

Methods of Systematic Upscaling

Achi Brandt

Department of Computer Science and Applied Mathematics
The Weizmann Institute of Science
Rehovot 76100, Israel

Abstract

Systematic Upscaling (SU) is a new multiscale computational methodology for the accurate derivation of equations (or statistical relations) that govern a given physical system at increasingly larger scales. Starting at a fine (e.g., atomistic) scale where first-principle laws (e.g., differential equations) are known, SU advances, scale after scale, to obtain suitable variables and operational rules for simulating the system at any large scale of interest. SU combines the complementary advantages of two multilevel computational paradigms that have emerged over the last 35 years: multigrid in applied mathematics and renormalization group in theoretical physics. It includes systematic procedures to iterate back and forth between all the scales of the physical problem, with a general criterion for choosing appropriate variables that operate at each level, and general techniques to derive their governing relations. Indefinitely large systems can in this way be simulated, with computation at each level being needed only within limited representative windows. No scale separation is assumed; unlike conventional ad-hoc multiscale modeling, SU is in principle quite generally applicable, free of slowdowns and bears fully-controlled accuracy.

Fields that can be greatly impacted by SU range from elementary particle physics and quantum chemistry to molecular and macromolecular dynamics, material science, nano-technology, bio-technology, and others. Cutting across such diverse fields, the present article does not focus on any specific application but on the generic principles of upscaling various *types* of systems, such as: problems defined on grids on one hand, and moving particles on the other hand; from ensembles of single-atom molecules to macromolecules in solution; local interactions as well as long-range ones; dynamical systems, both deterministic and stochastic; equilibrium calculations, including special procedures for low temperatures; energy minimization, particularly for functionals afflicted with multiscale nested attraction basins; etc. A suite of related upscaling techniques connects all these cases into one unified body of study.

1 Introduction

Despite their dizzying speed, modern supercomputers are still incapable of handling many most vital scientific problems. This is primarily due to *the scale gap*, which exists between the microscopic scale at which physical laws are given and the much larger scale of phenomena we wish to understand.

This gap implies, first of all, a huge number of *variables* (e.g., atoms or gridpoints or pixels), and possibly even a much larger number of *interactions* (e.g., one force between every pair of atoms). Moreover, computers simulate physical systems by moving few variables at a time; each such move must be very small, since a larger move would have to take into account all the motions that should in parallel be performed by all other variables. Such a computer simulation is therefore *extremely slow* in changing large-scale features of the system, and is particularly incapable of moving the system across large-scale *energy barriers*, which can each be crossed only by a large, *coherent* motion of very many variables.

This type of computational obstacles makes it impossible, for example, to calculate the properties of nature's building blocks (elementary particles, atomic nuclei, etc.), or to computerize chemistry and materials science, so as to enable the design of materials, drugs and processes, with enormous potential benefits for medicine, biotechnology, nanotechnology, agriculture, materials science, industrial processing, and so on. With current common methods the amount of computer processing often increases so steeply with problem size, that even much faster computers will not do.

Past studies have demonstrated that scale-born slownesses can often be overcome by multiscale algorithms. Such algorithms have first been developed in the form of fast *multigrid solvers* for discretized PDEs [2], [3], [5], [6], [27], [22], [38]. These solvers are based on two processes: (1) classical *relaxation* schemes, which are generally slow to converge but fast to *smooth* the error function; (2) approximating the smooth error on a *coarser grid* (typically having twice the meshsize), by solving there equations which are derived from the PDE and from the fine-grid residuals; the solution of these coarse-grid equations is obtained by using recursively the same two processes. As a result, large scale changes are effectively calculated on correspondingly coarse grids, based on information gathered from finer grids. Such multigrid solvers yield *linear complexity*, i.e., the solution cost is just proportional to the number of variables in the system.

In many years of research, the range of applicability of these methods has steadily grown, to cover most major types of linear and nonlinear large systems of equations appearing in sciences and engineering. This has been accomplished by extending the concept of “smoothness” in various ways, finally to stand generally for any poorly locally determined solution component, and by correspondingly diversifying the types of coarse representations, to include for instance grid-free solvers (algebraic multigrid [12], [13], [7], [32]), non-deterministic problems ([14], [29], [24], [9], [15]) and multiple coarse-level representations [16].

It has been shown (see the general survey [11]) that the inter-scale interactions can indeed eliminate all kinds of scale-associated difficulties, such as:

slow convergence (in minimization processes, PDE solvers, etc.); critical slowing down (in statistical physics); ill-posedness (e.g., of inverse problems); conflicts between small-scale and large-scale representations (e.g., in wave problems, bridging the gap between wave equations and geometrical optics); numerousness of long-range interactions (in many body problems or integral equations); the need to produce many fine-level solutions (e.g., in optimal control, design and data assimilation problems), many eigenfunctions (in electronic structure calculation, for example) or very many fine-level independent samples (in statistical physics); and more. Since the local processing (relaxation, for example) in each scale can be done in parallel at all parts of the domain, the multiscale algorithms, based on such processing, proved ideal for implementation on massively parallel computers.

To obtain even further generality, there emerge however two basic reasons to go beyond these multigrid methods. First, they cannot perform well for *highly nonlinear cases*, where configurations cannot be decomposed into weakly-interacting local and non-local parts. Second, for many systems even the linear complexity is not good enough, since the number of variables is huge. Such systems on the other hand are typically *highly repetitive*, in the sense that the same small set of governing equations (or Hamiltonian terms) keep repeating itself throughout the physical domain. This opens the way to the possibility of having, at the coarse level too, a small set of governing equations that are valid everywhere, and that can be derived from fine-level processing conducted only in some small representative “windows” (as defined below).

These two basic reasons point in fact in the same direction. Instead of relaxing the given system of equations so as to obtain a smooth error that can be approximated on a coarse level, one should use coarse level variables that are little sensitive to relaxation (e.g., representing chosen *averages*, rather than a subset of individual fine-level values) and that represent the *full* solution rather than the correction to any given current approximation. Such coarse variables can be chosen (as described below) so that the coarse-level equations can be derived just by local processing. We use the term “*upscaling*” for this type of direct (full-solution) transition from a fine level to a coarser one. Such a transition is valid even in those highly nonlinear cases where all scales interact with each other so strongly that correction-based multileveling is inapplicable.

