Progress In Electromagnetics Research, PIER 75, 171–207, 2007

PHYSICAL THEORY FOR PARTICLE SWARM OPTIMIZATION

S. M. Mikki and A. A. Kishk

Center of Applied Electromagnetic Systems Research
Department of Electrical Engineering
University of Mississippi
University, MS 38677, USA

Abstract—We propose an inter-disciplinary approach to particle swarm optimization (PSO) by establishing a molecular dynamics (MD) formulation of the algorithm, leading to a physical theory for the swarm environment. The physical theory provides new insights on the operational mechanism of the PSO method. In particular, a thermodynamic analysis, which is based on the MD formulation, is introduced to provide deeper understanding of the convergence behavior of the basic classical PSO algorithm. The thermodynamic theory is used to propose a new acceleration technique for the PSO. This technique is applied to the problem of synthesis of linear array antennas and very good improvement in the convergence performance is observed. A macroscopic study of the PSO is conducted by formulating a diffusion model for the swarm environment. The Einstein's diffusion equation is solved for the corresponding probability density function (pdf) of the particles trajectory. The diffusion model for the classical PSO is used, in conjunction with Schrödinger's equation for the quantum PSO, to propose a generalized version of the PSO algorithm based on the theory of Markov chains. This unifies the two versions of the PSO, classical and quantum, by eliminating the velocity and introducing position-only update equations based on the probability law of the method.

1. INTRODUCTION

There is a close connection between the evolution of the dynamical variables in physical systems and optimization. It has been known since two hundred years that the law of motion in Newtonian mechanics can be obtained by minimizing certain functional called the "action."

Moreover, with the invention of the path integral formalism by Richard Feynman in the last century, we now know that quantum phenomena can also be described by the very same approach [1]. It turns out that most of the known physical processes can be intimately related to some sort of "critical" or deeper variational problem: The optimum of this problem leads to the equations of motion in an elegant and neat way.

Although this has been known for a long time, optimization is still considered in physics, applied mathematics, and engineering, as a *tool* used to solve some difficult practical or theoretical problems. For example, the use of optimization codes in engineering problems is mainly due to our inability to find a desired solution in reasonable time. Optimization methods provides then a faster "search" for the best performance of the system of interest. However, optimization and physical systems are really two different ways to describe the same thing. If any particle moves in the trajectory that minimizes the action of the problem, then the search for the optimum of the functional (objective function) is equivalent to find the equations of motion of the particle. Our point of view is that if a more formal analogy between physical systems and optimization algorithms can be constructed on the fundamental level, then deeper insights on both theoretical physics and the algorithm itself can be obtained.

The main purpose of this paper is twofold. First, we take the particle swarm optimization (PSO) algorithm as a case study, or a toy model, to study how a Lagrangian formulation of this global optimization strategy can be constructed. This will motivate some of the tuning parameters that were introduced by the heuristic optimization community based on an intuition that is not directly related to the physical nature of the problem. Also, the Lagrangian formalism will reveal new hidden aspects in the algorithm. For example, we will show that the basic PSO algorithm lacks any "electromagnetic" nature. Particles are not electrically charged and hence do not radiate. However, the general formalism can provide some insights on how to add this electromagnetic behavior to the algorithm in future studies. The most interesting insight, however, is the possibility of looking to both the classical and quantum versions of the PSO algorithm as different manifestations of a single underlying Markov process, a view that is currently revived in theoretical physics.

Second, this work aims to introduce new inter-disciplinary perspectives for the artificial intelligence (AI) and evolutionary computing (EC) communities. While EC methods, generally referred to as heuristic, work very well with complicated problems [2–5], still little is known about the fundamental mechanisms responsible for the satisfactory performance of the technique. Although the social

intelligence point of view has been emphasized considerably in the PSO literature, it seems that the treatment so far is still conventional in the sense of working with interpretations based on the previously established literature of the GA [7]. The main thrust behind EC and AI strategies is the inspiration by nature, where in the case of AI nature is the human mind, while in EC it is the evolutionary paradigm. Therefore, it looks plausible to continue following this original impulse by allowing for further analogies inspired by similarities with other non-biological natural phenomena. It is the hope of the authors of this work that the utilization of concepts that are totally outside the traditional AI literature may illuminates new routes in studying the problem of foundations of EC methods.

In this paper, we introduce a new view on the PSO algorithm that is based on physics, rather than artificial intelligence (AI) or evolutionary computing (EC). The purpose of this analogy is to further enhance the understanding of how the algorithm works and to provide new tools for the study of the dynamics of the method. The basic idea is to observe the close similarity between a swarm of particles, communicating with each other through individual and social knowledge, and a collection of material particles interacting through a classic Newtonian field. In particular, molecular dynamics (MD) will be studied in connection to the PSO environment. Various relations and analogies will be established and illustrated through this article.

Particle swarm optimization (PSO) is a global search strategy that can handle difficult optimization problems. Since its introduction at 1995 [6], the PSO algorithm was refined and extended [3–5] to enhance its performance in many practical problems, including engineering electromagnetics [10–12]. The method proved to work successfully with standard test functions and practical engineering problems. It has been claimed that the PSO represents a new trend in artificial intelligence and soft computing, in which intelligence appears as an emergent behavior from the social dynamic interactions between swarm members [7]. To date, the work in [7] is still the only major study that providing comprehensive and in-depth analysis of why and how the algorithm works. However, the main theme introduced there is the evolutionary perspective and the new paradigm of social psychology, in which the individual behavior is understood through the collective interactions level. Little has been done in investigating the possible connections between the PSO algorithms, in its social intelligence context, with fundamental physics, like Newtonian mechanics and quantum theory. Of special importance is the recent work on quantum versions of the PSO in which the laws of physics, Schrodinger's equation in this case, were employed directly in the formulation [13–15]. While the

incorporation of physics in the quantum version was self-evident and fundamental, the situation with the classical PSO algorithm looks very different. Indeed, the laws of physics were not introduced as indispensable means to understand how the algorithm works. Provided that it is still possible to keep the social knowledge point of view in the picture, an analysis of the problem, which is inspired by fundamental physics, can reveal some hidden aspects and open the door for new perspective in the problem of foundations of evolutionary computing.

This paper is organized as follow. First, we review the basic procedure of the classical PSO in order to prepare for the mathematical treatment of the next sections. Second, a formal analogy between the PSO and MD is established by deriving the corresponding equivalent mechanical force acting on the particle in the PSO. The resulting equations are discretized to produce the equations of motion. Third, based on this analogy, the thermodynamic performance of the method is studied by defining effective macroscopic dynamic observables, like the swarm temperature, and then investigating their behavior. Fourth, a new acceleration technique is proposed based on the thermodynamic analysis presented before. The new method is applied to the problem of synthesis of linear array antennas. Fifth, a fully macroscopic formulation of the classical PSO is proposed by constructing a diffusion model for the particle swarm environment. Sixth, based on the knowledge of the pdf of the particle trajectory, a generalized version of the PSO method is proposed based on the theory of Markov chains. Finally, conclusions and suggestions for future work are given.

2. REVIEW OF THE PSO ALGORITHM

We start with an N-dimensional vector space $\mathbf{R}^N$. A population of M particles is assumed to evolve in this space such that each particle is assigned the following position and velocity vectors, respectively

$$\mathbf{r}^i(t) = \left[\begin{array}{cccccc} r^1(t) & r^2(t) & . & . & . & r^N(t) \end{array} \right]^T \tag{1}$$

$$\mathbf{v}^i(t) = \left[\begin{array}{cccccc} v^1(t) & v^2(t) & . & . & . & v^N(t) \end{array} \right]^T \tag{2}$$

where T is the transpose operator and $i \in \{1, 2, ..., M\}$.

In addition to these dynamic quantities, we postulate two memory locations encoded in the variables $p_n^{i,L}$, the local best of the ith particle, and p_n^g, the global best, both for the nth dimension. The basic idea of the classical PSO algorithm is the clever exchange of information about the global and local best values. This exchange is accomplished

by the following two equations

$$v_n^i(t+\Delta t) \;=\; w\, v_n^i(t)+c_1\varphi_1\left[p_n^{i,L}-x_n^i(t)\right]\Delta t+\mathrm{c}_2\varphi_2\left[p_n^g-x_n^i(t)\right]\Delta t \tag{3}$$

$$r_n^i(t+\Delta t)=r_n^i(t)+\Delta t\; v_n^i(t) \tag{4}$$

where $n \in \{1,2,...,N\}$; Δt is the time step; c_1 and c_2 are the cognitive and social factors, respectively; φ_1 and φ_2 are two statistically independent random variables uniformly distributed between 0 and 1; w is the inertia factor. More in-depth details on the PSO method can be found in [7–9].

It has been shown in [8] that in order for the basic PSO algorithm ($w=1$) to converge, all particles must approach the location $\mathbf{P}^i$ given by

$$P_n^i=\frac{1}{\varphi_1+\varphi_2}\left(\varphi_1 p_n^{i,L}+\varphi_2 p_n^g\right) \tag{5}$$

It is understood in equations (3)–(5) that for each dimension n the random number generators of φ_1 and φ_2 are initialized with different seeds. The convergence of the PSO algorithm to the location (5) can be guaranteed by proper tuning of the cognitive and social parameters of the algorithm [8]. To prevent explosion of the particles in the classical PSO, a maximum velocity $V_{\max}$ is introduced in each dimension to confine swarm members inside the boundary walls of the domain of interest. Also, the inertia factor w is decreased linearly during the algorithm run in order to enhance the convergence performance [7].

3. MOLECULAR DYNAMICS FORMULATION

The main goal of molecular dynamics (MD) is to simulate the state of a system consisting of very large number of molecules. A true consideration of such systems requires applying quantum-theoretic approach. However, the exact description of the quantum states requires solving the governing Schrödinger's equations, a task that is inherently impossible for many-particle systems [16]. Alternatively, MD provides a short cut that can lead to accurate results in spite of the many approximations and simplifications implied in its procedure [18].

