

Multi-objective particle swarm optimisation for molecular transition state search

Hettenhausen, Jan; Lewis, Andrew; Chen, Stephen; Randall, Marcus; Fournier, Rene

Published in:

Evolve: A bridge between probability, set oriented numerics, and evolutionary computation II

DOI:

[10.1007/978-3-642-31519-0_27](https://doi.org/10.1007/978-3-642-31519-0_27)

Licence:

Unspecified

[Link to output in Bond University research repository.](#)

Recommended citation(APA):

Hettenhausen, J., Lewis, A., Chen, S., Randall, M., & Fournier, R. (2012). Multi-objective particle swarm optimisation for molecular transition state search. In O. Schutze, CAC. Coello, AA. Tantar, E. Tantar, P. Bouvry, P. DelMoral, & P. Legrand (Eds.), *Evolve: A bridge between probability, set oriented numerics, and evolutionary computation II : Advances in intelligent systems and computing* (Vol. 175, pp. 415-430). (Advances in Intelligent Systems and Computing; Vol. 175). Springer. https://doi.org/10.1007/978-3-642-31519-0_27

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

For more information, or if you believe that this document breaches copyright, please contact the Bond University research repository coordinator.

Multi-Objective Particle Swarm Optimisation for Molecular Transition State Search

Jan Hettenhausen, Andrew Lewis, Stephen Chen, Marcus Randall and René Fournier

Abstract This paper describes a novel problem formulation and specialised Multi-Objective Particle Swarm Optimisation (MOPSO) algorithm to discover the reaction pathway and Transition State (TS) of small molecules. Transition states play an important role in computational chemistry and their discovery represents one of the big challenges in computational chemistry. This paper presents a novel problem formulation that defines the TS search as a multi-objective optimisation (MOO) problem. A proof of concept of a modified multi-objective particle swarm optimisation algorithm is presented to find solutions to this problem. While still at a prototype stage, the algorithm was able to find solutions in proximity to the actual TS in many cases. The algorithm is demonstrated on a range of molecules with qualitatively different reaction pathways. Based on this evaluation, possible future developments will be discussed.

Jan Hettenhausen

Institute for Integrated and Intelligent Systems, Griffith University, Australia e-mail: j.hettenhausen@griffith.edu.au

Andrew Lewis

Institute for Integrated and Intelligent Systems, Griffith University, Australia e-mail: a.lewis@griffith.edu.au

Stephen Chen

School of Information Technology, York University, Canada e-mail: sychen@yorku.ca

Marcus Randall

Faculty of Business, Technology and Sustainable Development, Bond University, Australia e-mail: mrandall@bond.edu.au

René Fournier

Department of Chemistry, York University, Canada e-mail: renef@yorku.ca

1 Introduction

Many chemical compounds can exist in a number of different forms, or isomers, each having the same molecular formula but different physical and chemical properties, due to their different molecular structure. The process of isomerisation, changing a molecule from one structure to another, involves passing through an intermediate or *transition state*. This transition state is the unstable intermediate state at which the molecule has the highest potential energy between the two, stable, low-energy isomers. Figure 1 illustrates a reaction path from reactant via transition state to product on the example of hydrogen cyanide.

TS play an important role in a variety of areas of theoretical chemistry. Accurate knowledge of geometry and potential energy of TS is, for example, required to understand chemical reactivity, reaction rates and mechanisms as well as analysing barrier lowering in catalysed reactions [17, 20]. For example, the Arrhenius law, $k = Ae^{-E_a/RT}$, gives a good fit to the observed rate constant $k(T)$ of chemical reactions. The parameters A and E_a are readily obtained if one knows the transition state: E_a is the energy difference between the saddle point and the local minimum, and A can be estimated from the curvature (energy second derivatives) at the saddle point. As transition states cannot be directly observed and even their approximation by experimental means is difficult, determining transition states computationally has been a field of ongoing research in computational chemistry.

Algorithmically this problem is usually approached by approximating the reaction pathway on the Potential Energy Surface (PES) of the molecule [35]. While generally not known in its entirety, individual points on the PES can be acquired by solving the Schrödinger equation for a specific conformation of the nuclei of the molecule in question. The PES can be pictured as a hilly landscape, in which the

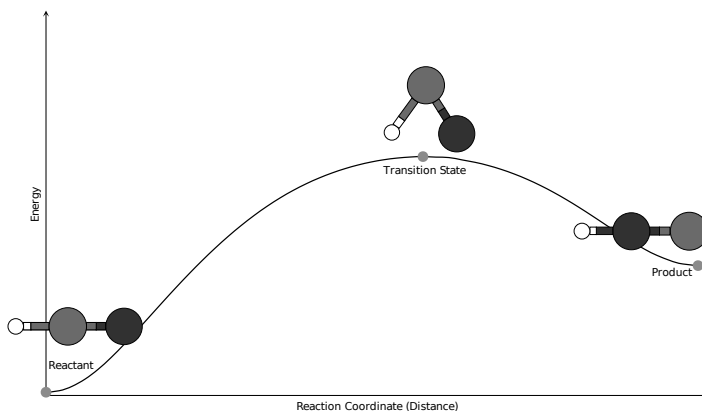


Fig. 1 Approximate illustration of a reaction path showing corresponding conformations of hydrogen cyanide.

stable isomers are valley bottoms (local minima). The transition states then are the passes, i.e. the lowest crossing point over the mountains between two valleys [35].

