.

- For the RI-MP2 method, better understanding of the fitting basis set is required, so that reliable results can be obtained without a large number of expensive "calibration" runs.

- For the OI-MP2 method, effective use of sparsity is crucial, but little is known about how sparsity grows in for real chemical systems or how to design parallel algorithms to use it efficiently.

- For any routinely used method, analytic gradients and other first-order properties are a practical necessity. They allow efficient optimization of structures and can be differentiated numerically to produce vibrational frequencies.

C.7.1 RI-MP2 Fitting Basis Sets

In addition to some work towards fitting basis sets for use in RI-DFT [27–29] and RI-SCF and -MCSCF [87–89] calculations, there has been some preliminary work on RI-MP2 fitting sets. Weigend and Häser [93] have employed fitting basis sets created without any optimization in conjunction with the SVP AO basis [83] for a series of calculations of reaction energies. A more detailed study, partly carried out in collaboration with Robert Harrison, has lead to RI-MP2 fitting basis sets [6, 10] for use in conjunction with the correlation consistent AO basis sets [23, 50, 103, 104] cc-pVDZ, cc-pVTZ, and aug-cc-pVTZ. The "-fit2-1" fitting sets (i.e. cc-pVDZ-fit2-1 and aug-cc-pVTZ-fit2-1), are designed to reproduce a set of 32 test cases involving molecules comprised of zeroth- and first-row atoms to better than 1% on average and a maximum of 2% error in any individual test case. These tests were chosen to represent a range of valance electron properties, including reaction energies involving single and multiple bonds, hydrogen bonding, van der Waals

Table C.1: Performance RI-MP2 calculations with the "-fit2-1" fitting basis sets. "Reactions" refer to the error (compared to exact MP2) results for the 32 test cases covering a range of interactions an properties. "Total Energies" refer to the error (compared to exact MP2) of the 55 separate calculations which contribute to the test cases.

		AO and Fitting Basis Set		
		cc-pVDZ cc-pVDZ-fit2-1	cc-pVTZ cc-pVTZ-fit2-1	aug-cc-pVTZ aug-cc-pVTZ-fit2-1
AO Descr.	H-He (size)	$2s1p$ (5)	$3s2p1d$ (15)	$4s3p2d$ (25)
	Li-Ne (size)	$3s2p1d$ (15)	$4s3p2d1f$ (35)	$5s4p3d2f$ (55)
Fit Descr.	H-He (size)	$9s6p4d$ (51)	$7s6p5d4f$ (95)	$7s6p5d4f$ (95)
	Li-Ne (size)	$10s5p4d1f$ (59)	$11s6p4d3f1g$ (98)	$11s6p6d3f2g$ (125)
Reactions	Avg. Error (%)	0.13	0.06	0.11
	Max. Error (%)	1.3	0.4	0.5
Total Energies	Min. Error (μE_h)	0.3	< 0.1	< 0.1
	Avg. Error (mE_h)	0.2	0.3	0.3
	Max. Error (mE_h)	0.6	1.5	1.4

interactions, and electrostatic properties. Table C.1 briefly describes the fitting sets and their numerical performance.

The practical experience of this work agrees with the theoretical expectation that as the AO basis increases in size, the AO product space grows increasingly redundant, and therefore the fitting basis required to fit it adequately is smaller in relative terms. This helps to alleviate concerns that calculations involving large fitting basis sets will suffer from linear dependencies in the fitting basis itself. Small eigenvalues were observed in the aug-cc-pVTZ case, but were easily handled without effecting the quality of the calculation. While our work thus far has not addressed the use of this method on potential energy surfaces, preliminary calculations by Weigend and Häser [93] are favorable.

The next logical step in the development of RI-MP2 fitting basis sets is to develop a more detailed and systematic understanding of the role of the fitting basis–both where errors arise and how to improve the approximation in a cost-effective manner. This will involve direct comparison of exact and approximate integrals, categorization according to the number of unique atomic centers involved in the integrals, and numerical optimization procedures as well as theoretical examination of the approximations targeted at bounding their errors.

This work will consider not only energies and finite field property calculations, but also gradients, potential energy surfaces, and a broader range of other properties in conjunction with the availability of relaxed density matrices (as a by-product of the gradient implementation).

In addition, we will investigate the effectiveness of ideas for constructing fitting basis sets beyond the usual atom-centered approach. On the one hand, it may be possible to draw on the work of Beebe and Linderberg [5] and O'Neal and Simons [66] to develop numerically-oriented approaches to the creation of fitting sets. On the other hand we will consider the use of off-atom fitting functions, and more physically-based approaches – for example, fitting each product using only a local neighborhood of functions, which would effectively lead to a fitting basis of constant size, thus eliminating a factor of N_{fit} from the cost of the method.

The work described here on understanding and developing fitting basis sets will result both in improvements to the simple recipes, developed in the preliminary work, as well as entirely new approaches to the construction of fitting basis sets which I expect to be more accurate and more efficient. The detailed numerical analysis of the fitting procedure will also help define the error bars for this method in comparison with the exact MP2. Moreover, much of the proposed work is largely independent of the particular electronic structure method to which the RI approximation is being applied.

C.7.2 Locality and Sparsity in OI-MP2

Orbital invariant formulations of correlated methods facilitate taking advantage of the sparsity or locality of interactions in large quantum mechanical systems. Theoretically, we understand that the number of two-electron integrals and double excitation amplitudes must asymptotically approach $O(N^2)$; practically, information about how real chemical systems approach this is quite limited and spotty. Because of the importance of sparsity and localized representations in quantum chemical methods (not just at the MP2 level), this area needs to be examined systematically and in some detail.