MD is based on calculating at each time step the positions and velocities, for all particles, by direct integration of the equations of motions. These equations are constructed basically from a classical Lagrangian formalism in which interactions between the particles are assumed to follow certain potential functions. The determination of the

specific form of these potentials depends largely on phenomenological models and/or quantum-theoretic considerations.

Besides MD, there are in general three other methods to study the evolution of large number of particles: Quantum mechanics, statistical mechanics, and Mont Carlo [18]. However, MD is preferred in our context over the other methods because it permits a direct exploitation of the structural similarities between the discrete form of the update equations of the PSO algorithm and the Lagrangian formalism.

3.1. Conservative PSO Environments

In order to formally construct the analogy between PSO and Newtonian mechanics, we consider a set of M identical particles, each with mass m, interacting with each other. We start first by a *conservative* system described by a Lagrangian function given by

$$L\left(\mathbf{r}^i, \dot{\mathbf{r}}^i\right) = \sum_{i=1}^{M} \frac{1}{2} m\, \dot{\mathbf{r}}^i \cdot \dot{\mathbf{r}}^i \;-\; U\left(\mathbf{r}^1, \mathbf{r}^2, \cdots, \mathbf{r}^M\right) \tag{6}$$

where $\mathbf{r}^i$ and $\dot{\mathbf{r}}^i = \mathbf{v}^i$ are the position and velocity of the ith particle, respectively. U is the potential function, describing the intrinsic strength (energy) of the spatial locations of the particle in the space. The equations of motion can be found by searching for the critical value for the action integral

$$S = \int_{t_1}^{t_2} L\left(\mathbf{r}^1, \mathbf{r}^2, ..., \mathbf{r}^M; \dot{\mathbf{r}}^1, \dot{\mathbf{r}}^2, ..., \dot{\mathbf{r}}^M\right) dt \tag{7}$$

where t_1 and t_2 are the initial and final times upon which the boundary of the trajectory is specified. The solution for this "optimization" problem is the Euler-Lagrange equation

$$\frac{\partial L}{\partial x^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i} = 0 \tag{8}$$

Equations (6) and (8) lead to

$$\mathbf{a}^i = \dot{\mathbf{v}}^i = \ddot{\mathbf{r}}^i = \frac{\mathbf{F}^i}{m} \tag{9}$$

where $\mathbf{F}^i$ is the mechanical force acting on the ith particle in the swarm and $\mathbf{a}^i$ is its resulted acceleration. The mechanical force can be expressed in terms of the potential function U as follows

$$\mathbf{F}^i = -\frac{\partial}{\partial \mathbf{r}^i} U\left(\mathbf{r}^1, \mathbf{r}^2, ..., \mathbf{r}^M\right) \tag{10}$$

Equations (6)–(10) represents a complete mechanical description of the particle swarm evolution; it is basically a system of continuous ordinary differential equations in time. To map this continuous version to a discrete one, like the PSO in (3) and (4), we consider the Euler-Cauchy discretization scheme [19]. It is possible to write the equations of motion in discrete time as

$$\mathbf{v}\left(k\Delta t\right) = \mathbf{v}\left(\left(k-1\right)\Delta t\right) + \Delta t\,\mathbf{a}\left(k\Delta t\right) \tag{11}$$

$$\mathbf{r}\left(k\Delta t\right) = \mathbf{r}\left(\left(k-1\right)\Delta t\right) + \Delta t\,\mathbf{v}\left(k\Delta t\right) \tag{12}$$

where $\triangle t$ is the time step of integration.

By comparing equation (3) and (4) to (11) and (12), it can be concluded that the PSO algorithm corresponds exactly to a swarm of classical particles interacting with a field of conservative force only if $w = 1$, which corresponds to the basic form of the PSO algorithm originally proposed in [6]. The acceleration is given by the following compact vector form

$$\mathbf{a}^i\left(k\Delta t\right) = \bar{\bar{\mathbf{\Phi}}}\left[\mathbf{P}^i - \mathbf{r}^i\left(k\Delta t\right)\right] \tag{13}$$

where $\bar{\bar{\mathbf{\Phi}}}$ is a diagonal matrix with the non-zero elements drawn from a set of mutually exclusive random variables uniformly distributed from 0 to 1.

The integer $k \in \{1, 2, ..., N_{\text{itr}}\}$, where N_{itr} is the total number of iterations (generations), represents the time index of the current iteration. Thus, the force acting on the ith particle at the kth time step is given by

$$\mathbf{F}^i\left(k\Delta t\right) = m\,\bar{\bar{\mathbf{\Phi}}}\left[\mathbf{P}^i - \mathbf{r}^i\left(k\Delta t\right)\right] \tag{14}$$

Equation (14) can be interpreted as Hooke's law. In the PSO algorithm, particles appear to be driven by a force directly proportional to the displacement of their respective positions with respect to some center given by (5). Thus, for each particle there exists an equivalent mechanical spring with an anisotropic Hooke's tensor equal to $m\bar{\bar{\mathbf{\Phi}}}$. Figure 1 illustrates this analogy.

The mass of the particle appearing in equation (14) can be considered as an extra parameter of the theory that can be chosen freely. This is because the basic scheme of the PSO method assumes point-like particles. Therefore, we will choose the mass m such that the simplest form of the quantities of interest can be obtained.

3.2. Philosophical Discussion

The main idea in the proposed connection between the PSO algorithm and the physical system descried by the Lagrangian (6) is to shift all

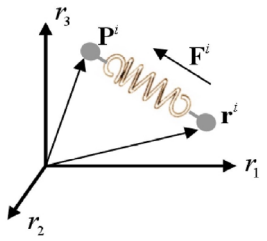


Figure 1. A 3-D illustration of the mechanical analogy to the PSO. The ith particle will experience a mechanical force identical to a spring with one end attached to the position and the other end attached to the particle's itself.

of the 'randomness' and social intelligence (represented by p) to the law of force acting on particles in a way identical to Newton's law in its discrete-time form. This is as if there exists a hypothetical Grand Observer who is able to monitor the motion of all particles, calculates the global and local best, average them in a way that looks random only to other observers, and then apply this force to the particles involved. Obviously, this violates relativity in which nothing can travel faster than the speed of light but Newtonian mechanics is not local in the relativistic sense. The assumption of intelligence is just the way the mechanical force is calculated at each time instant. After that, the system responds in its Newtonian way. There is no restriction in classical mechanics to be imposed on the force. What makes nature rich and diverse is the different ways in which the mechanical law of force manifests itself in Newton's formalism. In molecular dynamics (MD), it is the specific (phenomenological) law of force what makes the computation a good simulation of the quantum reality of atoms and molecules.

However, on a deeper level, even this Grand Observer who monitor the performance of the PSO algorithm can be described mathematically. That is, by combining equations (10) and (14), we get the following differential equation for the potential U

$$\frac{\partial U}{\partial \mathbf{r}^i} + m\,\bar{\bar{\mathbf{\Phi}}}\left(\mathbf{P}^i - \mathbf{r}^i\right) = 0 \tag{15}$$

The reason why we did not attempt to solve this equation is the fact that P, as defined in (5), is a complicated function of the positions of all particles, which enforces on us the many-body interaction theme. Also, the equation cannot be solved uniquely since we know only the discrete values of P, while an interpolation/extrapolation to the continuous limit is implicit in the connection we are drawing between

the PSO algorithm and the physical system. Moreover, since P is not a linear function of the interaction between particles, then equation (15) is a complex many-body nonlinear equation. While it can be solved in principle, the nonlinearity may produce very complicated patterns that look random (chaos) because of the sensitivity to the precision in the initial conditions. However, all of this does not rule out the formal analogy between the PSO and the Lagrangian physical system, at least on the qualitative level. In the remaining parts of this paper, we will take U to represents some potential function with an explicit form that is not known to us, and actually not very important in relation to the conclusions and results presented here.

Regarding to the appearance of random number generators in the law of force (13), few remarks are given here. It should be clear that a Newtonian particle moves in a deterministic motion because of our infinite-precision knowledge of the external force acting on it, its mass, and the initial conditions. According to $F = ma$, if m is known precisely, but F is random (let us say because of some ignorance in the observer state), then the resulting trajectory will look random. However, and this is the key point, the particle does not cease to be Newtonian at all! What is Newtonian, anyhow, is the dynamical law of motion, not the superficial, observer-dependent, judgment that the motion looks random or deterministic with respect to his knowledge.

This can be employed to reflect philosophically on the nature of the term 'intelligence.' We think that any decision-making must involve some sort of ignorance or state of lack of knowledge. Otherwise, AI can be formulated as a purely computational system (eventually by a universal Turing machine). However, the controversial view that such automata can ultimately describe the human mind was evacuated from our discussion at the beginning by pushing the social intelligence P to the details of the external mechanical force law, and then following the consequences that can be derived by considering that the particles respond in a Newtonian fashion. We are fully aware that no proof about the nature of intelligence was given here in the rigorous mathematical sense, although something similar has been attempted in literature [21].

To summarize, the social and cognitive knowledge are buried in the displacement origin $\mathbf{P}^i$, from which the entire swarm will develop its intelligent behavior and search for the global optimum within the fitness landscape. The difficulty in providing full analysis of the PSO stems from the fact that the displacement center $\mathbf{P}^i$ is varying with each time step and for every particle according to the information of the entire swarm, rendering the PSO inherently a many-body problem in which each particle interacts with all the others.

3.3. Nonconservative (Dissipative) PSO Environments

It is important to state that for the transformation between the PSO and MD, as presented by equation (13), to be exact, the inertia factor w was assumed to be unity. However, as we pointed out in Section 2, for satisfactory performance of the PSO algorithm in real problems one usually reverts to the strategy of linearly decreasing w to a certain lower value. In this section, we will study in details the physical meaning of this linear variation.

