C. Project Description

We propose development and implementation of evolutionary algorithms and their integration with density functional theories to investigate stable conformations of heteronuclear clusters containing metallic elements. Our specific aims include (i) optimization of reproduction operators and fitness criteria in evolutionary algorithms for searching stable conformations of clusters; (ii) development of a modular evolutionary algorithm program for geometry optimization which can be easily interfaced with several commonly used computer program suites for electronic structure calculations; (iii) optimization of reproduction operators and fitness criteria in evolutionary algorithms for electronic structure calculations; (iv) development of a co-evolutionary algorithm for micro scale (electronic structure) and macro scale (nuclear conformation search). Our objectives, once accomplished, will impact both the fields of evolution programming and physical/theoretical chemistry, and will significantly enhance our ability to characterize clusters and nanoclusters. It is multi-disciplinary and multi-institutional. A significant component of this proposal is the involvement of graduate and undergraduate students.

C.1. Significance and frame of problems

Structural, optical, and electronic properties of clusters have been the research focus in the fields of chemistry, physics and nanotechnology because clusters offer a unique opportunity to approach size-dependent phenomena. By systematically varying the sizes of clusters, researchers are able to explore and understand the emergence of collective bulk properties. Interests in clusters are further fueled by recent recognition that clusters form the building blocks of nanocrystals and quasi-crystals. Our knowledge in the properties of clusters would aid the rational design of nanocrystalline materials.

Computational simulations play a critical role in the investigations of clusters because simulations provide a direct link between an observable property and a well-defined model system. A fundamental problem in cluster research is the identification of stable conformations. Interactions between the constituents (atoms, molecules, or ions) of a cluster give rise to a complex energy "landscape" as a function of nuclear coordinates. The local and global minima on this energy landscape correspond to the stable conformations of a cluster. The number of stable conformations rises extremely rapidly with increasing cluster size, and it has been shown an NP-hard problem.[1-3]

Search the global minimum of a potential energy landscape has received significant attention in the past decade because of it is an obstacle shared by biology, chemistry, and physics.[4] Among the methods attempted to tackle this problem is genetic algorithm which is a variation of evolution programming, and much effort has been made to tailor genetic algorithms for geometry optimization of clusters and molecules. Evolution algorithms have been shown to be superior than the genetic algorithms. Our first objective is therefore to optimize evolutionary algorithms specifically for searching conformation of molecules and clusters.

Another significant, if not equally formidable, contribution to the computational cost is the representation of the energy landscape. The most accurate approaches are the first principle methods, for example, density functional theories, which treat electrons explicitly and therefore incorporate the effects of electron-electron correlation. Semiempirical and empirical potentials approximate the interactions as functions of the nuclear coordinates only, thereby reducing the computational cost. They are economical

for nano-scale systems, but inappropriate for investigating electronic and optical properties, because electron-electron interactions are not treated explicitly.

A common approach to the aforementioned problem is the “composite” approach in which conformations at the local minima of a semiempirical or an empirical potential are determined using optimization techniques such as genetic algorithms; these conformations subsequently serve as starting geometries in a DFT calculation and further optimized using gradient-based methods. This approach is limited by several inherent shortcomings. Semiempirical and empirical potentials must be carefully calibrated for the specific system of interest. Currently only the ones for homonuclear clusters are well characterized. Constructing a many-body potential for heteronuclear clusters is a daunting obstacle. The semiempirical and empirical potentials are by definition reductions of the energy landscapes derived from the first principle calculations. As such, there is not necessarily a one-to-one correspondence between the two types of energy landscape. In particular, the energetic ordering of the stable conformations derived from a semiempirical or empirical potential may not be consistent with that derived from DFT calculations.

We approach this problem from two directions. First and the most obvious is to interface evolutionary algorithms for geometry optimization with programs for electronic structure calculations. That is, an electronic structure calculation for a given nuclear conformation is carried out “on the fly” during a geometry optimization by evolutionary algorithms. Our objective is to implement evolutionary algorithms in the most modular form that can be easily interfaced with commonly used program suites for electronic structure calculations.

The second approach which we believe is novel is to “couple” the electronic and nuclear parts of a problem instead of a two-step process in the first approach.

Solving the electronic structures of a molecule or a cluster, by itself, is an optimization problem which requires solutions to differential equations with boundary conditions. In typical formulations of electronic structure theory, the working equations are nonlinear and the solutions asymptotically approach exponential functions. It has been demonstrated that genetic algorithm is efficient for solving stiff ordinary differential equations. We attempt to formulate electronic structure theory to be solved by evolutionary algorithms and to tailor the reproduction operators and fitness criteria in evolutionary algorithms for electronic structure calculations.

Simultaneous optimization of electronic structure and nuclear conformation may be considered a “co-evolution” process involving changes in both the micro and macro scales. This presents an opportunity to advance the methodology of evolutionary algorithms. Since only the global minimum is of interest, the convergence of electronic structure calculations at intermediate nuclear conformations need not be stringent. In other word, the fitness criteria need not be fixed, and can be adjusted during the optimization process. Consequently one expects that the “coupled” optimization should be more efficient.

