



Distance Geometry Algorithms in Molecular Modelling of Polymer and Composite Systems

R. V. N. MELNIK*

Mads Clausen Institute, University of Southern Denmark
DK-6400, Denmark
rmelnik@mci.sdu.dk

A. UHLHERR AND J. HODGKIN

CSIRO Molecular Science
Bag 10, Clayton South, VIC 3169, Australia

F. DE HOOG

CSIRO Mathematical and Information Science
GPO Box 664, Canberra, ACT 2601, Australia

Abstract—In this paper, microscopic atomistically detailed models of polymer and composite systems are studied in the context of distance geometry algorithms. The development of such algorithms in the general framework of molecular modelling is closely associated with problems in computational optimisation and is pursued here as a key to a better understanding of properties of polymer and composite systems at the macroscopic level. Applications of two distance geometry algorithms to topological optimisation of alkane chains and bulk decane structures are discussed with results of numerical simulations. © 2003 Elsevier Science Ltd. All rights reserved.

Keywords—Distance geometry algorithms, Polymer and composite systems, Computational optimisation.

1. INTRODUCTION

Development in polymers and composites is at the heart of many new industrial materials technologies. Recent advances in this field include applications of environmentally stable high temperature thermoplastics, new thermoset materials, photoresists for fabrication of VLSI, flexible electronic conductors, and highly polarised materials such as ferroelectrics, piezoelectrics, and nonlinear optical materials. In such areas as aerospace and automotive industries in particular, further progress in this field is often impeded by inadequate understanding of how atomistically detailed molecular structures lead to macroscopic properties. Because the polymers and composites (in fact, many commercial polymers are composites, e.g., thermosetting resins containing fillers) applied in these areas of industry are generally disordered and heterogeneous (at least at the microscopic level), the current experimental methods can generally provide only hazy

* Author to whom all correspondence should be addressed. The work was done while the author was with CSIRO Mathematical and Information Sciences.

The authors are thankful to Dr. D. Jenkins for his comments on the preliminary version of this paper. Suggestions of the referees leading to the improvement of the paper are also gratefully acknowledged.

guidelines concerning local structures. Various properties of these materials are interrelated at the molecular level, and changing a single *variable*, say, the extent of crosslinking, the plasticizer content, or the polymer molecular weight, will have mutual effects on the use response [1]. In many cases it is important to know for any given material property the effect caused by the whole set of fabrication-controlled material variables or molecular characteristics because the advantages that composite or polymer materials have to offer have to be balanced against their undesirable properties (difficult fabrication techniques, complex rheological behaviour, etc.). Therefore, modelling polymer and composite structures at the molecular level constitutes a natural approach to the solution of complex industrial engineering problems in the field where applications of mathematical and computational tools are both beneficial for industry and mathematically challenging. In this paper, we deal with such tools applied to molecular modelling, i.e., to the generation, manipulation, and representation of realistic 3D molecular structures with the purpose of understanding the physicochemical properties and macroscopic phenomena at the molecular level.

Molecular modelling has become a powerful approach in science and engineering in general, and in computational (bio)chemistry in particular. It is widely used for searching conformational space in order to find stable structures, for examining the effect of thermal motions on the structure of a molecule, for the analysis of the dynamics to obtain macroscopic materials' properties as a function of temperature and pressure, etc. Since any molecule can be in an infinite number of spatial states, i.e., configurations, in practical applications we have to limit ourselves by subsets of such configurations, known as conformations, which share common (physico)chemical properties [2].

In principle, using methods of quantum mechanics, we can compute first principles force fields and then, using the latter, we can apply molecular dynamics techniques to calculate trajectories which will allow us to extract the information on the macroscopic properties of the material. However, if we take into account that to describe even partially crystalline polymers one needs $\sim 10^6$ atoms per unit cell, it becomes clear that the above classical *ab initio* procedure (and other methods of quantum chemistry such as pseudospectral or cell multipole methods) can hardly be applied to realistic polymers and composites. Typically, we must not only average electronic degrees of freedom into force fields and charges, but also impose "hard" constraints on bonds and angles to eliminate those degrees of freedom which are only lightly excited, and hence, have negligible effect on conformational (i.e., long-scale) changes in polymer or composite molecules. In particular, it is customary to assume that all bonds and angles are constrained and only the torsional degrees of freedom in a chain are relevant in determining its overall conformations [3]. Due to such simplifications, molecular modelling methods cannot be perfectly accurate, but they can

- (a) reasonably indicate whether it might be worthwhile to synthesise the material, and
- (b) suggest alternative structures to consider [4].