The velocity equation for the PSO is given by

$$\mathbf{v}^i(k\Delta t) = w\mathbf{v}^i((k-1)\Delta t) + \bar{\bar{\mathbf{\Phi}}}\left[\mathbf{P}^i((k-1)\Delta t) - \mathbf{x}^i((k-1)\Delta t)\right]\Delta t \tag{16}$$

Thus, by rearranging terms we can write

$$\begin{aligned}\frac{\mathbf{v}^i(k\Delta t) - \mathbf{v}^i((k-1)\Delta t)}{\Delta t} = &-\frac{1-w(t)}{\Delta t}\mathbf{v}^i((k-1)\Delta t) \\ &+\bar{\bar{\mathbf{\Phi}}}\left[\mathbf{P}^i((k-1)\Delta t) - \mathbf{x}^i((k-1)\Delta t)\right]\end{aligned} \tag{17}$$

Here $w = w(t)$ refers to a general function of time (linear variation is one common example in the PSO community). By taking the limit when $\Delta t \longrightarrow 0$ we find

$$\mathbf{a}^i(t) = \beta\mathbf{v}^i(t) + \bar{\bar{\mathbf{\Phi}}}\left[\mathbf{P}^i(t) - \mathbf{x}^i(t)\right] \tag{18}$$

where

$$\beta = \lim_{\Delta t\to 0}\frac{w(t)-1}{\Delta t} \tag{19}$$

Comparing equation (13) with (18), it is clear that the conservative Lagrangian in (6) cannot admit a derivation of the PSO equation of motion when w is different from unity; there exists a term proportional to velocity that does not fit with Newton's second law as stated in (9).

This problem can be solved by considering the physical meaning of the extra terms. For a typical variation of w starting at unity and ending at some smaller value (typically in the range 0.2–0.4, depending on the objective function), then we find from (19) that β is negative. That is, since the *total* force is given by the product of the acceleration $\mathbf{a}^i$ in (18) and the mass m, then it seems that the term $\beta\mathbf{v}^i$ counts for "friction" in the system that tends to lower the absolute value of the velocity as the particles evolve in time. In other words, w amounts to a dissipation in the system with strength given by the factor β. This explains why the conservative Lagrangian failed to produce the

equations of motion in this particular case since the system is actually nonconservative.

Fortunately, it is still possible to employ a modified version of the Lagrangian formalism to include the dissipative case in our general analogy to physical systems. To accomplish this, we split the Lagrangian into two parts. The first is L', which represents the conservative part and consists of the difference between kinetic and potential energy as in (6) and repeated here for convenience

$$L' = \frac{1}{2}\sum_i m\,\dot{\mathbf{x}}^i \cdot \dot{\mathbf{x}}^i - U \tag{20}$$

where the potential energy is function of the positions only.

The second part, L'', accounts for the nonconservative or dissipative contribution to the system. Following [22], we write

$$L'' = \frac{1}{2}\sum_i \beta_i \dot{\mathbf{x}}^i \cdot \dot{\mathbf{x}}^i + \frac{1}{2}\sum_i \gamma_i \ddot{\mathbf{x}}^i \cdot \ddot{\mathbf{x}}^i \tag{21}$$

Here β_i represents the Rayleigh losses in the system and is similar to friction or viscosity. The second term that contains γ_i accounts for radiation losses in the system. For example, if the particles are electrically charged, then any nonzero acceleration will force the particle to radiate an electromagnetic wave that carries away part of the total mechanical energy of the system, producing irreversible losses and therefore dissipation. Both of m and γ are assumed to be constants independent of time but this is not necessary for β_i. Moreover, the conservative Lagrangian L' has the dimensions of energy while the nonconservative part L'' has the units of power.

The equation of motion for the modified Lagrangian is given by [22]

$$\left\{\frac{\partial L''}{\partial \dot{\mathbf{x}}^i} - \frac{d}{dt}\frac{\partial L''}{\partial \ddot{\mathbf{x}}^i}\right\} - \left\{\frac{\partial L'}{\partial \mathbf{x}^i} - \frac{d}{dt}\frac{\partial L'}{\partial \dot{\mathbf{x}}^i}\right\} = 0 \tag{22}$$

By substituting (20) and (21) to (22), we get

$$-\gamma_i \frac{d}{dt}\ddot{\mathbf{x}}^i + m\ddot{\mathbf{x}}^i + \beta_i \dot{\mathbf{x}}^i = -\frac{\partial U}{\partial \mathbf{x}^i} \tag{23}$$

Comparing (23) with (18), we immediately find

$$\gamma_i = 0 \tag{24}$$

and

$$\beta_i = -m\beta = m \lim_{\Delta t \to 0} \frac{1 - w\,(t)}{\Delta t} \tag{25}$$

where equation (15) has been used.

The result (24) means that there is no electromagnetic radiation in the PSO environment. One can say that the idealized particles in the PSO algorithm do not carry electric charge. However, it is always possible to modify the basic PSO to introduce new parameters. One of the main advantages of constructing a physical theory is to have an intuitive understanding of the meaning of tuning parameters in the algorithm. We suggest considering the idea of adding an electric charge to the particles in the algorithm and investigate whether this may lead to an improvement in the performance or not.

Equation (25) amounts to a connection between the physical parameter β, the Rayleigh constant of friction, and a tuning parameter in the PSO algorithm, w. This motivates the idea behind slowly decreasing the inertia w during convergence. As $w(t)$ decreases, $\beta(t)$ increases, which means that the dissipation of the environment will increase leading to faster convergence. From this discussion we see that the common name attributed to w is slightly misleading. From the mechanical point of view, inertia is related to the mass m while w controls the dissipation or the friction of the system. We suggest then calling w the 'friction constant' or the 'dissipation constant' of the PSO algorithm.

The fact that the system is dissipative when w is less than unity may lead to some theoretical problems in the next sections, especially those related to thermodynamic equilibrium. However, a careful analysis of the relative magnitudes of the 'friction force' and the 'PSO force' in (23) shows that for a typical linear variation of w from 0.9 to 0.2, the friction becomes maximum at the final iteration, most probably when the algorithm has already converged to some result. At the beginning of the run, the friction is minimum while particles are expected to be 'reasonably' away from the local and global bests; that is, the difference $\varphi(x - p)$ is large compared to βv. This means that as the algorithm evolves to search for the nontrivial optima, the friction is not very significant. It becomes so only at the final iterations when β in (19) is already large compared to the PSO force (14). This explains why we have to be careful in choosing w; specifically, if the dissipation is increasing faster than the "intelligent" PSO force, immature convergence will occur and the algorithm will not be able to catch the global optimum. Based on this discussion, we will always assume that $w(t)$ is chosen to vary with time "wisely" (i.e., by insuring that immature convergence is avoided). Therefore, we will ignore the effect of dissipation and treat the PSO as a collection of perfectly Newtonian particles interacting in conservative environment. Further results and studies in the remaining sections confirm this assumption.

4. EXTRACTION OF INFORMATION FROM SWARM DYNAMICS

The dynamic history of a swarm of particles is completely defined either by the system of equations (3) and (4) for the PSO, or (6)–(13) for MD. The two representations are equivalent, where the transformation between one to the other is obtained by the mapping in (13). After finishing the simulation, we will end up with a set of M trajectories, each describing the successive velocities (momentums) and positions of one particle during the course of time. Any trajectory is assumed to be one-dimensional surface in an abstract $2MN$-dimensional vector space. We assume that this space forms a manifold through the adopted coordinate system and call it the phase space of the swarm system.

Let us write the trajectory of the ith particle as

$$\mathbf{\Gamma}_i(t) = \left\{ \left(\mathbf{r}^i(t), \mathbf{p}_m^i(t) \right), \forall t \in I_t \right\} \tag{26}$$

in the continuous form and

$$\mathbf{\Gamma}_i(k) = \left\{ \left(\mathbf{r}^i(k\Delta t), \mathbf{p}_m^i(k\Delta t) \right) \right\}_{k=1}^{N_{\text{itr}}} \tag{27}$$

for the discrete case. Here I_t is the continuous time interval of the simulation and N_{itr} is the number of iterations. The *swarm dynamic history* of either the MD or the PSO can be defined as the set of all particles trajectories obtained after finishing the simulation

$$\bar{\mathbf{\Gamma}}(t) = \left\{ \mathbf{\Gamma}^i(t) \right\}_{i=1}^{M} \tag{28}$$

One of the main objectives of this article is to provide new insight on how the classical PSO algorithm works by studying different quantities of interest. We define any dynamic observable of the swarm dynamics to be a smooth function of the swarm dynamic history $\bar{\mathbf{\Gamma}}(t)$. That is, the general form of a dynamic property (observable) will be written as

$$A_\Phi(t) = \Phi \bar{\mathbf{\Gamma}}(t) \tag{29}$$

where Φ is a sufficiently smooth operator.

The PSO is inherently a stochastic process as can be inferred from the basic equations (3) and (4). Moreover, we will show later that a true description of MD must rely on statistical mechanical considerations where the particle's trajectory is postulated as a stochastic process. Therefore, single evaluation of a dynamic property, as defined in (29), cannot reflect the qualitative performance of interest. Instead, some sort of averaging is required in order to get reasonable

results. We introduce here two types of averages. The first one is the time average defined as

$$\langle A_\Phi \rangle = \lim_{T\to\infty} \frac{1}{T} \int\limits_{t_0}^{t_0+T} A_\Phi\left(t\right) dt \tag{30}$$

and

$$\langle A_\Phi \rangle = \frac{1}{N_{\text{itr}}} \sum_{k=1}^{N_{\text{itr}}} A_\Phi\left(k\right) \tag{31}$$

for continues and discrete systems, respectively. An important assumption usually invoked in MD simulations is that an average over finite interval should estimate the infinite integration/summation of (31) [16]. Obviously, this is because any actual simulation interval must be finite. The validity of this assumption can be justified on the basis that the number of time steps in MD simulations is large. However, no such restriction can be imposed in the PSO solutions because in engineering applications, especially real-time systems, the number of iterations must be kept as small as possible to facilitate good performance in short time intervals. Therefore, instead of time averages, we will invoke the ensemble average, originally introduced in statistical mechanics. We denote the ensemble average, or expected value, of a dynamic observable A_Φ as $E[A_\Phi]$. It is based on the fact that our phase space is endowed with a probability measure so meaningful probabilities can be assigned to outcomes of experiments performed in this space. Moreover, if the system that produced the swarm dynamics in (28) is ergodic, then we can equate the time average with ensemble average. In other words, under the ergodic hypothesis we can write [16, 20]

$$\langle A_\Phi \rangle = E\left[A_\Phi\right] \tag{32}$$

Therefore, by assuming that the ergodicity hypothesis is satisfied, one can always perform ensemble average whenever a time average is required. Such assumption is widely used in MD and statistical physics [16, 18]. However, in the remaining parts of this paper we employ ensemble average, avoiding therefore the controversial opinion of whether all versions of the PSO algorithm are strictly ergodic or not. Moreover, the use of ensemble average is more convenient when the system under consideration is dissipative.