Thus far, studies of locality have focused largely on SCF-level quantities, as in the recent work by Maslen and coworkers [57] (though this particular work does briefly consider sparsity of the MP2 density matrix as well), or have relied solely on physical arguments to enforce a locality *a priori*, as in the local correlation work of Sæbø and Pulay [72, 79, 81], and Maslen and Head-Gordon [58]. In this work, we propose to study sparsity and locality in correlated methods, starting with the MP2 level, in order to connect the physically-based approximations with a numerical understanding of the sparsity in integrals, excitation amplitudes and related quantities. The goal of this work is error bounds for the physically-based local correlation approximations, and approximations, controllable by numerical thresholds and tolerances, which can bridge between exact MP2 and the physically-based local correlation approximations. The development of such new approximations will involve controlled neglect of integrals, amplitudes, and orbital pairs, and will involve a combination of theoretical work and numerical experiments. This work will be extended to include gradients, once that capability is implemented for OI-MP2 (see Section C.7.3).

Collateral to this study of the physical and numerical aspects of sparsity are considerations of how it can be incorporated into high-performance (parallel) algorithms efficiently, so that storage requirements are minimized while performance is maximized. Sparsity is present in a wide variety of problems in computational science, though typically with somewhat different characteristics than that in chemical problems – logically, a kind of “block sparsity” fits the structure of many electronic structure applications, but this matrix structure is not commonly used in other fields.

Work will initially focus on efficient storage and manipulation of block sparse matrices, and later expand to emphasize improving algorithms for computations on such matrices (i.e.

matrix multiplication, diagonalization, etc.) and providing interfaces to other commonly used sparse data structures in order to take advantage of existing packages. This work will be done within the context of the Global Array (GA) Toolkit [63–65], the parallel programming model used in NWChem, and in close collaboration with researchers at EMSL. This allows our work to take advantage of the extensive capabilities of the GAs for dense linear algebra and facilitates rapid adoption of this work in a broad range of other applications, both within computational chemistry and in other areas, which already use GAs as their programming model. It will also be possible to use the lessons learned in this work to further the development of a general, efficient parallel programming language because of the conceptual similarity between the GAs and High Performance Fortran [43, 44], which has been accepted by the computer industry as the *de facto* standard version of Fortran for parallel computers.

C.7.3 Analytic Gradients in OI- and RI-MP2

These days, for a method to be practical for routine application to chemical problems, it must have analytic gradients and other first-order properties available. This capability allows the efficient determination of structures as well as offering the possibility of computing vibrational spectra by finite difference of gradients. Theory and implementation of traditional, exact MP2 gradients are well established and widely used. A number of parallel implementations have been developed as well [21, 33, 62]. This proposal includes the corresponding development and implementation which is required for both OI- and RI-MP2.

The theory of OI-MP2 gradients was outlined by Pulay and Saebø [73], and has been implemented in the Jaguar software package from Schrödinger, Inc. in the context of the pseudospectral approximation, but neither detailed equations nor a description of the implementation have been published [61]. I have previously derived detailed equations for the orbital-invariant form of the coupled-perturbed Hartree-Fock (CPHF) equations [7], which are required for MP2 gradients, and am currently in the process of producing detailed equations for the OI-MP2 gradient itself. There is reason to believe that OI-MP2 gradients may in fact be more efficient than the traditional formulation even without taking advantage of sparsity. This stems from the fact that in the traditional formulation, a four-index back-transformation of the two-particle density is usually performed in order to contract with the AO derivative integrals, while in the OI-MP2 framework, we can use the AO space representation of the virtual space directly which should simplify this step. Our implementation of OI-MP2 gradients will accommodate the use of UHF reference functions for calculations on open-shell systems, and will run efficiently on parallel machines.

As with the RI-MP2 energy, the RI-MP2 gradient equations can be obtained straightforwardly by inserting the RI approximation into the familiar exact gradient equations, though some care must be taken to distinguish quantities arising from the RI-MP2, which involve approximate integrals, from those arising from the SCF reference, which require exact integrals. This has been done recently by Weigend and Häser [93], who have also reported a preliminary implementation, as well as by myself [6]. Our implementation will be based on NWChem's parallel semidirect MP2 energy and gradient module [21], and will be distinguished from the work of Weigend and Häser by the ability to handle open-shell systems via a UHF reference, and extremely large systems through a parallel implementation.

C.7.4 RI-MP2 Applications

Because of the performance improvements they offer over the traditional MP2 formulation, the RI- and OI-MP2 methods make feasible calculations and entire studies which would otherwise be quite impractical. For instance, the RI-MP2 method has already demonstrated an order of magnitude performance improvement in calculations of the binding energy of K^+ to 12-crown-4 ether [9].

In order to showcase the full capabilities of the RI-MP2 method on a problem of environmental interest, I plan to extend the earlier K^+:12-crown-4 calculations [9, 31]. Crown ether systems are interesting because of their selective binding to metal cations, and have been proposed for use in separating high-level nuclear wastes such as ^{90}Sr and ^{137}Cs [45] among other applications. In order to design appropriately selective crown ether systems for specific situations, it is necessary to have a thorough understanding of the nature and origin of the selectivity. Recently it was shown that theory and experimental results for the gas-phase binding energy of the K^+ and Cs^+:18-crown-6 complexes were in marked disagreement [35, 48, 49]. This kind of disagreement must be resolved if we are to have any hope of achieving a detailed understanding of the nature of the preferential binding and ultimately of designing new chemical systems tailored to specific applications. The preferred theoretical treatment of these systems involves the use of progressively larger basis sets, allowing extrapolation to the complete basis set (CBS) limit for the level of theory applied. Because of the size of these systems, only relatively small basis set MP2 calculations can be performed, making credible extrapolation difficult. A comparison of DFT, MP2 and RI-MP2 treatments of the analogous 12-crown-4 system [31] showed the MP2 approaches were preferable to DFT, and that the RI-MP2 calculation was capable of reproducing the binding energy in a small fraction of the time required for exact MP2.

Figure C.2: Representative structures of bare 12-crown-4 and 18-crown-6 ethers and their complexes with K^+.