While for macro-molecules, such as proteins, the search for the various local minima of a molecule is still a challenging problem, it is considerably more difficult to obtain the saddle points even for small molecules. This is in particular due to the fact that TSs are first order saddle-points of the PES and therefore require significantly more complicated approaches than are necessary to identify local minima.

The process of identifying the TS for a given local minimum of a molecule (reactant) comprises five distinct steps (see, for example, Fournier *et al.* [17]):

1. Locate the reactant on the PES. Depending on the problem definition it may also be necessary to identify the second local minimum (product) which the reaction will result in.
2. Determine a good approximation for the TS. This is often based on finding an approximation to the reaction pathways and identifying a candidate TS on it.
3. Refine the TS by performing a constraints minimisation of the norm of gradient. Appropriate constraints need to be chosen to ensure that the search does not converge to either of the local minima.
4. Calculate the Hessian matrix and frequencies of the candidate TS to verify that the point is, in fact, a first order saddle-point.
5. Map out the intrinsic reaction coordinate.

Step 2 usually represents the most challenging aspect of identifying a TS. Once an approximate TS is identified, it is usually straightforward to find the actual TS. The algorithm presented in this paper addresses the problem of finding such an approximate TS with sufficient accuracy.

A large percentage of TS finding algorithms can be classified as either Shallowest Ascent (SA) or Path Interpolation (PI) methods. SA methods start at the reactant and iteratively try to follow the shallowest ascent out of the valley until the norm of gradient becomes zero and the TS is found. One of the first practical approaches to finding the TS was developed by Cerjan and Miller [5] in 1981. Other approaches were published by, for example, Baker [1] and Maeda and Kono [31]. To identify the shallowest ascent the second derivative matrix of the PES needs to be calculated in every step making SA approaches computationally more expensive than PI methods. Their advantage, however, is that only the first minimum needs to be known. SA approaches therefore solve a more general and more difficult problem than PI approaches, which require *a priori* knowledge of both minima, i.e. the reactant and the product. PI approaches start with a coarse approximate path between the two minima and gradually refines it using constraint optimisation. Common PI methods include nudged elastic band [15, 22, 23, 41], the string method (SM)[13], Growing String Method (GSM) [3, 4, 34, 38] and modified GSM [19, 20]. Further methods outside these two categories include the Fast Marching Method (FFM) [2, 11, 12] and Dewar's method [10]. Comprehensive overviews of existing methods can be found in del Campo and Köster [9] and Schlegel [35].

To the authors' knowledge, the first non-deterministic method was developed by Fournier *et al.* [17]. Their approach combines the ideas of SA with those of

common metaheuristics by using a small group of “climbers” that traverse an approximate reaction pathway towards the TS. Climbing direction and movements are thereby guided by a set of parameters acquired from the global information of the path gathered by all climbers. The performance of each climber is assessed with an aggregated fitness function based on energy, distance and changes in bonds. An interesting feature of this approach is that the climbers can continue after finding an approximate TS, allowing the algorithm to discover successive minima and TS without having to restart the algorithm from a new position.

Based on the idea that the reaction pathway follows the shallowest path out of the valley around the minimum, this paper presents a novel problem formulation for TS search treating the problem as a multi-objective optimisation problem of the following form:

$$\begin{aligned} \text{minimise } \mathbf{f}(\mathbf{x}) &= \{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})\} \\ f_k : \mathbb{R}^n &\rightarrow \mathbb{R}, \forall \mathbf{x} \in \mathfrak{F}, \mathfrak{F} \subseteq \mathfrak{S} \subset \mathbb{R}^n \end{aligned} \quad (1)$$

where \mathfrak{F} is the feasible space within the search space \mathfrak{S} based on a set of given constraints. A detailed discussion on this formulation and its justification will be provided in Section 2. Using this problem formulation then allows finding the reaction pathway and TS using multi-objective metaheuristics, a class of algorithms highly successful on a wide range of problems in science and engineering (e.g. [21, 24, 25, 37]). While non-deterministic, metaheuristics have the advantage of being inherently parallel and do not require structural knowledge of the problem domain.

As a solution technique, a prototype of a modified multi-objective particle swarm optimisation algorithm will be used. Particle Swarm Optimisation (PSO) takes its inspiration from the behaviour of swarming and flocking animals [29]. PSO models candidate solutions as a populations of particles that can either freely or with constraints move in the parameter space of the problem. The velocity of the particles is thereby governed by a velocity equation usually based on momentum, a particle specific memory of the best solution visited by that particle and a social component giving each particle knowledge of the best solution known to the swarm. Sometimes a turbulence component is used to improve diversity [39]. A variety of PSO adaptations exist to allow simultaneous optimisation of multiple objectives [6, 16, 32]. The primary difference to single-objective PSO is that multi-objective PSO (MOPSO) incorporates an archive that holds knowledge of a set of known optimal compromise solutions. Optimality in this context is defined on the basis of the *dominance* relation where one parameter vector \mathbf{x}_1 is said to dominate another parameter vector \mathbf{x}_2 if and only if

$$\begin{aligned} f_k(\mathbf{x}_1) &\leq f_k(\mathbf{x}_2), \forall k \in 1, \dots, n_k \\ \exists k \in 1, \dots, n_k &: f_k(\mathbf{x}_1) < f_k(\mathbf{x}_2) \end{aligned} \quad (2)$$