5. THERMODYNAMIC ANALYSIS OF THE PSO ENVIRONMENT

5.1. Thermal Equilibrium

Thermal equilibrium can be defined as the state of a system of molecules in which no change of the macroscopic quantities occurs with time. This means that either all molecules have reached a complete halt status (absolute zero temperature), or that averages of the macroscopic variables of interest don't change with time. To start studying the thermal equilibrium of a swarm of particles, we will employ the previous formal analogy between the PSO and classical Newtonian particle environments to define some quantities of interest, which will characterize and enhance our understanding of the basic mechanism behind the PSO.

In physics, the temperature of a group of interacting particles is defined as the average of the kinetic energies of all particles. Based on kinetic theory, the following expression for the temperature can be used [16]

$$T(t) = \frac{1}{M} \frac{m}{Nk_B} \sum_{i=1}^{M} |\mathbf{v}_i(t)|^2 \tag{33}$$

where k_B is Boltzmann's constant and for the conventional 3-dimensional space we have $N = 3$. Analogously, we define the particle swarm temperature as

$$T(t) = E\left[\frac{1}{M} \sum_{i=1}^{M} |\mathbf{v}_i(t)|^2\right] \tag{34}$$

where without any loss of generality the particle's mass is assumed to be

$$m = Nk_B \tag{35}$$

Since there is no exact a *priori* knowledge of the statistical distribution of particles at the off-equilibrium state, the expected value in (34) must be estimated numerically. We employ the following estimation here

$$T(t) = \frac{1}{MB} \sum_{j=1}^{B} \sum_{i=1}^{M} \left|\mathbf{v}_i^j(t)\right|^2 \tag{36}$$

where B is the number of repeated experiment. In the jth run, the velocities $\mathbf{v}_i^j(t), i = 1, 2, .., M$, are recorded and added to the sum. The total result is then averaged over the number of runs. The initial positions and velocities of the particles in the PSO usually

follow a uniformly random distribution. At the beginning of the PSO simulation, the particles positions and velocities are assigned random values. From the kinetic theory of gases point of view, this means that the initial swarm is not at the thermal equilibrium. The reason is that for thermal equilibrium to occur, particles velocities should be distributed according to Maxwell's distribution [16], which is, strictly speaking, a Gaussian probability distribution. However, since there is in general no prior knowledge about the location of the optimum solution, it is customary to employ the uniformly random initialization.

In general, monitoring how the temperature evolves is not enough to determine if the system had reached an equilibrium state. For isolated systems, a sufficient condition to decide whether the system has reached macroscopically the thermal equilibrium state is to achieve maximum entropy. However, the calculation of entropy is difficult in MD [16]. Usually MD simulations produce automatically many macroscopic quantities of interest; by monitoring all of them, it is possible to decide whether the system had reached equilibrium or not. However, in the basic PSO algorithm there are no corresponding quantities of interest. What we need is really one or two auxiliary quantities that can decide, with high reliably, if the system has converged to the steady thermal state. Fortunately, such a measure is available in literature in a form called the α-factor. We employ the following thermal index [17]

$$\alpha\left(t\right)=\frac{1}{B}\sum_{j=1}^{B}\frac{\frac{1}{M}\sum_{i=1}^{M}\left\|\mathbf{v}_{i}^{j}\right\|^{4}}{\left[\frac{1}{M}\sum_{i=1}^{M}\left\|\mathbf{v}_{i}^{j}\right\|^{2}\right]^{2}} \tag{37}$$

where the usual definition of the Euclidean norm for a vector $\mathbf{v}$ with length N is given by

$$\left\|\mathbf{v}\right\|=\sqrt{\sum_{n=1}^{N}v_{n}^{2}} \tag{38}$$

Here, $\mathbf{v}_i^j$ is the velocity of the ith particle at the jth experiment. At equilibrium, the index above should be around 5/3 at isothermal equilibrium [17].

5.2. Primary Study Using Benchmark Test Functions

In order to study the qualitative behavior of the swarm thermodynamics, we consider the problem of finding the global minimum of the

following N-dimensional standard test functions

$$f_1(x) = \sum_{n=1}^{N} x_n^2 \tag{39}$$

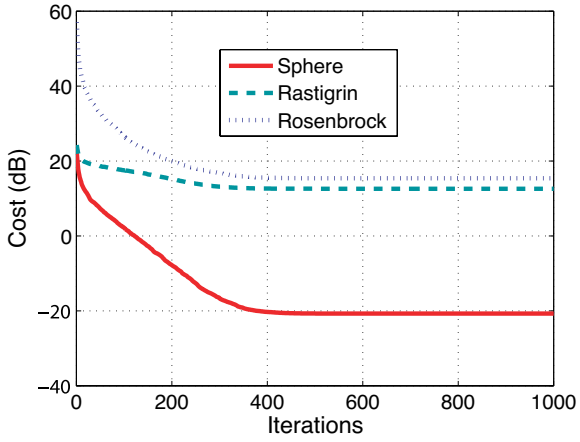
$$f_2(x) = \sum_{n=1}^{N} \left[x_n^2 - 10\cos(2\pi x_n) + 10\right] \tag{40}$$

$$f_3(x) = \sum_{n=1}^{N-1} \left[100\left(x_{n+1} - x_n^2\right)^2 + (x_n - 1)^2\right] \tag{41}$$

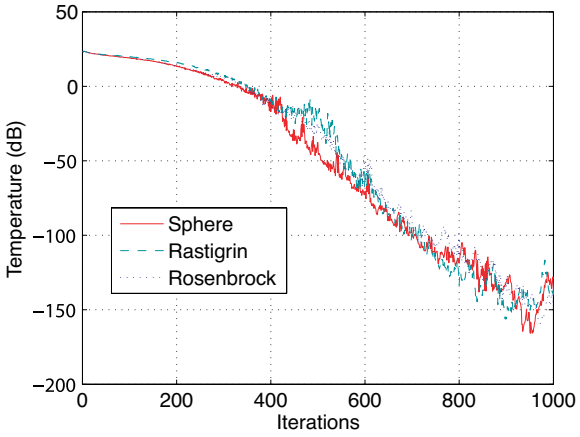
The sphere function in (39) has a single minimum located at the origin. The function defined in (40) is known as Rastigrin function with a global minimum located at the origin. This is a hard multimodal optimization problem because the global minimum is surrounded by a large number of local minima. Therefore, reaching the global peak without being stuck at one of these local minima is extremely difficult. The problem in (41), known as Rosenbrock function, is characterized by a long narrow valley in its landscape with global minimum at the location $[1, 1, ..., 1]^T$. The value of the global minimum in all of the previous functions is zero.

Figure 2 shows the cost evolution together with the time history of the swarm temperature obtained from the PSO simulation of the three test functions (39)–(41). A swarm of 20 particles searching in a 10-dimensional space is considered for each case. Figure 2(a) indicates that only the sphere function has successfully converged to the global optimum. On the other hand, by examining the thermal behavior of the system as indicated by Figure 2(b), it is clear that the swarm temperature drops rapidly, suggesting that the swarm is "cooling" while evolving with time. Eventually, the temperature goes to zero, indicating that the swarm has reached the state of thermal equilibrium. Notice that in thermodynamics zero temperature is only a special case. Convergence to a constant non-zero temperature can also be characterized as thermal equilibriums.

It is interesting to observe from Figure 2 that the thermal behaviors of the three solutions look indistinguishable. This is the case although the convergence curves of Figure 2(a) shows that only the sphere function has converged to its global optimum. This demonstrates that convergence to thermal equilibrium does not mean that convergence to the global optimum has been achieved. Rather, convergence to a local optimum will lead to thermal equilibrium. The swarm temperature cannot be used then to judge the success or failure of the PSO algorithm.



(a)



(b)

Figure 2. Cost function and temperature evolution for a swarm of 20 particles searching for the minimum in a 10-dimensional space. The parameters of the PSO algorithm are $c_1 = c_2 = 2.0$, and w is decreased linearly from 0.9 to 0.2. The search space for all the dimensions is the interval $[-10, 10]$. A maximum velocity clipping criterion was used with $V_{\max} = 10$. The algorithm was run 100 times and averaged results are reported as (a) Convergence curves. (b) Temperature evolution.

The conclusion above seems to be in direct contrast to the results presented in [23]. This work introduced a quantity called average velocity defined as the average of the *absolute* values of the velocities in the PSO. Although the authors in [23] did not present a physical analogy between MD and PSO in the way we provided in Section 3, the average velocity will qualitatively reflect the behavior of the swarm temperature as defined in (36). The main observation in [23] was that convergence of the 'average velocity' (temperature) to zero indicates global convergence. Based on this, they suggested an adaptive algorithm that forces the PSO to decrease the average velocity in order to guarantee successful convergence to global optimum. However, it is clear from the results of Figure 2 that this strategy is incorrect. Actually, it may lead the algorithm to be trapped in a local optimum.