With a better understanding of the fitting basis and analytic gradient, as proposed here, RI-MP2 will be in a position to replace the traditional MP2, allowing investigations of these crown ether systems to be expanded to larger basis sets and larger systems. In particular, I propose to use RI-MP2 calculations to complete the CBS extrapolation of the structures and energetics of K^+:18-crown-6 (see Fig. C.2). Calculations on the Cs^+ system, utilizing relativistic effective core potentials, will be the second phase, and extension to other cations is planned for the future, as well as more complex crowns. This work will be coordinated with related activities at EMSL to insure that this work extends theirs. The key advantage of using the RI-MP2 is that it will allow calculations using larger basis sets, and on larger crowns, than are possible with the traditional methods being used by EMSL researchers.

C.7.5 OI-MP2 Applications

Photoluminescent polymers are of great interest for use in organic light emitting diodes and related applications. In these materials, the role of interchain interactions is of great importance. A recent paper by Blatchford and coworkers [11] reports for three pyridine-based polymers, the photoluminescence (PL) spectrum of samples cast as films is strongly red-shifted (~0.5 eV) relative to the solution and powder form spectra and the films show a much lower quantum yield. These effects are attributed to interchain interactions, which would be much stronger in the film than in either the solution or powder sample, and for practical applications of such polymers, which typically involve a solid-state form, need to be controlled or eliminated. Blatchford and coworkers propose a mechanism involving the formation of *aggregate states* which tend to delocalize the ground and excited states across several chains. The energy of the ground state is lowered, and the excited state is split into several distinct levels. The OI-MP2 method is well-suited to a theoretical investigation of this effect, as the systems involved are large, and the effect in question is large enough that extremely high-level (i.e. large basis set) calculations should not be required to model it. We will consider various oligomers of poly(p-pyridine) and poly(p-pyridyl vinylene), and the effect of bringing two or more oligomers into proximity on the ground and excited electronic states of the system. An important question to be answered is how many chains must interact in order to produce this effect, as the answer may contribute to the development of ways to avoid the formation of aggregate states through design and processing of the polymers.

C.8 Educational Aspects

The proposed work touches on a number of very significant areas of correlated electronic structure methods and high-performance computing. It is rare to find candidates for post-doctoral research positions with the particular combination of experience required for this project; this is not a problem, but rather an excellent opportunity for the researcher to extend their training into new areas. It is perhaps most likely that the researcher will have little or no background in parallel computing. Our experience in the High Performance Computational Chemistry group at EMSL indicates that the Global Array programming model is fairly easy to learn and use – postdoctoral researchers in that group were quickly able to make substantive contributions to the NWChem package.

In addition, my own unique position of being affiliated with both the Northeast Parallel Architectures Center and the Chemistry Department of Syracuse University (which are housed in the same building), will provide the researcher with a flourishing interdisciplinary environment including both chemists and computational scientists in other fields. In the course of my own work at NPAC, I have already gotten a number of computer science graduate students and research staff interested in problems on the computational side of computational chemistry. I also have active or nascent collaborations with a number of the Chemistry faculty, coupling theoretical and experimental work.

The nature of the proposed work and the environment here at Syracuse provide an excellent opportunity for the postdoctoral researcher to broaden their background and learn new skills, such as parallel computing, which can be applied well outside the confines of the proposed work.

C.9 Summary

This proposal describes a plan to develop the OI-MP2 and RI-MP2 methods, placing them on a solid footing as viable alternatives to traditional MP2 calculations on large chemical systems. The impact of the work, however, will extend far beyond these two methods. The resolution of the identity approximation has already been applied to a variety of methods, and is especially promising for higher-level correlated methods. The ideas of sparsity and locality associated with orbital invariant formulations are also much more widely applicable than just the MP2 method. In addition, the work of this proposal will be implemented for parallel computers, and will result in tools (i.e. NWChem and the Global Array Toolkit), algorithms, and practical experience that will benefit the entire field of computational chemistry in the course of adapting to the latest directions in high performance computing. At the same time, this proposal builds on my previous work in the areas of correlated electronic structure methods and high-performance (parallel) computing, and would provide a clear step forward in my overall research goal of making large-scale correlated electronic structure calculations more efficient and more accessible to the chemistry community.

D. Bibliography

[1] Jan Almlöf. private communication.

[2] Jan Almlöf. Elimination of energy denominators in Møller-Plesset perturbation theory by a Laplace transformation approach. *Chem. Phys. Lett.*, 181(4):319–320, 28 June 1991.

[3] C. Van Alsenoy. Ab initio calculations on large molecules: The multiplicative integral approximation. *J. Computat. Chem.*, 9(6):620–626, 1988.

[4] Charles W. Bauschlicher, Jr. An efficient two-electron integral transformation for vector-concurrent computer architectures. *Theoret. Chim. Acta*, 76:187–193, 1989.

[5] Nelson H. F. Beebe and Jan Linderberg. Simplification in the generation and transformation of two-electron integrals in molecular calculations. *Int. J. Quantum Chem.*, 12:683–705, 1977.

[6] David E. Bernholdt. to be published.

[7] David E. Bernholdt. unpublished work.

[8] David E. Bernholdt and Robert J. Harrison. Orbital invariant second-order many-body perturbation theory on parallel computers: An approach for large molecules. *J. Chem. Phys.*, 102(24):9582–9589, 22 June 1995.

[9] David E. Bernholdt and Robert J. Harrison. Large-scale correlated electronic structure calculations: The RI-MP2 method on parallel computers. *Chem. Phys. Lett.*, 250:477–484, 8 March 1996.

[10] David E. Bernholdt and Robert J. Harrison. Fitting basis sets for the RI-MP2 approximate second-order many-body perturbation theory method. *J. Chem. Phys.*, submitted.

[11] J. W. Blatchford, S. W. Jessen, L.-B. Lin, T. L. Gustafson, D.-K. Fu, H.-L. Wang, T. M. Swager, A. G. MacDiarmid, and A. J. Epstein. Photoluminescence in pyridine-based polymers: Role of aggregates. *Phys. Rev. B*, 54(13):9180–9189, 1 October 1996.