Hence, with the spotlight fixed on these strengths of molecular modelling methods together with essential computational cost savings compared to the quantum mechanics methods, we conclude that the molecular modelling approach is a natural choice in simulating complex polymer and composite structures.

To get started with most of the molecular modelling procedures, including molecular dynamics (MD) and Monte-Carlo (MC), one has to determine an initial topological structure of the material under consideration. For complex polymer and composite systems, this task is far from trivial and, moreover, the solution to this task keeps a key to success in the whole molecular modelling exercise. Indeed, since polymers and composites are characterised by a wide-range hierarchy of different length and time scales, in molecular simulations of complex materials such as long-chain polymers, the large-scale conformational characteristics of the system are given too little time to evolve. As a result, if the initial configuration is poorly chosen, the system often remains trapped within its neighbourhood, which leads to insufficient samplings of the configu-

ration space and unreliable estimations of the dynamic and structural properties of the material. Therefore, a major challenge in the field of molecular modelling of polymer and composite systems is to build realistic and *computationally inexpensive initial configurations* which can then be relaxed effectively by the available molecular modelling methodologies [5]. In this paper, we show that this challenge can be effectively addressed with distance geometry algorithms followed by molecular modelling simulations. Since we consider these two powerful tools as complementary, the paper consists of two closely connected parts. First, we describe the general methodology of molecular modelling, pointing out why, where, and how the computational cost should and could be reduced when working with a hierarchy of models for simulation of polymers and composites (Sections 2–4). Then, we focus on mathematical aspects of the distance geometry approach allowing us to provide an efficient starting point for improving the existing molecular modelling codes; we give an algorithmic description of a new computational procedure and show its effectiveness on the results of several computational experiments (Sections 5–7). A detailed discussion of these issues will follow the structure according to which the rest of the paper is organised.

- In Section 2, we describe a general framework for modelling and simulation of polymers and composites and the place of distance geometry in this framework.
- In Section 3, we provide the reader with the mathematical foundations of the distance geometry methodology considered here as a starting point in molecular modelling.
- In Section 4, we focus on principal criteria for the design of distance geometry algorithms and discuss the effect of topological constraints on macroscopic properties of material structures. Other methods for finding starting geometries are also discussed in this section.
- In Section 5, the general problem of distance geometry is split into a series of problems in computational optimisation.
- In Section 6, we apply potential-function-smoothing procedures to simulation of short-chain polyethylene isolated in space and discuss options for improvements of such procedures.
- Section 7 is devoted to the description of a distance geometry algorithm capable of dealing with bulk material structures. Results of numerical simulations of long alkane chains and bulk decane structures are also presented in this section.
- Conclusions and future directions are discussed in Section 8.

2. DISTANCE GEOMETRY METHODOLOGY IN A HIERARCHY OF MODELS FOR SIMULATION OF POLYMERS AND COMPOSITES

A better understanding of molecular structures of polymer and composite systems and their intrinsic connection with macroscopic properties of these systems requires the application of tools from chemical physics, molecular biology, applied mathematics together with sophisticated measurement, and computational techniques. Remarkably, the situation might look relatively simple, at least in principle. Indeed, given

- (a) nuclear coordinates,
- (b) the number of electrons the system contains, and
- (c) a set of atomic orbital forms assigned to each atom,

the most rigorous *ab initio* methodologies to atomistic simulation of structural, mechanical, electrical, and optical properties of materials are those that are based on first principles quantum mechanics methods in which the electronic states and structures are calculated directly from the Schrödinger equation, which governs the motions of electrons and nuclei in atoms and molecules [6]. Such methodologies, known for almost a century¹, remain very restricted in size

¹In 1929 Paul Dirac noted, “The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus, completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble.”

and time scales and are usually computationally impractical in modelling realistic polymer and composite systems. Typically, quantum mechanical calculations can provide information on potential surfaces for small molecules, but moving towards more complex polymer and composite systems, characterised by a wide-range hierarchy of different length and time scales, *empirical* energy functions of the molecular mechanics type become the only practical source of such information [7,8]. In principle, however, using coarse graining and averaging procedures we can devise a hierarchy of models for simulation of polymers and composites, briefly summarised below.