In Section C.2 we give a brief overview of the general methodology of evolutionary algorithm. In Sections C.3 — C.6 we discuss our four objectives, in particular, their significance, current challenges and specific procedures. In Section C.7 we discuss the validation and assessment of our objectives which are carried out by applying the new methods to several prototypical systems in cluster research and

nanotechnology. In Section C.8 we outline a timetable and discuss briefly how this proposal will impact the education and training of undergraduate and graduate students.

C.2. Background: Evolution algorithm

Evolution strategies (ES) belong to a group of optimization methods called *evolutionary algorithm* (EA) which are based upon the principle of adaptive selection found in the natural world. In optimization, minimum and/or maximum values of a property E which depends on some parameters, $\{\chi_i\}$, are sought after,

$$E = E(\{\chi_i\}). \tag{1}$$

In evolutionary algorithms the search for the minimum value of E emulates the process of natural selection found in nature. In the natural world organisms contain a specific set of genes, which is the genotype. The interaction of the genotype with its environment determines the organisms morphology and behavior — i.e., its phenotype. This phenotype translates into an organism's fitness, which ultimately determines the likelihood of survival.

Figure 1 shows the similarity between evolutionary algorithms and evolution. In evolutionary algorithms the optimization problem parameters (genes) are encoded in a data structure (genotype). The solution defined by the problem parameters is qualitatively defined by the problem parameters is qualitatively evaluated (phenotype). Good solutions are retained (survive) and stochastically altered (reproduced) to create new populations (new solutions). This process is repeated until either a fixed number of iterations (generations) have been processed or an acceptable solution has been found. More details on evolutionary algorithms can be found in various references (e.g. see Bäck[5]).

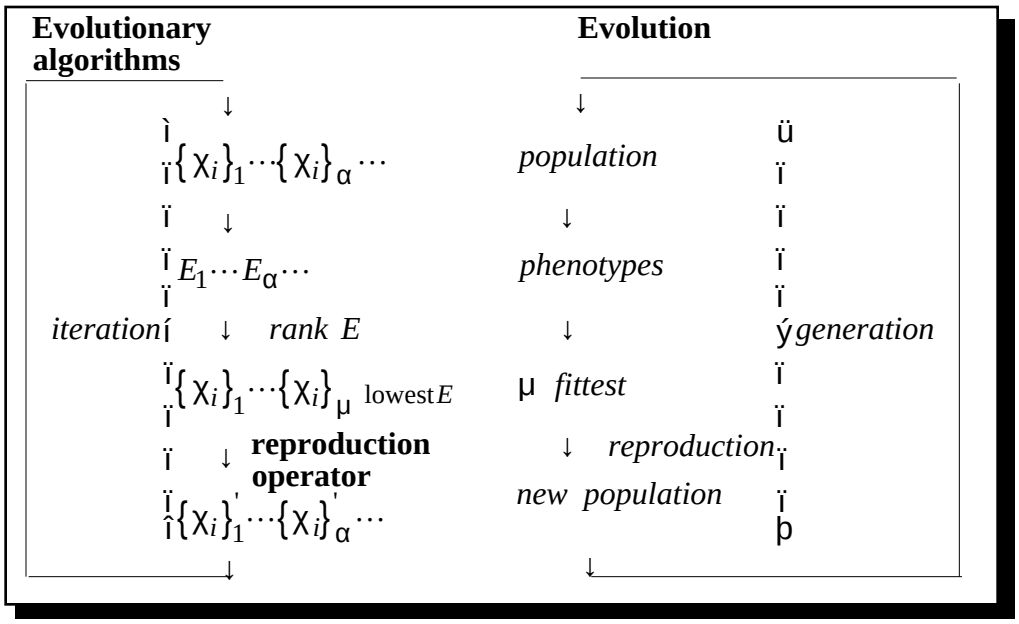

Figure 1. Comparison between evolutionary algorithms and natural selection

Evolutionary algorithms have emerged as an extremely powerful heuristic search technique. These algorithms attempt to emulate the principles of Darwinian evolution to search vast problem spaces for good solutions. The three best known evolutionary algorithms are genetic algorithms (GAs), evolution strategies (ES) and evolutionary programming (EP). GAs have been used in the past for cluster studies and the results indicate they are clearly superior to simulated annealing approaches.[6-9]

Since new solutions are generated by perturbation, the effectiveness and efficiency of evolution strategies depend very much on how the reproduction operator introduces the perturbation. A common practice is to add to χ_i a "noise" from a gaussian distribution:

$$\chi'_{i\alpha} = \chi_{i\alpha} + N_{i\alpha}(0, \sigma_{i\alpha}) \tag{2}$$

where $N(0, \sigma)$ is a random variable obtained from a gaussian distribution with zero mean and standard deviation σ. Subscripts i and α collectively denote the i-th parameter (χ_i) of the α-th set ($\{\chi_i\}_\alpha$) in the population respectively (see Figure 1). A new $N_{i\alpha}(0, \sigma_{i\alpha})$ for each χ_i is generated in each iteration. The values for the standard deviation $\{\sigma_i\}$ may evolve as well. In a process known as *intermediate recombination*, the standard deviation $\sigma'_{i\alpha}$ of $\chi'_{i\alpha}$ is obtained as follows,

$$\sigma'_{i\alpha} = \frac{1}{2}(\sigma_{i\alpha} + \sigma_{i\beta}) \tag{3}$$

where β is randomly selected from the μ sets of best solutions in the current iteration. This procedure has been shown to accelerate convergence of optimization [???]. Further improvement of convergence rates may be possible by combining intermediate recombination with mutation (in σ's).

