

Simulated annealing for topological solitons

Mark Hale

Centre for Particle Theory, University of Durham, Durham DH1 3LE, United Kingdom

Oliver Schwindt and Tom Weidig

Department of Physics, UMIST, Manchester M60 1QD, United Kingdom

(Received 4 February 2000)

The search for solutions of field theories allowing for topological solitons requires that we find the field configuration with the lowest energy in a given sector of topological charge. The standard approach is based on the numerical solution of the static Euler-Lagrange differential equation following from the field energy. As an alternative, we propose to use a simulated annealing algorithm to minimize the energy functional directly. We have applied simulated annealing to several nonlinear classical field theories: the sine-Gordon model in one dimension, the baby Skyrme model in two dimensions and the nuclear Skyrme model in three dimensions. We describe in detail the implementation of the simulated annealing algorithm, present our results and get independent confirmation of the studies which have used standard minimization techniques.

PACS number(s): 02.70.-c, 12.39.Dc

I. INTRODUCTION

Solitons have played an ever increasing role in the description of physical phenomena since their discovery by Russell [1] in 1834. Generally speaking, a soliton is a stable localized solution of a nonlinear partial differential equation which propagates at a constant speed and stays localized even after an interaction with another soliton (see Ref. [2] for an introduction to solitons). They are particlelike extended objects. Well-known soliton models in one dimension (1D) are the Korteweg–de Vries (KdV) and the sine-Gordon models. The stability of the KdV soliton is due to the dynamical balance between the nonlinear and the dispersive terms in the KdV equation. This differential equation belongs to the class of integrable models that can be solved exactly. The sine-Gordon model is also integrable and the stability of its soliton is also based on the balance between nonlinearity and dispersion. However, it also belongs to the wider class of models whose solitons are stable by conservation of a topological charge or winding number, as discussed in Sec. IV A. In topology, the field is interpreted as a mapping from physical space to field space and a field configuration with given topological charge cannot dynamically change into a field configuration with a different charge. In this paper, when we discuss topological solitons, we shall consider only the field configuration with lowest energy in a nonzero topological charge sector.

Topological solitons arise in many areas of physics: field theory (e.g., vortices, monopoles, and instantons, as discussed in Ref. [3], and Hopf solitons [4,5]), condensed matter (e.g., baby Skyrmions [6]), nuclear physics (Skyrmions, see Ref. [7]), cosmology (e.g., cosmic strings [8]), and string theory or M-theory (e.g., Olive-Montonen duality [9]). They possess many interesting properties. The conservation of topological charge can be used to model particle conservation and annihilation: the number of solitons is conserved and two solitons with opposite charge can annihilate. Since the stability of solitons is assured by topology, there exists considerable freedom in constructing appropriate Lagrangians

for physical systems. The common constraint of Lorentz invariance is easily imposed by using covariant terms. However, the use of topological models has a major disadvantage. The models are generally not integrable and few analytical techniques are available (only *Ansätze*, topological bounds, etc.). It is therefore crucial to study the models using reliable and efficient numerical methods.

We are looking for the (static) lowest-energy field configuration in a given topological sector. Thus, we need to minimize the energy functional E of the field theory, the integral of an energy density \mathcal{E} over a manifold M :

$$E[C] = \int_M d^n x \mathcal{E}(x, f(x), f'(x)), \quad (1)$$

where $f(x)$ is a field configuration C . The topology typically imposes some boundary conditions $f(\partial M)$. We are searching for the function $f_{\min}(x)$ that gives the lowest value for E . There are two possible approaches: to solve the Euler-Lagrange equation of the functional E with respect to the function $f(x)$ or to minimize E through some other means.

Hitherto only the first approach, via the Euler-Lagrange equation, has been used with topological systems. We shall review the standard numerical techniques that apply shooting or relaxation methods and discuss their reliability and ease of use. In this paper, we show how to minimize the energy functional directly by using the simulated annealing (SA) algorithm, as proposed in Refs. [10,11]. SA is based on the fact that a solid which is slowly cooled down, assuring thermal equilibrium at each temperature, reaches its ground state. The SA algorithm describes the cooling process and a Metropolis subalgorithm brings a system into thermal equilibrium. SA has been applied to minimization problems in such diverse areas as combinatorial optimization (such as the traveling salesman problem), circuit design, finance, physics, and military warfare: see Refs. [12,13], Sec. 10.9. We give a detailed introduction to SA and describe our implementations. We find the topological solitons of the sine-Gordon model in 1D, the baby Skyrme model in 2D, and the nuclear

Skyrme model in 3D and compare our results to those obtained using standard minimization techniques.

II. MINIMIZATION VIA THE EULER-LAGRANGE EQUATION

The standard procedure uses the Euler-Lagrange equation resulting from the variation of the functional E with respect to the function $f(x)$ to find the minimal energy solution $f_{\min}(x)$. In the 1D case, for example, the problem is a two-point boundary value problem satisfying the differential equation

$$\frac{d}{dx} \left(\frac{d\mathcal{E}}{df'} \right) - \frac{d\mathcal{E}}{df} = 0. \quad (2)$$

It is a second order ordinary differential equation (ODE), and a partial differential equation in two or more dimensions, which is equivalent to a set of first order ODEs. Let $f(a)$ and $f(b)$ represent the boundary conditions a field configuration has to satisfy over the interval $[a, b]$. There are two standard approaches: the shooting and the relaxation methods (as discussed in [13], Chap. 17).

A. The shooting method

The shooting method is usually based on integration from one boundary to the other. The value of the function at the point $x=a$ is taken to be $f(a)$ and an initial guess α for its derivative is made. A numerical integration, for example with a Runge-Kutta method, up to the other boundary point $x=b$ then gives an estimate $f_{\alpha}(b)$ for f at b . This value is compared to the known boundary value $f(b)$ and α is adjusted to match $f_{\alpha}(b)$ more closely to $f(b)$. This procedure is repeated until the desired accuracy is achieved. The shooting method is unrivaled in speed and accuracy, but applicable only in one dimension.

B. Relaxation methods

Gauss-Seidel over-relaxation (SOR) is commonly used to solve the boundary problem directly. A time-dependent differential equation, a diffusion equation, is constructed out of the 1D ODE (2),

$$\frac{df(x,t)}{dt} = \omega dx^2 \left[\frac{d}{dx} \left(\frac{d\mathcal{E}}{df'} \right) - \frac{d\mathcal{E}}{df} \right]. \quad (3)$$

If the system reaches equilibrium, i.e., $df/dt=0$, this configuration is a solution to Eq. (2). One starts out with a configuration satisfying the boundary conditions. The coefficient ω of the leading term, which has the form d^2f/dx^2 , is dimensionless and determines the speed of convergence. The choice of integration method is not very sensitive; we can use Euler integration, the Runge-Kutta method, or the Crank-Nicholson method. The standard SOR uses the Euler method with updated information from already computed field values at lattice points and ensures better convergence. It is possible to neglect the derivative of the energy functional with respect to the derivative of f , i.e., we consider only the change with respect to the degree of freedom itself; see, e.g., [4].

In many cases one also studies the time evolution of the models. Computer codes developed for this purpose can be adapted to find minimal energy solutions by adding a damping term to the equation of motion. For example, in [14], two baby Skyrmions are put in an attractive configuration and form an oscillating bound state. As the system has been made dissipative, the energy of the system decreases with time until a minimum is reached. Thus, finding a minimal energy solution is translated into a damped time evolution. The same effect can often be reached by working with a finite box and absorbing any outward propagating radiation on the boundaries.

Relaxation techniques are well documented and applicable in any dimension (see Ref. [13]). The SOR method has theoretically the best rate of convergence but might be less than optimal since the best choice of ω can rarely be determined for a nonlinear system and must be made by trial and error. Using damping in a time-evolution problem is convenient, but one first has to set up the time-evolution code. It is impossible to estimate the error on an integration step and one needs to monitor conserved quantities. This is especially important if, as is often the case, the field has to satisfy a constraint. Furthermore, if the initial configuration is far from the global minimum, we might end up in a local minimum. Moreover, the derivation of the corresponding Euler-Lagrange equation becomes increasingly difficult when higher-order terms are added to the Lagrangian or when complicated constraints on the field space are present.

We have come to the conclusion that the weaker points of iterative minimization techniques via the Euler-Lagrange equation are: uncertainty about the global nature of the minimum obtained; lack of direct control over the integration errors (important for constrained fields), and tedious derivation for complicated Lagrangians.

III. MINIMIZATION VIA SIMULATED ANNEALING

Minimizing the energy functional directly is a more straightforward approach than solving the equations of motion, and we propose to use the flexible and easy-to-implement simulated annealing technique.

A. Metropolis principle

In 1953 Metropolis *et al.* [15] proposed an algorithm, now called the Metropolis or $M(RT)^2$ algorithm, that can be used to bring a statistical system into thermal equilibrium. The $M(RT)^2$ is most commonly used to evaluate thermal averages $\langle \mathcal{F} \rangle$ of a quantity $\mathcal{F}(C)$,

$$\langle \mathcal{F} \rangle = \frac{\int \mathcal{F}(C) P(C) dC}{\int P(C) dC}. \quad (4)$$

Here $P(C)$ is a probability distribution for configurations C . It must satisfy $P(C) \geq 0$ and $\int P(C) dC < \infty$ in order to be normalizable. For example, $\langle \mathcal{F} \rangle$ can stand for the thermal average of our energy functional $E[C]$ in Eq. (1). In fact, the Metropolis algorithm is only one of the possible sampling methods for the Monte Carlo evaluation of the integral (see [16], Sec. 3.7).