[12] S. F. Boys. Localized orbitals and localized adjustment functions. In P.-O. Löwdin, editor, *Quantum Theory of Atoms, Molecules, and the Solid State*, pages 253–262. Academic Press, New York, 1966.

[13] J. C. Burant, M. C. Strain, G. E. Scuseria, and M. J. Frisch. Analytic energy gradients for the gaussian very fast multipole method (GvFMM). *Chem. Phys. Lett.*, 248:43, 1996.

[14] J. C. Burant, M. C. Strain, G. E. Scuseria, and M. J. Frisch. Kohn-sham analytic energy second derivatives with the gaussian very fast multipole method (GvFMM). *Chem. Phys. Lett.*, 258:45–52, 1996.

[15] M. Challacombe, E. Schwegler, and J. Almlöf. Fast assembly of the Coulomb matrix: A quantum chemical tree code. *J. Chem. Phys.*, 104:4685, 1996.

[16] M. Challacombe, E. Schwegler, and J. Almlöf. Modern developments in Hartree-Fock theory: Fast methods for computing the Coulomb matrix. In J. Leczszynski, editor, *Computational Chemistry: Review of Current Trends*, pages 53–107, Singapore, 1996. World Scientific.

[17] Jaeyoung Choi, Jack J. Dongarra, and David W. Walker. PUMMA: Parallel universal matrix multiplication algorithms on distributed memory concurrent computers. Technical Report ORNL/TM-12252, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA, May 1993. URL http://www.netlib.org/scalapack/pumma.ps.

[18] E. Clementi, S. Chin, and D. Logan. Large scale computations on the loosely coupled array of processors. In M. Dupuis, editor, *Supercomputer Simulations in Chemistry*, pages 130–164. Springer-Verlag, 1986.

[19] E. Clementi, G. Corngiu, and S. Chakravorty. KGNMOL: A program for large molecules and molecular interactions. In Enrico Clementi, editor, *Modern Techniques in Computational Chemistry: MOTECC–90*, chapter Chapter 7, pages 343–434. ESCOM Science Publishers, Leiden, The Netherlands, 1990.

[20] Lawrence A. Covick and Kenneth M. Sando. Four-index transformation on distributed memory computers. *J. Computat. Chem.*, 11(10):1151–1159, 1990.

[21] Miles J. O. Deegan and Robert J. Harrison. in perparation.

[22] B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin. On some approximations in applications of $x\alpha$ theory. *J. Chem. Phys.*, 71(8):3396–3402, 15 October 1979.

[23] Thom H. Dunning, Jr. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90(2):1007–1023, 15 January 1989.

[24] Michel Dupuis, Steve Chin, and Antonio Marquez. Modern tools for including electron correlation in electronic structure studies: HONDO and CHEM-STATION. In G. L. Malli, editor, *Relativistic and Electronic Correlation Effects in Molecules and Clusters*, volume 318 of *NATO ASI Series B (Physics)*, pages 315–338. Plenum Press, New York, 1994.

[25] Volker Dyczmons. No n^4-dependence in the calculation of large molecules. *Theoret. Chim. Acta*, 28:307–310, 1973.

[26] Clyde Edmiston and Klaus Ruedenberg. Localized atomic and molecules orbitals. *Rev. Mod. Phys.*, 35(3):457–465, July 1963.

[27] Karin Eichkorn, Oliver Treutler, Holger Öhm, Marco Häser, and Reinhart Ahlrichs. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.*, 240:283–290, 1995.

[28] Karin Eichkorn, Oliver Treutler, Holger Öhm, Marco Häser, and Reinhart Ahlrichs. Erratum: Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.*, 242:652–660, 1995.

[29] Karin Eichkorn, Florian Weigend, Oliver Treutler, and Reinhart Ahlrichs. Auxiliary basis sets for main row atoms and transition metals and their use to approximate coulomb potentials. *Theoret. Chem. Acct.*, 97(1–4):119–124, 1997.

[30] David Elwood, George Fann, and Richard Littlefield. Peigs parallel eigensystem solver, version 2.1, 1995.

[31] David Feller, Edoardo Aprà, Jeff A. Nichols, and David E. Bernholdt. The structure and binding energy of K^+-ether complexes: A comparison of MP2, RI-MP2 and density functional methods. *J. Chem. Phys.*, 105(5):1940–1950, 1 August 1996.

[32] Martin Feyereisen, George Fitzgerald, and Andrew Komornicki. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. *Chem. Phys. Lett.*, 208(5,6):359–363, 1993.

[33] Graham D. Fletcher, Alistair P. Rendell, and Paul Sherwood. A parallel second-order Moller-Plesset gradient. *Mol. Phys.*, 91(3):431, June 20 1997.

[34] Herbert A. Früchtl, Rick A. Kendall, Robert J. Harrison, and Kenneth G. Dyall. An implementation of RI-SCF on parallel computers. *Int. J. Quantum Chem.*, 64(1):63–69, 1997.

[35] Eric D. Glendenning, David Feller, and Mark A. Thompson. An ab initio investigation of the structure and alkali metal cation selectivity of 18-crown-6. *J. Am. Chem. Soc.*, 116:10657–10669, 1994.

[36] Leslie F. Greengard. *The Rapid Evaluation of Potential Fields in Particle Systems.* ACM Distinguished Dissertations. The MIT Press, Cambridge, Massachusetts, USA, 1988.

[37] Claudia Hampel and Hans-Joachim Werner. Local treatment of electron correlation in coupled cluster theory. *J. Chem. Phys.*, 104(16):6286–6297, 22 April 1996.

[38] R. J. Harrison. Ab initio chemistry in parallel. Technical Report CCSF-14-92, Caltech Concurrent Supercomputing Facility, Pasadena, CA, 1992.

[39] Robert J. Harrison. private communication.

[40] Marco Häser. Møller-Plesset (MP2) perturbation theory for large molecules. *Theoret. Chim. Acta*, 87:147–173, 1993.