In fact, upscaling, under the name “renormalization”, was first introduced into exactly such systems where all scales interact most strongly: the highly nonlinear systems of statistical mechanics at the critical temperature of phase transition. The *renormalization group* (RG) method (see, e.g., [39], [36], [37]) was developed contemporaneously with, but independently of the multigrid method, its chief purpose having been to investigate the behavior of such critical systems at the limit of very large scales. The RG method has thus focused on analyzing, theoretically and computationally, the fixed point of the group of successive renormalization steps, and various universal asymptotic power laws associated with it. Little has been done on the other hand to upscale systems without a fixed point, which is the prevalent situation in many practical problems. Also, the RG computational efficiency remained very limited, due to the lack of a

systematic coarse-to-fine transition, which is needed either for accelerating simulations at all levels (as in multigrid solvers) and/or for confining them to small representative windows (as described below).

This article reviews the principles and typical tools of a new computational methodology, called *Systematic Upscaling (SU)*. Building on the complementary advantages of multigrid and RG described above, SU is a methodical derivation of numerical equations (or statistical rules) that govern a given physical system at increasingly larger scales, starting at a microscopic scale where first-principle laws are known, and advancing, scale after scale, to obtain suitable variables and operational rules for processing the system at much larger scales. No scale separation is assumed; in fact, small ratio (typically 1:2 to 1:4) between successive scales is essential: it ensures slowdown-free computations that at each scale can be confined to certain representative “windows” (subdomains selected by the coarser scale, each containing relatively few variables).

The development of SU aims at providing necessary tools for surmounting extreme computational bottlenecks in many areas of science and engineering, such as: statistical mechanics, especially at phase transition; elementary-particle physics; electronic structure of molecular system (for deriving the inter-atomic force fields); molecular dynamics of fluids, condensed matter, nano structures and macromolecules, including proteins and nucleic acids; turbulent fluid dynamics; and global optimization of systems with multiple-scale energy barriers.

The SU approach should not be confused with various methods of multiscale *modelling* (MSM, also called “multiscale simulation”) being developed in several fields (e.g., materials science). Those methods study a physical system by employing several different ad-hoc models, each describing a very different scale of the system. Their basic approach is the *fine-to-coarse parameter passing*, in which data obtained from simulating a finer scale model, often coupled with experimental observations, are used to determine certain parameters of a larger scale model, regarding the latter as a coarse graining of the former. Except for the values of those parameters, the larger-scale model is assumed to be known, including its degrees of freedom and the general form of its equations — while in SU they are all systematically derived. In particular, a basic feature missing in MSM (as in RG) is the accurate transition from coarse levels back to finer ones, and the use of this transition as a systematic vehicle for choosing an adequate set of coarse variables and for supplying boundary values for the finer-levels windows. Successful as MSM methods are in various cases, they lack generality, are often inapplicable (requiring wide scale separation) or inaccurate (based on questionable large-scale models), and involve much slowdown due to large scale gaps. SU, by contrast, is built to inherit from multigrid and RG general and methodical procedures to construct and iteratively employ all intermediate scales, with slowness-free efficiency and fully-controlled coarse-level accuracy.

Unlike MSM and classical macroscopic physical models, the SU coarse-level operational rules are not closed-form equations; they are typically more like open-ended numerical tables, that can always be extended to achieve ever higher accuracy.

2 Example: Simple macromolecule

Various examples of systematic upscaling will be surveyed below. To give a very concrete idea, an example of particular relevance to biomolecules will be discussed in more details. This is a simple polymer model, whose systematic upscaling has been studied in [1], and described also in Sec. 14.6 in [11].

Simulation of long polymers (and generally all macromolecules) is one of the most computationally intensive tasks. This is due mostly to the large variation in time scales (10^{-15} seconds to several hours) and length scales ($1\text{\AA}$ – $1000\text{\AA}$) involved in each problem and the many energy barriers and attraction basins found at all scales. While much of the interesting behavior occurs at longer time (or length) scales, the shorter scales, where the basic equations are given, constrain the size of steps in simulations. By applying systematic upscaling these constraints can hopefully be removed, as different physical scales are resolved on corresponding computational levels.

The model. The simple polymer is a non-branching long chain of n repeat units of CH_2 called monomers; see the comprehensive survey [23]. We use the united-atom model of [30], where each monomer is considered as a single unit (“atom”) and details of its internal structure and interactions are ignored, so mathematically the polymer is represented as a chain of n points in $\mathbb{R}^3$, located at positions $u_j = (x_j, y_j, z_j)$. The internal coordinates employed to describe interactions are the distances $r_{i,j} = |u_i - u_j|$, the angles θ_i (angle between the lines $\overline{u_{i-1}u_i}$ and $\overline{u_i u_{i+1}}$) and the dihedral (or torsion) angles $\phi_{i+1/2}$ (angle between the planes $\overline{u_{i-1}u_i u_{i+1}}$ and $\overline{u_i u_{i+1} u_{i+2}}$). The overall Hamiltonian (energy) functional is

$$\begin{aligned}
 H(u) = & \sum_{i=1}^{n-1} K_r (r_{i,i+1} - r_0)^2 && \text{bond-length potentials} \\
 & + \sum_{i=2}^{n-1} K_\theta (\cos \theta_i - \cos \theta_0)^2 && \text{bond-angle potentials} \\
 & + \sum_{i=2}^{n-2} F_\phi(\phi_{i+1/2}) && \text{bond-dihedral (torsion) potentials} \\
 & + \sum_{|i-j| \geq 4} \epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right] && \text{Lennard-Jones potentials}
 \end{aligned} \tag{2.1}$$

where $r_0 = 1.52 \text{ \AA}$, $K_r = 250 \text{ Kcal/mol/\AA}^2$, $\theta_0 = 110^\circ$, $K_\theta = 60 \text{ Kcal/mol}$, $\epsilon = 0.09344 \text{ Kcal/mol}$, $\sigma = 4.5 \text{ \AA}$ and

$$F(\phi) = \frac{1}{2} [K_\phi^1(1 - \cos \phi) + K_\phi^2(1 - \cos 2\phi) + K_\phi^3(1 - \cos 3\phi)] ,$$

with $K_\phi^1 = 1.6 \text{ Kcal/mol}$, $K_\phi^2 = -0.867 \text{ Kcal/mol}$, $K_\phi^3 = 3.24 \text{ Kcal/mol}$. It is easy to see that F_ϕ is a function featuring three local minima with 1.5 to

2 Kcal/mol energy barriers between them. Also observe that the bond-length potentials are much stronger than the bond-angle potentials which in turn are much stronger than the torsion potentials. Therefore the latter are the main active degrees of freedom: large-scale changes in the shape (folding) of the polymer depend mainly on dihedral changes.