For the system to reach thermal equilibrium it needs to satisfy the condition of detailed balance,

$$K(C_2|C_1)P(C_1) = K(C_1|C_2)P(C_2). \quad (5)$$

Here $P(C)$ is the probability of finding the system in the configuration, or state, C , and $K(C_2|C_1)$ is the conditional probability of moving from C_1 to C_2 . The conditional probability K is usually decomposed as

$$K(C_2|C_1) = A(C_2|C_1)T(C_2|C_1), \quad (6)$$

where the transition probability $T(C_2|C_1)$ can be chosen to be any normalized distribution. It is used to select a random trial move from C_1 to C_2 . The complications are kept in $A(C_2|C_1)$, which gives the probability of accepting this move and is the correction to the arbitrarily chosen $T(C_2|C_1)$. The key element of the algorithm is the evaluation of the function $A(C_2|C_1)$ by a rejection technique. Thus the function $T(C_2|C_1)$ is sampled, and the resulting configuration is accepted or rejected depending on the value of $A(C_2|C_1)$. One usually defines

$$q(C_2|C_1) = \frac{T(C_1|C_2)P(C_2)}{T(C_2|C_1)P(C_1)} \geq 0 \quad (7)$$

and

$$A(C_2|C_1) = \min(1, q(C_2|C_1)). \quad (8)$$

If the configuration C_2 has a lower energy than C_1 , it is accepted. Otherwise, it is accepted with the probability $q(C_2|C_1)$. This procedure is repeated a large number of times and eventually the system reaches an equilibrium. Here we define an equilibrium to be the ensemble of states where the average of the energy does not show systematic changes.

After L steps, equilibrium is established and the system fluctuates around $\langle \mathcal{F} \rangle$. The thermal average is approximated by the sum

$$\langle \mathcal{F} \rangle = \frac{1}{N} \sum_{i=L+1}^{L+N} \mathcal{F}(C_i), \quad (9)$$

where C_i is a state at thermal equilibrium and N is the number of iterations over which we compute the average. We can interpret every trial move as representing a unit of quasitime having passed. This cannot be converted to real units of time, but it is possible to average thermodynamic properties over quasitime when a system is in equilibrium. It is possible to compute the mean of the quantity because the unit of measurement, which is the number of trial moves, cancels. If a trial move is rejected, the old configuration has to be counted in any averages.

For all the examples to be discussed, $P(C) = e^{\beta E[C]}$, where the temperature is defined by $\beta = (k_B T)^{-1}$. If the transition probability $T(C_2|C_1)$ is chosen to be uniform, $q(C_2|C_1) = e^{\beta(E[C_2] - E[C_1])}$, or

$$q(C|C_{\text{new}}) = e^{\beta(E - E_{\text{new}})}. \quad (10)$$

Here, E is the energy of the system in the present configuration C and E_{new} is the energy of the new configuration C_{new} that was obtained through a random change in the state of the system (sampled from the distribution T). If the energy of the new configuration C_{new} is lower, the change is accepted. If

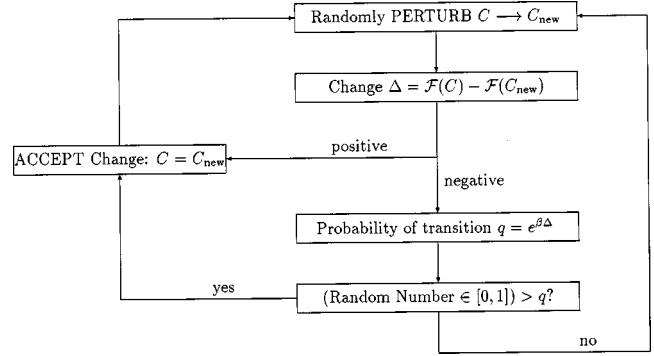


FIG. 1. The Metropolis algorithm: Scheme for thermal equilibrium.

the energy is higher, the system accepts this upward step with a transition probability q . Thus the system can escape local minima and achieve thermal equilibrium. In Fig. 1 we show a flow diagram representing the Metropolis algorithm. In this diagram, \mathcal{F} denotes the functional being minimized.

B. Simulated annealing

In 1982, Kirkpatrick, Gelatt, and Vecchi [17] observed a deep analogy between annealing of solids and optimization or minimization problems. A solid that is sufficiently slowly cooled down, i.e., is at thermal equilibrium at each temperature, will reach its ground state. If the energy is the functional to be optimized or minimized, the SA scheme should find the minimum energy function of this functional according to a statistical proof by Geman and Geman [18]. One starts out with a configuration at a high temperature and runs the Metropolis algorithm. The size of change of the functional is proportional to the temperature. Once we have reached thermal equilibrium, the temperature is decreased according to a cooling schedule and the procedure repeated as often as necessary. SA is a conceptually easy-to-understand minimization technique.

There are several varieties of SA algorithms, each designed to speed up the minimization of a particular problem. The application to a minimization of a continuous problem deserves some reflection on the discretization of the derivatives. The most important question is the cooling schedule; a sufficiently slow cooling is crucial for the “statistical proof of convergence.” Quite often, a slow cooling is not needed to reach the global minimum and a faster cooling schedule can be used: this SA version is called simulated quenching. For more information on SA, we recommend reading Refs. [12,13], Sec. 10.9. There is no unique way of implementing the SA scheme and there exists ample opportunity to improve the code. However, every SA implementation faces the same issues.

Which initial guess? Unlike the case of the relaxation method, the initial configuration is not important as the system should be able to jump out of local minima. However, an initial guess close to the global minimum solution can lead to a reduction of the running time.

What sampling method to use? The changes to C should be made such that the configuration space is well sampled. As the temperature decreases, so should the size of the changes. Usually the choices made are random in the con-

figuration space with a Gaussian or Lorentzian distribution. However, since we are dealing with functionals rather than functions such steps are expensive to compute. Therefore, we restrict ourselves to changes at individual gridpoints.

At what initial temperature to start? A high temperature puts the system in thermal equilibrium quickly, but at too high temperature the soliton unwinds. Too low a choice for the initial temperature can leave the system in a local minimum. The best choice is found by trial and error.

When is equilibrium reached? The determination of the equilibrium position is crucial and a statistical study of the changes of the system is essential. Equilibrium is reached when the energy of the system fluctuates but does not show any systematic trend. However, metastable states have been observed (like glasses), where the state of the system changes so slowly that one runs the risk of interpreting it as being constant.

What choice of cooling schedule? The temperature should decrease according to a logarithmic rule to assure convergence to the global minimum (see Ref. [18]). However, it takes a long time to reach thermal equilibrium with such a cooling schedule. Often, the cooling is speeded up by an exponential cooling schedule using big temperature decreases or a weaker equilibrium condition.

Discretization of the continuous functional \mathcal{F} ? It is important not to use the central difference, because it does not depend on the function at the center point. However, this problem can be overcome by computing the derivatives midway between two gridpoints. We discuss this later.

Use of constraints? Constraints are no problem, because we use random changes that satisfy the constraints.

IV. SIMULATED ANNEALING IN ONE DIMENSION

We have used the sine-Gordon model for our 1D SA implementation, because it is one of the simplest field theories exhibiting extended structures and it is exactly solvable (see Ref. [19]). The sine-Gordon model is also a very good toy model for solitonic quantum field theories, for the quantum mass correction and the S matrix are exactly known. Further, Coleman [20] has shown that the quantum sine-Gordon model and the massive Thirring model are dual to each other: the bosonic soliton in the sine-Gordon is a fermion in the massive Thirring model. Finally and most importantly for this paper, the soliton solution is known exactly and we can compare it to our SA results.

A. The sine-Gordon model

The sine-Gordon model is described by the Lagrangian density

$$\mathcal{L} = \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - (1 - \cos \phi). \quad (11)$$

For simplicity, we have set the mass and the coupling to 1. The Lagrangian is invariant under $\phi \rightarrow \phi + 2\pi n$, $n \in \mathbf{Z}$. Here $\phi(x, t)$ is an angle in field space, the circle S^1 . The field has to go to the vacuum sufficiently fast for the soliton to be localized and of finite energy. Therefore, we can identify the spatial infinity in each direction with one single point and

compactify the one-dimensional space \mathbf{R}^1 to S^1 . The field theory of the sine-Gordon model can be described by the map

$$\phi(t): S^1 \rightarrow S^1 \quad (12)$$

at a given time t . This nontrivial mapping gives us the possibility to partition the space of all possible field configurations into equivalence classes having the same topological charge or winding number. We can visualize this concept with a belt. We can trivially close it or we can twist one side by 180° and close it or we can antitwist it by 180° , i.e., twist it by -180° , and close it. The twist in the belt cannot be undone unless one opens the belt. Topological solitons can be thought of as twisted field configurations. The homotopy group $\Pi_1(S^1) = \mathbf{Z}$ describes the twists in the map. For example, if we twist the belt twice and then antitwist it twice, we get back to an untwisted belt, very much like an annihilation process in particle physics. The ‘‘twist,’’ i.e., the topological charge, is fixed by boundary conditions and conserved.