[41] Marco Häser and Jan Almlöf. Laplace transformation techniques in Møller-plesset perturbation theory. *J. Chem. Phys.*, 96(1):489–494, 1 January 1992.

[42] High Performance Computational Chemistry Group, Pacific Northwest National Laboratory, Richland, Washington 99352, USA. NWChem, a computational chemistry package for parallel computers, version 3.1, 1997.

[43] High Performance Fortran Forum. *High Performance Fortran Language Specification (Version 1.1)*, November 1994. URL http://www.crpc.rice.edu/HPFF/hpf1.

[44] High Performance Fortran Forum. *High Performance Fortran Language Specification (Version 2.0.δ)*, October 1996. URL http://www.crpc.rice.edu/HPFF/hpf2.

[45] E. Philip Horwitz, Mark L. Deitz, and Dan E. Fisher. SREX: A new process for the extraction and recovery of strontium from acidic nuclear waste streams. *Solvent Exchange Ion Extraction*, 9(1):1–25, 1991.

[46] Jonathan N. Hurley, David L. Huestis, and William A. Goddard, III. Optimized two-electron-integral transformation procedures for vector-concurrent computer architectures. *J. Phys. Chem.*, 92:4880–4883, 1988.

[47] Steven Huss-Lederman, Elaine M. Jacobson, Anna Tsao, and Guodong Zhang. Matrix multiplication on the Intel Touchstone Delta. *Concurrency: Practice and Experience*, 6(7):571–594, October 1994.

[48] A. R. Katritzky, N. Malhotra, R. Ramanathan, R. C. Kemeriat, Jr., J. A. Zimmerman, and J. A. Eyler. Measurement of gas-phase binding energies of crown ethers with metal ions by Fourier-transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.*, 6:25, 1992.

[49] Alan R. Katritzky, Clifford H. Watson, Zofia Dega-Szafran, and John R. Eyler. Collisionally activated dissociation of n-alkylpyridinium cations to pyridine and alkyl cations in the gas phase. *J. Am. Chem. Soc.*, 112(7):2471, March 28 1990.

[50] Rick A. Kendall, Thom H. Dunning, Jr., and Robert J. Harrison. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.*, 96(9):6796–6806, 1 May 1992.

[51] Rick A. Kendall and Herbert A. Früchtl. The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development. *Theoret. Chem. Acct.*, 97(1–4):158–163, 1997.

[52] R. Kutteh and J. B. Nicholas. A generalized fast multipole approach for Hartree-Fock and density functional computations. *Chem. Phys. Lett.*, 238:173, 1995.

[53] Jean-Marc Langlois, Richard P. Muller, Terry R. Coley, III William Goddard, Murco N. Ringnalda, Youngdo Won, and Richard A. Friesner. Pseudospectral generalized valance-bond calcuulations: Application to methylene, ethylene, and silylene. *J. Chem. Phys.*, 92(12):7488–7497, 15 June 1990.

[54] Ajay C. Limaye and Shridhar R. Gadre. A general parallel solution to the integral transformation and second-order moller-plesset energy evaluation on distributed memory parallel machines. *J. Chem. Phys.*, 100(2):1303–1307, 15 January 1994.

[55] Todd J. Martinez and Emily A. Carter. Pseudospectral Møller-Plesset perturbation theory through third order. *J. Chem. Phys.*, 100(5):3631–3638, 1 March 1994.

[56] Todd J. Martinez, Aseem Mehta, and Emily A. Carter. Pseudospectral full configuration interaction. *J. Chem. Phys.*, 97(3):1876–1880, 1 August 1992.

[57] P. E. Maslen, , C. Ochsenfeld, C. A. White, M. S. Lee, and M. Head-Gordon. Locality and sparsity of ab initio one-particle density matrices and localised orbitals. *J. Phys. Chem.*, submitted.

[58] P. E. Maslen and M. Head-Gordon. Non-iterative local second order Møller-Plesset theory. *Chem. Phys. Lett.*, in press.

[59] Wilfried Meyer. Theory of self-consistent electron pairs. An iterative method for correlated many-electron wavefunctions. *J. Chem. Phys.*, 64(7):2901–2907, 1 April 1976.

[60] J. W. Mintmire and B. I. Dunlap. Fitting the Coulomb potential variationally in linear-combination-of-atomic-orbitals density-functional calculations. *Phys. Rev. A*, 25(1):88–95, 1982.

[61] Robert B. Murphy. private communication.

[62] Ida M. B. Nielsen. A new direct MP2 gradient algorithm with implementation on a massively parallel computer. *Chem. Phys. Lett.*, 255:210–216, 7 June 1996.

[63] Jaroslaw Nieplocha and Robert J. Harrison. Global array tools library. URL ftp://ftp.pnl.gov/pub/global.

[64] Jaroslaw Nieplocha, Robert J. Harrison, and Richard J. Littlefield. Global arrays: A portable "shared-memory" programming model for distributed memory computers. In *Supercomputing'94*, pages 340–349, Los Alamitos, California, USA, 1994. Institute of Electrical and Electronics Engineers and Association for Computing Machinery, IEEE Computer Society Press.

[65] Jaroslaw Nieplocha, Robert J. Harrison, and Richard J. Littlefield. Global arrays: A non-uniform-memory-access programming model for high-performance computers. *J. Supercomputing*, 10(2):169, 1996.

[66] Douglas W. O'Neal and Jack Simons. Application of Cholesky-like matrix decomposition methods to the evaluation of atomic orbital integrals and integral derivatives. *Int. J. Quantum Chem.*, 34:673–688, 1989.

[67] Itai Panas and Jan Almlöf. A fragment multipole approach to long-range Coulomb interactions in Hartree-Fock calculations on large systems. *Int. J. Quantum Chem.*, 42:1073–1089, 1992.

[68] H. Petersen, E. R. Smith, and D. Soelvason. Error estimates for the fast multipole method. I. The two-dimensional case. *Proc. Roy. Soc. A*, 448:389, 1994.