There are two general versions of evolution strategies. Let there be λ new sets of $\{\chi_i\}'$ generated from each $\{\chi_i\}$ in the μ best sets of solutions by perturbation. In the $(\mu+\lambda)$-ES version, the new solutions and the current ones form a new population (thus $\mu\lambda + \lambda$ in total); and their corresponding values of E are compared to derive another μ best sets of solutions. In the (μ,λ)-ES a population is formed exclusively from the $\{\chi_i\}'$ (thus $\mu\lambda$ in total). That is, each best solution survive only one iteration (generation). While this may result in short phases of recession, it does avoid long stagnation phases, it has been suggested that the (μ,λ)-ES works better in multimodal search spaces.

From the foregoing discussion, it is seen that the reproduction operator for ES is in essence stochastic. This does not mean, however, that the ES is merely conducting a random search. The deterministic selection of the best solutions during each iteration favors those regions of the search space where good solutions have been previously found.

C.3. Specific aim: development of evolutionary algorithms for nuclear conformation optimization

Stability of a conformation is in part determined by its energy relative to that of the constituents separated infinitely apart. Interactions between the constituents (atoms, molecules, or ions) of a cluster give rise to a complex energy "landscape" as a function of nuclear coordinates. The local and global minima on this energy landscape correspond to the stable conformations of a cluster. One method of solving this optimization problem is to explore the potential energy surface (PES) composed of all possible cluster

conformations. As the cluster size increases so does the number of degrees of freedom in the placement of the atoms. This characteristic produces a PES with numerous local optima [10]. In fact, determining the ground state energy level of atomic clusters is known to be NP-hard [1, 3]. Hence heuristic search techniques are the only viable alternative.

A commonly used method for searching conformations is the simulated annealing method.[2] For example it has been applied to geometry optimization of polypeptides (in which CH and CH_2 groups are represented by united atom approach to reduce the dimension s of the problem).[11] Whether SA can locate the global minimum — given no limits on execution time — depends on the cooling schedule which in the simplest form depends on three parameters: starting temperature, cooling rate and number of cooling steps. A variation of SA is to use the Liouville equation rather than the Newton's equation.[12, 13]

Many attempts have been made to tailor GA for finding the global minimum conformation. A modified genetic algorithm has been applied to atomic Ar clusters and molecular water clusters up to $(H_2O)_{13}$.[14] In that particular work, water is treated as rigid molecules without internal degrees of freedom. Genetic algorithm has been applied to atomic clusters of rare gas.[15-17] Attempted has been made to optimize the GA procedure for search atomic cluster conformations,[18, 19] including the twinning mutations, add-and-etch processes, and seeding of the initial parental population with selected structural motifs.

Our preliminary work with silicon clusters suggests ES may prove to be an even better algorithm.[20] Consequently, we intend to use an ES in this research effort. Implementation of ES in our preliminary work[20] is shown in Figure 1. During each generation (iteration of the ES algorithm) a population of individuals or potential solutions are evolved to produce new solutions. Only the highest fit solutions survive to become parents for the next generation. The ES terminates after a sufficient number of generations have been processed. Figure 1 shows the ES algorithm used to search for low energy conformations.

1. Create an initial population of μ individuals. Randomly place each atom in Euclidean space.
2. For each individual, generate λ > μ offspring by applying a reproduction operator (describe below).
3. Determine the fitness of each offspring by computing its potential energy.
4. Select the m fittest individuals for survival. Discard the other individuals.
5. Proceed to step 2 unless Γ generations have been processed.

Figure 1. The evolutionary algorithm for finding low energy conformations

Each individual represents a cluster conformation. The mutation operator used to produce offspring (new conformations) is quite simple. A bond between two atoms in the cluster is randomly chosen. Let r_i denote its current length. Then the new value of the bond length is r'_i which is given by

$$r'_i = r_i + \sigma_i^r N_i \tag{4}$$

where N produces a normally distributed random variable with zero mean and unit variance. The notation N_i indicates that a new random variable is generated for each r'_i. Offspring may have the same bond lengths as their parent except with a different angle that an atom makes with a coordinate axis. In a similar manner, offspring can be produced by randomly selecting an angle and perturbing it by

$$\phi'_i = \phi_i + \sigma_i^{\phi} N_i \qquad (5)$$

Individuals in a population are composed of both object parameters (the r_i's and ϕ_i's) and strategy parameters (the σ's). Offspring are produced by perturbing these parameters through the process differs depending on the parameter type. Object parameters are mutated by adding Gaussian noise while strategy parameters are adapted. A log normal form of self-adaptation is commonly used since it has been shown to be effective over a wide variety of search space topologies.[21] We choose the following log format

$$\sigma' = \sigma \ \exp(\tau \ N) \qquad (6)$$

where $\tau \quad N$.