Equilibrium statistics. Before studying the polymer dynamics (i.e., its evolution in time), our first interest is its equilibrium statistics, i.e., its probable features after a sufficient long time, which are usually the most important features. We remark that, more generally, the *multiscale* study of the dynamics of a system should normally follow the development of a multiscale description of its equilibria, because, at any large time interval, fine-scale details of the system cannot be calculated deterministically; they are effectively in a local equilibrium slave to larger-scale degrees of freedom (see Sec. 5.2).

In the long run, according to the theory of statistical mechanics, the probability density $P(u)$ to find the polymer in a particular configuration $u = (u_1, u_2, \dots, u_n)$ is proportional to $\exp(-H(u)/k_B T)$, where T is the absolute temperature and k_B is the Boltzmann constant. Properties of interest are typically averages and variances of all kinds of *observables* $O(u)$, such as, for example, the end-to-end observable $O(u) = |u_1 - u_n|$. The average of any observable O is defined by

$$\langle O \rangle = \int P(u) O(u) du ,$$

and its variance is

$$\langle O^2 \rangle - \langle O \rangle^2 .$$

Monte-Carlo methods at equilibrium. To calculate equilibrium statistics, an atom-by-atom Monte-Carlo (MC) process is usually performed. In this process, each atom in its turn changes position stochastically, according to the probability density distribution $P(u)$. Making repeated sweeps of this process, one can calculate the desired statistics on the sequence of produced configurations.

To calculate accurate averages of some observable in any extensive system, however, an extremely long sequence of configurations should be generated. There are two general causes for this complexity: (1) Due to the local nature of the Monte-Carlo process, only very slowly it affects large-scale conformational features, hence extremely many Monte-Carlo sweeps are needed to produce each new, statistically independent configuration. (2) Many such independent samples are needed to average out the deviation observed at each of them.

In addition, in the case of macromolecules like our polymer, such MC simulations are particularly slow, because in sampling the main active degrees of freedom, the dihedrals, the simulation stepsizes are constrained by the stronger bond-length and angle potentials. More important, each MC switch of the local minimum around which one dihedral lives is only probable provided suitable (but unknown) similar switches are simultaneously performed at a substantial number of neighboring dihedrals.

For short chains ($n \leq 30$, say), very efficient Monte Carlo simulations can be performed by changing at each step one *internal* degree of freedom (such as $r_{i,i+1}$, θ_i or $\phi_{i+1/2}$), keeping all other internal degrees of freedom fixed, instead of changing the position of one atom at a time; indeed, each such move changes the positions of many atoms. Sampling the different attraction basins of each dihedral is easily accomplished this way. For longer chains, however, these moves are highly inefficient: Each dihedral move corresponds to a global folding of the chain that most probably at some place or another brings far atoms to be too close to each other, increasing substantially the Lennard-Jones potentials, hence having very low MC acceptance probability.

The challenge, then, is to devise a Monte-Carlo process that will calculate the equilibrium statistics of *long* polymers. As we will see, systematic upscaling not only dramatically reduce the number of degrees of freedom to be simulated, but also eliminate the slowness associated with the dihedral attraction basins.

3 Systematic Upscaling (SU): An outline

Local equations and interactions. Computationally we will always deal with a discrete system, whose n variables (or unknowns) $u_1, u_2, \dots, u_n$ will typically be either discrete values of functions (grid values, or finite elements, etc.), or locations of particles. An equation in the d -dimensional physical space (usually $1 \leq d \leq 4$) is called *local* if it involves only $O(1)$ neighboring unknowns. A discretized partial differential equation, for example, is a system of local equations. Similarly, an “*interaction*”, i.e., an additive term in an energy functional or Hamiltonian H , is called *local interaction* if it involves only $O(1)$ neighboring variables. (The Lennard Jones potentials in (2.1) can be considered local, since they become negligible for far particles — unlike electrostatic interactions, for example.) In equilibrium calculations we will assume below that H already includes the $(k_B T)^{-1}$ factor, so that the probability density $P(u)$ of each configuration $u = (u_1, u_2, \dots, u_n)$ is proportional to $\exp(-H(u))$.

For simplicity of discussion we describe SU first for deterministic systems of local equations (including energy minimization with local interactions) or stochastic systems of local interactions at equilibrium. We will point out later natural extensions to long-range equations or interactions, and to dynamic and non-equilibrium systems.

Coarsening. Similar to multigrid, SU is based on two processes: The usual *local processing* (relaxation in deterministic problems, Monte Carlo (MC) in stochastic ones) and repeated *coarsening*, creating increasingly coarser descriptions of the same physical system. At each coarsening stage, one constructs from a current level of description (the fine level) a coarser level, employing the following general principles.

To each fine-level configuration $u = (u_1, \dots, u_n)$ one defines (using the general coarsening criterion described below) a unique coarse-level configuration $u^c = (u_1^c, \dots, u_m^c)$, which is a vector with a reduced number of variables; typically $.1n < m < .6n$. Different fine level configurations u and $\hat{u}$ may well give

rise to the same coarse level configuration $u^c = \hat{u}^c$.

Examples of such fine-to-coarse transformations:

(i) For discretized continuous (e.g., PDE) problems — each coarse variable is an average of several neighboring fine variables.