The corresponding Euler-Lagrange equation for the sine-Gordon model is

$$\ddot{\phi} - \phi'' + \sin \phi = 0. \quad (13)$$

One can find the minimal energy solution by solving the static version using theoretical or numerical methods. The 1-soliton, i.e., minimal energy solution of topological charge 1, can be derived from the Bogomolnyi equation (see [19], Sec. 2.5):

$$\phi' = \pm \sqrt{2(1 - \cos \phi)}. \quad (14)$$

Rewriting this in terms of $\sin(\phi/2)$, integrating, and inverting the resulting relation, one finds

$$\phi_{s,t}(x) = 4 \arctan[\exp(x + x_0)]. \quad (15)$$

We can derive the solitons with higher charge via a Bäcklund transformation [19]. The static minimal energy solution satisfies the boundary condition $\phi(-\infty) = 0$ and $\phi(\infty) = 2\pi$; the field winds around the field sphere S^1 once. The expression for the energy density is

$$\varepsilon(x) = 4 \sin^2[\phi_{st}(x)]. \quad (16)$$

The energy goes to zero at spatial infinity and the integral is finite. The total energy is $\int \varepsilon(x) dx = 8$ sine-Gordon energy units. In the next section, we discuss the 1-soliton, Eq. (15), and calculate its total energy using the SA scheme.

B. Implementation of simulated annealing

Three aspects of the SA implementation are crucial for successful minimization: the derivatives, the sampling method, and the cooling schedule with thermal equilibrium.

1. Derivatives

The most accurate discretized derivative is the centered difference. However, this causes problems with derivative terms as it does not depend on the function at the point where the energy is being evaluated. This results in a decou-

pling between neighboring points, which gives rise to two independent sublattices. The configuration becomes spiky since the values jump between the two sublattices. To avoid this problem, the energy is computed between the grid points rather than at the grid points. The value of the function between the gridpoints is taken to be the average of the values at the surrounding points:

$$\phi(x_{i+1/2}) = \frac{\phi(x_i) + \phi(x_{i+1})}{2}, \quad (17)$$

$$\frac{\partial \phi(x_{i+1/2})}{\partial x} = \frac{\phi(x_{i+1}) - \phi(x_i)}{dx}. \quad (18)$$

2. Sampling

Typically, SA is used to minimize a function $f(x)$ with x being a vector. The general form of a change to a configuration is

$$x_i \rightarrow x_i + M_{ij} U_j,$$

where M_{ij} is a matrix, and U_i is a vector of random numbers satisfying an appropriate probability distribution (see Ref. [21]). The matrix M_{ij} needs to be chosen such that the configuration space is well sampled. Information from the cooling process can be used to dynamically adjust M_{ij} . We are interested in minimizing energy functionals on a lattice of N gridpoints so our x vector will have N components. This makes calculating a new configuration quite an intensive process. To simplify matters we sweep across the grid changing individual points one at a time. The random numbers U_i are taken from a Lorentzian distribution, rather than a Gaussian distribution. This is a quite common modification to the original SA algorithm, as the Lorentzian has a longer tail. The mean and width of the distribution need to be chosen so that we get a good sampling of configuration space. A narrow distribution will sample only the local neighborhood while a wide distribution will spend too much time probing irrelevant configurations. To achieve a good balance the width is adjusted so that 50% of all the proposed new configurations are accepted (this is called the acceptance rate). If the mean is taken to be linearly dependent on the temperature, then the acceptance rate will remain roughly constant throughout the cooling. This leaves the constant of proportionality to be determined at the start of the cooling process.

3. Cooling schedule and thermal equilibrium

We use an exponential cooling schedule; the temperature is decreased by a fixed ratio at each cooling step. This violates Geman and Geman's statistical guarantee of reaching the minimum solution. Since we do not expect many local minima in 1D, this should not be a problem here.

There are several approaches to determining whether the configuration is in equilibrium. A popular one is based on a sliding average, also known as binning, where the mean energy calculated over a number of iterations is monitored to see whether it has converged to a fixed value. We employ a simpler, related condition that monitors the lowest energy obtained during a set sequence or chain of iterations until no new low from one chain to the other is found. Since equilib-

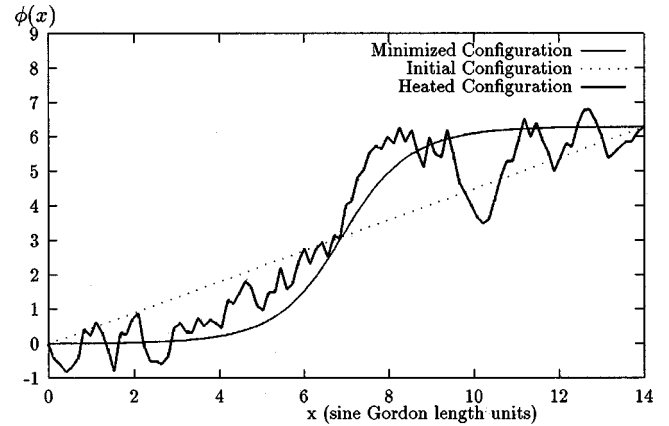


FIG. 2. 1D SA cooling for the sine-Gordon model.

rium is by its very nature statistical it is important to know how many iterations need to be sampled. A ballpark figure seems to be 10–15 samples per point on average. We chose the total number of points in the chain to be $10N$, where N is the number of grid points, and checked empirically that this was enough.

C. Results of simulated annealing

Specifically, we look here at the sine-Gordon model where we need to minimize the following energy functional:

$$E[\phi] = \int_{x=-a}^{x=a} dx \left(\frac{1}{2} \partial_i \phi \partial_i \phi + 1 - \cos \phi \right). \quad (19)$$

We impose a winding or topological charge of 1 by setting $\phi(-a) = 0$ and $\phi(a) = 2\pi$. We could use a 2D constrained field to represent S^1 , i.e., $\vec{\phi} = (\phi_1, \phi_2)$ with $\vec{\phi} \cdot \vec{\phi} = 1$. However, we opted for an angle representation, because it allows us to use fixed boundaries and the soliton cannot unwind. In the 2D constrained coordinates, the winding is a twist in the configuration over the whole grid and a very big fluctuation induced by a high temperature undoes the twist.

We use different grid sizes. In Figs. 2 and 3, we represent different aspects of the cooling of a sine-Gordon soliton. We start out with an initial field configuration, here a straight line satisfying the boundary conditions. We then heat up the configuration until thermal equilibrium is reached (thick solid

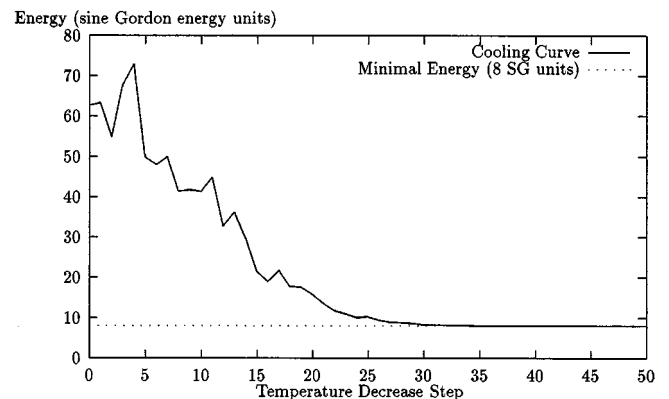


FIG. 3. Typical cooling curve for 1D SA in the sine-Gordon model.

TABLE I. Energy (sine-Gordon energy units) versus number of points used for the sine-Gordon model.

Points	Energy
51	8.035
101	8.009
201	8.002
301	8.001

line in Fig. 2). We cool it down by slowly decreasing the temperature after reaching thermal equilibrium. Finally, we obtain a minimal energy solution close to Eq. (15).

We set the acceptance rate to 50% and take $10N$ for the length of the monitored chain to test thermal equilibrium. A change in these values affects the speed and quality of minimization. We have rerun the minimization under the same setting and find the same energy value. This is a good indicator that the monitored chain is long enough to obtain thermal equilibrium. Our box size is 20 sine-Gordon length units. The more points we use the closer the result becomes to the exact soliton energy, namely, 8 (see Table I).

To conclude, we find our SA code to be a very convenient minimization technique in 1D. We have successfully tested our 1D SA code on many different models. The implementation of the SA search for solitons is faster, for we did not have to derive the Euler-Lagrange equation.

V. SIMULATED ANNEALING IN TWO DIMENSIONS

We have used the baby Skyrme model [14] for our 2D SA implementation, because there are exact and numerical solutions available that we can compare to our SA results. The baby Skyrme model is used to study some aspects of the quantum Hall effect (see Ref. [6]) and is a convenient $(2+1)$ D toy model for the $(3+1)$ D nuclear Skyrme model (see Ref. [7]), which requires much more computational resources.