There are two general versions of an ES. In the (m + l)-ES version m parameters produce l offspring; parents and offspring compete equally for survival. In the (m,l)-ES each parent produces multiple offspring but only the m best of the offspring survive. Notice that here parents survive only one generation regardless of their fitness. While this may result in short phases of recession, it does avoid phases of long stagnation. (In this work we will initially use the (m,λ) version with l = 7m.)

More specifically, we will employ a simplified model that treats the atom as a point mass in 3-dimensional space. Spherical coordinates will be used to specify each point mass location. Hence for an N atom cluster, an individual in the population of an ES encodes

$$\{r_1, \varphi_1, \theta_1; r_2, \varphi_2, \theta_2; \cdots; r_N, \varphi_N, \theta_N\} \qquad (7)$$

and the mutation operator selects a single element of this set and perturbs it to produce an offspring (c.f. Eqs. (4) and (5)).

The fitness of an individual is the energy where lower energy values denote higher fitness.

C.4. Specific aim: integration of evolutionary algorithms with electronic structure calculations

A prerequisite for any geometry optimization is a potential energy function, analytic or numerical, that provides the energy for a given nuclear conformation. Semiempirical and empirical potentials approximate the interactions as functions of the nuclear coordinates only, thereby reduce the computational cost. Currently several approaches provide the empirical and semiempirical potential energy functions for metallic systems The constant-volume pair potentials[22] are pair-potential models with explicit volume-dependent terms and are best suited for bulk uniform metallic systems. The embedded-atom method belong to the general category of pair functional methods (a comparison of various parameterization and interpretations is given by Raeker and Depristo[23]). This method takes into account the variation of bond strength with respect to the coordination numbers, and can therefore be applied to inhomogeneous metallic systems including clusters and nanoclusters. There are, however, serious limitations. The embedded-atom methods give the best results for metals with completely empty

or filled d orbitals but not as reliable for the transition metals with partially filled d orbitals. In part this is due to the assumed spherical symmetry of the pair functionals.[24] It is also caused by the average nature of the empirical data used in fitting the potentials.[24] Several other approaches have been developed to circumvent the limitations for the transitional metals. These include , for example, adding angular dependent terms in the pair functionals and taking into consideration the variation of the density of state s.

General-purpose molecular mechanics force fields for nonmetals are typically developed and parameterized under the constraint of fixed coordination numbers. They are therefore not appropriate for heteronuclear clusters of metallic and non-metallic elements. In general potential energy functions for metal-nonmetal interactions are less well developed. In order to investigate heteronuclear clusters with mixed metallic and nonmetallic elements, one faces the daunting task of constructing a potential energy function based on either first principle calculations or empirical data which is often for bulk systems and may not be appropriate for parameterizing potential energy functions for clusters and nanoclusters.

Moreover an inherent limitation of semiempirical and empirical potential energy functions and functionals is that they often can treat only ground electronic states. Many of the important observations and applications in nanotechnology concern electronic transitions and excited state properties. A "composite" approach is commonly employed to circumvent this problem. Conformations at the local minima of a semiempirical or an empirical potential are determined using heuristic stochastic optimization techniques such as simulated annealing and genetic algorithms; these conformations subsequently serve as starting geometries in a first principle calculation and further "relaxed" (i.e., optimized) using gradient-based methods such as Newton-Ralphson and conjugated gradient methods. The composite approach is applied to for example the structure and kinetics of silicon oxide interface where preliminary structures are first obtained using Monte Carlo simulated annealing using an analytic empirical potential and then refined using plane-wave expansion DFT calculations;[25] the lowest energy structures of gold nanclusters (Au_n, n =38, 55, and 75) using a modified GA with a Gupta n-body potential first and then relaxed using conjugate-gradient method with DFT-LDA.[26] The composite approach is limited by several inherent shortcomings. Semiempirical and empirical potentials must be carefully calibrated for the specific system of interest. Currently only the ones for homonuclear clusters are well characterized. Constructing a many-body potential for heteronuclear clusters is a daunting obstacle. The semiempirical and empirical potentials are by definition reductions of the energy landscapes derived from the first principle calculations. As such, there is not necessarily a one-to-one correspondence between the two types of energy landscape. In particular, the energetic ordering of the stable conformations derived from a semiempirical or empirical potential may not be consistent with that derived from DFT calculations.