A better perspective on the thermal behavior of the system can be obtained by studying the evolution of the α-index. Figure 3 illustrates an average of 1000 run of the experiments of Figure 2. Careful study of the results reveals a different conclusion compared with that obtained from Figure 2(b). It is clear from Figure 2(a) that the PSO solutions applied to the three different functions converge to a steady-state value at different number of iterations. The sphere function, Rastigrin function, and Rosenbrock valley converge roughly at around 400, 300, and 350 iterations, respectively. By examining the time histories of the α-index in Figure 3, it is seen that the sphere function's index is the

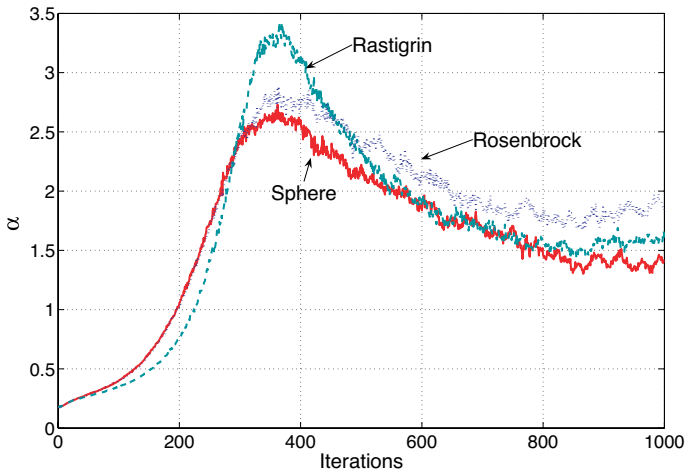


Figure 3. The time evolution of the α-index for the problem of Figure 2. The PSO algorithm was run 1000 times and averaged results are reported.

fastest in approaching its theoretical limit at the number of iterations where convergence of the cost function has been achieved. For example, the thermal index of the Rastigrin case continued to rise after the 300th iteration until the 385th iteration where it starts to fall. Based on numerous other experiments conducted by the authors, it seems more likely that the α-index can be considered an indicator to decide if the PSO algorithm has converged to the global optimum or not. It looks that convergence to global optimum in general corresponds to the fastest convergence to thermal equilibrium as indicated by our studies of the associated α-index. Moreover, it is even possible to consider that the return of the α-index to some steady state value, i.e., forming a peak, as a signal that convergence has been achieved. However, we should warn the reader that more intensive work with many other objective functions is required to study the role of the α-index as a candidate for global optimum convergence criterion.

The fact that the α-index is not converging exactly to 5/3 can be attributed to three factors. First, very large number of particles is usually required in MD simulations to guarantee that averaged results will correspond to their theoretical values. Considering that to reduce the optimization cost we usually run the PSO algorithm with small number of particles, the obtained curves for the α-index will have noticeable fluctuations, causing deviations from the theoretical limit. Second, it was pointed out in Section 4 that because of the presence of an inertia factor w in the basic PSO equations with values different from unity, the analogy between the PSO and MD is not exact. These small deviations from the ideal Newtonian case will slightly change the theoretical limit of the α-index. Our experience indicates that the new value of α at thermal equilibrium is around 1.5. Third, the theoretical value of 5/3 for the α-index is based on the typical Maxwell's distribution at the final thermal equilibrium state. However, the derivation of this distribution assumes weakly interacting particles in the thermodynamic limit. It is evident, however, from the discussion of Section 3 that the PSO force is many-body global interaction that does not resemble the situation encountered with rarefied gases, in which the kinetic theory of Maxwell applies very well. This observation forces us to be careful in drawing quantitative conclusions based on the formal analogy between the PSO and the physical system of interacting particles.

5.3. Energy Consideration

Beside the basic update equations (3), (4), (11), and (12), it should be clear how to specify the particles behavior when they hit the boundaries of the swarm domain. Assume that the PSO system under

consideration is conservative. If we start with a swarm total energy E, the law of conservation of energy states that the energy will stay constant as long as no energy exchange is allowed at the boundary. From the MD point of view, there are two possible types of boundary conditions (BC), dissipative and non-dissipative. Dissipative boundary conditions refer to the situation when the special treatment of the particles hitting the wall leads to a drop in the total energy of the swarm. An example of this type is the absorbing boundary condition (ABC), in which the velocity of a particle hitting the wall is set to zero [16, 18]. This means that the kinetic energy of a particle touching one of the forbidden walls is instantaneously "killed." If a large number of particles hit the wall, which is the case in many practical problems, there might be a considerable loss of energy in the total swarm, which in turns limits the capabilities of the PSO to find the optimum solution for several problems. The reflection boundary condition (RBC) is an example of non-dissipative BCs. However, for this to hold true reflections must be perfectly elastic; that is, when a particle hits the wall, only the sign of its vector velocity is reversed [16, 18]. This will keep the kinetic energy of the particle unchanged, leading to a conservation of the total energy of the swarm.

In the latter case, it is interesting to have a closer look to the energy balance of the PSO. From equation (6), we consider the total mechanical energy

$$E = K + U \tag{42}$$

where K and U are the total kinetic and potential energies, respectively. It is possible to show that this energy is always constant for the conservative Lagrangian defined in (6) [22].

According to the thermodynamic analysis of Subsection 5.1, when particles converge to the global optimum, the temperature drops rapidly. This means that when the swarm evolves, its kinetic energy decreases. Since the total energy is conserved, this means that convergence in the PSO can be interpreted as a process in which the kinetic energy is continually converted to potential energy. The final thermodynamic state of the swarm at the global optimum represents, therefore, a unique spatial configuration in which the particles achieve the highest possible potential energy in the searched region in the configuration space. This conclusion applies only approximately to the case when the boundary condition is dissipative since the energy loss at the wall is small compared to the total energy of the swarm (provided that large number of particles is used).

5.4. Dynamic Properties

Autocorrelation functions measure the relations between the values of certain property at different times. We define here a collective velocity autocorrelation function in the following way

$$\Psi(t) = \frac{\left\langle \frac{1}{M}\sum_{i=1}^{M} \mathbf{v}_i(t_0)\cdot\mathbf{v}_i(t_0+t)\right\rangle}{\left\langle \frac{1}{M}\sum_{i=1}^{M} \mathbf{v}_i(t_0)\cdot\mathbf{v}_i(t_0)\right\rangle} \tag{43}$$

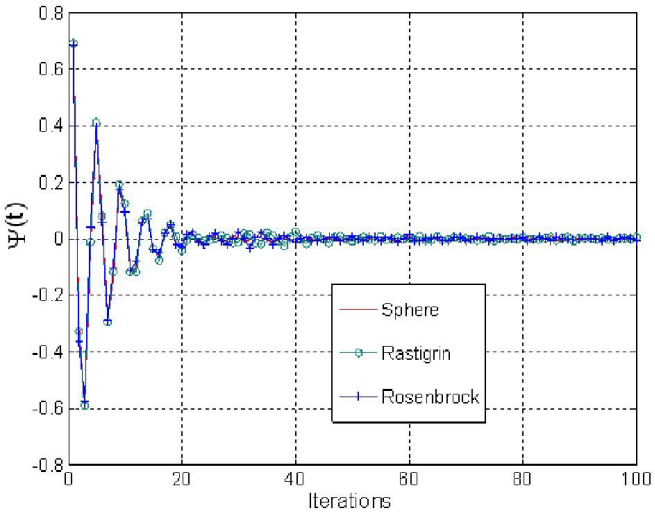
where t_0 is the time reference. The time average in (43) can be estimated by invoking the ergodicity hypothesis in the expected value operator in (29). The final expression will be

$$\Psi(t) = \frac{1}{B}\sum_{j=1}^{B}\left\{\frac{\sum_{i=1}^{M} \mathbf{v}_i^j(t_0)\cdot\mathbf{v}_i^j(t_0+t)}{\sum_{i=1}^{M} \mathbf{v}_i^j(t_0)\cdot\mathbf{v}_i^j(t_0)}\right\} \tag{44}$$

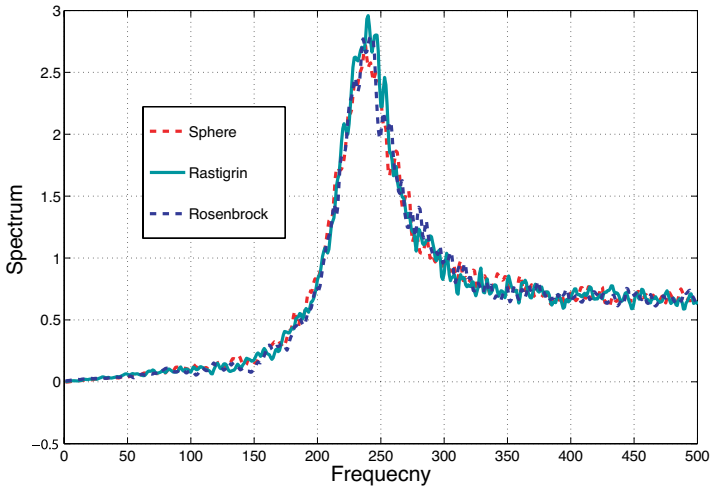
where B is the number of experiments.

Figure 4(a) illustrates the calculations of the autocorrelation function in (44) with the time reference $t_0 = 0$. The behavior of this curve has a very strong resemblance to curves usually obtained through MD simulations. Such curves may provide a lot of information about the underlying dynamics of the swarm environment. First, at the beginning of the simulation there are strong interactions between particles, indicating that the velocity of each particle will change as compared to the reference point. This is because, according to Newton's first law, if there are no forces acting on a particle the velocity will stay the same. However, in solids each molecule is attached to a certain spatial location. For this, it is hard to move a molecule far away from its fixed position. What happens then is that the particle will oscillate around its initial position. This explains the strong oscillations of $\Psi(t)$ in Figure 4(a), which suggest that the PSO environment resembles a solid state. The oscillations will not be of equal magnitude, however, but decay in time, because there are still perturbative forces acting on the atoms to disrupt the perfection of their oscillatory motion. So what we see is a function resembling a damped harmonic motion.

In liquids, because there are no fixed positions for the molecules, oscillations cannot be generated. Thus, what should be obtained is



(a)



(b)

Figure 4. The velocity autocorrelation function for the problem of Figure 2 with $B = 100$ (a) Time domain results (b) Frequency domain results with frequency given in terms of $1/\Delta t$.

a dammed sinusoid with one minimum. For gases, all what can be observed is a strong decay following an exponential law.

The observation that for the three different cost functions we get very similar early time response reflects probably that the randomly uniform initialization of the PSO algorithm construct a similar crystal-like solid structures, making early interactions of the particles similar. However, after 20 iterations the responses become different, as we expect, since particles start to develop different inter-molecular forces according to the fitness landscape encountered.