The set of all globally non-dominated parameter vectors form the Pareto-optimal set or

$$\mathfrak{P}^* = \{\mathbf{x}^* \in \mathfrak{F} \mid \nexists \mathbf{x} \in \mathfrak{F} : \mathbf{x} \prec \mathbf{x}^*\} \quad (3)$$

and the set of corresponding objective vectors is called Pareto-front or

$$\mathfrak{PF}^* = \{\mathbf{f}(\mathbf{x}^*) \mid \mathbf{x}^* \in \mathfrak{P}^*\}. \quad (4)$$

which could be described as the set of optimal compromise solutions. For each solution no objective can be further improved without worsening another.

On the basis of the MOPSO algorithm a method will be presented that can successfully converge to the minimum energy pathway. However, a number of adaptations and additions were necessary to address the specific challenges of identifying the reaction on the PES in the context of a multi-objective optimisation problem. These will be discussed in Sections 2 and 3.

The algorithm was tested on a set of small molecules with different reaction pathways. The results of these tests will be discussed in Section 4. Concluding remarks and a discussion of future work and extensions to the method presented can be found in Section 5.

2 The Multi-Objective Optimisation Problem Formulation

This paper introduces a novel way of describing the TS search problem by defining the minimum energy pathway section-wise as Pareto-fronts, transforming it into a multi-objective optimisation problem. Based on this formulation, algorithms can solve the TS search problem without knowledge of the second minimum as required by PI approaches. It further makes the problem formulation of the more general TS search problem independent of knowledge of the gradient.

Essential to the problem formulation presented is that the reaction pathway will always follow the path of the shallowest ascent and points on the path will have the lowest potential energy of all conformations with the same distance from minimum 1. Furthermore, the potential energy is subject to a steady increase along the reaction path [17]. This allows defining the reaction path as a Pareto-front in a multi-objective optimisation problem (Equation 1) that minimises potential energy and maximises distance. In particular, it allows solving the more generic TS search problem requiring only *a priori* knowledge of the first minimum, i.e. the reactant. Based on the reaction pathway the approximate TS can then be identified in a second step. The metric for the first objective, molecular potential energy, is derived from solutions of the electronic Schrödinger equation [36] using quantum chemistry software (Gaussian 03) [18]. The details of this computation are beyond the scope of this paper. The metric for the second objective is the distance between the conformation of a trial solution and that of the original reactant, "minimum 1".

A common definition for the distance between two molecular conformations with identical atoms is the least root mean square deviation (IRMSD). The IRMSD can be calculated by shifting each conformation to its geometrical centroid, finding its optimal alignment and calculating the root mean square deviation (RMSD), defined

as

$$RMSD(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{1}{N_{atoms}} \sum_{i=1}^N |\mathbf{x}_i - \mathbf{y}_i|^2} \quad (5)$$

for corresponding sets of atoms \mathbf{x} and \mathbf{y} of magnitude N_{atoms} . Each atom is represented by its Cartesian coordinates. The optimal alignment of the molecules can be efficiently computed with either the Kabsch algorithm [28, 27] or using quaternions [7].

Defining distance as a scalar value, such as IRMSD, practically allows quantifying all combined atom movements in a way meaningful to the problem domain. However, it does not provide any insight into the direction of the movement, i.e. whether it is towards or away from the TS. Furthermore, for any distance an infinite number of conformations with different (or possibly identical) potential energies can be constructed. While this does not affect the definition of the reaction path as Pareto-front as such, it poses a challenge to algorithms trying to converge towards the Pareto-front by recombining individuals from its population. To ensure convergence, several counter-measures were developed as part of this research, using the specific knowledge available in TS search. Rather than random initialisation of parameters, the initial population is generated by conducting a Monte-Carlo inspired search around the minimum. During this stage, a set of candidate solutions is created by applying small random movements to some atoms of minimum 1. To ensure sufficient diversity in the population, a larger than required set of candidates is generated and its fittest members, by energy and distance are chosen. Candidate solutions c for which

$$RMSD(c, nn_i) \gg \|RMSD(c, min1) - RMSD(nn_i, min1)\|, i \in [1, 2] \quad (6)$$

with nn_1 and nn_2 being the nearest neighbours as seen from minimum 1, are rejected from the approximate front. This further ensures fast convergence towards the reaction pathway by removing solutions likely to be on a different pathway. In particular for symmetric or partly symmetric molecules, it further proved beneficial to align candidate molecules to each other at regular intervals but particularly at the initialisation stage. This is to ensure that identical but rotated conformations do not slow down the convergence of the algorithm. Since rotations are calculated as part of determining the IRMSD, this process does not incur any additional computational costs and can therefore easily be performed in each iteration.