A. The baby Skyrme model

The nonlinear σ model is described by the Lagrangian

$$\mathcal{L} = \frac{1}{2} \partial_\mu \vec{\phi} \cdot \partial^\mu \vec{\phi}, \quad (20)$$

where $\vec{\phi}$ is a three-dimensional field vector on the sphere S^2 , i.e., $\vec{\phi} \cdot \vec{\phi} = 1$. The field at a time t is a map

$$\phi(t): S^2 \rightarrow S^2 \quad (21)$$

and the associated homotopy group is $\Pi_2(S^2) = \mathbf{Z}$. The existence of the topological charge, i.e., the twisted field configuration representing a soliton, is ensured by topology. It is given by

$$B = \frac{1}{8\pi} \int d^2x \epsilon_{\mu\nu} \vec{\phi} \cdot (\partial^\mu \vec{\phi} \times \partial^\nu \vec{\phi}). \quad (22)$$

However, we also need to make sure that the soliton has a stable scale. From Derrick's theorem [22], the energy functional corresponding to Eq. (20) is scale invariant. A change of scale does not change the energy and therefore numerical

errors can significantly change the scale of the soliton. It is therefore necessary to add extra terms to stabilize the soliton, fixing the scale. If we were to extend the model to $(3+1)$ D, the σ model term would lead to an expanding soliton, and a balancing term needs to be added to ensure stability.

The baby Skyrme model is a modified version of the S^2 σ model and the Lagrangian is

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \partial_\mu \vec{\phi} \cdot \partial^\mu \vec{\phi} - \theta_S [(\partial_\mu \vec{\phi} \cdot \partial^\mu \vec{\phi})^2 \\ & - (\partial_\mu \vec{\phi} \cdot \partial_\nu \vec{\phi})(\partial^\mu \vec{\phi} \cdot \partial^\nu \vec{\phi})] - \theta_V V(\vec{\phi}). \end{aligned} \quad (23)$$

The addition of a potential and a Skyrme term to the Lagrangian ensures stable solitonic solutions. The Skyrme term has its origin in the nuclear Skyrme model and the baby Skyrme model can therefore be viewed as its $(2+1)$ D analog. Further, in $(2+1)$ D, a potential term is necessary in the baby Skyrme models to ensure stability of Skyrmions; this term is optional in the $(3+1)$ D nuclear Skyrme model. One drawback of the model is that the potential term is free for us to choose. The most common choices are $V = (1 + \phi_3)^4$ (the holomorphic model has an exact one-Skyrmion solution; see Ref. [23]), $V = (1 + \phi_3)$ (a one-vacuum potential studied in Ref. [24]), and $V = (1 - \phi_3)(1 + \phi_3)$ (a two-vacua potential studied in Ref. [14]). Except for the first choice, no closed form minimal energy solutions are known. The baby Skyrme model is a nonintegrable system, and explicit solutions to its resulting differential equations are nearly impossible to find. Numerical methods are the only way forward.

B. Implementation

Our 2D and 3D SA implementations originate from a more general framework, the study of phase transitions in topological systems at finite temperatures and densities [25]. The thermodynamic partition function describes a system at a given temperature and can only be evaluated numerically in the Skyrme models. The evaluation of thermal averages, as discussed before, can be done with Monte Carlo techniques and the Metropolis principle is one of the possible sampling techniques. Conveniently, the thermal average of the energy at zero temperature is equivalent to the minimal energy of the energy functional to be minimized.

We start out with the grand canonical partition function [26]

$$\begin{aligned} \mathcal{Z}(\beta, V, \mu) = & \int_{\text{all } x_1, \dots, x_N} \int_{\text{all } p_1, \dots, p_N} d^n x_1 d^n p_1 \cdots d^n x_N d^n p_N \\ & \times \sum_{i=0}^N \exp[\beta(\mu_i - E_i)], \end{aligned} \quad (24)$$

where E_i is the energy of the i th particle system at temperature $\beta = (k_B T)^{-1}$ and V is the integration range. The integral ranges over all phase space and is $2nN$ dimensional, where n is the number of space dimensions. The thermodynamic partition function for the baby Skyrme model has the form

$$\mathcal{Z}(\beta, V, \mu) = \int \prod_k \delta(\vec{\phi}_k \cdot \vec{\phi}_k - 1) d^3 \phi_k \sqrt{\left(\frac{1 + \partial_i \vec{\phi}_k \cdot \partial_i \vec{\phi}_k}{\det(\mathcal{M}_k)} \right)} \times \exp[-\beta(\mathcal{V}_k - \mu \mathcal{B}_k)]. \quad (25)$$

Here \mathcal{M} is the mass density matrix, \mathcal{V} the potential energy density, and \mathcal{B} the topological charge density. The input parameters of the thermodynamic partition function are the temperature β , the volume of the system V , and the chemical potential μ . The δ function is required due to the constraint on the $\vec{\phi}$ field.

At zero temperature, the factor in front of the exponent becomes irrelevant and, if we further set $\mu=0$, the thermodynamic partition function reduces to the integral

$$Z = \int \prod_k \delta(\vec{\phi}_k \cdot \vec{\phi}_k - 1) d^3 \phi_k \exp(-\beta \mathcal{V}_k), \quad (26)$$

where \mathcal{V}_k is the potential energy density. The implementation of \mathcal{Z} is similar to the implementing of Z , but \mathcal{Z} contains information that is not necessary for finding minimal energy solutions. The value of Z is not of interest to us, but the $\vec{\phi}$ distribution as $\beta \rightarrow \infty$ is. The probability density function that is sampled at every point k when applying the Monte Carlo technique is the sum over all neighbors of k ,

$$P_k = \exp\left(-\beta \sum_{i=1}^{\text{number of neighbors}} \mathcal{V}_i\right). \quad (27)$$

Examples of Monte Carlo calculations using a grand canonical ensemble are given in Refs. [27,28]. A good discussion of possible errors and how to deal with them is given in Ref. [28].

1. Monte Carlo simulation for the baby Skyrme model

We apply the Metropolis principle in the simplest possible way and select a new vector $\vec{\phi}_{\text{new}}(x_k)$ at a grid point k from a uniform probability distribution function over the unit sphere, as the integration measure with the δ function implies. We choose each of the components ϕ_a uniformly between -1 and 1 . If the sum of the squares of the components $\phi_1^2 + \phi_2^2 + \phi_3^2$ is larger than 1 the sample is rejected. All accepted vectors are scaled to obtain unit length [16]. The transition probability of this simple method is

$$T(C_2|C_1) = \begin{cases} \frac{1}{(\text{surface area of sphere})} & \text{on the unit sphere} \\ 0 & \text{otherwise,} \end{cases} \quad (28)$$

and therefore

$$q(C_2|C_1) = \frac{P(C_2)}{P(C_1)}, \quad (29)$$

where the present vector $\vec{\phi}_{\text{present}}(x_k) \in C_1$ and the newly selected vector $\vec{\phi}_{\text{new}}(x_k) \in C_2$. The quantity q is the new inte-

grand of the partition function (27) divided by the present one. It is easiest to test a trial move for one grid point at a time, although other methods will be discussed. The acceptance probability defined by Eq. (8) is calculated by

$$A(C_2|C_1) = \min(1, \exp[-\beta(\mathcal{V}_{\text{new}} - \mathcal{V}_{\text{present}})]). \quad (30)$$

A change to the vector $\vec{\phi}$ at lattice point k modifies the potential energy on the grid point k and its neighbors only (using a linear approximation for the derivatives). This is all the information needed to apply the Metropolis method. The quantities of interest to measure are the potential energy E and the topological charge B , which should be conserved and is a check on the numerics.

A uniform sampling of the distribution has an extremely low acceptance rate; too many vectors are rejected. We therefore use a biased sampling technique where a new vector $\vec{\phi}_{\text{new}}$ is sampled near $\vec{\phi}_{\text{present}}$, the vector that it is supposed to replace. We sample vectors in an intrinsic frame where the z axis corresponds to the present vector. The vector

$$\vec{n}^{\text{int}} = (n_1^{\text{int}}, n_2^{\text{int}}, n_3^{\text{int}}) = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta) \quad (31)$$

gives the components of the new vector in Cartesian coordinates in the intrinsic frame. The Euler angles that define the rotation from the intrinsic frame to the laboratory frame are given by

$$\vec{k}^{\text{lab}} = (\cos \beta \sin \alpha, \sin \beta \sin \alpha, \cos \alpha). \quad (32)$$

The angles α and β are used to rotate the new vector \vec{n}^{int} to the laboratory frame,

$$\vec{n}^{\text{lab}} = \begin{pmatrix} n_1^{\text{lab}} \\ n_2^{\text{lab}} \\ n_3^{\text{lab}} \end{pmatrix} = \begin{pmatrix} \cos \alpha \cos \beta & -\sin \beta & \sin \alpha \cos \beta \\ \cos \alpha \sin \beta & \cos \beta & \sin \alpha \sin \beta \\ -\sin \alpha & 0 & \cos \alpha \end{pmatrix} \begin{pmatrix} n_1^{\text{int}} \\ n_2^{\text{int}} \\ n_3^{\text{int}} \end{pmatrix}. \quad (33)$$

In the previous terminology, the new unit field vector at the grid point k is $\vec{\phi}_{\text{new}}(x_k) = \vec{n}^{\text{lab}}$, and it is a vector selected from a particular probability distribution function that is rotationally symmetric about the present vector $\vec{\phi}_{\text{present}}(x_k)$ ($= \vec{k}^{\text{lab}}$). The vectors $\vec{\phi}_{\text{new}}(x_k)$ and $\vec{\phi}_{\text{present}}(x_k)$ are inserted into the acceptance probability. The angle θ is sampled uniformly on $[0, A)$, where $A \leq \pi$, and ϕ is sampled uniformly on $[0, 2\pi)$. The corresponding transition probability is

$$T(C_2|C_1) = \begin{cases} \frac{1}{2\pi A} & \text{if } 1 - \cos \theta < A \\ 0 & \text{otherwise.} \end{cases} \quad (34)$$

No importance sampling is imposed and the probability distribution function is uniform, so that the acceptance rate (30) can be applied directly. The optimal value of A , which we

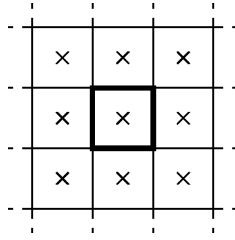


FIG. 4. A picture of the plaquette, where the fields are evaluated at the intersections of the lines and the measured quantities calculated at the midpoints \times .

are free to choose, depends on the particular configuration, especially on the topological charge and the temperature β . No A dependency was discovered in any ensemble averages other than the acceptance rate. Therefore, we believe that this method is reliable and efficient. We allow A to vary automatically to achieve an acceptance rate near 40%. The choice of A influences the rate at which equilibrium is reached, which is defined as the absence of change in the average energy over a large number of steps. At each temperature, the system was required to reach equilibrium before being cooled further. This ensures that cooling does not occur too quickly.