Model Type	Methodologies	Space Scales	Time Scales
Quantum Level	Quantum chemistry, Quantum Monte Carlo, Band theory	1 Å–10 Å	10^{-15} – 10^{-12} s
Atomic/Molecular Level	Classical dynamics, Statistical mechanics	10 Å–100 Å	10^{-12} – 10^{-6} s
Phenomenological	Nanotechnological, microstructural	100 Å–1 μm	10^{-6} – 10^{-3} s
Continuum/Discrete	Microstructural, Semimacrostructural, Macrostructural	1 μm → cms	10^{-3} s → mins
Physical/Engineering	Engineering design	Expanding limits	Expanding limits

Some entries in the above table are overlapping and could be attributed to the same group under other classification schemes. For example, in the context of this paper, it is worthwhile noting that models and methodologies related to quantum chemistry and molecular dynamics are referred to generically as atomistic simulation. Our interest in atomistic simulation of polymer and composite systems is limited to the assumption that the force field² has been derived empirically using the molecular mechanics approach [9]. In other words, by assuming that a model and a force field have been already chosen, we put ourselves in the framework of molecular modelling where the ultimate goal is to find the geometry with the minimum strain energy by rearrangement of the nuclei. Provided electronic effects are negligible, this approach is considered as a predictive tool leading to the possibility of evaluating macroscopic properties of the material from its microscopic structure.

Much progress has recently been achieved along this path with MC procedures applied to polymer and composite systems (in particular, with modifications of configuration-bias MC, concerted rotation, and bridging moves methods [10]), where the focus has been placed on the algorithms for which the time of simulation depends *weakly* on the length of the chain. Although it is sometimes argued that this feature gives a substantial advantage to MC procedure compared to the classical MD procedures, we think that it would be proper to say that the former could be considered as a real alternative to the latter only for *static* thermodynamic properties (after all, one has to deal with a fast increase in the longest relaxation time when the chain length increases, in one way or another!). For the purpose of this paper, it is important to emphasise that most of the available molecular modelling procedures, including both MD and MC, exhibit a high sensitivity to parameterisation processes related to the choice of model and force field in a specific engineering problem, as well as the choice of the initial geometry which then is relaxed with such procedures.

Therefore, at the initial stage of molecular modelling, we need an efficient method that requires neither force field parameters, nor a starting conformation, but at the same time is capable of predicting correctly the overall topological structure of the molecular system. To be viable for the analysis of complex polymer and composite systems, this method has to be computationally efficient and relatively simple. The developing, testing, and putting such a method in the context of molecular modelling procedures would be an important contribution to the field.

²That is, the set of functions that defines an approximation to the strain energy together with the collection of terms that parametrises this approximation.

Distance geometry algorithms (DGAs) are the most obvious candidates for forming the basis of such a method, because

- (a) distances are coming naturally from chemical bonds, bond angles, and torsional angles, and
- (b) the complete conformational space of the system is contained within the set of minimum and maximum interatomic distances between all pairs of atoms in the molecule system.

Since their first appearance in computational biochemistry in the late 1970s, DGAs have been applied not only in nuclear magnetic resonance structure determination and in conformational analysis of small molecules, where they performed as well as other sampling procedures, but also in such technologically important fields as constructing genetic maps [11]. These algorithms have recently attracted the attention of researchers as a potentially powerful method for building approximate models of complex polymer and composite structures in conformational analysis [12]. By developing efficient methods for generating and minimising coordinates of each atom within the structure against the distance error function, we open a way to a substantial increase in the effectiveness of molecular dynamics codes and molecular modelling procedures that allow us to extract physicomechanical properties of technologically important polymers and composites [6,13].