(ii) For the above-described simple polymer, which consists of a chain of n atoms at the three-dimensional locations $(u_1, u_2, \dots, u_n)$ — each coarse-level “atom” location u_j^c is at the average location of q , say, consecutive real atoms:

$$u_j^c = \frac{1}{q}(u_{qj-q+1} + u_{qj-q+2} + \dots + u_{qj}) , \quad (j = 1, \dots, m; \quad m = \frac{n}{q}) . \quad (3.1)$$

(iii) For a simple atomistic fluid, described by the positions u in space of its n molecules — the coarse level variables are defined at the points of a lattice placed over the flow domain, with each variable u_j^c summarizing a property of the set of molecules around that lattice point (e.g., their total mass, or density, total dipole moment, etc.). Or each u_j^c may be a vector which summarizes several such properties. Indeed, at lower temperatures, as the fluid starts to solidify, additional types of coarse variables must enter (i.e., u_j^c must be a longer vector), accounting for an increasing number of larger-scale order parameters (see Sec. 5.6).

(iv) For a lattice of Ising spins — each coarse variable is again an Ising spin, standing for the *sign* of the sum of fine-level spins over a certain block (coarsening by “majority rule” [37]).

Interpolation. To any given coarse configuration $U = (U_1, \dots, U_m)$, there are generally many fine-level configurations u *compatible* with U (i.e., such that coarsening u gives $u^c = U$). The interpolation (transition from U to one specific fine configuration u) is created by *compatible Monte Carlo* (CMC) (or compatible relaxation, in the deterministic case), i.e., by the local processing, restricted to configurations compatible with U . The interpolation is completed once the CMC has practically reached its equilibrium (or the compatible relaxation has converged). (CMC was first introduced in [18].)

For instance, in the case of polymer (Example (ii) above), if the coarse variables are defined by (3.1), each step in a CMC would offer a simultaneous change of two consecutive atomic positions, u_k and u_{k+1} , such that $u_k + u_{k+1}$ is kept unchanged ($qj - q + 1 \leq k \leq qj - 1; 1 \leq j \leq m$).

The general coarsening criterion. *The fine-to-coarse transformation is said to be adequate if (and to the extent that) the compatible Monte Carlo equilibrates fast (or the compatible relaxation converges fast).*

In our polymer example, for instance, with the Hamiltonian (2.1), it experimentally turns out that the coarsening (3.1) is adequate for $q = 2$ or 3 , but a larger q yields much slower CMC equilibration. (The reason is that, at coarsening ratio $q \leq 3$, the fixed coarse values u_j^c implicitly nearly fix the dihedral values.)

Such a coarsening criterion, based on the convergence rate of compatible relaxation, was first introduced, and is already widely used, in the framework of solving linear systems by algebraic multigrid (AMG); see Sec. 3 in [10].

A major problem in coarsening any system is to find a suitable set of coarse variables. The above criterion gives a general and very effective tool for developing such a set. The adequacy of that set implies practically local dependence of every fine variable on neighboring coarse variables, and hence the feasibility to construct, just by local processing, a set of “*equations*” (in the form of numerical tables) that will govern correct simulations at the coarse level.

In highly repetitive systems (defined above), this local processing need not be done everywhere: the coarse-level equations can iteratively be derived by comparing coarse-level with fine level simulations, where the latter need be performed only in some relatively small representative regions. To start, these regions are independent pieces of fine-level domains, such as a small grid with periodic boundary conditions (e.g., in Examples (i) or (iv)), or a small piece of the chain (in Example (ii)), or a flow domain with a given number of molecules and periodic boundary conditions (in Example (iii)). Later, once approximate equations have already been derived for the next coarser level, the representative regions are chosen to be *windows*, i.e., relatively small subdomains, on the boundaries of which the fine level is kept compatible with the coarse level. Simulations inside these windows supply corrections to the coarse-level equations. The coarse level chooses where and when to establish a new window; normally this is done wherever coarse-level neighborhoods have surfaced for which no previous fine-level simulations have supplied enough statistics (see Sec. 4 below for details.)

Thus, fine level simulations supply (or correct) the governing equations (or operation tables) for the next coarser level. On the other hand, the coarse level selects the windows where these fine-level simulations should take place. Iterating back and forth between all the levels quickly settles into a self-consistent multilevel equilibrium and compatibility; as in multigrid, if the coarsening ratio n/m is not large, no slowdown should occur. More important, at each level the computations need extend only over a collection of small windows, whose number depends on the diversity of local situations, not on the size of the entire domain.

Preliminary experiments with the four simple examples mentioned above have already revealed the very high potential of the SU approach. For instance, in Example (ii), in which conventional simulations run into extreme slowdowns, even the single coarsening level (2.1), with $q = 3$, already accelerates the simulation by at least two orders of magnitude, while accurately reproducing all the relevant statistics (using the coarse Hamiltonian described in Sec. 4 below). This is due to the fact that the dihedral attraction basins, the main reason for slowness of fine-level simulations, are already averaged over at the coarse level.

4 Derivation of coarse equations/interactions

Basic hypothesis: localness of coarsening. The solution to a system of local equations cannot be determined locally: it depends on *all* equations, near

and far. However, *what can essentially be determined just from local information are the coarser-level equations (or interactions)*. More precisely: provided the coarse set of variables is adequate (satisfying the above general coarsening criterion), a coarse system of local equations (or interactions) equivalent to the fine-level system (in the sense that a coarse solution/equilibrium would yield the fine solution/equilibrium by a brief compatible local processing) is iteratively obtainable *locally* (i.e., by iteratively processing only a fine-level neighborhood comparable in size to the typical distance between coarse variables), with an error that decreases exponentially as a function of the total work. (This work grows with enlarging the number of variables involved in each coarse equation, extending the size of the fine-level neighborhood and increasing the number of iterations invested in the local processing.) This hypothesis has emerged from the long and diverse experience with both RG and multigrid solvers.