The choice (34) samples only a portion rather than the whole of the unit sphere. This is a valid method because we are modeling a continuous system, and therefore the vectors can reach any region in a number of steps. As A is varied automatically, the whole unit sphere is sampled for high temperatures. If the region that is unsampled for low temperatures were sampled, the vectors selected there would have virtually zero probability of being accepted. For generality, we discuss a more rigorous method using importance sampling in the Appendix. There, new vectors are chosen from a Gaussian (or other) distribution centered around the present vector. Importance sampling allows vectors from all over the unit sphere to be selected at any temperature, and this might be necessary for some systems, especially when calculating thermal averages. The disadvantage of importance sampling over restricting the transition probability is the increased amount of computing time.

2. Calculation of field derivatives for a field on S^2

We calculate derivatives in a similar way to the 1D implementation, where measurable quantities are calculated in the center of the plaquettes. An illustration of a plaquette is given in Fig. 4, where the field vectors are evaluated at the intersections of the lines and all measurable quantities are calculated at the midpoints. If a field vector is altered, using the Monte Carlo method as described above, the effect of a change has to be calculated on the four surrounding midpoints.

All field vectors at each of the four grid points lie on a unit sphere and the average of the four field vectors must also be of unit length, because the topology requires unitarity everywhere. A simple average fails this criterion unless all vectors point in identical directions. We still use the average, corrected by scaling it to unit length.

This also impacts the calculation of derivatives. For example, the error in taking the x derivatives is minimized if

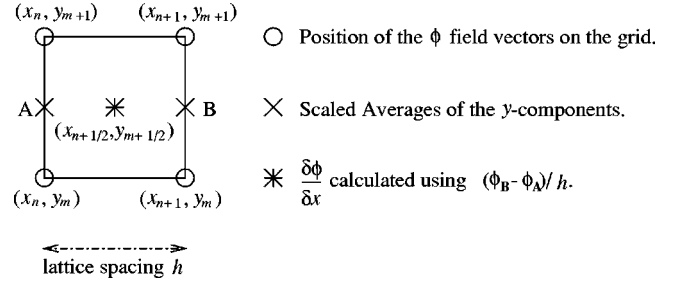


FIG. 5. Illustration of the scaling of the center derivative.

$\frac{1}{2} \sum_{b=\{b_1, b_2\}} \phi_b(x_{n+1})$ and $\frac{1}{2} \sum_{a=\{a_1, a_2\}} \phi_a(x_n)$ are scaled to unit length before the latter is subtracted from the former to obtain the x derivative. Here $\{a_1, a_2\}$ are the two grid points with the coordinates (x_n, y_m) and (x_n, y_{m+1}) , and $\{b_1, b_2\}$ are the two grid points with the coordinates (x_{n+1}, y_m) and (x_{n+1}, y_{m+1}) . These vectors are on the corners of the plaquette (see Fig. 5). Thus, the derivative is calculated by

$$\left. \frac{\partial \phi}{\partial x} \right|_{x_{n+1/2}} = \text{Scaled} \left(\frac{1}{2} \sum_{b=\{b_1, b_2\}} \phi_b(x_{n+1}) \right) - \text{Scaled} \left(\frac{1}{2} \sum_{a=\{a_1, a_2\}} \phi_a(x_n) \right). \quad (35)$$

This derivative works very well in practice. At very high temperatures the numerics may break down, because the derivative (35) is by definition an underestimate. The 3D analog of this formula is obtained by replacing $1/2$ by $1/4$ and summing over the four components of a and the four components of b .

3. Updating mechanisms

In our 1D simulations, we have randomly selected which grid point should be sampled. In our 2D and 3D implementation, we sweep over the grid points sequentially. The new vectors are stored and the changes to the field are updated only after a complete sweep over the entire grid to avoid unwanted sequential correlations. We have split the grid into four independent subgrids, each labeled by a different symbol in Fig. 6. The subgrids are chosen at random and, at each sweep, only one of the possible four composing vectors of the derivatives and field averages at the midpoints is changed. This avoids the creation of fluctuations between neighboring vectors for high acceptance rates, which produce an unphysical increase in energy. Unfortunately, the

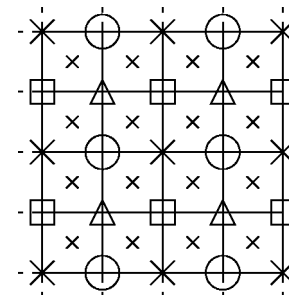


FIG. 6. The four subgrids for single point changes, each labeled by a symbol.

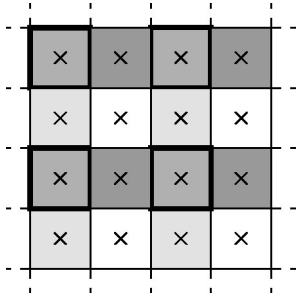


FIG. 7. The four subgrids for plaquette changes, each shaded differently.

change of single grid points at a time does not favor collective motion, where a localized energy distribution moves in one direction. Changing regions of grid points at a time has proved to be more efficient.

We have successfully used a plaquette updating mechanism. For a given plaquette, we sample a vector in the intrinsic frame [see Eq. (31)]. Then this frame is rotated as in Sec. V B for each of the four vectors on the plaquette separately. Figure 4 illustrates a plaquette surrounded by the nine affected midpoints. The total change in the energy density of all these midpoints is now calculated and all four vectors are accepted or they are all rejected. Again, we have split the grid into four subgrids, each shaded differently in Fig. 7. However, some midpoints are affected by two or four plaquettes from the same subgrid. Therefore, we choose the subgrids randomly rather than sequentially to avoid unwanted correlations.

C. Results and comparison

We have investigated the three different baby Skyrme models mentioned before. To that end, we need to minimize the energy functional

$$E[\vec{\phi}] = \int d^2x \left\{ \frac{1}{2} \partial_i \vec{\phi} \cdot \partial_i \vec{\phi} + \theta_S [(\partial_i \vec{\phi} \cdot \partial_i \vec{\phi})^2 - (\partial_i \vec{\phi} \cdot \partial_j \vec{\phi})(\partial_i \vec{\phi} \cdot \partial_j \vec{\phi})] + \theta_V V(\vec{\phi}) \right\}. \quad (36)$$

First, we looked at the simplest holomorphic model with the potential $V=(1+\phi_3)^4$. There exists an explicit one-soliton solution,

$$W = \sqrt[4]{\theta_V/2\theta_S}(x+iy), \quad (37)$$

where the W field is the stereographic projection of $\vec{\phi}$ on the complex plane, given by $W=2(\phi_1+i\phi_2)(1-\phi_3)^{-1}$. We choose $\theta_S=\theta_V=\frac{1}{2}$, where the total energy equals $4\pi(1+8/3\sqrt{2})\approx 36.2618$ Skyrme energy units. Since the soliton profile has a polynomial decay, we need a large lattice. With a 350×350 grid and lattice spacing $h=0.05$, we obtain $E=36.4890$ Skyrme energy units and $B=0.9999$. Here, the energy is slightly higher than the exact solution because of the finite lattice effects. The holomorphic baby Skyrmion has the slowest decay of any of the models discussed, and therefore can be seen as the worst case scenario.

We have also looked at the baby Skyrme models with one vacuum, where $V=1+\phi_3$, and two vacua, where $V=1-\phi_3^2$. The parameters for the one-vacuum model have been

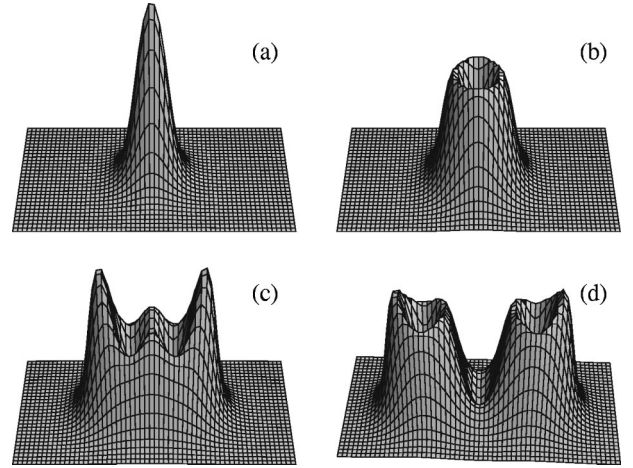


FIG. 8. Baryon density plot with identical vertical scales. We show the one-vacuum Skyrmions of charge 1 (a), 2 (b), 3 (c), and 4 (d). The mesh spacing is 0.4 Skyrme length units, and the total area plotted is 2.4×1.2 Skyrme length units squared in each figure.

fixed to $\theta_S=\frac{1}{4}$ and $\theta_V=0.1$ and for the two-vacua to $\theta_S=0.44365$ and $\theta_V=0.05$ in agreement with existing literature (see Ref. [14]). We use an 80×80 grid with periodic boundary conditions and lattice spacing $h=0.4$ Skyrme length units. The minimal energy solution in the first four topological sectors is shown in Figs. 8 and 9. We compare the energies per charge with the calculations from Ref. [14] in Table II.