Though it requires more computational resources, direct use of energy calculated by the first principle in geometry optimization may be less time consuming and less ambiguous. Searching stable conformations directly with ab initio energy has also been attempted by Monte Carlo simulations using ab initio energy as input. For example a Monte Carlo simulated annealing method was applied to $HCl(H_2O)n$, n = 3 and 4 clusters.[27] Also by using a variation of SA, i.e., instead of Newton's equation the Liouville's equation is used.[28] Our objective is to implement ES in a modular form that can be interfaced with several commonly used program suites for electronic structure calculations such as Gaussian 98[29] and Gamess.[30]

C.5. Specific aim: development of evolutionary algorithms for electronic structure calculation

Solving the electronic structures of a molecule or a cluster, by itself, is an optimization problem which requires solutions to differential equations with boundary conditions. In typical formulations of electronic structure theory, the working equations are nonlinear and the solutions asymptotically approach exponential functions. An advantage of EA is that it does not share the numerical instability of iterative procedures for solving a stiff ordinary differential equation. This is because, unlike iterative methods, differential operators are used to assess fitness but not suggest a better guess.[31] This has been demonstrated using GA in several stiff and non-stiff ordinary differential equation problems.[31] This is particularly ideal for electronic structure problems because the wavefunction asymptotically approaches an exponential function and numerical instability. GA has been used to solve the electronic structure of a helium atom.[32]

With nearly half a century of development in theory[33] and algorithms for electronic structure calculations, most computer program suites, whether they are based on density functional theory or molecular orbital theory, employ sophisticated numerical methods that are developed and optimized specifically for electronic structure problems. The performance of an algorithm depends in part on the particular hardware architecture where it is implemented. It is therefore critical to design and test new algorithms to adapt to new development in computer architecture. For example, extensive effort has been devoted to developing the quantum Monte Carlo methods for electronic structure calculations.[34, 35]

We attempt to formulate electronic structure theory to be solved by evolutionary algorithms and to tailor the reproduction operators and fitness criteria in evolutionary algorithms for electronic structure calculations. Our objective is to solve the following optimization problem:

> *Given a molecule or a cluster of N electrons which are subject to coulombic forces, find the electronic wavefunction within the Born-Oppenheimer approximation.*

Since this is a interdisciplinary proposal, we review below briefly the density functional theory. In particular, we focus on the nonlinearity and the many-body nature of the problem.

Because of the disparity between the masses and, consequently, the kinetic energies of an electron and a nucleus, the Born-Oppenheimer approximation assumes that electronic configurations adjust instantaneously with respect to the nuclear motion. That is, nuclei appear standing still relative to the electrons. This leads to the separation of the nuclear and electronic parts of the Schrödinger equation, and allows one to solve the electronic Schrödinger equation for a fixed nuclear arrangement:

$$\hat{H}_{el}\,\Psi(\{\mathbf{r}\};\{\mathbf{R}\}) = E_{el}(\{\mathbf{R}\})\Psi(\{\mathbf{r}\};\{\mathbf{R}\}) \qquad (1)$$

where the electronic wavefunction $\Psi(\{\mathbf{r}\};\{\mathbf{R}\})$ is a function of the electronic coordinates $\{\mathbf{r}\}$ and depends parameterically on the nuclear coordinates $\{\mathbf{R}\}$. The electronic hamiltonian in atomic units is defined as

$$\begin{aligned}\hat{H}_{el} &= \hat{T} + \hat{V} \\ &= -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{a=1}^{N_a}\sum_{i=1}^{N}\frac{Z_a}{r_{ai}} + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{1}{r_{ij}} + \frac{1}{2}\sum_{a=1}^{N_a}\sum_{b\neq a}^{N_a}\frac{Z_a Z_b}{r_{ab}} \qquad (2)\end{aligned}$$

where subscripts i and j designate electrons; subscripts a and b designate nuclei; N is the number of electrons; and N_a is the number of nuclei. The eigenvalue, $E_{el}(\{\mathbf{R}\})$, defines the potential energy landscape which governs the nuclear motions.

In density functional theory, the electronic energy $E_{el}(\{\mathbf{R}\})$ is a functional ($E_{el}[\rho]$) of electron density (ρ) with contributions from the kinetic energy ($T[\rho]$), the nuclear-electron interactions ($V_{ne}[\rho]$) and electron-electron interactions ($V_{ee}[\rho]$):

$$E_{el}[\rho] = T[\rho] + V_{ne}[\rho] + V_{ee}[\rho] \tag{3}$$

The Hohenberg-Kohn theorem states that E_{el} derived from any electron density cannot be less than that derived from the true ground state electron density for a given nuclear arrangement. This allows solutions of $E_{el}(\{\mathbf{R}\})$ by variationally minimizing $E_{el}[\rho]$ with respect to ρ:

$$\delta \{E_{el}[\rho] - \mu \int \rho(\mathbf{r})\, d\mathbf{r}\} = 0 \tag{4}$$

$$\frac{\delta E_{el}[\rho]}{\delta \rho(\mathbf{r})}\bigg|_{\rho=\rho_0} = \mu \tag{5}$$

where ρ_0 is the ground-state electron density, μ is a Lagrange undetermined multiplier which is defined as the chemical potential of the electronic system.