The augmented, multi-objective algorithm is essentially capable of climbing up the reaction path towards the TS, since the reaction path corresponds with the Pareto front while the search is concentrated between minimum 1 and the TS. In the region of the TS, the search must be modified to successfully identify the TS. The algorithm first needs to detect when it has reached the region of the TS and restrict further expansion of the front accordingly. One approach is to test candidate conformations for a norm of gradient close to zero [17]. However, this approach is successful only with solutions *very* close to the TS and the diversity inherent in the MOPSO swarm means many particles may be beyond the TS before this test can confirm its location. This may lead to potential solutions being prematurely removed from the archive.

Instead, particles can detect when they are near or beyond the TS by observing changes in gradient direction, which generally points coarsely towards the nearest local minimum. If the direction of the gradient changes by angle $\alpha \geq 90^\circ$, then it can be assumed that the gradient is pointing towards a new local minimum. The algorithm prevents such solutions entering the archive and corrective measures can be taken to pull solutions back into the feasible region. Refinement of solutions close to the TS then usually leads to a candidate TS within very few iterations.

A steepest-descent search (or comparable algorithm) starting from the conformation with gradient pointing towards the second minimum will quickly discover the second half of the reaction pathway. Some refinements to this method will have to be made as it occasionally fails and terminates the search prematurely. While included in the current algorithm, this feature was not used for the evaluation discussed in Section 4.

A final consideration in defining TS approximation as a multi-objective optimisation problem is an appropriate coordinate system for the parameter space. A range of alternative coordinate systems exist to describe molecules. Apart from defining the location of individual atoms by their Cartesian coordinates, some coordinate systems are based on bond lengths, valence angles etc. Such non-Cartesian coordinate systems are particularly relevant for macro-molecules, where the “curse of dimensionality” makes the use of Cartesian coordinates infeasible, as $3 \cdot N_{Atoms}$ coordinates are necessary to describe a molecule. For small molecules, as are common in the TS search literature, Cartesian coordinates provide a suitable coordinate system for algorithmic analyses of reaction pathways. For the purposes of evaluating the proposed method, molecules with between three and seven atoms were chosen.

As the algorithm described in this paper is, at its current stage, mainly targeted at small molecules, Cartesian coordinates are the natural choice to represent the parameter space. Cartesian coordinates provide a strong coupling to the potential energy surface and allow analysing a wide range of molecules easily and without any adaptation [35]. This allows definition of the feasible space of Equation 1 as $\mathcal{S} \subset \mathbb{R}^{3N_{Atoms}}$. However, in principle the algorithm described in this paper could be adapted to other coordinate systems with only minor adjustments to the algorithm’s configurable parameters.

In the following Section MOPSO-specific aspects of the proposed TS search approach will be discussed.

3 Application of MOPSO

As outlined in Section 1, a MOPSO algorithm comprises a set number of particles that remains constant throughout the optimisation process. Particles normally start at random positions in parameter space and are allowed to move in parameter space based on a velocity equation. This velocity is based on the particle’s previous velocity, a selected globally non-dominated solution and a solution from the particle’s memory of non-dominated solutions. Each of these is weighted with a constant and

a random factor, generally aiming to over- or undershoot the respective component with about equal probability. In each iteration, newly found non-dominated solutions are added to the archive, solutions no longer non-dominated are removed. By pulling the particles towards well performing parameter vectors, the swarm usually converges towards the Pareto-front.

However, the nature of the TS search problem renders attempts to naïvely apply MOPSO or the most common multi-objective genetic algorithm, NSGA-2 [8], unsuccessful. Even given a starting population within a small range around minimum 1, neither method could converge to the minimum energy pathway within a reasonable number of iterations. Therefore, based on the considerations discussed in Section 2, a number of changes were made to MOPSO to allow it to move along the reaction path and identify the TS. These modifications comprise a Monte-Carlo inspired initialisation of the population, an adjusted guide particle selection mechanism, an adapted velocity update and the inclusion of the available knowledge of the gradient for each conformation.

To create a viable starting population a set of candidate conformations is created based on the known minimum 1. The size of this set is chosen as twice the number of particles in the swarm. Candidate conformations are created by adding a random movement to the position of the atoms of minimum 1. With each candidate created the magnitude of random dispersion increases slightly, leading to a set of candidate particles with a range of distances. From these particles the best half, measured by non-dominance and, if not sufficient for the swarm size, proximity to non-dominated solutions, is chosen as the initial population. Factors are chosen sufficiently small to keep the population within a range significantly smaller than the anticipated distance to the TS.