Another way to look into the dynamics of the PSO is to calculate the Fourier spectrum of the time autocorrelation function as defined in (43). This gives some information about the underlying resonant frequencies of the structure. Figure 4(b) illustrates the calculated spectrum. The shape looks very similar to results obtained for solid materials in MD [18] where a profound single peak is apparent. The Fourier transform of (43) is also called the generalized phonon density of state [18].

We conducted numerous experiments with several test functions where it seems that the classification of the PSO environment into gas, liquid, solid is strongly dependent on the boundary conditions, population size, and number of dimensions. Table 1 summarizes the shape of the velocity autocorrelation function corresponding to the three cases.

Table 1. Velocity autocorrelation function corresponding to different states for the PSO environment.

State	Velocity Autocorrelation
Gas	Damped exponential
Liquid	One period damped sinusoid with single minimum
Solid	Damped sinusoid with possibly multiple-oscillations

6. ACCELERATION TECHNIQUE FOR THE PSO ALGORITHM

In this section, we will provide an application for the thermodynamic analysis of Section 4. The idea is to employ the natural definition of temperature, as given by equation (36), in a technique similar to the simulated annealing. The idea is the observation that the decay of temperature is a necessary condition for convergence to the global optimum. This means that if the PSO has enough recourse to search

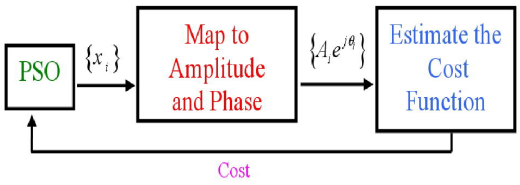


Figure 5. Block diagram representation of the linear array synthesis problem. Here the particles are $x_i \in \{0 \leq x_i \leq 1, i = 1, 2, ..., 2N\}$. Amplitudes and phases are ranged as $A_i \in \{A_{\min} \leq A_i \leq A_{\max}, i = 1, 2, ..., N\}$ and $\theta_i \in \{0 \leq \theta_i \leq 2\pi, i = 1, 2, ..., N\}$, respectively.

for the best solution, then temperature decay may be enhanced by artificially controlling the thermal behavior of the system. We employ a technique used in some MD simulations to push the swarm to equilibrium. Mainly, the velocity of each particle is modified according to the following formula

$$\mathbf{v}_i(k+1) = \sqrt{\frac{T_0}{T(k)}}\mathbf{v}_i(k) \tag{45}$$

where T_0 is a user-defined target temperature. The meaning of (45) is that while the swarm is moving toward the thermal equilibrium of the global optimum, the velocities are pushed artificially to produce a swarm with a pre-specified target temperature. In simulated annealing, the artificial cooling of the algorithm is obtained by a rather more complicated method, in which the probability is calculated for each iteration, to decide how to update the particles. Moreover, there the biggest obstacle is the appearance of new control parameters, the temperature, which has no clear meaning. This adds more burdens on the user in choosing the suitable value for the tuning parameters. However, it seems that the technique in (45) may provide an easier alternative.

The proposed acceleration technique will be demonstrated with a practical application from electromagnetics. We consider the design of linear array antenna to meet certain pre-defined goal. Suppose that there is an N-elements linear array with uniform separation between the elements of d. The normalized array factor is given by [24]

$$AF(u) = \frac{1}{\max_u \{|AF(u)|\}} \sum_{n=1}^{N} I_n e^{j2\pi ndu/\lambda} \tag{46}$$

where $I_n, n = 1, 2, ..., N$, are the current distribution (generally

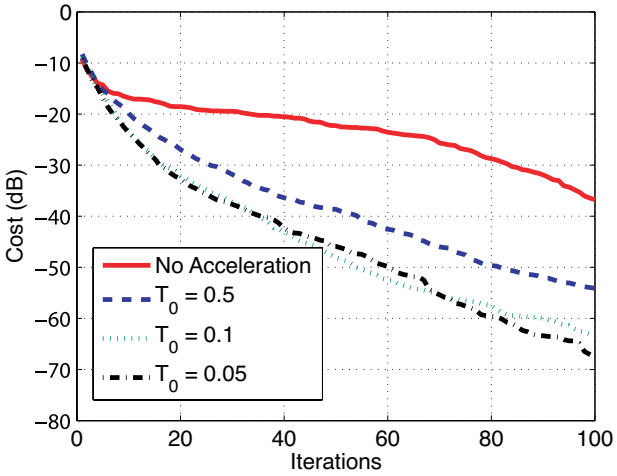


Figure 6. Convergence curves for the synthesis of symmetric pattern by using a linear array of 14 elements. The population size is 8 particles. The inertia factor w is decreased linearly from 0.9 to 0.2. $c_1 = c_2 = 2.0$.

complex); $u = \sin\theta$, where θ is the angle from the normal to the array axis; $d = \lambda/2$.

The "Don't exceed criterion" is utilized in the formulation of the objective function. That is, an error will be reported only if the obtained array factor exceeds the desired sidelobe level. The information about the amplitudes and phases are encoded in the coordinates of the particle, which are normalized between 0 and 1. Figure 5 illustrates the relation between the abstract PSO optimization process and the physical problem.

The PSO algorithm is applied to synthesize an array antenna with a tapered side lobe mask that decreases linearly from $-25\,\text{dB}$ to $-30\,\text{dB}$. The beamwidth of the array pattern is 9°. The total number of sampling points is 201. Figure 6 illustrates the convergence curves results where it is clear that the proposed artificial cooling can successfully accelerate the convergence of the method. The final obtained radiation pattern for $T_0 = 0.05$ is shown in Figure 7.

7. DIFFUSION MODEL FOR THE PSO ALGORITHM

The development of the PSO algorithm presented so far was based on drawing a formal analogy between MD and the swarm environment. This means that a *microscopic* physical treatment is employed in

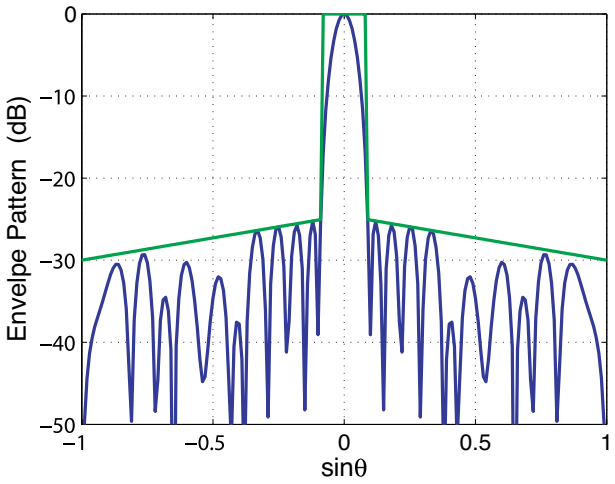


Figure 7. Radiation pattern of the linear array antenna corresponding to best case of Figure 6 ($T_0 = 0.05$).

modeling the PSO. In this section we develop an alternative view that is based on *macroscopic* approach. This new analogy can be constructed by postulating that particles in the PSO environment follow a diffusion model. The use of diffusion models is now central in solid-state physics and molecular dynamics since the pioneering works of Boltzmann and Einstein more than one hundred year ago. Moreover, since Einstein work on the origin of Brownian motion, the diffusion equation is used routinely as a theoretical basis for the description of the underlying stochastic dynamics of molecular processes.

The justification of the Diffusion model can be stated as follow. From (18), we can write the equation of motion of a particle in the PSO algorithm as

$$\ddot{\mathbf{x}}(t) = \beta \mathbf{v}(t) + \mathbf{\Gamma}(t) \tag{47}$$

where $\mathbf{\Gamma}(t) = \bar{\mathbf{\Phi}}[\mathbf{P}(t) - \mathbf{x}(t)]$. From the definition of $\mathbf{P}$ in (5), one can see that even in the continuous limit $\Delta t \to 0$ the term $\mathbf{\Gamma}(t)$ can be discontinuous because of the sudden update of the global and local positions. Thus, $\mathbf{\Gamma}(t)$ can be easily interpreted as a random fluctuations term that arises in the PSO environment in a way qualitatively similar to random collision. Notice that the particles in the PSO are originally thought of as soft spheres (i.e., there is no model for physical contacts between them leading to collisions). This interpretation brings the PSO immediately to the realm of the Brownian motion, where the dynamics is described by an equation similar to (47).

We start from the basic diffusion equation [16]

$$\frac{\partial \tilde{N}\left(\mathbf{x},t\right)}{\partial t}=D\nabla^{2}\tilde{N}\left(\mathbf{x},t\right) \tag{48}$$

where $\tilde{N}$ is the concentration function of the molecules, $\mathbf{x}$ is the position of the particle, and D is the diffusion constant.

Einstein obtained (48) in his 1905 paper on Brownian motion [25]. Direct utilization of this equation leads to solutions for the concentration of particles as functions of time and space. However, a revolutionary step taken by Einstein was to give the particle concentration different interpretations, deviating strongly from the classical treatment. To illustrate this, let us consider a one-dimensional diffusion equation as given below

$$\frac{\partial \tilde{N}}{\partial t}=D\frac{\partial^{2}\tilde{N}}{\partial x^{2}} \tag{49}$$

Einstein's interpretation of the physical meaning of the concentration function, $\tilde{N}$, started by considering a short time interval, Δt, in which the position changes from x to $x+dx$. A very important step was to refer each particle to its center of gravity [25]. Here, we propose to refer each particle to the location p_{n}^{i} , as defined in (5), where we consider this to be the *natural* center of gravity for particles in the PSO environment. Therefore, it is possible to define the new position variable

$$r_{k}^{i,n}=x_{k}^{i,n}-p_{k-1}^{i,n} \tag{50}$$

where we have modified the notation to allow for the iteration index k to appear as the subscript in (50) while the indices of the particle number i and the dimension n are the superscripts.