The actual derivation of the coarse equations, incorporating RG techniques or their RMG (renormalization multigrid) modification [18], is based on fine-level simulations in relatively small representative regions. The simulations produce a sequence of fine-level configurations $u^{(1)}, u^{(2)}, \dots$, which is readily translated into a sequence of the corresponding coarse configurations $u^{(1)c}, u^{(2)c}, \dots$. There exist several approaches as to how and in what form to derive governing coarse-level rules from this latter sequence. We briefly describe two basic approaches with which experience has already been gained.

1. Dependence table. In this approach, the sequence of coarse configurations calculated by fine-level simulations is used to accumulate statistics of the dependence of each coarse variable (called, each in its turn, the “*pivot*”) on a certain set of ν neighboring coarse variables (the “*neighborhood*”). For this purpose the set of possible values of the neighborhood (each being a point in $\mathbb{R}^\nu$) is partitioned into bins. In fully deterministic problems, the average value of the pivot in each bin is accumulated and then tabulated. From such a table, the pivot value for each individual neighborhood can be interpolated, which is all one needs in order to operate (e.g., perform relaxation) at the coarse level. In stochastic problems, additional statistics (such as variation and higher moments) of the pivot over each neighborhood bin are tabulated, enough to enable accurate Monte Carlo simulations at the coarse level. Successful experience with simple versions of this approach, including the cases of Examples (iii) and (iv) above, are reported in [18], [19], [20] and [34].

2. Coarse Hamiltonian. In this approach the sequence of coarse configurations $\{u^{(j)c}\}_j$ is used to calculate averages of quite many coarse-level observables $O_1, O_2, \dots, O_\mu$. Denote the average of O_i by $\langle O_i \rangle_f$, ($i = 1, \dots, \mu$), the subscript f indicating that this average of a coarse observable has been extracted from the *fine*-level simulations ($u^{(j)c}$ extracted from $u^{(j)}$). The coarse level itself is intended to be governed by a (yet to be calculated) Hamiltonian-like functional $H^c(u^c)$, that is, the probability of a coarse configuration u^c will be proportional to $\exp(-H^c(u^c))$. For any given approximate H^c , one can run simulations at the *coarse* level during which $\langle O_i \rangle_c$, the average of O_i according

to H^c , can be calculated. The aim is to construct H^c such that

$$\langle O_i \rangle_c = \langle O_i \rangle_f, \quad (i = 1, \dots, \mu). \quad (4.0)$$

For this purpose H^c is written in the general form

$$H^c(u^c) = \sum_{k=1}^K a_k H_k(u^c), \quad (4.1)$$

where each H_k is a known functional of u^c (see example below) and $\{a_k\}$ is a set of coefficients that need to be found. A crude approximation to H^c , possibly with a reduced number of terms (reduced K), can inexpensively be obtained from small-scale fine-level calculations, by assuming independence of various quantities (as in the example below). The approximation is then improved in few Newton-like iterations, during which K may increase. Specifically, in each iteration H^c is changed by adding to it $\delta H^c = \sum_k \delta a_k H_k$. Using the first-order relation (used for example by K.G. Wilson and R.H. Swendsen; cf., e.g., [26])

$$\delta \langle O \rangle = \langle O \rangle \langle \delta H^c \rangle - \langle O \cdot \delta H^c \rangle, \quad (4.2)$$

one gets a system of μ equations

$$\sum_k (\langle O_i \rangle \langle H_k \rangle - \langle O_i H_k \rangle) \delta a_k = \langle O_i \rangle_f - \langle O_i \rangle_c, \quad (i = 1, \dots, \mu) \quad (4.3)$$

from which δa_k can be calculated by least squares. (Usually the sequence of observables will include $H_1, H_2, \dots, H_K$ and possibly some others.)

An observable O for which the discrepancy $\langle O \rangle_f - \langle O \rangle_c$ remains particularly large can be added to the list of Hamiltonian terms (thus increasing K) to facilitate decrease of that discrepancy in the next iterations. This may in particular apply to some of the “second moment” observables $O_i H_k$, whose averages, needed in (4.3), are anyway being calculated. Adding to the Hamiltonian the observables with largest discrepancies usually cause the discrepancies to sharply decrease in other observables as well, including still-higher-moment observables.

4.1 Example.

A test of the coarse Hamiltonian approach has been carried out with the polymer case (Example (ii) in Sec. 3, employing the united-atom model of Sec. 2), using the coarsening (3.1) with $q = 3$. (A preliminary description, but with $q = 4$, has appeared in [1].) The first approximation to H^c is chosen in a general form similar to the fine-level Hamiltonian (2.1),

$$H^c(u^c) = \sum_i F_1(|r_i^c - r_{i+1}^c|) + \sum_i F_2(\theta_i^c) + \sum_i F_3(\tau_i^c) + \sum_{|i-j|>2} F_4(|r_i^c - r_j^c|), \quad (4.4)$$

where $r_i^c = u_i^c$ is the location of the i -th coarse “atom”, $|r_i^c - r_{i+1}^c|$ is the distance between two successive coarse “atoms”, θ_i^c is the angle $(r_{i-1}^c, r_i^c, r_{i+1}^c)$, τ_i is the

torsion $(r_{i-1}^c, r_i^c, r_{i+1}^c, r_{i+2}^c)$ and F_4 is a Lennard-Jones-like interaction. Each of the initially-unknown functions F_ℓ can be expanded in the form

$$F_\ell(\xi) = \sum_j a_{\ell,j} w_j(\xi) , \quad (\ell = 1, 2, 3, 4) \quad (4.5)$$

with unknown coefficients $a_{\ell,j}$ and known basis function $w_j(\xi)$; e.g., local basis functions (finite elements). Upon collecting (over the relevant $\sum_i$ in (4.4)) all terms that include the same unknown $a_{\ell,j}$, the coarse Hamiltonian (4.4) obtains the general form (4.1). A reasonable first approximation to F_4 is the given Lennard-Jones interactions (see (2.1), with $|r_i^c - r_j^c|$ replacing r_{ij}), multiplied by q^2 . A first approximation to F_1 (similarly: F_2, F_3) can be calculated with a short polymer chain (e.g., $n = 24$ and $m = 8$) from the bare distribution of the distances $\{|r_i^c - r_{i+1}^c|\}_{i=2}^{m-2}$, omitting (or calculating separately) the exceptional distances at the ends ($i = 1$ and $i = m - 1$), and ignoring all correlations. The iterations described above will then automatically correct for those correlations, introducing on the way some new explicit correlation terms into H^c (with increasing K).