The results for the two-vacua model are the same for both studies within an accuracy of a few parts in 10^4 . The results from Ref. [14] were obtained via the shooting method. We can apply this accurate method, because the Skyrmions are radially symmetric and the minimization reduces to a 1D problem. There is a slight discrepancy in energy when comparing the one-vacuum model results. The energies in Ref. [14] were obtained on a 2D lattice using a damped time evolution. The energy of the 1-Skyrmion calculated using the shooting method is $E=19.65$ Skyrme energy units. Our SA result agrees well with this value. The inaccuracies in Ref. [14] arise due to a different derivative approximation, which

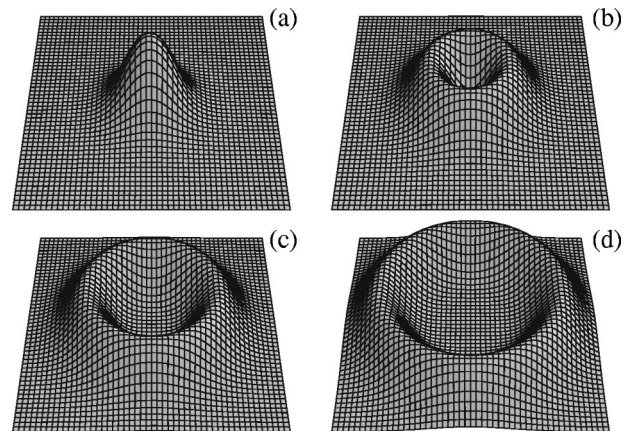


FIG. 9. Baryon density plot with identical vertical scales. We show the two-vacua Skyrmions of charge 1 (a), 2 (b), 3 (c), and 4 (d). The mesh spacing is 0.4 Skyrme length units, and the total area plotted is 2.0×1.0 Skyrme length units squared in each figure.

TABLE II. Baby Skyrme models: Comparison of our SA results with Euler-Lagrange results. The energy E is given in Skyrme energy units and the baryon number B is dimensionless.

Topological charge	One-vacuum model			Two-vacua model		
	B	SA E/B	Ref. [14] E/B	B	SA E/B	Ref. [14] E/B
1	0.999 78	19.6505	19.47	0.999 89	19.6572	19.65
2	1.999 73	18.4452	18.27	1.999 84	17.6530	17.65
3	2.999 62	18.5257	18.34	2.999 83	17.2259	17.22
4	3.999 52	18.4014	18.22	3.999 89	17.0677	17.07

tends to reduce the energy. The errors due to the fixed boundaries are also greater in Ref. [14], although this tends to increase the energy of the 1-Skyrmion. We believe that periodic boundary conditions, which have been used for our SA result, have the advantage that the tails of the Skyrmions can spread out further. The finite lattice effects are still present as Skyrmions could then interact with themselves over the boundaries. For our SA model, a study [10] on the 1-Skyrmion case shows that the grid size used induces an error of the order 0.01%. The error due to not having relaxed the system properly is 0.01%. The largest error is due to finite difference effects and has a possible size of 0.1%. Because of the successful agreement of our SA result for the 1-Skyrmion with the result of the shooting method, it is believed that the SA solutions with higher topological charge are also more accurate than the results quoted in Ref. [14]. Finally, we show an example of the cooling schedule for its 3-Skyrmion in Fig. 10.

We studied three different baby Skyrme models. Changing from one potential to the other could not have been easier. In the case of the iterative techniques, changing the differential equation is in itself conceptually easy, but, in practice, a lot of time is spent on getting the coefficients right and checking the derivation. For future research, we intend to use SA to do a systematic check on the multi-Skyrmion

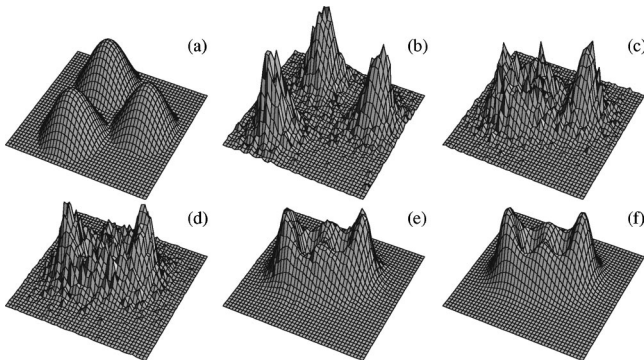


FIG. 10. Baryon density plot of identical vertical scales. We show a SA cooling for the 3-skyrmion in the one-vacuum model: (a) the starting configuration; (b) system is heated to $\beta=500$ and Skyrmions repel each other because their isospins are initially in the same direction; (c) isospins rotate relative to each other and Skyrmions attract each other; (d) equilibrium has been reached for $\beta=500$; (e) system is cooled to $\beta=5000$; (f) minimal energy solution at $\beta=\infty$. The mesh spacing is 0.4 Skyrme length units, and the total area plotted is 1.8×1.6 Skyrme length units squared in each figure.

structure. Specifically, we are interested in the class of potentials that leads to radially symmetric multi-Skyrmions.

VI. SIMULATED ANNEALING IN THREE DIMENSIONS

We have chosen the nuclear Skyrme model [7] for our 3D SA implementation, because we believe that SA is a flexible tool for exploring the multi-Skyrmion structure.

In the 1960s, Skyrme constructed an effective field theory of mesons where the baryons are the topological solitons of the theory. Research by 't Hooft and Witten has established that the nuclear Skyrme model shows important similarities to the low-energy effective Lagrangian of QCD [29,30]. The 1-Skyrmion can be interpreted as a nucleon with reasonable success [31]. The numerical work by Braaten, Townend, and Carson [32] and Battye and Sutcliffe [33] on the structure of classical multi-Skyrmions supports the idea that an appropriate quantization of these minimal energy solutions for a given topological sector could possibly lead to an effective description of atomic nuclei. However, the calculation of quantum properties of multi-Skyrmions is very difficult. Partly this is due to the fact that these minimal energy solutions are not radially symmetric and the theory is nonrenormalizable. This is rather frustrating, for the claim that the Skyrme model, descending from a large- N QCD approximation, models mesons, baryons, and higher nuclei is a very attractive one. Numerical methods are probably the only way forward and the SA scheme might be useful in exploring further the multi-Skyrmion structure for different versions of the Skyrme model.

A. The nuclear Skyrme model

The nuclear Skyrme Lagrangian

$$\mathcal{L} = \frac{1}{2} \partial_\mu \vec{\phi} \cdot \partial^\mu \vec{\phi} - \frac{1}{4} [(\partial_\mu \vec{\phi} \cdot \partial^\mu \vec{\phi})^2 - (\partial_\mu \vec{\phi} \cdot \partial_\nu \vec{\phi})(\partial^\mu \vec{\phi} \cdot \partial^\nu \vec{\phi})] \quad (38)$$

is a straightforward extension of the nonlinear σ model containing an additional fourth order term called the Skyrme term. We need to include this extra term to ensure stability of the soliton. The mapping becomes

$$\phi(t): S^3 \rightarrow S^3. \quad (39)$$

More realistic Lagrangians should probably include higher-order correction terms. The SA scheme is especially useful, because the extra term can be included trivially.

B. Implementation

The 3D implementation is very similar to the 2D code and we will only mention new issues relevant to the 3D case. On a 4D unit sphere, the integration measure is given by

$$\begin{aligned}
& \int_{\text{unit sphere}} dx dy dz dw \\
&= \int_{\theta=0}^{2\pi} \int_{\phi=0}^{\pi} \int_{\chi=0}^{\pi} \sin \theta \sin^2 \chi d\phi d\theta d\chi \\
&= \int_{\theta=0}^{2\pi} \int_{\cos \theta=-1}^1 \int_{\chi/2-(1/4)\sin(2\chi)=0}^{\pi/2} d\phi d \cos \theta \\
&\quad \times d\left(\frac{\chi}{2} - \frac{1}{4} \sin(2\chi)\right) \\
&= \int_{\theta=0}^{2\pi} \int_{u=-1}^1 \int_{v=0}^{\pi/2} d\theta du dv. \tag{40}
\end{aligned}$$

In order to rotate the new vector from the intrinsic frame to the laboratory frame, the angle χ must be evaluated from v . The equation

$$v = \frac{\chi}{2} - \frac{1}{4} \sin(2\chi) \tag{41}$$

cannot be rewritten in terms of χ , and therefore it must be solved numerically. The observation that most solutions will be in the region of small v due to the importance sampling implies that a small- χ approximation might be useful. For

small χ , $v \approx \chi^3/3$. Therefore, the inversion is performed numerically by tabulating χ uniformly on $[0, \pi]$ against $v^{1/3} = [\chi/2 - \frac{1}{4} \sin(2\chi)]^{1/3}$. The χ value corresponding to the selected $v^{1/3}$ is found in this table, and a linear interpolation is applied between the two nearest values of $v^{1/3}$ to give a better approximation to χ . Using 1000 precalculated entries the error on the whole region $\chi \in [0, \pi]$ is then less than 10^{-7} . On the region $\chi \in [0, 0.1]$, where almost the entire selection of χ lies, the error is less than 10^{-14} , which corresponds to the precision of double real numbers. This method does not create any significant errors and is time consuming. No faster method seems to be possible.