In the Kohn-Sham density functional theory, the correlated many-body problem of an N-electron system is mapped to a problem of N non-interacting fermions moving under the influence of an "effective potential." The effective potential is constructed so that the ground-state electron densities of the two systems are identical. A new energy functional due to the effective potential is defined as

$$E_{el}[\rho] = T_s[\rho] + V_{ne}[\rho] + J[\rho] + V_{xc}[\rho] \tag{6}$$

The wavefunction of N non-interacting fermions may be written as a Slater determinant of natural spin orbitals $\psi_i(\mathbf{r},s)$:

$$\Psi_s = \frac{1}{\sqrt{N!}} \det [\psi_1\psi_2\psi_3\cdots\psi_i] \tag{7}$$

and the associated electron density is

$$\rho(\mathbf{r}) = \sum_{i=1}^{N} |\psi_i(\mathbf{r},s)|^2 \tag{8}$$

Given the wavefunction, the kinetic energy functional $T_s[\rho]$ may be evaluated exactly. It should be noted however that the variation of $T_s[\rho]$ with respect to ρ is different from that of the true kinetic energy functional $T[\rho]$ for a *correlated* many-body system. Common to both the correlated and the non-interacting fermion systems are $V_{ne}[\rho]$, the electrostatic interactions between electrons and nuclei, and $J[\rho]$, the classical coulombic interactions between electrons. The differences between the correlated and the non-interacting fermion systems (including the differences between $T[\rho]$ and $T_s[\rho]$) are included in the exchange-correlation functional, $V_{xc}[\rho]$. The effective potential for the non-interacting fermions is

$$\upsilon_{eff}(\mathbf{r}) = \frac{\delta V_{ne}[\rho]}{\delta \rho(\mathbf{r})} + \frac{\delta J[\rho]}{\delta \rho(\mathbf{r})} + \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})}$$

$$= \upsilon(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + \upsilon_{xc}(\mathbf{r}) \quad (9)$$

where $\upsilon(\mathbf{r})$ is the electrostatic potential due to nuclei, and $\upsilon_{xc}(\mathbf{r})$ is called the exchange-correlation potential.

In the Kohn-Sham density functional theory, minimization of $E[\rho]$ with respect to ρ is carried out indirectly by varying the spin orbitals subject to the constraint that the spin orbitals are orthonormal. This procedure leads to a set of Kohn-Sham equations:

$$\hat{h}_{eff}\,\psi_i = [-\frac{1}{2}\nabla_i^2 + \upsilon_{eff}]\,\psi_i = \varepsilon_i\psi_i \quad (10)$$

Because υ_{eff} in the effective hamiltonian varies with the electron density (equation (9)) and, therefore, the spin orbitals (equation(8)), the Kohn-Sham equations are solved iteratively.

In the implementation of the Kohn-Sham density function theory, the spatial part of spin orbitals, $\psi_i(\mathbf{r},s)$ are represented by linear combinations of basis function φ_μ:

$$\psi_i = \sum_{\mu}^{M} c_{i\mu}\,\varphi_\mu \quad (11)$$

where M is the number of basis functions. The density is related to the basis functions by

$$\rho(\mathbf{r}) = \sum_i \sum_{\mu\nu} c_{i\mu}\,c_{i\nu}\,\varphi_\mu(\mathbf{r})\,\varphi_\nu(\mathbf{r}) \quad (12)$$

In terms of the basis functions, equation (10) becomes

$$\sum_{\mu} \hat{h}_{eff}\,\varphi_\mu c_{i\mu} = \sum_{\mu} \varepsilon_i\,\varphi_\mu c_{i\mu} \quad (13)$$

or secular equations,

$$\sum_{\mu} (h_{\mu\nu} - S_{\mu\nu}\varepsilon_i)\,c_{i\mu} = 0 \quad (14)$$

where $h_{\mu\nu}$ equals $\langle\varphi_\nu|\hat{h}|\varphi_\mu\rangle$ and $S_{\mu\nu}$ is the overlap integral $\langle\varphi_\nu|\varphi_\mu\rangle$. The coefficients $\{c_{i\mu}\}$ are obtained by diagonalizing an M-by-M matrix with matrix elements $h_{\mu\nu} - S_{\mu\nu}\varepsilon_i$. In practice, energy minimization begins with an initial guess of $\{c_{i\mu}\}$. The effective hamiltonian, in particular, $\upsilon_{eff}]$ is then constructed based on the initial $\{c_{i\mu}\}$. With the effective hamiltonian, $h_{\mu\nu}$ and $S_{\mu\nu}$ are calculated for each pair of basis functions. A new set of $\{c_{i\mu}\}$ is obtained by solving the secular equations (equation (13)) and used to update the effective hamiltonian. This process is repeated until the coefficients $\{c_{i\mu}\}$, and therefore, the wavefunction converge.

C.6. Specific aim: development of a co-evolutionary algorithm for coupled micro scale (electronic structure) and macro scale (nuclear conformation search)

In Sections C.4 and C.5 geometry optimization is presented as a two step process: first a first principle calculation is carried out for a given nuclear conformation, and then the fitness is evaluated followed by

reproduction and mutation. Such division is artificial. By coupling the two steps with an adjustable fitness criteria that varies depending on which region of a potential energy landscape is explored, convergence may be achieved much sooner. This coupled optimization problem may be considered analogous to the co-evolution process that spans both the micro and macro scales. Although co-evolutionary concepts are well studied aspects of evolutionary algorithms, they have never been used in this context. Coupled optimization problems are not unique to chemistry. Therefore the new development in methodology will contribute to many other fields.