Note that these corrections are first still being done with the *short* polymer chain, where very efficient fine-level Monte-Carlo simulations can be performed (as explained in Sec. 2). By the basic hypothesis above (and confirmed in this case by test calculations), the shortness of the chain used initially only marginally affects the accuracy of the coarse interactions located sufficiently deep inside (several atomic distances from the ends) of the chain. Then, with the improved H^c one can already make simulations at the coarse level, with much longer chains (e.g., $n = 120$). One can now further improve H^c by making additional iterations, now making the fine-level simulations in windows within that longer chain. Then H^c can be similarly used to derive Hamiltonians at still coarser levels, whose simulations will be performed on still longer chains. The longer chains may produce new situations (e.g., contact points due to folding) that require some new, window-within-window calculations at all finer levels to further correct H^c (including for example new terms that correspond to the contact points). The formulation is very flexible, allowing introducing new Hamiltonian terms to account for new situations.

In the case of more complicated macromolecules (not made of one single repeat unit) there should of course be employed more windows, one window per each local molecular structure. But since many of these structures typically have many intra-molecular and inter-molecular repetitions, a limited number of windows may represent many long chains. (No less important for efficiency is of course the result, mentioned at the end of Sec. 3 above, that the coarse-level simulations are much faster than the fine-level ones, even when the latter are being done with the full macromolecule.)

5 Extensions

Many important extensions of the upscaling techniques described above, to diverse physical situations, can be developed. The following is a survey of some principal directions that have already been explored to some extent.

5.1 Long range interactions

Long-range (e.g., electrostatic) interactions can each be decomposed into the sum of a smooth interaction and a local one (“smooth” and “local” being meant on the scale of the next coarse level). All familiar physical interactions, even with oscillatory kernels, can be decomposed this way (see [8] and examples in [17], [35] and [31]). To any desired accuracy, the smooth part can directly be represented at the coarse level, e.g., by aggregated charges and dipoles moving with the coarse level “atoms” (in Example (ii) above) or by adjoint interpolation of charges to the coarse-level lattice (in Example (iii)). The local part is essentially transferred, together with all other local interactions, using the fine/coarse iterations described in Sec. 4 above. Effectively, the amount of work invested per charge involves only calculating its local interactions, and, even more importantly, only charges within selected windows need be treated. It can be shown that this is possible due to the smoothness of the non-local interactions, which makes them little sensitive to the (properly designed) local MC moves. (This, incidentally, is not true with the type of decompositions used by the Fast Multipole Method [25], which therefore cannot be used in the framework of multiscale particle movements.)

5.2 Dynamical systems

For time-dependent systems, the general coarsening criterion of Sec. 3 is replaced by the analogous requirement that differences between the fine-scale statistics of two evolving configurations, that are suitably adjusted to be compatible at the coarse scale, practically disappear within few time steps. Dependence tables (e.g., in kinetic Monte Carlo computations) have been derived in the form of flux dependence on both current-time and previous-time neighboring coarse variables [33]. A computational criterion has as well been formulated for the size of the time steps to increase with the spatial coarsening level, so as to maintain full efficiency. A Hamiltonian-like functional that governs every time step can also be developed analogously to the one described in Sec. 4.

For Hamiltonian systems (e.g., corresponding to Examples (ii) and (iii) in Sec. 3), the multiscale structure allows a natural combination of temperature-accurate statistical simulation at small scales with time-accurate dynamics at large scales. Assuming that after any given time interval the fine-scale degrees of freedom settle into a local equilibrium slave to the coarse-level averages (where that scale and that level increase with the size of the time interval), the general criterion for choosing the coarse variables can directly be applied as in Sec. 3, using the *equilibrium* CMC, for appropriate temperatures. (Note that the

local temperature may well serve as a local coarse-level variable, corresponding to a local averaging of kinetic energy.) Large time steps, based on implicit discretization of Newton law, can then be made, using a multigrid-like solver where the relaxation at fine levels is replaced by CMC (cf. [11, §14.8]). This approach yields two benefits in performing very large time steps: firstly, it allows much easier handling of local minima. Secondly, it avoids the killing of highly-oscillatory modes (those vibrations that are not resolved by the time step), which would occur if the implicit equations of a large time step were imposed at all scales. Instead, these modes assume stochastic amplitudes according to their equilibrium probability distribution. The desired temperature, and, in particular, regional temperature variations when relevant, are introduced very directly in this way, avoiding the need for fabricating Langevin stochastic forces.

Another possible approach is to first develop at equilibrium a coarse-level Hamiltonian $H^c(u^c)$ such that the relation (deviating here from the above notation) $\text{Probability}(u^c) \sim \exp(-H^c(u^c)/k_B T)$ will simultaneously hold for a full range of temperatures T . (This can be achieved by adding several moments $(H^c)^m$ to the list of observables (O_i) used in (4), and constructing a *joint Hamiltonian* (see Sec. 5.4 below) for different temperatures in the range of interest). Then use this H^c in Newtonian dynamics at the coarse level, where effective coarse-level masses (and their possible dependence on the coarse coordinates) are determined by comparing (in windows of fine-level dynamic simulations) coarse-level accelerations with the gradient of H^c .

Still another approach, relevant for fluids, is a Boltzmann-type upscaling in the 6D space of positions and velocities. Starting with dynamic simulations at the individual-particle level, increasingly coarser spatial levels will describe velocity distributions at progressively higher resolutions.

5.3 Stochastic coarsening

Our studies (e.g., [33]) have shown that averaging upon coarsening should often better be stochastic. The added stochasticity is important to create smoother coarse dynamics, hence simpler dependence table or easier H^c , especially for a fine level with discrete-state (e.g., integer-valued) variables or highly oscillating Hamiltonian (as in real-time Feynman path integrals). One general way is to modify a deterministic averaging (or antepolation — the adjoint of interpolation) by adding to each coarse variable a small stochastic increment, where the field of increments is in equilibrium governed by a Hamiltonian-like functional H_p . A corresponding CMC process has been developed, and the general coarsening criterion then effectively checks that H_p has been properly designed, i.e., it prohibits increment fields that correspond to averaging long-range fine-level moves.