3. MATHEMATICAL FOUNDATIONS OF MOLECULAR MODELLING: DISTANCE GEOMETRY AS A STARTING POINT OF SIMULATIONS

The foundations of the distance geometry methodology in computational (bio)chemistry can be traced back to the Born-Oppenheimer approximation of the Schrödinger equation (for mathematical history of distance geometry the reader can consult [14–16]), where the nuclei are viewed as executing small oscillations about an equilibrium conformation which is a result of time-averaged or steady-state electronic configuration. If we assume that the (dynamical) behaviour of a molecular system is well predictable using classical mechanics, then this approximation allows us, at least in principle, to calculate time-dependent movement of each atom in a molecule by solving Newton's equations of motion for all degrees of freedom

$$\frac{d^2x_i}{dt^2} = \frac{\mathbf{F}_i}{m_i}, \quad (3.1a)$$

$$\mathbf{F}_i = -\frac{\partial \mathcal{E}}{\partial x_i}, \quad i = 1, \dots, 3N, \quad (3.1b)$$

given the energy surface \mathcal{E} and the derivative of energy in terms of nuclear coordinates x_i which determines the instantaneous force on any atom \mathbf{F}_i [4]. Assuming that from (3.1b) we can derive good approximations of forces between the atoms, we expect to have a good approximation of the geometry of the molecule by solving $3N$ equations (3.1a). Note that deriving such approximations is not a trivial task, because the force law in molecular dynamics requires accounting not only for the two-body force, but also for three-body force (by the bond angles), and four-body force (by the torsion angles). If the molecular system under consideration is Hamiltonian, we identify the total energy of the system with the Hamiltonian function

$$H(\mathbf{p}, \mathbf{q}, t) \equiv \mathcal{E} = K + P, \quad (3.2)$$

where kinetic energy, K , and the potential (strain) energy, P , are represented by

$$K = \frac{1}{2} \langle M\mathbf{p}, \mathbf{p} \rangle, \quad (3.3a)$$

$$P(\mathbf{x}, t) = \sum_{\forall \text{ molecules}} (\mathcal{E}_b + \mathcal{E}_\theta + \mathcal{E}_\phi + \mathcal{E}_{nb}), \quad (3.3b)$$

with $\mathbf{x} = (x_1, \dots, x_{3N})$ and $\mathbf{p} = (p_1, \dots, p_{3N})$ being the corresponding positions and momenta of the atoms, $M \in \mathbb{R}^{3N \times 3N}$ being the diagonal mass matrix $\text{diag}(m_1, \dots, m_{3N})$, and $\langle \cdot, \cdot \rangle$ being the Euclidean inner product in \mathbb{R}^{3N} , so that

$$K \equiv \frac{1}{2} \mathbf{p}^T M^{-1} \mathbf{p} = \sum_{i=1}^{3N} \frac{p_i^2}{2m_i}. \quad (3.4)$$

The summations with respect to each term in the potential energy function (3.3b) denote the total bond deformation (stretching) energy, the total valence angle deformation (bending) energy, the total torsional (dihedral) angle deformation energy, and the total nonbonded interaction (separated by at least two atoms, i.e., van der Waals) energy, respectively.

Molecular dynamics simulations are initiated by the kinetic energy of the atoms (3.4) chosen from a random velocity distribution, and we assume that

$$\mathbf{p}(t_0) = \mathbf{p}_0. \quad (3.5)$$

The resulting distortions from the equilibrium configuration are opposed by a force commensurate with the gradient of H . Since H can be only an approximation [17], obtained, for example, from the consistent valence force field theory (see, for example, [18] and references therein), it is often convenient to split its “potential” part (3.3b) into the sum of weak and strong interactions [19]

$$P = P_{\text{weak}} + P_{\text{strong}}, \quad P_{\text{strong}} = \epsilon^{-2} U, \quad (3.6)$$

with appropriate scaling factor (or a singular perturbation parameter) ϵ and function U . In what follows, we assume only that all energy terms participating in the definition of H can be effectively expressed as a function of the coordinates of the atoms constituting the molecule system and a set of parameters computed from experimental data. Having obtained an approximation to the system Hamiltonian, we can reduce numerical integration of the equations of motions (3.1) to the solution of the following Hamilton’s equations:

$$\frac{\partial H}{\partial p_i} = \frac{dx_i}{dt}, \quad \frac{\partial H}{\partial x_i} = -\frac{dp_i}{dt}. \quad (3.7)$$