C.7. Validation and assessment

The success of numerical method development could be measured by (i) timing comparison, (ii) accuracy of the calculated results (iii) applications to problems that cannot be treated with existing methodologies. Timing comparison requires implementation of methods to be compared It is beyond the scope of this proposal. We focus on the last two criteria in our assessment.

1. Application to transition-metal clusters

As discussed in Section C.4, transition metals with partially filled d orbitals are less reliably described by the embedded-atom pair potential functionals.[24] They are therefore ideal systems for demonstrating the range of applicability of the methods we will develop, in particular, the ones described in Sections C.4, C.5 and C.6. Transition metal clusters of different sizes (i.e., containing different number of atoms or ions) have been investigated experimentally and the trends of various properties with respect to cluster size . Typical properties investigated include ionization potentials, bond dissociation energies, magnetic moments, electron affinities, and photoelectron spectra In most experiments, the conformations of the clusters are unresolved, and therefore theoretical investigations can make significant contributions to the understanding of these clusters.

For clusters consist of Ni, in particular, ionization potential has been measured for Ni_3 up to Ni_{90}, and unexpected trend has bee observed[36] The conformations of Ni_3 – Ni_5 have been studied experimentally and theoretically.[3738] They are therefore ideal for the initial stage of our proposed research, i.e., for optimizing the reproduction operators and fitness criteria used in the evolutionary algorithms. Previous theoretical calculations on these small Ni clusters will serve as the main test cases for our new development detailed in Section C.5 and C.6 involving optimization of the electronic degrees of freedom. Once we are confident that our methodologies are optimized and free of error, we will investigate larger Ni clusters and compare the calculated ionization potential with experiment.

2. Application to metal-carbide clusters and nanocrystals

Heteronuclear clusters containing metallic and nonmetallic elements present a challenge to our current ability to simulate them. Heteronuclear clusters containing carbons and transition metals (so-called met-cars[39-41]) are particularly fascinating because they form specific stoichiometries (fixed ratios): M_8C_{12} and $M_{14}C_{13}$ (where M denotes transition metals and C denotes carbons). Theoretical calculations can shed light on the underlying causes of favored stoichiometry of these met-car and nanocrysals. By comparing the calcualted conformations and vibrational spectra, we can compare our results with limited experimental observations and gauge the robustness of our our methods.

C.8. Timetable and educational impact

The timetable and strategy for carrying out our research plan is given below.

References cited

(1) Greenwood, G. “Revisiting the complexity of finding globally minimum energy configurations in atomic clusters,” *Z. Physik. Chemie* **1999**, *211*, 105 - 114.

(2) Andricioaei, I.; Straub, J. E. “Finding the needle in the haystach: algorithms for conformational optimization,” *Computers in Physics* **1996**, *10*, 449 - 454.

(3) Wille, L.; Vennik, J. “Computational complexity of the ground-State determination of atomic clusters,” *J. Phys. A.* **1985**, *18*, L419 - L422.

(4) Wales, D. J.; Scheraga, H. A. “Global optimization of clusters, crystals, and biomolecules,” *Science* **1999**, *285*, 1368 - 1372.

(5) Bäck, T. *Evolutionary algorithms in theory and practice*; Oxford University Press: New York, 1996.

(6) Hartke, B. “Global geometry optimization of clusters using genetic algorithms,” *J. Phys. Chem.* **1993**, *97*, 9973 - 9976.

(7) Zeiri, Y. “Prediction of the lowest energy structure of clusters using a genetic algorithm,” *Phys. Rev. E* **1995**, *51*, R2769 - R2772.

(8) Judson, R.; Colvin, M.; Meza, J.; Huffer, A.; Gutierrez, D. “Do intelligent configuration search techniques outperform random search for large molecules?,” *Int. J. Quantum Chem.* **1992**, *44*, 277 - 290.

(9) Tuffery, P.; Etchebest, C.; Hazout, S.; Lavery, R. “A critical comparison of search algorithms applied to the optimization of protein side-chain conformations,” *J. Comput. Chem.* **1993**, *14*, 790 - 798.

(10) Hoare, M. “Structure and dynamics of simple microclusters,” *Adv. Chem. Phys.* **1979**, *40*, 49 - 135.

(11) Moret, M. A.; Pascutti, P. G.; Bisch, P. M.; Mundim, K. C. “Stochastic molecular optimization using generalized simulated annealing,” *J. Comput. Chem.* **1998**, *19*, 647 - 657.

(12) Ma., J. e. a. *J. Chem. Phys.* **1993**, *99*, 4024.

(13) Ma., J. e. a. *J. Chem. Phys.* **1994**, *101*, 533.

(14) Niesse, J. A.; Mayne, H. R. “Global optimization of atomic and molecular clusters using the space-fixed modified genetic algorithm method,” *J. Comp. Chem.* **1997**, *18*, 1233 - 1244.

(15) Zeiri, Y. “Study of the lowest energy structure of atomic clusters using a genetic algorithm,” *Comput. Phys. Comm.* **1997**, *103*, 28 - 42.

(16) Michaelian, K. “A symbiotic algorithm for finding the lowest energy isomers of large clusters and molecules,” *Chem. Phys. Lett.* **1998**, *293*, 202 - 208.