5.4 Joint H^c

The same coarse functional H^c should sometimes simultaneously satisfy (4.0) for several different MC situations, such as: (a) under different external fields; (b) at

different temperatures (cf. Sec. 5.2); (c) in different energy basins (cf. Sec. 5.6). Generally, this can be achieved by adding in (4.1) terms H_k that are particularly sensitive to the differences between the different simulated situations.

5.5 Complex fluids

More elaborate coarse Hamiltonians are needed for fluids with more complex molecules of one or several species, such as water with methanol, or glycerol, etc. A gradual construction plan is due, starting for example with H^c constructed for atomistic equilibrium in a periodic domain containing only two molecules. Adding then to the simulation one molecule at a time, the coefficients of H^c are updated by (4.3), with additional terms H_k that correspond at each iteration to correlation observables that are still ill approximated.

5.6 Low temperatures

At high temperatures, the coarse variables for a simple fluid in equilibrium are gridpoint values, each standing for some local averaging of $m(\mathbf{x})$, the masses m of particles at various positions $\mathbf{x} = (x_1, x_2, x_3)$ near the gridpoint. Suppose that at lower temperatures the fluid starts to crystallize, roughly with periods $\mathbf{u}^{(\ell)} = (u_1^{(\ell)}, u_2^{(\ell)}, u_3^{(\ell)})$, ($\ell = 1, 2, 3$), say. Three new coarse-level fields should then enter, standing for local averaging of $\exp(2\pi i \mathbf{w}^{(\ell)} \cdot \mathbf{x}) * m(\mathbf{x})$, ($\ell = 1, 2, 3$), where $\mathbf{w}^{(\ell)} \cdot \mathbf{u}^{(\ell)} \simeq \delta_{k\ell}$. If the crystal is perfect and $\mathbf{w}^{(\ell)}$ are exactly known, these coarse variables will turn out constant. When $\mathbf{w}^{(\ell)}$ are only approximate, these variables will oscillate smoothly. Similar averaging at the next coarser levels will then describe these oscillations, effectively correcting the erroneous $\mathbf{w}^{(\ell)}$. If the crystal is not perfect, meaningful averaging of this type will persist only up to a certain scale; usually, the lower the temperature the larger the scale [11, §14.7.3].

Also upon lowering the temperature, energy barriers emerge at increasingly larger scales. By insisting on constructing, level after level, *joint* H^c (see Sec. 5.4), statistically correct transitions between different energy basins can efficiently be simulated.

5.7 Multiscale annealing

Quite generally, as in the above example (Sec. 5.6), as a system is gradually cooled, increasingly larger-scale degrees of freedom are identified. This identification of increasingly larger collective moves makes such a computation extremely more effective than simple simulated annealing [28] for *minimizing* the energy (the limit $T \rightarrow 0$), especially in the physically common situation of multiscale nested attraction basins. The multiscale annealing can provide an efficient solver to very difficult global optimization problems. (See much more in [11, §18.2], [21].)

5.8 Coarse-level transparency of fine observables

Often, an observable of interest is not directly expressed in terms of the coarse-level variables. Similar to the procedures for deriving coarse-level Hamiltonian (see Sec. 4), a general procedure can be developed for computing a functional dependence of a quantity of interest upon the coarse variables, based on suitable statistics accumulated during the fine-level simulations.

5.9 Determinism and stochasticity

Much of the discussion above is written in terms of stochastic systems, but can be extended to deterministic ones. Moreover, a stochastic system at the fine level often yields a deterministic system at large enough scales. The opposite exists too: A deterministic fluid flow at the small viscous scale can acquire stochastic features at the large scales reigned by turbulence. The coarsening approaches described above can accommodate such transitions.

References

- [1] Bai, D. and Brandt, A., Multiscale computation of polymer models. In: Brandt, A., Bernholc, J. and Binder, K. (Eds.), *Multiscale Computational Methods in Chemistry and Physics*. NATO Science Series: Computer and System Sciences, Vol. 177, IOS Press, Amsterdam (2001), pp. 250–266.
- [2] Brandt, A., Multi-level adaptive technique (MLAT) for fast numerical solutions to boundary value problems, in *Proc. 3rd Int. Conf. on Numerical Methods in Fluid Mechanics* (Cabannes, H. and Temam, R., eds.), Lecture Notes in Physics 18, Springer-Verlag, 1973, pp. 82–89.
- [3] Brandt, A., Multi-level adaptive solutions to boundary value problems, *Math. Comp.* **31** (1977) 333–390.
- [4] Brandt, A., Multi-grid solvers on parallel computers, in: *Elliptic Problem Solvers* (Schultz, M., ed.), Academic Press, New York, 1981, pp. 39–84.
- [5] Brandt, A., Guide to multigrid development, in *Multigrid Methods* (Hackbusch, W. and Trottenberg, U., eds.), Springer-Verlag, 1982, pp. 220–312.
- [6] Brandt, A., *Multigrid Techniques: 1984 Guide, with Applications to Fluid Dynamics*, 1984, 191 pages, ISBN-3-88457-081-1; GMD-Studien Nr. 85; Available from GMD-AIW, Postfach 1316, D-53731, St. Augustin 1, Germany, 1984.
- [7] Brandt, A., Algebraic multigrid theory: The symmetric case, in *Preliminary Proc. Int. Multigrid Conf.*, Copper Mountain, Colorado, April 6–8, 1983; *Appl. Math. Comp.* **19** (1986) 23–56.