[69] H. Petersen, E. R. Smith, and D. Soelvason. Error estimates for the fast multipole method. II. The three-dimensional case. *Proc. Roy. Soc. A*, 448:401, 1994.

[70] H. Petersen, D. Soelvason, J. W. Perram, and E. R. Smith. The very fast multipole method. *J. Chem. Phys.*, 101:8870, 1994.

[71] János Pipek and Paul G. Mezey. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *J. Chem. Phys.*, 90(9):4916–4926, 1 May 1989.

[72] Peter Pulay. Localizability of dynamic electron correlation. *Chem. Phys. Lett.*, 100(2):151–154, 2 September 1983.

[73] Péter Pulay and Svein Saebø. Orbital-invariant formulation and second-order gradient evaluation in Møller-Plesset perturbation theory. *Theoret. Chim. Acta*, 69:357–368, 1986.

[74] Péter Pulay, Svein Saebø, and Wilfried Meyer. An efficient reformulation of the closed-shell self-consistent electron pair theory. *J. Chem. Phys.*, 81(4):1901–1905, 15 August 1984.

[75] Guntram Rauhut and Peter Pulay. Considerations regarding the local treatment of Laplace transformation MPPT. *Chem. Phys. Lett.*, 248:223–227, 12 January 1996.

[76] Alistair P. Rendell and Timothy J. Lee. Coupled-cluster theory employing approximate integrals: An approach to avoid the input/output and storage bottlenecks. *J. Chem. Phys.*, 101(1):400–408, 1 July 1994.

[77] Murco N. Ringnalda, Mahfoud Belhadj, and Richard A. Friesner. Pseudospectral hartree-fock theory: Applications and algorithmic improvements. *J. Chem. Phys.*, 93(5):3397–3407, 1 September 1990.

[78] Svein Saebo. The effect of electron correlation on the trans-cis energy difference of glyoxal. *Chem. Phys.*, 113:383–389, 1987.

[79] Svein Sæbø and Peter Pulay. Local correlation interaction: An efficient approach for larger molecules. *Chem. Phys. Lett.*, 113(1):13–18, 4 January 1985.

[80] Svein Saebø and Peter Pulay. Fourth-order Møller-Plessett perturbation theory in the local correlation treatment. I. Method. *J. Chem. Phys.*, 86(2):914–922, 15 January 1987.

[81] Svein Saebø and Peter Pulay. Local treatment of electron correlation. *Annu. Rev. Phys. Chem.*, 44:213–236, 1993.

[82] V. R. Saunders, C. Freyria-Fava, R. Dovesi, L. Salasco, and C. Roetti. On the electrostatic potential in crystalline systems where the charge density is expanded in Gaussian functions. *Mol. Phys.*, 77(4):629–665, 1992.

[83] Ansgar Schäfer, Hans Horn, and Reinhart Ahlrichs. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.*, 97(4):2571–2577, 15 August 1992.

[84] Martin Schütz and Roland Lindh. An integral direct, distributed data, parallel MP2 algorithm. *Theoret. Chim. Acta*, 95:13–34, 1997.

[85] Harrell Sellers, Svein Sæbø, and Peter Pulay. The ring puckering potential of oxetane: Local correlation results. *Chem. Phys. Lett.*, 132(1):29–31, 28 November 1986.

[86] M. C. Strain, G. E.Scuseria, and M. J. Frisch. Achieving linear scaling for the electronic quantum Coulomb problem. *Science*, 271:51, 1996.

[87] Seiichiro Ten-no and Suehiro Iwata. Three-center expansion of electron repulsion integrals with linear combination of atomic electron distributions. *Chem. Phys. Lett.*, 240:578–584, 1995.

[88] Seiichiro Ten-no and Suehiro Iwata. Multiconfiguration self-consistent field procedure employing linear combination of atomic-electron distributions. *J. Chem. Phys.*, 105(9):3604–3611, 1 September 1996.

[89] Seiichiro Ten-no and Suehiro Iwata. On approximating electron repulsion integrals with linear combination of atomic-electron distributions. *Int. J. Quantum Chem.*, 60(7):1319–1324, 1996.

[90] O. Vahtras, J. Almlöf, and M. W. Feyereisen. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.*, 213(5,6):514–518, 1993.

[91] Robert A. van de Geijn and Jerrell Watts. SUMMA: Scalable universal matrix multiplication algorithm. Technical Report TR-95-13, Department of Computer Sciences, Univ. of Texas, Austin, TX, June 1995. URL file://net.cs.utexas.edu/pub/techreports/tr95-13.ps.

[92] J. D. Watts and M. Dupuis. Vector and parallel implementations of a fourth-order Moller-Plesset perturbation theory (MP4) program. Technical Report KGN-197, IBM Corporation, Data Systems Division, Kingston, NY 12401, August 16 1989.

[93] Florian Weigend and Marco Häser. RI-MP2: First derivatives and global consistency. *Theoret. Chem. Acct.*, 97(1–4):331–340, 1997.

[94] C. A. White and M. Head-Gordon. Derivation and efficient implementation of the fast multipole method. *J. Chem. Phys.*, 101:6593–6605, 1994.

[95] C. A. White and M. Head-Gordon. Fractional tiers in fast multipole method calculations. *Chem. Phys. Lett.*, 257:647–650, 1996.

[96] C. A. White and M. Head-Gordon. Rotating around the quartic angular momentum barrier in fast multipole method calculations. *J. Chem. Phys.*, 105:5061–5067, 1996.

[97] C. A. White, B. Johnson, P. Gill, and M. Head-Gordon. The continuous fast multipole method. *Chem. Phys. Lett.*, 230:8–16, 1994.

[98] C. A. White, B. G. Johnson, P. M. W. Gill, and M. Head-Gordon. Linear scaling density functional calculations via the continuous fast multipole method. *Chem. Phys. Lett.*, 253:268–278, 1996.