(17) Chaudhury, P.; Bhattacharyya, S. P. “Locating critical points on multi-dimensional surfaces by genetic algorithm: test cases including normal and perturbed argon clusters,” *Chem. Phys.* **1999**, *241*, 313 - 325.

(18) Deaven; Ho *Phys. Rev. Lett.* **1995**, *75*, 288.

(19) Wolf, M. D.; Landman, U. “Genetic algorithms for structural cluster optimization,” *J. Phys. Chem. A* **1995**, *102*, 6129 - 6137.

(20) Greenwood, G.; Liu, Y.-P. “Finding low energy conformations of atomic clusters using evolution strategies,” *Proc. Evol. Prog.* **1998**, *VII*, 493 - 502.

(21) Bäck, T.; Hoffmeiser, F. “Basic aspects of evolution strategies,” *Statistics & Computing* **1994**, *4*, 51 - 63.

(22) Vitek, V. "Pair potentials in atomistic computer simulations," *MRS Bulletin* **1996**, *21*, 20 - 23.
(23) Raeker, T. J.; DePristo, A. E. *Intern. Rev. Phys. Chem.* **1991**, *93*, 1.
(24) Foiles, S. M. "Embedded-atom and related methods for modeling metallic systems," *MRS Bulletin* **1996**, *21*, 24 - 28.
(25) Ng, K.-O.; Vanderbilt, D. "Structure and oxidation kinetics of the Si(1000)-SiO_2 interface," *Physical Review B* **1999**, *59*, 10132 - 10137.
(26) Gazón, I. I.; Machaelian, K.; Beltrán, M. R.; Posada-Amarillas, A.; Ordejón, P.; Artacho, E.; Sánchez-Portal, D.; Soler, J. M. "Lowest energy structures of gold nanoclusters," *Phys. Review. Lett.* **1998**, *81*, 1600 - 1603.
(27) Bacelo, D. E.; Binning, R. C. J.; Ishikawa, Y. "Ab initio Monte Carlo simulated annealing study of $HCl(H_2O)n$(n = 3, 4) clusters," *J. Phys. Chem. A* **1999**, *103*, 4631 - 4641.
(28) Field, M. J. "Global optimization using ab initio quantum mechanical potentials and simulated annealing of the classical Liouville equation," *J. Chem. Phys.* **1995**, *103*, 3621 - 3628.
(29) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Liu, K. G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gompers, R.; Martin, R. L. F., D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. "Gaussian 98," In Gaussian, Inc.: Pittsburgh, 1998; pp .
(30) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. J.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. "GAMESS," *J. Comput. Chem.* **1993**, *14*, 1347 - 1363.
(31) Diver, D. A. "Applications of genetic algorithms to the solution of ordinary differential equations," *J. Phys. A* **1993**, *26*, 3503 - 3513.
(32) Zeiri, Y.; Fattal, E.; Kosloff, R. "Application of genetic algorithm to the calculation of bound states and local density approximations," *J. Chem. Phys.* **1995**, *102*, 1859 - 1862.
(33) Department, I. "The 1998 Nobel Prize in Chemistry in the area of quantum chemistry to The 1998 Nobel Prize in Chemistry in the area of quantum chemistry to Walter Kohn and John Pople," In *Royal Swedish Academy of Sciences*; Stockholm, Sweden, 1998; pp .
(34) Anderson, J. B. "Monte Carlo methods in electronic structure," *Annu. Rev. Phys. Chem.* **2000**, *51*,
(35) Luchow, A. "Monte Carlo methods in electronic structure," *Annu. Rev. Phys. Chem.* **2000**, *51*,
(36) Knickelbein, M. B.; Yang, S.; Riley, S. J. "Near-threshold photoionization of nickel clusters: Ionization potentials for Ni_3 to Ni_{90}," *J. Chem. Phys.* **1990**, *93*, 94 - 104.
(37) Castro, M.; Jamorski, C.; Salahub, D. R. "Structure, bonding, and magnetism of small Fe_n, Co_n, and Ni_n clusters, n < o r= 5," *Chem. Phys. Lett.* **1997**, *271*, 133 - 142.
(38) Weber, S. E.; Jena, P. "Isomerism of Ni cluster anions and its role on photodetachment spectroscopy," *Chem. Phys. Lett.* **1997**, *281*, 401 - 406.
(39) Guo, B. C.; Kerns, K. P.; Castleman, A. W., Jr "$Ti_8C_{12}^+$-metallo-carbohedrenes: a new class of molecular clusters?," *Science* **1992**, *255*, 1411 - 1413.
(40) Guo, B. C.; Kerns, K. P.; Castleman, A. W., Jr "Metallo-carbohedrenes ($M_8C_{12}^+$ - (M=V, Zr,Hf, and Ti)): a class of stable molecular cluster ions," *Science* **1992**, *256*, 515 - 516.

(41) Pilgrim, J. S.; Duncan, M. A. “Beyond metallo-carbohedrenes: growth and decomposition of metal-carbon nanocrystals,” *J. Am. Chem. Soc.* **1993**, *115*, 9724 - 9727.