- [8] Brandt, A., Multilevel computations of integral transforms and particle interactions with oscillatory kernels, *Comp. Phys. Comm.* **65** (1991) 24–38.
- [9] Brandt, A., Multigrid methods in lattice field computations, *Nucl. Phys. B* (Proc. Suppl.) **26** (1992) 137–180.
- [10] Brandt, A., General highly accurate algebraic coarsening schemes, *Electr. Trans. Num. Anal.* **10** (2000) 1–20.
- [11] Brandt, A., Multiscale scientific computation: review 2001. In Barth, T.J., Chan, T.F. and Haimes, R. (eds.): *Multiscale and Multiresolution Methods: Theory and Applications*, Springer Verlag, Heidelberg, 2001, pp. 1–96. URL: www.wisdom.weizmann.ac.il/~achi/review00.ps.
- [12] Brandt, A., McCormick, S. and Ruge, J., Algebraic multigrid (AMG) for automatic multigrid solution with application to geodetic computations, Institute for Computational Studies, POB 1852, Fort Collins, Colorado, 1982.
- [13] Brandt, A., McCormick, S. and Ruge, J., Algebraic multigrid (AMG) for sparse matrix equations, in *Sparsity and its Applications* (Evans, D.J., ed.), Cambridge University Press, Cambridge, 1984, pp. 257–284.
- [14] Brandt, A., Ron, D. and Amit, D.J., Multi-level approaches to discrete-state and stochastic problems, in *Multigrid Methods, II* (Hackbusch, W. and Trottenberg, U., eds.), Springer-Verlag, 1986, pp. 66–99.
- [15] Brandt, A., Galun, M. and Ron, D., Optimal multigrid algorithms for calculating thermodynamic limits, *J. Stat. Phys.* **74** (1994) 313–348.
- [16] Brandt, A. and Livshits, I., Wave-ray multigrid method for standing wave equations, *Electronic Trans. Num. An.* **6** (1997), 162–181.
- [17] Brandt, A. and Venner, C.H., Multilevel evaluation of integral transforms with asymptotically smooth kernels, *SIAM J. Sci. Comp.* **19** (1998), 468–492.
- [18] Brandt, A. and Ron, D., Renormalization multigrid (RMG): Statistically optimal renormalization group flow and coarse-to-fine Monte Carlo acceleration, *J. Stat. Phys.* **102** (2001) 231–257.
- [19] Brandt, A. and Iliyn, V., Multilevel approach in statistical physics of liquids. In: Brandt, A., Bernholc, J. and Binder, K. (Eds.), *Multiscale Computational Methods in Chemistry and Physics*. NATO Science Series: Computer and System Sciences, Vol. 177, IOS Press, Amsterdam (2001), pp. 187–197.
- [20] Brandt, A. and Ilyin, V., Multilevel Monte Carlo methods for studying large-scale phenomena in fluids. *J. of Molecular Liquids*, **105** (2003) 253–256.

- [21] Brandt, A. and Ron, D., Multigrid solvers and Multilevel Optimization Strategies. In: *Multilevel Optimization and VLSICAD*, (J. Cong and J.R. Shinnerl, eds.), Kluwer Academic Publishers, Boston, 2003, pp. 1–69. URL: www.wisdom.weizmann.ac.il/~achi/ucla.pdf
- [22] Briggs, W.L., Henson, V.E., and McCormick, S.F., *A Multigrid Tutorial*, 2nd Ed., SIAM, 2000.
- [23] Flory, P. J., *Statistical Mechanics of Chain Molecules*, January 1989, ISBN 1569900191.
- [24] Goodman, J. and Sokal, A.D., Multigrid Monte Carlo methods for lattice field theories. *Phys. Rev. Lett.*, 56 (1986) 1015–1018.
- [25] Greengard, L., *The Rapid Evaluation of Potential Fields in Particle Systems*, MIT Press, Cambridge, 1988.
- [26] Gupta, R., Open problems in Monte Carlo Renormalization Group: Application to critical phenomena, *J. Appl. Phys.* **61** (1987) 3605–3611.
- [27] Hackbusch, W., *Multigrid Methods and Applications*. Springer, Berlin, 1985.
- [28] Kirkpatrick, S., Gelatt Jr., C.D. and Vecchi, M.P., Optimization by simulated annealing, *Science* **220** (1983), 671.
- [29] Mack, G. and Pordt, A., Convergent perturbation expansions for Euclidean quantum field theory., *Comm. Math. Phys.* **97** (1985) 267. Also: Mack, G. in: *Nonperturbative Quantum Field Theory*, G. t’Hooft *et al.*, eds. Plenum Press, New York, 1988, p. 309.
- [30] Paul, W., Yoon, D.Y. and Smith, G.D., An optimized united atom model for simulations of polymethylene melts, *J. Chem. Phys.* **103** (1995) 1702–1709.
- [31] Ramirez, I.H., *Multilevel Multi-Integration Algorithm for Acoustics*, Ph.D. Thesis, University of Twente, Enschede, The Netherlands, 2005.
- [32] Ruge, J. and Stüben, K., Algebraic multigrid. In *Multigrid Methods* (McCormick, S. F., ed.), SIAM, Philadelphia, 1987, pp. 73–130.
- [33] Saad, N., *Multiscale Algorithm for Time-Dependent One-Dimensional System*, Feinberg Graduate School, Weizmann Institute of Science, Rehovot, Israel, 2005.
- [34] Shmulyian, S., *Toward Optimal Multigrid Monte Carlo Computation in Two-Dimensional $O(N)$ Non-Linear σ -Models*, Ph.D. Thesis, Weizmann Institute of Science, 1999.

- [35] Suwan, I., Multiscale Methods in Molecular Dynamics, Ph.D. Thesis, Feinberg Graduate School, Weizmann Institute of Science, Rehovot, Israel, 2005.
- [36] Swendsen, R.H., Renormalization group Monte Carlo methods, *Phys. Rev. Lett.* **42** (1979) 859–862.
- [37] Swendsen, R.H., Monte Carlo renormalization-group studies of the $d = 2$ Ising model, *Phys. Rev. B* **20** (1979) 2080–2087.
- [38] Trottenberg, U, Oosterlee, C.W., and Schüller, A., *Multigrid*. Academic Press, London, 2000.
- [39] Wilson, K.G., The Renormalization Group and Critical Phenomena. *Rev. Mod. Phys.*, 55 (1983) 583–600. (1982, Nobel Prize Lecture)