[99] Robert A. Whiteside, J. Stephen Binkley, Michael E. Colvin, and Henry F. Schaefer III. Parallel algorithms for quantum chemistry. I. integral transformations on a hypercube multiprocessor. *J. Chem. Phys.*, 86(4):2185–2193, 15 February 1987.

[100] R. Wiest, J. Demuynck, M. Bénard, M. Rohmer, and R. Ernenwein. A program system for ab initio MO calculations on vector and parallel processing machines. III integral reordering and four-index transformation. *Comp. Phys. Comm.*, 62:107–124, 1991.

[101] Theresa L. Windus, Michael E. Schmidt, and Mark S. Gordon. Parallel algorithm for integral transformation and GUGA MCSCF. *Theoret. Chim. Acta*, 89:77–88, 1994.

[102] Adrian T. Wong, Robert J. Harrison, and Alistair P. Rendell. Parallel direct four-index transformation. *Theoret. Chim. Acta*, 93:317–331, 1996.

[103] David E. Woon and Thom H. Dunning, Jr. cc-pvdz and cc-pvtz basis sets for li, be. to be published.

[104] David E. Woon and Thom H. Dunning, Jr. Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. *J. Chem. Phys.*, 100(4):2975–2988, 15 February 1994.

E. Biographical Sketches

David E. Bernholdt

Northeast Parallel Architectures Center
Syracuse University
111 College Place
Syracuse, NY 13244-4100

Phone: 315 443 3857
Fax: 315 443 1973
E-mail: bernhold@npac.syr.edu

Education

Ph.D. in Chemistry, minors in Physics and Mathematics, and Certification in Chemical Physics, Univ. of Florida, Gainesville, FL, 1993.
B.S. in Chemistry *Cum Laude* with Highest Distinction, Univ. of Illinois, Urbana, IL, 1986

Professional Experience

1996– Research Assistant Professor, Department of Chemistry, Syracuse University, Syracuse, NY. **1995–** Alex G. Nason Fellow and Research Scientist, Northeast Parallel Architectures Center, Syracuse University, Syracuse, NY. **1995–** Affiliate Staff Scientist, Pacific Northwest National Laboratory, Richland, WA. **1993–1995** AWU Postdoctoral Fellow, High Performance Computational Chemistry Group, Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA. **1986–1993** Teaching or Research Assistant, University of Florida, Gainesville, FL. **1985–1986** Teaching Assistant, University of Illinois, Urbana, IL.

Related and Significant Publications

[1] J. Anchell, E. Apra, D. Bernholdt, P. Borowski, T. Clark, D. Clerc, H. Dachsel, M. Deegan, M. Dupuis, K. Dyall, G. Fann, H. Früchtl, M. Gutowski, R. Harrison, A. Hess, J. Jaffe, R. Kendall, R. Kobayashi, R. Kutteh, Z. Lin, R. Littlefield, X. Long, B. Meng, J. Nichols, J. Nieplocha, A. Rendell, M. Stave, T. Straatsma, H. Taylor, G. Thomas, K. Wolinski, and A. Wong. *NWChem, A Computational Chemistry Package for Parallel Computers, Version 3.1.* Pacific Northwest National Laboratory, Richland, Washington 99325-0999 USA, 1997.

[2] D. E. Bernholdt, E. Aprà, H. A. Früchtl, M. F. Guest, R. J. Harrison, R. A. Kendall, R. A. Kutteh, X. Long, J. B. Nicholas, J. A. Nichols, H. L. Taylor, A. T. Wong, G. I. Fann, R. J. Littlefield, and J. Nieplocha. Parallel computational chemistry made easier: The development of NWChem. *Int. J. Quantum Chemistry: Quantum Chem. Symposium*, 29:475–483, 1995.

[3] David E. Bernholdt and Robert J. Harrison. Orbital invariant second-order many-body perturbation theory on parallel computers: An approach for large molecules. *J. Chem. Phys.*, 102(24):9582–9589, 22 June 1995.

[4] David E. Bernholdt and Robert J. Harrison. Large-scale correlated electronic structure calculations: The RI-MP2 method on parallel computers. *Chem. Phys. Lett.*, 250:477–484, 8 March 1996.

[5] David E. Bernholdt and Robert J. Harrison. Fitting basis sets for the RI-MP2 approximate second-order many-body perturbation theory method. *J. Chem. Phys.*, submitted.

[6] David Feller, Edoardo Aprà, Jeff A. Nichols, and David E. Bernholdt. The structure and binding energy of K^+-ether complexes: A comparison of MP2, RI-MP2 and density functional methods. *J. Chem. Phys.*, 105(5):1940–1950, 1 August 1996.

[7] M. F. Guest, E. Aprà, D. E. Bernholdt, H. A. Früchtl, R. J. Harrison, R. A. Kendall, R. A. Kutteh, X. Long, J. B. Nicholas, J. A. Nichols, H. L. Taylor, A. T. Wong, G. I. Fann, R. J. Littlefield, and J. Nieplocha. Advances in parallel distributed data software; computational chemistry and NWChem. In *Applied Parallel Computing. Computations in Physics, Chemistry and Engineering Science*, volume 1041 of *Lecture Notes in Computer Science*. Springer, Heidelberg, 1996.

[8] Robert J. Harrison, Martyn F. Guest, Rick A. Kendall, David E. Bernholdt, Adrian T. Wong, Mark Stave, James Anchell, Anthony Hess, Rik Littlefield, George I. Fann, Jarek Nieplocha, Greg S. Thomas, David Elwood, Jeff Tilson, Ron L. Shepard, Albert F. Wagner, Ian T. Foster, Ewing Lusk, and Rick Stevens. High performance computational chemistry. II. A scalable SCF program. *J. Computat. Chem.*, 17:124, 1995.

[9] Meenakshi A. Kandaswamy, Mahmut T. Kandemir, Alok N. Choudhary, and David E. Bernholdt. An experimental study to analyze and optimize Hartree-Fock application's I/O with PASSION. *Int. J. Supercomputer Appl.*, in press.