As the cosines and sines of ϕ , θ , and χ are known, the rotation can be performed using a rotation matrix similar to Eq. (33). The new choice of vector in the intrinsic frame is given by

$$\begin{aligned}
\vec{n}^{\text{int}} &= (n_1^{\text{int}}, n_2^{\text{int}}, n_3^{\text{int}}, n_4^{\text{int}}) \\
&= (\sin \chi \sin \theta \cos \phi, \sin \chi \sin \theta \sin \phi, \sin \chi \cos \theta, \cos \chi). \tag{42}
\end{aligned}$$

The z axis in the intrinsic frame coincides with $\vec{k}^{\text{lab}} = \vec{\phi}_{\text{present}}(x_k)$ in the laboratory frame, as in the baby Skyrme model. The rotation angles γ , α , and β (in that order) are defined by

$$\vec{k}^{\text{lab}} = (\cos \beta \sin \alpha \sin \gamma, \sin \beta \sin \alpha \sin \gamma, \cos \alpha \sin \gamma, \cos \gamma). \tag{43}$$

The transformation \vec{n}^{int} to \vec{n}^{lab} is performed using the matrix

$$\begin{pmatrix} n_1^{\text{lab}} \\ n_2^{\text{lab}} \\ n_3^{\text{lab}} \\ n_4^{\text{lab}} \end{pmatrix} = \begin{pmatrix} \cos \alpha \cos \beta & -\sin \beta & \sin \alpha \cos \beta \cos \gamma & \sin \alpha \cos \beta \sin \gamma \\ \cos \alpha \sin \beta & \cos \beta & \sin \alpha \sin \beta \cos \gamma & \sin \alpha \sin \beta \sin \gamma \\ -\sin \alpha & 0 & \cos \alpha \cos \gamma & \cos \alpha \sin \gamma \\ 0 & 0 & -\sin \gamma & \cos \gamma \end{pmatrix} \begin{pmatrix} n_1^{\text{int}} \\ n_2^{\text{int}} \\ n_3^{\text{int}} \\ n_4^{\text{int}} \end{pmatrix}. \tag{44}$$

For similar reasons to those already discussed previously, we use a restricted transition probability to be uniform over $[0, A)$, where $A \leq \pi$. The method of importance sampling has also been investigated for the nuclear Skyrme model and is given in the Appendix. The transition probability is therefore

$$T(Y|X) = \begin{cases} \frac{1}{4\pi A} & \text{if } \frac{\chi}{2} - \frac{1}{4} \sin(2\chi) < A \\ 0 & \text{otherwise.} \end{cases} \tag{45}$$

The angles are sampled by

$$v = \frac{\chi}{2} - \frac{1}{4} \sin(2\chi) = A \xi_1,$$

$$u = \cos \theta = 2\xi_2 - 1,$$

$$\phi = 2\pi \xi_3, \tag{46}$$

where ξ_1 , ξ_2 , and ξ_3 are three random variables uniformly sampled from 0 to 1. The new vector $\vec{\phi}_n(x_k) = \vec{n}^{\text{lab}}$ is a vector that has been selected from a uniform probability distribution centered around the previous vector $\vec{\phi}_p(x_k) (= \vec{k}^{\text{lab}})$ where only the angle χ between these vectors has an upper limit. Finally, $\vec{\phi}_n(x_k)$ and $\vec{\phi}_p(x_k)$ are inserted into Eq. (30) to find the acceptance rate. As in the baby Skyrme models, the value of A is automatically chosen to have an acceptance rate near 40%. The cooling is also controlled in the same manner as described in 2D.

C. Results and comparison

The 3D implementation of SA is computationally much more intensive than the 2D case. The accuracy of our numerical simulations is therefore reduced due to limited re-

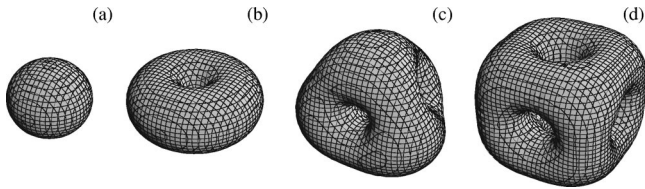


FIG. 11. Plot of the same constant baryon density surface. We show the first four multi-Skyrmion solutions (from left to right). All plots are to the same scale. The mesh spacing of the plotted objects is 0.12 Skyrme length units.

sources in computation time and memory. The intergrid spacing used is 0.12 Skyrme length units and is close to the upper limit where the numerics break down (at the reasonably high temperature required to do SA sufficiently fast). The maximum grid size that can be used to obtain results in a reasonable time is $80 \times 80 \times 80$. The finite volume causes an increase of energy for the 1-Skyrmion because it repels itself over the periodic boundary, and therefore induces an error of 1% [10]. The error due to not having relaxed the system properly is 0.1%. The error due to finite difference effects has a maximum of 0.3%.

The skyrmions of topological charge 1–4 are shown in Fig. 11. We show an example of the cooling schedule for the 4-Skyrmion in Fig. 12. These energies per charge are contrasted in Table III with those results obtained by Batty and Sutcliffe [33]. It is very difficult to compare the results. The 1-Skyrmion solution gives more information for comparison. It is spherically symmetric and the shooting method in the hedgehog *Ansatz* can be used. The energy of a 1-Skyrmion minimized in the 1D SA code is $E=73.12$ Skyrme energy units (in the continuous limit). We are not sure if the result for the 1-Skyrmion in Ref. [33] is truly more accurate or just a coincidence. Their topological charge, an indicator for the discretization error, is certainly

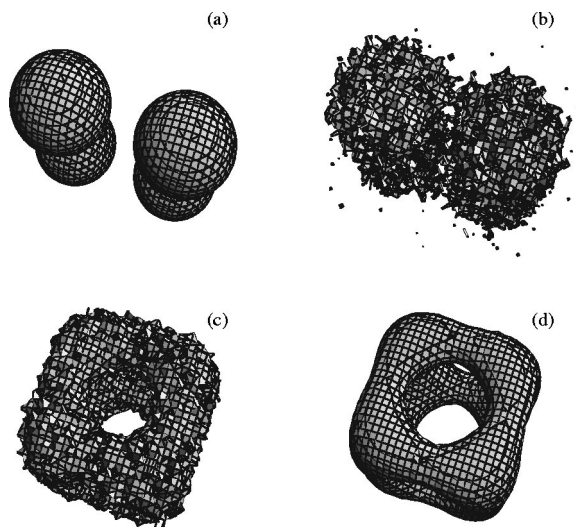


FIG. 12. Plot of constant baryon density surface. We show a SA search of the $B=4$ Skyrmion: (a) the starting configuration of four 1-Skyrmions; (b) the system heated to $\beta=500$ where the Skyrmions fuse into one; (c) the system in equilibrium at $\beta=500$ where the structure emerges; (d) the minimal energy solution at $\beta=\infty$. All plots are to the same scale. The mesh spacing of the plotted objects is 0.1 Skyrme length units.

TABLE III. The nuclear Skyrme model: SA versus published results. The energy E is given in Skyrme energy units and the baryon number B is dimensionless.

Topological charge	SA		Ref. [33]	
	B	E/B	B	E/B
1	1.0015	73.75	0.984	72.96
2	2.0030	70.31	1.972	69.34
3	3.0042	68.52	2.960	67.69
4	4.0048	66.30	3.948	66.09

less accurate. The finite lattice effect increases the energy of the 1-Skyrmion and therefore the energy obtained should be larger than 73.12 Skyrme energy units. Unfortunately, we do not know which lattice parameters they have used, making a good comparison impossible. However, we get the same minimal energy structure.

VII. CONCLUSION

We have shown that SA is an alternative way of finding the minimal energy solution in a given topological charge sector. We independently confirmed the validity of the studies using the standard minimization techniques. It is very hard to objectively compare the different approaches. However, we have found SA to be a more convenient and flexible minimization technique. The implementation and fine-tuning of our SA codes took a fair amount of time due to a lack of prior research in this area. In comparison to other methods, we are confident that future implementations will take us considerably less time. We did not find any significant differences in speed of minimization. The SA codes can be made faster by fine-tuning the cooling parameters. We prefer SA minimization because of its ease of use. Speed considerations are irrelevant in 1D and 2D and we can use parallel computing in the 3D case.