With the initial population within a small distance from minimum 1, the swarm will have to move up the reaction path and thereby trade off convergence towards the actual path, i.e. the Pareto-front, and progress towards the TS. Particularly with the inclusion of a gradient component, which will be discussed below, it is necessary to ensure that the swarm does not collapse back towards the minimum. This can be achieved by employing an appropriate selection mechanism for the global best guide. A wide range of guide particle selection mechanisms exist [16, 33, 6, 30]. However, all these approaches target algorithms approaching the Pareto-front top down and aim at a good diversity and spread of the solutions. The algorithm presented is designed to move up the reaction pathway, requiring a swarm pushing forward while at the same time attempting not to diverge away from the front. Adaptive grid and crowding distance based methods were thereby less successful in terms of convergence speed than roulette wheel selection of archived solutions with a strong bias towards solutions in the foremost third of the Pareto-front by distance. The reason for this is that the initial population is already reasonably close to a small section of the Pareto-front. From there it needs to follow the Pareto-front towards the TS. Adaptive grid and crowding distance based selection mechanisms are highly successful at achieving good diversity and spread across the entire Pareto-front, when the swarm moves top down towards it. This particular variation of MOPSO, however, requires a swarm that moves along the Pareto-front without diverging from

it but with a constant pressure to move forward. Roulette wheel selection provides a good mechanism to achieve this pressure without moving so fast that the swarm would diverge from the front. For the purposes of this paper non-dominated solutions not in the foremost third of the front were chosen with a probability of 10 percent.

The size of the archive is limited to 40 solutions. If this size is exceeded, solutions with the highest crowding distance are removed. However, the crowding distance for solutions in the foremost third is biased with a factor of two, effectively allowing denser crowding for particles with large distances.

An interesting aspect of the TS search problem is the availability of the gradient. While a numerical approximation of the Hessian is infeasible due to its inherent computational complexity, incorporation of the gradient allows performing correcting steps to move particles and archived solutions closer to the Pareto-front. To achieve this, the velocity update equation was amended with a gradient-based term, resulting in the following velocity equation:

$$\mathbf{v}_{t+1} = w_t \cdot \mathbf{v}_t + c_1 r_1 (\hat{\mathbf{y}}_{\text{pbest}} - \mathbf{x}_t) + c_2 r_2 (\hat{\mathbf{y}}_{\text{gbest}} - \mathbf{x}_t) + c_3 r_3 \mathbf{g}_{\text{comp},t} \quad (7)$$

where \mathbf{v}_t represents the previous velocity (momentum), $\hat{\mathbf{y}}_{\text{pbest}}$ and $\hat{\mathbf{y}}_{\text{gbest}}$ denote the personal and global best respectively and \mathbf{x}_t stands for the particle's parameter vector at the current iteration t . The factors w_t , $c_{1,2,3}$ and $r_{1,2,3}$ represent the constant and random weights. The gradient based component is represented by \mathbf{g}_{comp} . It is defined as

$$\mathbf{g}_{\text{comp},t} = \dot{\mathbf{x}}_t - \frac{\dot{\mathbf{x}}_t \cdot (\mathbf{x}_t - \mathbf{min}_1)}{|\mathbf{x}_t - \mathbf{min}_1|^2} \cdot (\mathbf{x}_t - \mathbf{min}_1) \quad (8)$$

and serves the purpose of pushing particles towards the Pareto-front while minimising the gradient's drag towards the minimum. An additional normalisation step is added should the magnitude of the \mathbf{g}_{comp} become greater than the magnitude of the combined other velocity components. The total velocity is furthermore subjected to clamping [14], scaling all velocities that exceed a set magnitude to a predefined maximum velocity. This step proved to improve convergence towards the TS by keeping particles close to the Pareto-front. A value close to the maximum value for random movements in the initialisation step is chosen as the maximum velocity.

In combination, these modifications allow the swarm to quickly move up the reaction path without diverging too far from it. However, streamlining the swarm in this way bears the risk of losing diversity in the population. Such a loss in diversity can significantly slow down the progress of the search. To counteract that, a turbulence component [29, 40] was added which, with a certain probability, adds a random mutation to particles. The probability was chosen as $\frac{2}{\text{dim}\mathbf{x}}$, where $\text{dim}\mathbf{x}$ represents the number of parameters, selecting on average two components of the parameter vector for mutation. The mutation for each selected component is uniformly randomly chosen from the interval $[-v_{\text{max}}, v_{\text{max}}]$.

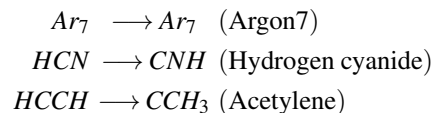
A final optional step was the possibility for archived solutions to perform a small number of gradient descent steps. When this option is included, archived solutions perform a small number of gradient descent steps (usually around 3), after each

of which the new candidate solution is added to the archive iff non-dominated. To minimise the computational costs, this step is only performed every few iterations and only on solutions that have either been newly added to the archive or where all the gradient descent steps in the previous round were successfully added to the archive. Doing these additional gradient descent steps only marginally improves the convergence and quality of the TS approximation. However, it improves the quality of the overall pathway found as part of the search.

In summary, MOPSO can be adapted to find an approximation to a chemical reaction pathway that leads up to a TS. Modifications are necessary to ensure that MOPSO stays in an area where the MOO problem definition is applicable. Primarily this is achieved by using available knowledge of the reactant to generate the initial population and applying clamping to ensure that the movements of the swarm remain relatively small. To improve the performance of the swarms movement along the path, a guide particle selection mechanism with a bias for solutions further away from the reactant is used. Other performance related changes include the incorporation of the gradient to improve convergence towards the minimum energy path and a turbulence component to improve diversity within the swarm while operating under these constraint conditions. The following section will discuss the specific parameters chosen for the algorithm as well as present the results of a number of test runs on a variety of molecules.