Clearly, the diffusion equation (49) is invariant under the linear transformation (50). During the short time interval Δt, particle locations will change according to a specific probability low. This means that particles trajectories are not continuous deterministic motions, but a stochastic path reflecting the irregular motion of particles in noisy environments. Therefore, the re-interpretation of the diffusion equation in (48) is that it gives us the probability density function (pdf) of the motion, not the classical density. Assuming the following boundary conditions

$$\tilde{N}\left(r,0\right)=\delta(r),\quad \lim_{r\rightarrow\infty}\tilde{N}\left(r,t\right)=0,\quad \int_{-\infty}^{\infty}\tilde{N}\left(r,t\right)dx=M \tag{51}$$

the solution of (49) can be written as

$$\tilde{N}(r,t) = \frac{M}{\sqrt{4\pi D}} \frac{e^{-\frac{r^2}{4Dt}}}{\sqrt{t}} \tag{52}$$

where this can be easily verified by substituting (52) back into (49).

Equation (52), which represents a Gaussian distribution for the particle's positions, was observed empirically before [26]. The development of this section presents, therefore, a theoretical derivation for the pdf of particles moving in the PSO algorithms that matches the experimentally observed results. However, the standard interpretation of (52) in the theory of diffusion equations is that it represents the Green's function of the medium. That is, depending on what is the boundary-condition imposed on the initial swarm different distributions will result.

It is very important for the subsequent developments of this paper to recognize that (52) should be interpreted as a *conditional* pdf. This can be inferred from the fact that the time variable appearing in (52) is referenced to the origin. As we mentioned above about Einstein's analysis of the particles trajectory, the total motion can be expressed as a succession of short intervals of time. This means that if we start at a time instant t_{k-1}, then the conditional pdf can be expressed as

$$f\left(x_k^{i,n},\ t_k | x_{k-1}^{i,n}, t_{k-1}\right) = \frac{1}{\sqrt{4\pi D\left(t_k - t_{k-1}\right)}} \exp\left\{-\frac{\left(x_k^{i,n} - p_{k-1}^{i,n}\right)^2}{4D\left(t_k - t_{k-1}\right)}\right\} \tag{53}$$

where f stands now for the pdf of the particles position. This final form is valid at arbitrary time instants.

It remains to see how the diffusion constant in (53) can be estimated. Here, we come to the main contribution of Einstein's paper in which he introduced the first connection between the macroscopic representation of the diffusion model and the microscopic representation of MD. Basically, the following relation was derived in [25]

$$D = \lim_{t\to\infty} \frac{\left\langle \left[\mathbf{r}(t) - \mathbf{r}(0)\right]^2 \right\rangle}{6t} \tag{54}$$

Clearly, this expression can be easily evaluated from the MD or PSO simulation. Therefore, it is possible to get the conditional pdf of particles motion by calculating the diffusion constant as given in (54). It should be mentioned that the pdf is in general different for varying boundary conditions. The well-established theory of diffusion

equations can be employed in solving for the exact pdf of PSO under these different boundary modes.

Few general remarks on the diffusion approach to the PSO is needed her. First, as we have just mentioned, the diffusion equation is subject to various boundary conditions that govern the constraints on the swarm evolution. The Gaussian pdf in (52) was stated only as a particular solution under some general conditions (51). If the PSO can be modeled as an evolution of particles under certain boundary modes, then one needs to construct the appropriate diffusion model and solve the boundary value problem. The formal implementation of this program requires further theoretical investigations and numerical studies, which are beyond the scope of the current work. The theory of diffusion is very rich and mature in theoretical physics; our sole purpose is to invite mathematically-oriented researchers in the optimization community to consider learning from physicists in that regard.

8. MARKOV MODEL FOR SWARM OPTIMIZATION TECHNIQUES

8.1. Introduction

In this section, a unified approach to the PSO techniques presented in this paper will be proposed based on the theory of Markov chains. One can visualize the Markov process as a hidden (deeper) layer in the description of dynamic problems. It does not present new behavior in the phenomenon under study, but may provide better insight on the way we write down the equations of motion. Since algorithms are sensitive to the various mathematical schemes used to implement them (e.g., some models are computationally more efficient), it is possible to gain new advantages, theoretical and conceptual, by investigating other deeper layers like Markov chains.

We start from the observation that the Gaussian pdf of particle positions lead to the possibility of eliminating the velocity from the algorithm and developing, instead, position-only update equations [26]. The recently introduced quantum version of the PSO has naturally no velocity terms [13–15]. This is because in quantum mechanics, according to Heisenberg's principle of uncertainty, it is not possible to determine both the position and the velocity with the same precision. It can be shown that all quantum mechanical calculations will produce a natural conditional pdf for the positions in the swarm environment as solutions of the governing Schrödinger's equation [15]. Therefore, one way to unify both the quantum and classical approaches is to consider their common theme: The existence of certain conditional probability distributions underlying the stochastic dynamics of the problem. Since

this pdf can be obtained either by solving the diffusion equation for the classical PSO, or Schrödinger's equation for the quantum PSO, then the update equation can be directly derived from the pdf without using any further physical argument. An example will be given in the next section.

8.2. Derivation of the Update Equations Using Probability Theory

The diffusion equation and Schrödinger's equation can be considered as "sources" for various probability density functions that govern the stochastic behavior of moving particles. Now, since the pdf of the particle at the kth iteration is known, we will simulate the trajectory by transforming a uniformly distributed random variable into another random variable with the desired pdf. Let the required transformation that can accomplish this be denoted as $y = h(x)$. Then, the cumulative density function (cdf) of both X and Y will be $F_X(x)$ and $F_Y(y)$, respectively. It is easy to show that for the transformation h to produce the distribution $F_Y(y)$, starting from the uniform distribution $F_X(x)$, one must have [20]

$$y = F_Y^{-1}(x) \tag{55}$$

As an example, consider the Laplace distribution that was obtained from the delta potential well version of the quantum PSO algorithm [13–15]. The pdf of the particle trajectory can be written as

$$f\left(r_k^{i,n} \middle| r_{k-1}^{i,n}\right) = \frac{g\ln\sqrt{2}}{\left|r_{k-1}^{i,n}\right|} \exp\left\{-2g\ln\sqrt{2}\frac{\left|r_k^{i,n}\right|}{\left|r_{k-1}^{i,n}\right|}\right\} \tag{56}$$

where $r_k^{i,n}$ is defined as in (50) and g is the only control parameter in the quantum PSO. By solving (55), denoting the uniform random variable by u, and the position r by y, one obtains

$$x_k^{i,n} = \begin{cases} p_{k-1}^{i,n} + \dfrac{\left|x_{k-1}^{i,n} - p_{k-2}^{i,n}\right|}{2}\ln(2u) & 0 < u \le \dfrac{1}{2} \\ p_{k-1}^{i,n} + \dfrac{\left|x_{k-1}^{i,n} - p_{k-2}^{i,n}\right|}{2}\ln\dfrac{1}{2(1-u)} & \dfrac{1}{2} < u \le 1 \end{cases} \tag{57}$$

This alternative derivation of the update equations was accomplished without reference to any argument in physics; the concept of collapsing the wave function, as presented in [15], has not been mentioned at all; only probability theory is needed in arriving

to (57). Moreover, notice that the control parameter g cancels out in the last expression. The new update equations above were tested numerically and demonstrated a performance as good as the original quantum PSO.

8.3. Markov Chain Model

A Markov process is defined as a stochastic process whose past has no influence on the future if its present is specified [20]. Mathematically, we state this definition as follow

$$P\left\{\mathbf{x}_n \leq x_n | \, \mathbf{x}_{n-1}, \mathbf{x}_{n-2}, ..., \mathbf{x}_1, \mathbf{x}_0\right\} = P\left\{\mathbf{x}_n \leq x_n | \, \mathbf{x}_{n-1}\right\} \tag{58}$$

where $\mathbf{x}_k$, $k = 1, 2, ..., n$, is a sequence of random variables defined over a time index k and P is the probability operator. It is understood that the sequence starts at the initial time $k = 0$, which can be considered the *cause*, and then "propagates" the *effect* through certain stochastic rules. Equation (58) tells us that the effect at a certain time instance k is dependent only on the information of the previous time step $k-1$. If we define the information of the sequence at k to be the state at that time instant, then it is possible to describe the Markov chain as a stochastic process in which the system can remember only the previous state.

From the analysis presented at the previous parts of this article, we have found that the PSO algorithm, being classic or quantum, involves a conditional pdf in which information about the current position depends only on information coming from the previous time step. In other words, by examining some of the distributions presented here, such as (53) and (56), it is clear that only the previous state (iteration) will be involved in the evaluation of the probability of the next iteration. Therefore, we classify all of the possible versions of the PSO algorithm, being classic or quantum, as stochastic Markov chains.

It is clear that the functional form of the conditional pdf does depend on the time index k. This is because the new center of gravity $p_k^{i,n}$ [see equation (50)] varies from iteration to iteration. A Markov process that satisfies such a property is called inhomogeneous Markov process [20].

From the chain rule in probability theory, one can write [20]

$$f\left(\mathbf{x}_k, \mathbf{x}_{k-1}, ..., \mathbf{x}_1, \mathbf{x}_0\right) = f\left(\mathbf{x}_k | \, \mathbf{x}_{k-1}\right) f\left(\mathbf{x}_{k-1} | \, \mathbf{x}_{k-2}\right) \cdots f\left(\mathbf{x}_2 | \, \mathbf{x}_1\right) f\left(\mathbf{x}_0\right) \tag{59}$$

where f stands for the pdf of the random variables presented at its argument.

Equation (59) implies that the joint statistics of all the positions, evaluated at all iterations up to the present time, can be expressed

in terms of the conditional pdf, at various previous times, plus the initial distribution at the first iteration $f(\mathbf{x}_0)$. This initial condition is independent of the algorithm and must be chosen by the user. It is customary in evolutionary methods to select the starting population as a uniform random distribution over the range of interest.

Of special importance in Markov chains is the following result, known as Chapman-Kolmogorov equation [20]

$$f\left(\mathbf{x}_n|\,\mathbf{x}_k\right) = \int_{-\infty}^{\infty} d^N x_m \, f\left(\mathbf{x}_n|\,\mathbf{x}_m\right) f\left(\mathbf{x}_m|\,\mathbf{x}_k\right) \tag{60}$$

where $n > m > k$. Here $d^N x_m$ stands for the volume element (scalar) in the abstract N-dimensional space $\mathbf{R}^N$. Equation (60) can be used to establish a causal link between any combinations of three iterations that need not to be successive.