[10] M.F.Guest, E.Apra, D.E.Bernholdt, H.A.Fruchtl, R.J.Harrison, R.A.Kendall, R.A.Kutteh, X.Long, J.B.Nicholas, J.A.Nichols, H.L.Taylor, A.T.Wong, G.I.Fann, R.J.Littlefield, and J.Nieplocha. High-performance computing in chemistry; nwchem. *Future Generation Computer Systems*, 12(4):273–289, December 1996.

Recent Collaborators

JL Anchell,[1] E Apra,[2] RJ Bartlett,[3] RR Birge,[4] AN Choudhary,[5] PD Ellis,[6] D Elwood,[6] GI Fann,[6] DF Feller,[6] IT Foster,[7] GC Fox,[4] HA Früchtl,[8] MF Guest,[9] RJ Harrison,[6] AC Hess,[6] MA Kandaswamy,[4] MT Kandemir,[4] RA Kendall,[6] RA Kutteh,[6] RJ Littlefield,[6] X Long,[6] E Lusk,[7] JB Nicholas,[6] JA Nichols,[6] J Nieplocha,[6] SA Perera,[3] RL Shepard,[7] MS Stave,[10] R Stevens,[7] HL Taylor,[6] GS Thomas,[6] JL Tilson,[7] AF Wagner,[7] AT Wong.[11]

Institutional Affiliations: [1]Schrödinger, Inc., [2]Center for Study of the Relations between Structure and Chemical Reactivity, Nat'l Research Council of Italy, [4]Syracuse Univ., [5]Northwestern Univ., [6]Pacific Northwest Nat'l Lab., [7]Argonne Nat'l Lab., [8]Fujitsu European Centre for Information Technology, [9]Daresbury Lab., UK, [3]Univ. of Florida, [10]Oxford Molecular, Inc., [11]Nat'l Energy Research Supercomputer Center.

Students and Postdoctoral Scholars

None

Advisors

Graduate: Rodney J. Bartlett, University of Florida
Postgraduate: Robert J. Harrison, Pacific Northwest National Laboratory

F. Summary Proposal Budget

Replace with NSF Form 1030s.

F.1 Personnel

The proposed research will require a significant amount of time and effort to bring to fruition. I have allocated funds to support a postdoctoral researcher to carry out the majority of the work, along with a month of my time for supervision and assistance. The salary amount budgeted allows for some flexibility in hiring; a new PhD would not be offered the full amount. This figure is in line with current practice in the Syracuse Chemistry Department, and is actually on the low end of salaries paid to junior researchers in NPAC.

The monies allocated to Technical Support provide for the day to day operation of the computer infrastructure which is crucial to this work. This frees the researchers to focus on the science, and is more cost-effective than if I had to do the work myself.

F.2 Equipment

I have budgeted funds in year 1 for the purchase of a desktop workstation to support the postdoctoral researcher. Other computing resources which will be used to carry out the proposed research are already available in NPAC or will be accessed remotely (see Section H).

F.3 Travel

The travel budget includes support for one major conference a year to present results of this work. It also includes partial support for one trip to the Pacific Northwest National Laboratory each year to facilitate collaboration. While e-mail, phone, and fax can be very effective communication tools in general, there is sometimes no substitute for face to face discussions and close collaboration. One particular example in this case will be work on incorporating sparsity in the Global Array programming model, which will require close cooperation with the author of the GAs. The cost of such trips will be shared with the Laboratory.

F.4 Supplies

A modest amount is requested for materials and supplies. This includes basic computer and office supplies, telephone expenses, paper, transparencies for presentations, diskettes, etc.

F.5 Publication Costs

Monies are requested to cover anticipated publication costs.

G. Current and Pending Support

Replace with NSF Form 1239.

H. Facilities, Equipment, and Other Resources

H.1 Local Computer Facilities

NPAC operates substantial computer facilities, including desktop workstations, file and compute servers, as well as experimental systems, and a leading edge network infrastructure. The PI's desktop system is an IBM RS/6000 43P which, in combination with NPAC-operated PC systems, provides the basic environment for daily use, including software development and data analysis. Other resources relevant to this project include a 12-node IBM SP2 parallel computer, an 8-node DEC Alpha cluster interconnected with a Gigaswitch, an 8-node Sun UltraSPARC cluster interconnected with an ATM (asynchronous transfer mode) network. These clusters are shared by all NPAC users, but available in dedicated parallel mode on a reservation basis. An 8-CPU SGI machine with R10000 processors is largely devoted to other projects, but provides a shared-memory parallel platform for testing purposes.

NPAC's local network includes a variety of high-speed connections among server machines (ATM, FDDI, 100-base-T ethernet) and wiring capable of 100-base-T to all offices.

Remote supercomputer centers which will be used for the larger calculations in this project are easily accessible via Syracuse's T3 connection to the Internet. Syracuse expects to join vBNS (national academic ATM network backbone) in the coming year, which will further enhance access to remote sites.

NPAC has a professional system support group which handles day to day operation and maintenance of the computer facilities.

H.2 Remote Computer Facilities

In addition to the various NSF-funded supercomputer centers, the PI's collaboration with the Pacific Northwest National Laboratory on this project provides access to several Dept. of Energy large-scale computer facilities which will be utilized for some of the larger calculations in this project. Sites include the Pacific Northwest National Laboratory, Argonne National Laboratory, and the National Energy Research Supercomputer Center.

H.3 Office

NPAC provides adequate office facilities for all participants.

H.4 Major Equipment

The computer systems described above constitute the major equipment for this effort.

H.5 Other Resources

NPAC provides basic secretarial services, duplicating facilities, and the like.

I. Letters of Cooperation and Support

J. A. Nichols, Technical Group Leader, High Performance Computational Chemistry, Environmental Molecular Science Laboratory, Pacific Northwest National Laboratory (2 pages).