In the stationary case at constant temperature T and no volume change, this system can be simplified by introducing the (stationary) separable canonical density associated with the Hamiltonian approximation H [2]

$$f(\mathbf{x}, \mathbf{p}) = \frac{1}{Z} \exp(-\beta H(\mathbf{x}, \mathbf{p})) = X(\mathbf{x})P(\mathbf{p}), \quad \beta = \frac{1}{k_B T}, \quad (3.8)$$

where k_B is the Boltzmann constant, and Z is the partition sum. This simplification becomes often a basis for MC computation where we typically keep those conformations where the Boltzmann factors $\exp(-\Delta E/k_B T)$ are larger than the random number from the interval $[0, 1]$. Our interest lies with the general dynamic case where we seek such positions of the atoms that satisfy the least action principle

$$\delta S = 0, \quad S = \int_{t_0}^{t_1} L dt, \quad L = K - P, \quad (3.9)$$

where L is the Lagrangian of the system, δS is the first variation of S , and $[t_0, t_1]$ is the interval of observation. This leads to system (3.1), or under appropriate assumptions, (3.7). In addition to an approximation of the Hamiltonian/Lagrangian of the molecular system, in order to start computation with a molecular modelling procedure, we have to compute good approximations to

the initial conditions of the system, defined by the initial impulse (3.5) and the initial geometry of the system

$$\mathbf{x}(t_0) = \mathbf{x}_0. \quad (3.10)$$

The choice of initial geometries is an important issue for all energy-based molecular modelling procedures, and the role of this choice increases even further with the concept of conformation meta-stability being put forward and efficient dynamics-based clustering methods being developed [2,20]. For successful simulations with MD procedures, which typically cover time-intervals of the order of tens of picoseconds up to a few nanoseconds [9], the quality of (3.10) is critical.

4. ACCOUNTING FOR THE EFFECT OF TOPOLOGICAL CONSTRAINTS AND OTHER CRITERIA IN THE DESIGN OF DISTANCE GEOMETRY ALGORITHMS

When designing new materials and dealing with complex polymer and composite systems, little information is available about the high-dimensional energy surface prior to minimisation by a molecular modelling code. However, it is known that the quality of this minimisation depends critically on the choice of the initial topology which, in turn, can be found using computational optimisation procedures based on

- deterministic (or grid search) methods that cover all areas of the potential energy surface systematically,
- stochastic methods (Monte-Carlo, DGAs, genetic algorithms, etc.), and
- molecular dynamics.

The most powerful (although still computationally very demanding) approach to the solution of many practical problems in materials science is the MD approach. While the MD methods are highly competitive in searching local conformational space, they often fail in generating starting geometries due to their inefficiency in overcoming energy barriers. The first two groups above are not influenced by high energy barriers and can be used for scanning large areas of the potential energy function. However, deterministic methods are becoming less feasible computationally and their efficiency decreases with the technology moving towards designing new complex material structures. Hence, only the stochastic group of methods is left as the most practical one for our purpose of building realistic initial configurations. Such initial configurations can not only then be relaxed effectively with MD, but efficient methodologies have already emerged to calculate such characteristics of the material as specific heat, thermal expansion tensor, compliance tensor, and others directly from thermodynamic response relations, using MD simulations. Therefore, a good initial approximation becomes a key to success of the whole mathematical modelling exercise in this field. In choosing an algorithm for its construction, we have to take into account several important issues.

Since the equilibration time for dense polymers is in many orders of magnitude larger than currently available MD simulation times, we have to be able to build an initial topological structure that resembles the equilibration one with little computational effort. Closely connected to the *computing time* issue is the algorithm *efficiency*. Recall that the mean square end-to-end distance, an important relaxation characteristic of polymers, is approximated according to the Flory hypothesis by

$$\langle r^2 \rangle = NCl^2, \quad (4.1)$$

where C is the characteristic ratio of the polymer with the number N of skeletal bond lengths l . In order to achieve (4.1) with standard approaches, we need a simulation time proportional to $O(N^3)$.