There are several areas we want to look at next. First of all, we will optimize SA by using more sophisticated update and cooling mechanisms and by parallelization. We are also currently investigating the possibility of doing time evolution via SA minimization of the action. At the same time, we intend to look at a wide range of models. We shall investigate the multi-Skyrmion structure of several baby Skyrme models. Research is also underway in the use of symmetry breaking terms for the nuclear Skyrme model. Moreover, the 2D and 3D codes will be used to study phase transitions in the baby and nuclear Skyrme models at finite temperatures and densities [25]. To conclude, SA is a flexible tool; all we really need is an energy functional to minimize.

ACKNOWLEDGMENTS

First of all, the authors would like to thank Niels Walet for his support and advice. We would also like to acknowledge discussions with Klaus Gernoth, Laur Järvi, and Bernard Piette. The authors acknowledge PPARC (M.H.) and EPSRC (O.S. and T.W.) for research grants.

APPENDIX: IMPORTANCE SAMPLING

For importance sampling, the integral (4) is rewritten as

$$\langle \mathcal{F} \rangle = \int \left(\frac{\mathcal{F}(C)P(C)}{\tilde{P}(C)} \right) \tilde{P}(C) dC. \quad (\text{A1})$$

Here $P(C)$ can be normalized without loss of generalization (see Ref. [16]) and $\tilde{P}(C)$ is a different probability density function that satisfies

$$\tilde{P}(C) \geq 0, \quad \int \tilde{P}(C) dC = 1, \quad (\text{A2})$$

and

$$\frac{\mathcal{F}(C)P(C)}{\tilde{P}(C)} < \infty \quad (\text{A3})$$

except on a countable set of points. In this method one chooses a $\tilde{P}(C)$ that minimizes the variance, which is

$$\text{var}\{\langle \mathcal{F} \rangle\} = \int \frac{\mathcal{F}^2(C)P^2(C)}{\tilde{P}^2(C)} dC - \langle \mathcal{F} \rangle^2. \quad (\text{A4})$$

The measurement of statistical accuracy is given by $\text{var}\{\langle \mathcal{F} \rangle\}$. More samples reduce the variance. Alternatively, the same variance using fewer samples can be achieved with importance sampling. In practice, the closer $\tilde{P}(C)$ is to $\mathcal{F}(C)P(C)$, the smaller the variance becomes. It is known that if

$$\tilde{P}(C) = \frac{\mathcal{F}(C)P(C)}{\langle \mathcal{F} \rangle}, \quad (\text{A5})$$

then the integral is equal to \mathcal{F} with zero variance. However, we need to respect the constraints (A2), and, even worse, choosing a \tilde{P} requires knowledge of $\langle \mathcal{F} \rangle$ prior to evaluating the integral.

1. Baby Skyrme models

The application of importance sampling to the partition function (26) is complicated, because the range of integration is on a sphere of unit length. Therefore, we need to use a $\tilde{P}(\vec{\phi}(x_k))$ that is nonzero only on the unit sphere. Further, the maximum or most likely area of accepted values depends on the present vector, and therefore $\tilde{P}(\vec{\phi}_n(x_k))$ should not be restricted to a certain region of the sphere, but should depend on the present vector $\vec{\phi}_p(x_k)$. Looking at the results of a uniform probability distribution function on the sphere, a Gaussian distribution of the polar angle θ seems to be a good choice for \tilde{P} , where $\theta=0$ is in the direction of the present vector. This distribution is then rotated around the azimuths ϕ axis and therefore the ϕ distribution is uniform. The Gaussian-distributed θ and the uniformly distributed $\vec{\phi}$ define the probability for the new vector. This vector is inserted into $\mathcal{F}(C)$ as before, and the Metropolis algorithm accepts or rejects this particular choice. The quantity q is given by

$$q(C_2|C_1) = \frac{P(C_2)\tilde{P}(C_1)}{P(C_1)\tilde{P}(C_2)}. \quad (\text{A6})$$

To implement this method, a Gaussian is chosen centered along the z axis in the intrinsic frame. This Gaussian is, in terms of the z coordinate,

$$\tilde{f} = \exp[-A(1 - \cos \theta)^2] = \exp[-A(1 - z)^2]. \quad (\text{A7})$$

The integral

$$\int_{-1}^1 N \exp[-A(1 - z)^2] dz \quad (\text{A8})$$

satisfies Eq. (A2), where N is a normalization constant and A is an arbitrary parameter that changes the breadth of the Gaussian and thereby alters the acceptance rate. This is sampled using the Box-Müller method, by choosing

$$z = 1 - \frac{1}{\sqrt{A}} \sqrt{-\ln \xi_1} |\cos(2\pi\xi_2)|, \quad (\text{A9})$$

which is a shifted Gaussian so that the peak is at $z = \cos \theta = 1$. All values of $z < -1$ are rejected. The azimuth angle ϕ is sampled uniformly along $[0, 2\pi)$ by $\phi = 2\pi\xi_3$. The acceptance probability becomes

$$A(C_2|C_1) = \min(1, \exp[-\beta(\nu_n - \nu_p) + A(1 - z)^2]), \quad (\text{A10})$$

using Eqs. (A6) and (A7).

2. Nuclear Skyrme model

We do importance sampling by prioritizing small v values and selecting u and ϕ with uniform probability. The small- v region corresponds to the small- χ region. Importance sampling is used because the newly selected four-dimensional unit vector should be in the neighborhood of the previous vector. This new vector is selected in the intrinsic frame, as discussed in Sec. VB, where the previous vector is pointing in the $\chi=0$ direction. A good probability distribution is

$$\tilde{f} = e^{-Av}. \quad (\text{A11})$$

The quantity v is therefore selected using

$$v = -\frac{1}{A} \ln \left\{ \left[\exp\left(-\frac{\pi}{2}A\right) - 1 \right] \xi_1 + 1 \right\}, \quad (\text{A12})$$

where ξ is a uniform random variable on $(0,1)$. The other angles are sampled as in Eq. (46). The acceptance probability now becomes

$$A(C_2|C_1) = \min(1, \exp[-\beta(\mathcal{V}_n - \mathcal{V}_p) + Av]). \quad (\text{A13})$$

A similar probability distribution function can be used for the baby Skyrme model and is faster than the given Gaussian. In practice, importance sampling is not used, because it is computationally more time consuming than restricting the transition probability.

- [1] J. S. Russell, *Report of the 14th Meeting of the BAAS, York* (John Murray, London, 1844), p. 311.
- [2] P. G. Drazin and R. S. Johnson, *Solitons: An Introduction* (Cambridge University Press, Cambridge, 1996).
- [3] L. H. Ryder, *Quantum Field Theory* (Cambridge University Press, Cambridge, 1994), Chap. 10.
- [4] L. Faddeev and A. J. Niemi, *Nature* (London) **387**, 58 (1997).
- [5] R. A. Battye and P. M. Sutcliffe, *Phys. Rev. Lett.* **81**, 4798 (1998).
- [6] S. L. Sondhi, A. Karlhede, S. A. Kivelson, and E. H. Rezayi, *Phys. Rev. B* **47**, 16 419 (1993).
- [7] G. Holzwarth and B. Schwesinger, *Rep. Prog. Phys.* **49**, 825 (1986).
- [8] A. Vilenkin and E. P. S. Shellard, *Cosmic Strings and Other Topological Defects* (Cambridge University Press, Cambridge, 1994).
- [9] M. Duff, e-print hep-th/9805177.
- [10] O. Schwindt, UMIST report (unpublished).
- [11] T. Weidig, Ph.D. thesis, University of Durham, 1999.
- [12] L. Ingber, *J. Math. Comput. Model.* **18**, 29 (1993).
- [13] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C* (Cambridge University Press, Cambridge, 1992).
- [14] T. Weidig, *Nonlinearity* **12**, 1489 (1999).
- [15] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- [16] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods* (Wiley, New York, 1986).
- [17] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science* **220**, 671 (1983).
- [18] S. Geman and D. Geman, *IEEE Trans. Pattern Anal. Mach. Intell.* **26**, 721 (1984).
- [19] R. Rajarman, *Solitons and Instantons* (North-Holland, Amsterdam, 1996).
- [20] S. Coleman, *Phys. Rev. D* **11**, 2088 (1975).
- [21] D. Vanderbilt and S. G. Louie, *J. Comput. Phys.* **56**, 259 (1984).
- [22] G. H. Derrick, *J. Math. Phys.* **5**, 1252 (1964).
- [23] B. M. A. G. Piette and W. J. Zakrzewski, *Chaos, Solitons and Fractals* **5**, 2495 (1995).
- [24] B. M. A. G. Piette, B. J. Schoers, and W. J. Zakrzewski, *Z. Phys. C* **65**, 165 (1995).
- [25] O. Schwindt and N. Walet (unpublished).
- [26] F. Mandel, *Statistical Physics* (Wiley, New York, 1988).
- [27] K. Binder, *Monte Carlo Methods* (Springer-Verlag, Berlin, 1979).
- [28] D. Frenkel and B. Smith, *Understanding Molecular Simulation* (Academic, London, 1996).
- [29] G. 't Hooft, *Nucl. Phys. B* **72**, 461 (1973).
- [30] E. Witten, *Nucl. Phys. B* **160**, 55 (1979).
- [31] G. Adkins, C. Nappi, and E. Witten, *Nucl. Phys. B* **228**, 552 (1983).
- [32] E. Braaten, S. Townend, and L. Carson, *Phys. Lett. B* **235**, 147 (1990).
- [33] R. A. Battye and P. M. Sutcliffe, *Phys. Rev. Lett.* **79**, 363 (1997).