4 Results

For the evaluation of the method, three molecules with qualitatively different reaction pathways were chosen. The specific molecules chosen, the number of test runs and the metrics reported were chosen to allow comparison with related literature [17]. At the current stage most emphasis was put on the development of a suitable problem definition and MOPSO-based algorithm. Plans for qualitative comparisons with other algorithms as well as future work on making the approach more competitive will be discussed in Section 5. The following molecules were chosen for the evaluation:



The argon 7 cluster is a well understood molecule often used for benchmarking. As a noble gas, the potential energy and gradient of argon clusters can be calculated quite cheaply using the Lennard-Jones potential [26]. The Lennard-Jones potential is applicable to all molecules comprising only neutral atoms and provides a mean of accurately approximating the potential energy of a given conformation at negligible computational costs. The potential energy of the other molecules in this evaluation

was calculated using Gaussian 03 [18]. Gaussian is commercial software for computational quantum chemistry. Like the Lennard-Jones potential it allows calculating the potential energy and the gradient for a given molecular conformation, but at substantially higher computational costs. The distance objective for all molecules was calculated as outlined in Section 2. The results from all test runs are summarised in Table 1. Shown are the actual distance between minimum and TS, the difference in potential energy between the solution found and the actual TS as well as the differences in distance, measured by distance of the candidate to the TS and by the difference in distance of the TS and minimum and candidate and minimum. Following other publications [17], the energy is measured in *kcal/mol*, distances in angstrom (\AA). In general, the solutions achieved are quite competitive with respect to their potential energy and in many cases also with respect to their distance to the actual TS. The following paragraphs will briefly discuss the algorithm's behaviour on the different molecules.

Argon 7 – The argon 7 cluster starts with 5 of its atoms in a regular pentagon and the two remaining atoms above and below the centre of the plane spanned by the pentagon (Figure 2(a)). Towards the TS one of the pentagon atoms (due to the symmetry this can be any one of the five) moves away from the centre while the other pentagon atoms move closer towards each other, making the pentagon irregular and giving it a pointy shape. The two atoms above and below the centre take a slightly angled position towards each other (Figure 2(b)). Towards the second minimum the pentagon shape is broken up altogether and the atoms take the shape illustrated in Figure 2(c).

The symmetry of the molecule poses challenges to the algorithm until solutions align. On argon 7 the front will occasionally move away from the actual reaction pathway but usually converge back to its vicinity quickly. However, in one of the runs the algorithm followed a different reaction coordinate ending up with a path at a significant distance from the desired path (see table 1 run 3 of argon 7).

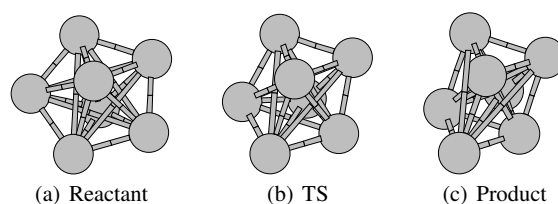


Fig. 2 Minima and TS of Argon 7 (Ar_7)

Hydrogen cyanide – Hydrogen cyanide starts with all three atoms in a line. Towards the transition state, Hydrogen and Nitrogen rotate around the carbon atom towards each other, forming a triangle and a bond forms between the hydrogen and nitrogen atoms. As the molecular structure approaches the second minimum, the bond between the hydrogen and carbon atoms breaks and the molecule straightens

out, forming a straight line again. The process is illustrated in Figure 3. Test runs on *HCN* generally achieved good performance. An approximate transition state could be acquired in usually less than 15 iterations and with a difference in energy of only 0.001605675 kcal/mol on average.

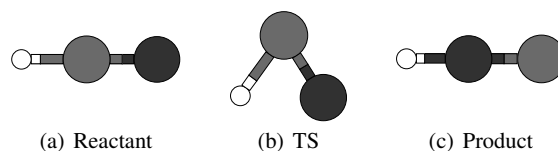


Fig. 3 Minima and TS of Hydrogen cyanide (*HCN*)

Acetylene – At minimum 1 of acetylene all four atoms are aligned in a straight line with the two carbon atoms in the middle. Towards the TS one of the hydrogen atoms rotates around the carbon atom it is bonded with, ending up in a position forming a triangle with the two carbon atoms. Towards the second minimum its bond changes over to the other carbon atom, eventually ending up in a Y-shape with both hydrogen atoms bonded with one of the carbon atoms. The three relevant states are illustrated in Figure 4. Acetylene proved challenging in the tests performed. The algorithm could successfully perform the rotation of the hydrogen but had a strong tendency to rotate both hydrogen atoms but one of them usually to a lesser degree. While the resulting solutions differ from the actual TS due to the second hydrogen atom, they could successfully perform the rotation and find a solution with only a small difference in potential energy to the TS. In the four test runs MOPSO successfully found one solution very close to the actual transition state and with the second hydrogen in the correct position. The three other runs found solutions with a similar potential energy but a significant distance away from the TS. Empirically this can be explained with the relatively small increase in potential energy arising from the second rotation.