Finally, the algorithm for the construction of initial geometries has to be capable of dealing with *crosslink structures* such as polymer networks (both randomly- and end-crosslinked [21,22]), composite materials such as interpenetrating polymer networks with organic and nonorganic

fillers [23], etc. A good benchmark example for this capability of the algorithms is provided by high deformability networks with essentially complete recovery. Such networks were traditionally attributed to the domain of “rubber-like elasticity” where the crucial assumption, which usually justifies the transition from the microscopic to the macroscopic level, is the *affine deformation* assumption [24]. This assumption states that points of crosslinkage move in such a way that components of the end-to-end length of each chain change in the same ratio as the corresponding dimensions of the bulk rubber during deformation. A network is seen then as a Gaussian (or ideal) “phantom” chain if it may pass freely through their neighbours and themselves and leads to the Gaussian distribution for the end-to-end vector of flexible polymers. More precisely, the number of possible confirmations for walks is assumed to be a function of the distance between the two endpoints taken as the Gaussian distribution for random walks. This means that an ideal “phantom” network assumes that the junctions may fluctuate freely without being restricted by topological constraints due to the existence of neighbouring chains (and hence, topological constraints caused by entanglements of network chains are disregarded in this consideration). The resulting change in entropy under deformation (the entropic elasticity), leads to the elastic shear modulus (the plateau modulus measured by shear experiment)

$$G^0 \propto \frac{k_B T}{N_s}, \quad (4.2)$$

with N_s being the characteristic of the chain length (the strand length between crosslinks). The point here is that for longer chains where $N_s > N_e$ (N_e is a characteristic entanglement length [25]), approximation (4.2) can deviate substantially from the real situation. Shear modulus should be corrected to account not only for the entropic contributions, but also for the trapping contributions due to noncrossability of the chain [21,22] since real networks are not ideal “phantom” chains

$$G^0 = \left[\frac{(\nu - h\mu)}{V} \right] k_B T + T_e G_N^0, \quad (4.3)$$

where V is the total volume of the molecular system, ν is the number of elastically active strands, μ is the number of elastically active cross-linked, h is an empirical parameter $0 \leq h \leq 1$, G_N^0 is the melt plateau modulus for the uncrosslinked system, and T_e is the trapping factor. There is evidence [26] that (4.3) converges to a limiting value around $k_B T/N_e$.

More generally, whenever $N_s > N_e$ the classical Rose-type models [27], where the dynamics of short chains can be well understood with models based on the Langevin dynamics of individual random walks, need correction due to the fact that the effect of topological constraints (entanglement effects) become markedly pronounced. For longer relaxation times (exceeding the Rouse relaxation time $\tau_e \propto N_e^2$), the constraints are supposed to become dominant and the chain moves along its own coarse grained contour. The transition from Rouse to reptation can be characterised by the diffusion coefficient $D(N_s)$, which is drastically decreasing during such a transition

$$D(N_s) \propto N_s^{-1} \implies D(N_s) \propto N_s^{-2}. \quad (4.4)$$

Dramatic qualitative changes are observed in other properties, including viscous, during such a transition.

Since the classical theories disregard topological constraints, and in real networks the strands are impenetrable and highly entangled, numerous attempts have been made to approximate topological constraints by local geometrical constraints [28]. However, only the reptation concept proposed by Edwards and developed by de Gennes takes the noncrossing of the chains explicitly into account. Incidentally, the idea of this concept originated from the investigation of the effect of topological constraints in polymer networks, and the observation that the topological constraints of each chain, as imposed by the surrounding, eventually cause a motion along the polymer’s own coarse-grained contour. “Tube” models are among the simplest examples of approximating

topological constraints by local geometrical constraints [29]. In particular, taking into account that the free ends of a melt would not significantly alter the behaviour for very long chains on scales much smaller than the chain diameter and for intermediate times, it is often assumed that they become confined to a tube with the diameter which coincides with the diameter of a subchain of length N_e , i.e.,