For the general case of k successive steps, it easy to calculate the marginal pdf of the kth state as follows

$$f\left(\mathbf{x}_k\right) = \int_{x_{k-1}} \int_{x_{k-2}} .... \int_{x_0} d^N x_{k-1} \, d^N x_{k-2} ... \, d^N x_0 \, f\left(\mathbf{x}_k, \mathbf{x}_{k-1}, ..., \mathbf{x}_1, \mathbf{x}_0\right) \tag{61}$$

By substituting (60) into (61) we get

$$\begin{aligned} f\left(\mathbf{x}_k\right) \; = \; & \int_{x_{k-1}} \int_{x_{k-2}} .... \int_{x_0} d^N x_{k-1} \, d^N x_{k-2} ... \, d^N x_2 \, d^N x_0 \, f\left(\mathbf{x}_k|\,\mathbf{x}_{k-1}\right) \\ & \cdots f\left(\mathbf{x}_{k-1}|\,\mathbf{x}_{k-2}\right) f\left(\mathbf{x}_2|\,\mathbf{x}_1\right) f\left(\mathbf{x}_0\right) \end{aligned} \tag{62}$$

This gives the pdf of the future kth iteration by causally accumulating causally the effects of all previous states (iterations) as originated from an initial distribution.

8.4. Generalized PSO Algorithm

It is clear from equation (62) that knowledge of the conditional pdf of the sequence of positions, together with the pdf of the initial distribution, is enough to give an exact characterization of the total pdf for any future iteration. Based on these facts, we propose a generalized PSO algorithm summarized as follow

(i) Specify an initial distribution with known pdf $f(\mathbf{x}_0)$.

(ii) Derive the conditional pdf $f\left(\mathbf{x}_k|\,\mathbf{x}_{k-1}\right)$, for arbitrary k, by solving the diffusion equation (classic PSO) or Schrödinger's equation (quantum PSO).

(iii) Calculate the marginal pdf $f(\mathbf{x}_k)$ using equation (62).

(iv) Update the position by realizing the random number generator with the pdf $f(\mathbf{x}_k)$ (for example, the method presented in Subsection 8.2).

The generalized PSO algorithm, as presented above, has the following advantages over the traditional classical or quantum versions

(i) The velocity terms are canceled. If the derived update equations of the position are not complicated, then this may considerably reduce the computational complexity of the new algorithm.

(ii) The algorithm is purely stochastic. The rich literature on stability and time evolution of stochastic processes can be applied to provide a deeper understanding of the convergence of the PSO algorithm.

(iii) The generalized PSO algorithm eliminates the variations in the physics behind the method; it shows that all PSO methods are special classes within the wider group of Markov chains.

8.5. Prospectus and Some Speculations

In particular, it can be inferred from the overall discussion of this section that the Markov approach unifies the classical and the quantum versions of the swarm mechanics, a problem that is fundamental in theoretical physics. It is interesting to refer to the work proposed, yet in a different context, by Nelson in 1966 [27]. The idea was to derive quantum mechanics from a classical model starting from the stochastic diffusion equation. Even though not all physicists have accepted the idea, very recently a general approach based on stochastic dynamics was proposed to provide a theoretical derivation of quantum mechanics by formulating the problem as a stochastic dynamical system obeying some reasonable energy-related principles [28]. If quantum mechanics can be derived from the diffusion equation, then the diffusion theory of Section 7 and the Markov model of this section are enough to theoretically derive any possible PSO algorithm that can be mapped to existing physical theory. A general approach to establish physical theory for natural selection was formulated by Lee Smolin [29] where it can be seen that evolution, as described in biology, could be linked, on the fundamental level, to the Standard Model of particle physics. Similar ideas, beside the approach presented in our work, are in conformity with each other and strongly suggest that the foundations of the swarm intelligence methods could be rooted ultimately in physics, rather than biology.

9. CONCLUSIONS

In this paper, establishing a formal analogy between the PSO environment and the motion of a collection of particles under Newtonian laws provided a new insight on the PSO algorithm. In particular, the well-established branch of physics molecular dynamics was exploited to study the thermodynamic behavior of the swarm. The thermodynamic analysis provided a new global criterion to decide whether a convergence of the algorithm is successful or not. A new acceleration technique for the classical PSO algorithm was proposed based on thermodynamic theory. The new technique was employed successfully in improving the convergence performance with the practical problem of synthesis of linear array antennas. A diffusion model for the PSO environment was proposed. Upon solving the corresponding diffusion equation, the familiar Gaussian distribution, observed only empirically before, was obtained. The new diffusion model was integrated with the quantum model, developed in other works, to propose a unified scheme using the theory of Markov chains. The generalized algorithm has several advantages, like the possibility of reducing the computational complexity and opening the door for plenty of stability analysis techniques to be applied for the study of the basic PSO algorithm.

ACKNOWLEDGMENT

This work was partially supported by the Department of Electrical Engineering, University of Mississippi and the National Science Foundation under Grant No. ECS-524293.

REFERENCES

1. Levin, F. S., *An Introduction to Quantum Theory*, Cambridge University Press, 2002.
2. Sijher, T. S. and A. A. Kishk, "Antenna modeling by infinitesimal dipoles using genetic algorithms," *Progress In Electromagnetics Research*, PIER 52, 225–254, 2005.
3. Mahanti, G. K., A. Chakrabarty, and S. Das, "Phase-only and amplitude-phase only synthesis of dual-beam pattern linear antenna arrays using floating-point genetic algorithms," *Progress In Electromagnetics Research*, PIER 68, 247–259, 2007.
4. Meng, Z., "Autonomous genetic algorithm for functional optimization," *Progress In Electromagnetics Research*, PIER 72, 253–268, 2007.

5. Riabi, M. L., R. Thabet, and M. Belmeguenai, "Rigorous design and efficient optimizattion of quarter-wave transformers in metallic circular waveguides using the mode-matching method and the genetic algorithm," *Progress In Electromagnetics Research*, PIER 68, 15–33, 2007.
6. Kennedy, J. and R. C. Eberhart, "Particle swarm optimization," *Proc. IEEE, Conf. Neural Networks IV*, Piscataway, NJ, 1995.
7. Kenedy, J. and R. C. Eberhart, *Swarm Intelligence*, Morgan Kaufmann Publishers, 2001.
8. Clerc, M. and J. Kennedy, "The particle swarm: explosion, stability, and convergence in a multi-dimensional complex space," *IEEE Trans. Evolutionary Computation*, Vol. 6, No. 1, 58–73, Feb. 2002.
9. Kadirkamanathan, V., K. Selvarajah, and P. J. Fleming, "Stability analysis of the particle swarm optimizer," *IEEE Trans. Evolutionary Computation*, Vol. 10, No. 3, 245–255, June 2006.
10. Ciuprina, G., D. Ioan, and I. Munteanu, "Use of intelligent-particle swarm optimization in electromagnetics," *IEEE Trans. Magn.*, Vol. 38, No. 2, 1037–1040, Mar. 2002.
11. Robinson, J. and Yahya Rahmat-Samii, "Particle swarm optimization in electromagnetics," *IEEE Trans. Antennas Progat.*, Vol. 52, 397–407, Feb. 2004.
12. Boeringer, D. and D. Werner, "Particle swarm optimization versus genetic algorithms for phased array synthesis," *IEEE Trans. Antennas Progat.*, Vol. 52, No. 3, 771–779, Mar. 2004.
13. Sun, J., B. Feng, and W. B. Xu, "Particle swarm optimization with particles having quantum behavior," *Proc. Cong. Evolutionary Computation, CEC2004*, Vol. 1, 325–331, June 2004.
14. Mikki, S. M. and A. A. Kishk, "Investigation of the quantum particle swarm optimization technique for electromagnetic applications," *IEEE Antennas and Propagation Society International Symposium*, 45–48, Vol. 2A, July 3–8, 2005.
15. Mikki, S. M. and A. A. Kishk, "Quantum particle swarm optimization for electromagnetics," *IEEE Trans. Antennas Progat.*, Vol. 54, 2764–2775, Oct. 2006.
16. Haile, J. M., *Molecular Dynamics Simulation*, Wiley, New York, 1992.
17. Schommers, W., "Structures and dynamics of surfaces I," *Topics in Current Physics*, W. Schommers and P. von Blanckenhagen (eds.), Vol. 41, Springer-Verlag, Berlin, Heidelberg, 1986.

18. Rieth, M., *Nano-Engineering in Science and Technology*, World Scientific, 2003.
19. Kreyszig, E., *Advanced Engineering Mathematics*, 8th edition, Wiley, 1999.
20. Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, 3rd edition, McGraw-Hill, 1991.
21. Penrose, R., *The Emperor's New Mind*, 1989.
22. Gossick, B. R., *Hamilton's Principle and Physical Systems*, Academic Press, 1967.
23. Iwasaki, N. and K. Yasuda, "Adaptive particle swarm optimization using velocity feedback," *International Journal of Innovative Computing, Information and Control*, Vol. 1, No. 3, 369–380, September 2005.
24. Bhattacharyya, A. K., *Phased Array Antennas*, Wiley-Interscience, 2006.
25. Einestien, A., "On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat," *Ann. Phys.* (Leipzig), Vol. 17, 549–560, 1905.
26. Kennedy, J., "Probability and dynamics in the particle swarm," *Proc. Cong. Evolutionary Computation, CEC2004*, Vol. 1, 340–347, June 19–23, 2004.
27. Nelsson, E., "Derivation of Schrödinger equation from Newtonian mechanics," *Phys. Rev.*, Vol. 150, 1079–1085, 1966.
28. Smolin, L., "Could quantum mechanics be approximation to another theory?" http://arxiv.org/abs/quant-ph/0609109, September 2006.
29. Smolin, L., *Life of the Cosmos*, Oxford University Press, 1999.