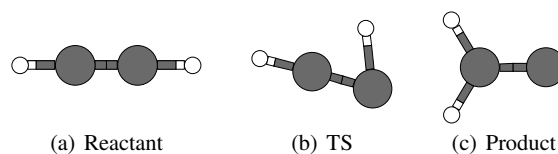


Fig. 4 Minima and TS of Acetylene (*HCCH*).

In summary, the algorithm could successfully navigate the potential energy surface, in most cases coming up with solutions close to the TS. Currently the algorithm may not reliably determine the transition reliably without the help of a human decision maker. However, user interaction could also be a pathway to signif-

Table 1 Overview of results achieved with the proposed algorithm on a range of molecules. A population of 40 particles was used for each run.

Molecule	Dist $min1 \rightarrow ts$	$\delta E_a(kcal/mol)$	$Dist_{ts}(\text{\AA})$	$\delta Dist_{min}(\text{\AA})$	Iterations
<i>Ar₇</i>	0.1911537	0.076588	0.018438	0.006651	48
<i>Ar₇</i>	0.1911537	0.028173	0.012577	0.000053	73
<i>Ar₇</i>	0.1911537	0.245090	0.117432	0.014742	89
<i>Ar₇</i>	0.1911537	0.039635	0.028680	0.008752	88
<i>HCN</i>	0.3666355	0.001007	0.005780	0.008611	22
<i>HCN</i>	0.3666355	0.001568	0.013504	0.003783	15
<i>HCN</i>	0.3666355	0.001341	0.031942	0.032877	16
<i>HCN</i>	0.3666355	0.001813	0.015113	0.032877	10
<i>HCCH</i>	0.463125	0.091507	0.639619	0.445840	10
<i>HCCH</i>	0.463125	0.002149	0.311172	0.083044	23
<i>HCCH</i>	0.463125	0.013743	0.204976	0.115168	31
<i>HCCH</i>	0.463125	0.000763	0.015012	0.121488	31

icant improvements to performance. Incorporating domain knowledge would allow dynamically constraining the search space and reducing the dimensionality of the problem. For example, both HCN and HCCH could easily be constrained to only planar movements, reducing the search space by one third. Future developments will therefore consist of a combination of improvements to the algorithm itself and the incorporation of domain knowledge through user interaction. In combination, it is hoped this will allow discovery of TS on significantly larger molecules.

5 Conclusion

This paper presented a novel formulation of the Transition Search (TS) search problem by treating it as a multi-objective optimisation problem and introduced a proof of concept MOPSO-based algorithm to solve this problem. At the current stage, a reasonable success rate on a range of molecules can be reported. In terms of performance, the proof of concept MOPSO algorithm still requires more objective function evaluations to find the TS than some of the other methods. This is partially alleviated by the inherent parallelism of the approach so that a comparison in wall clock time, parallel execution of the objective function assumed, is already performing similarly to some other algorithms. Some future work will be dedicated to improving the success rate and stability of the search as well as the computational costs. In particular, the incorporation of domain knowledge could greatly aid this goal. It is hoped the incorporation of domain knowledge and dynamic constraints will enable the MOPSO based approach to become a viable alternative for larger molecules which current techniques cannot process in a satisfactory amount of time.

References

1. Jon Baker. An algorithm for the location of transition states. *Journal of Computational Chemistry*, 7(4):385–395, 1986.
2. S.K. Burger, Y. Liu, U. Sarkar, and P.W. Ayers. Moving least-squares enhanced shepard interpolation for the fast marching and string methods. *The Journal of chemical physics*, 130:024103, 2009.
3. Steven K. Burger and Weitao Yang. Quadratic string method for determining the minimum-energy path based on multiobjective optimization. *The Journal of Chemical Physics*, 124(5):054109, 2006.
4. Steven K. Burger and Weitao Yang. Sequential quadratic programming method for determining the minimum energy path. *The Journal of Chemical Physics*, 127(16):164107, 2007.
5. Charles J. Cerjan and William H. Miller. On finding transition states. *The Journal of Chemical Physics*, 75(6):2800–2806, 1981.
6. C.A.C. Coello and M.S. Lechuga. MOPSO: A Proposal for Multiple Objective Particle Swarm Optimization. In *Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on*, volume 2. IEEE Computer Society, 2002.
7. Evangelos A. Coutsias, Chaok Seok, and Ken A. Dill. Using quaternions to calculate rmsd. *Journal of Computational Chemistry*, 25(15):1849–1857, 2004.
8. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *Evolutionary Computation, IEEE Transactions on*, 6(2):182–197, 2002.
9. J.M. Del Campo and A.M. Köster. A hierarchical transition state search algorithm. *The Journal of chemical physics*, 129:024107, 2008.
10. M.J.S. Dewar, E.F. Healy, and J.J.P. Stewart. Location of transition states in reaction mechanisms. *Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics*, 80(3):227–233, 1984.
11. Bijoy Dey, Stuart Bothwell, and Paul Ayers. Fast marching method for calculating reactive trajectories for chemical reactions. *Journal of Mathematical Chemistry*, 41:1–25, 2007. 10.1007/s10910-006-9060-6.
12. Bijoy K. Dey, Marek R. Janicki, and Paul W. Ayers. Hamilton-jacobi equation for the least-action/least-time dynamical path based on fast marching method. *The Journal of Chemical Physics*, 121(14):6667–6679, 2004.
13. Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. String method for the study of rare events. *Phys. Rev. B*, 66:052301, Aug 2002.
14. Russ Eberhart, Pat Simpson, and Roy Dobbins. *Computational intelligence PC tools*. Academic Press Professional, Inc., San Diego, CA, USA, 1996.
15. R. Elber and M. Karplus. A method for determining reaction paths in large molecules: Application to myoglobin. *Chemical Physics Letters*, 139(5):375 – 380, 1987.
16. J. Fieldsend and S. Singh. A multi-objective algorithm based upon particle swarm optimisation. *Proceedings of The UK Workshop on Computational Intelligence*, pages 34–44, 2002.
17. Rene Fournier, Satya Bulusu, Stephen Chen, and Jamie Tung. Using swarm intelligence for finding transition states and reaction paths. *The Journal of Chemical Physics*, 135(10):104117, 2011.
18. M. J *et al.* Frisch. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.
19. Anthony Goodrow, Alexis T. Bell, and Martin Head-Gordon. Development and application of a hybrid method involving interpolation and ab initio calculations for the determination of transition states. *The Journal of Chemical Physics*, 129(17):174109, 2008.
20. Anthony Goodrow, Alexis T. Bell, and Martin Head-Gordon. Transition state-finding strategies for use with the growing string method. *The Journal of Chemical Physics*, 130(24):244108, 2009.
21. S.K. Goudos, Z.D. Zaharis, D.G. Kampitaki, I.T. Rekanos, and C.S. Hilar. Pareto optimal design of dual-band base station antenna arrays using multi-objective particle swarm optimization with fitness sharing. *Magnetics, IEEE Transactions on*, 45(3):1522 –1525, march 2009.