$$\text{diam}_T \propto N_e^{1/2}. \quad (4.5)$$

Although the mathematical theory of entanglements in polymers is at the beginning of its development, it is already well understood that topologically trapped entanglements play an important role, particularly when the average strand length is much longer than the entanglement length of the uncrosslinked melt [21,22]. Having accounted for these entanglements via topological constraints, many physicochemical properties of polymer and composite materials, including elastic constants (Young's modulus, Poisson ratio, compressibility), thermodynamic properties and surface energies, and stress-strain curves for finite deformations [6], can be defined more exactly. This can be done effectively using tools of molecular modelling, provided a good approximation to the initial topology is obtained. Indeed, in designing new materials and modelling complex polymer and composite systems, a key to success is kept by topological interactions in networks, and it is advantageous to start simulations of such interactions with the distance geometry.

5. DISTANCE GEOMETRY OPTIMISATION AS A NATURAL APPROACH TO MODELLING COMPLEX POLYMER SYSTEMS AND COMPOSITES

In recent years classical distance geometry procedures, including metric matrix methods, embedding algorithms, torsional space methods, have provided important tools in molecular modelling applications ranging from conformations of small molecules, protein, and peptides to the time-average conformations of biological macromolecules (helping in the solution of the central problem of molecular biophysics), pharmacophore³ modelling, and drug-receptor docking [12,16,30,31]. In most of these applications the conformational space is searched by generating a large number of independent solutions within the constraints of the model, and each structure is then evaluated for the inclusion in the final ensemble of low-energy conformations, using a force field or other energetic evaluation. Using such classical algorithms, the answer to the question on the existence of a model that satisfies our experimental constraints comes at a cost of dealing with

- (a) $(1/2)(N - 1)N$ nonbonded terms (for structures of N atoms) [4],
- (b) $\sim 2N^2$ units of memory, and
- (c) computationally costly smoothing techniques (i.e., triangle inequalities, see further details in Section 6).

In order to be successful for complex polymer and composite systems, computational efficiency of distance geometry procedures used at the initial stage of molecular modelling should be improved. Since it becomes clear (see, for example, [12] and references therein) that such procedures have potential in the computer-aided molecular design where polymers and composites worthy of actual synthesis can be effectively predicted, in the last few years there has been an increasing interest in the development of such improved procedures. In the process of computer-aided molecular design, distance geometry and molecular dynamics can be considered as complementary methods [30] in a sense that efficient DG procedures can generate complex models rapidly at relatively small computational cost and produce structures that are appropriate starting points for molecular mechanics and dynamics calculations [32]. Although results on MD computations applied to polymers and composites are extensive, applications and analysis of improved DG procedures are still lacking in the literature.

³Set of atoms/groups that are required for bioactivity of a molecule.

Mathematically speaking, the problem of distance geometry applied to atomistic simulation of polymer and composite systems can be formulated as an optimisation problem.

PROBLEM 1. Since measurements (obtained from NMR data or otherwise) may introduce errors into the symmetric matrix (of measured distances) D with nonnegative entries $d_{ij} \geq 0$, $i, j = 1, \dots, n$, $i \neq j$, and such that $d_{ii} = 0$, $i = 1, \dots, n$, in the general case this matrix, called a dissimilarity or predistance matrix, will deviate from the Euclidean matrix [33]. Recall (e.g., [34]) that if for a predistance matrix $D_{\text{Eucl.}} = (\delta_{ij})_{i,j=1,\dots,n}$, there exist n points $\mathbf{x}^l \in \mathbb{R}^r$, $l = 1, \dots, n$, such that

$$\|\mathbf{x}^i - \mathbf{x}^j\|_2 = \delta_{ij}, \quad (i, j) \in \mathcal{S}, \quad (5.1)$$

is satisfied, we call such a matrix the Euclidean distance matrix. In (5.1), δ_{ij} is the given distance between atoms i and j , \mathcal{S} is a subset (say, $i, j = 1, \dots, n$, $n \leq N$) of all atom pairs ($i, j = 1, \dots, N$), $\mathbf{x}^k \in \mathbb{R}^r$, $k = 1, \dots, N$ are the sought-for positions of the atoms in the molecular system, and r is the embedding dimension. Therefore, the first problem of distance geometry can be formulated as follows. Having a predistance matrix D , we have to “complete” it to a Euclidean distance matrix $D_{\text{Eucl.}}$. This problem is equivalent to the problem of finding such a Euclidean distance matrix $D_{\text{Eucl.}}$ from a set of all admissible Euclidean matrices that solves the weighted Euclidean matrix optimisation problem