22. Graeme Henkelman and Hannes Jónsson. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *The Journal of Chemical Physics*, 113(22):9978–9985, 2000.
23. Graeme Henkelman, Blas P. Uberuaga, and Hannes Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of Chemical Physics*, 113(22):9901–9904, 2000.
24. Stefan Janson, Daniel Merkle, and Martin Middendorf. Molecular docking with multi-objective particle swarm optimization. *Applied Soft Computing*, 8(1):666 – 675, 2008.
25. Nanbo Jin and Yahya Rahmat-Samii. Particle swarm optimization for antenna designs in engineering electromagnetics. *J. Artif. Evol. App.*, 2008:9:1–9:10, January 2008.
26. J. E. Jones. On the determination of molecular fields. ii. from the equation of state of a gas. *Proceedings of the Royal Society of London. Series A*, 106(738):463–477, 1924.
27. W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 32(5):922–923, 1976.
28. W. Kabsch. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 34(5):827–828, 1978.
29. J. Kennedy and R. Eberhart. Particle swarm optimization. *Neural Networks, 1995. Proceedings., IEEE International Conference on*, 4, 1995.
30. X. Li. A non-dominated sorting particle swarm optimizer for multiobjective optimization. In *Genetic and Evolutionary Computation—GECCO 2003*, pages 198–198. Springer, 2003.
31. Satoshi Maeda and Koichi Ohno. Global mapping of equilibrium and transition structures on potential energy surfaces by the scaled hypersphere search method: applications to ab initio surfaces of formaldehyde and propyne molecules. *The Journal of Physical Chemistry A*, 109(25):5742–5753, 2005. PMID: 16833907.
32. J. Moore and R. Chapman. Application of particle swarm to multiobjective optimization. Technical report, Department of Computer Science and Software Engineering, Auburn University, 1999.
33. S. Mostaghim and J. Teich. Strategies for finding good local guides in multi-objective particle swarm optimization. In *Swarm Intelligence Symposium, 2003. SIS'03. Proceedings of the 2003 IEEE*, pages 26–33. IEEE Computer Society, 2003.
34. Baron Peters, Andreas Heyden, Alexis T. Bell, and Arup Chakraborty. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *The Journal of Chemical Physics*, 120(17):7877–7886, 2004.
35. H.B. Schlegel. Exploring potential energy surfaces for chemical reactions: an overview of some practical methods. *Journal of computational chemistry*, 24(12):1514–1527, 2003.
36. E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Physical Review*, 28(6):1049–1070, 1926.
37. A.M. Sharaf and A.A.A. El-Gammal. A multi objective multi-stage particle swarm optimization mopso search scheme for power quality and loss reduction on radial distribution system. In *International Conference on Renewable Energies and Power Quality, ICREPQ*, 2009.
38. Daniel Sheppard, Rye Terrell, and Graeme Henkelman. Optimization methods for finding minimum energy paths. *The Journal of Chemical Physics*, 128(13):134106, 2008.
39. Y. Shi and R. Eberhart. A modified particle swarm optimizer. *Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on*, pages 69–73, 1998.
40. A. Stacey, M. Jancic, and I. Grundy. Particle swarm optimization with mutation. In *Evolutionary Computation, 2003. CEC'03. The 2003 Congress on*, volume 2, pages 1425–1430. IEEE, 2003.
41. Semen A. Trygubenko and David J. Wales. A doubly nudged elastic band method for finding transition states. *The Journal of Chemical Physics*, 120(5):2082–2094, 2004.