$$f_1(D) := \|\Omega \circ (D - D_{\text{Eucl.}})\|_F \rightarrow \min, \quad (5.2)$$

where $\Omega = \{\omega_{i,j}\}_{(i,j) \in \mathcal{S}}$ is the weight matrix (see [35] for a special case $\Omega = I$), $\|\cdot\|$ denotes the Frobenius norm ($\|A\|_F = \sqrt{\text{tr} A^T A}$), and \circ denotes the Hadamard-Schur product (by definition, $A \circ B = (a_{ij} b_{ij})_{i,j=1,\dots,n}$; see, for example, [36,37]). In [33,35] (see also references therein) problem (5.2) was considered for the case $n = N$.

PROBLEM 2. In the general case we have to include the possibility of $n < N$, and this leads us to the problem of finding positions of atoms $\mathbf{x}^l \in \mathbb{R}^r$, $l = 1, \dots, N$, such that

$$f_2(\mathbf{x}) = \sum_{(i,j) \in \mathcal{S}} \omega_{i,j} H(d_{ij}) \rightarrow \min, \quad (5.3)$$

where

$$d_{ij} \equiv \|\mathbf{x}^i - \mathbf{x}^j\|_2 \quad \text{and} \quad H(d_{ij}) = (d_{ij}^2 - \delta_{ij}^2)^2, \quad (5.4)$$

with known values of δ_{ij} . Problem (5.3),(5.4) is solved by such $\mathbf{x} \in \mathbb{R}^{r \times N}$ that

$$f_2(\mathbf{x}) = 0. \quad (5.5)$$

We assume that the embedding dimension is 3 and $N \geq 4$ (since $r \leq N - 1$).

In reality, however, the distance matrix is given only approximately, and some elements are defined by lower and upper bounds that are not equal, as they are in (5.1). As a result, the process of converting the distance bonds matrix into Cartesian coordinates becomes much more tedious. Apart from the problem of whether the structure can be determined uniquely from incomplete but exact data given by inequalities (e.g., [38]), this leads to the problem of finding positions $\mathbf{x}^1, \dots, \mathbf{x}^N \in \mathbb{R}^3$ such that inequalities with the given lower, l_{ij} , and upper, u_{ij} , bounds on a set of distance constraints (5.1) with $\delta_{ij} \in [l_{ij}, u_{ij}]$ are satisfied, i.e.,

$$l_{ij} \leq \|\mathbf{x}^i - \mathbf{x}^j\| \leq u_{ij}, \quad (i, j) \in \mathcal{S}. \quad (5.6)$$

Assuming (5.6) and following the approach originally proposed in [16], we are looking for the solution(s) of the following problem:

$$\tilde{f}_2(\mathbf{x}) \rightarrow \min, \quad \mathbf{x} \in \mathbb{R}^{3N}, \quad (5.7)$$

where $\tilde{f}_2(\mathbf{x})$ can be thought of as a “potential” function, although, strictly speaking, we do not invoke here any energetic considerations directly. The choice of the potential function $\tilde{f}_2(\mathbf{x})$ is not unique and can influence the computational effectiveness of the DG algorithm associated with such a choice. An important remark should be made at this point. In a number of mathematical biology applications, it is often sufficient to perform a conformational search for the molecular system in order to isolate only a small number of conformation(s) (usually, by using the minimum energy principle). The situation in the materials science, condensed matter physics, chemical physics, and polymer science might be essentially different where we often need to characterise a compound and its properties by a representative subset of all the accessible conformational states. Additional constraints, coming naturally from the essence of the problem (e.g., for bulk materials such constraints are imposed by dense packing of chains), greatly increase the time required to sample the accessible configuration space of the molecular system [39]. In order to succeed in such situations, current MD simulation procedures should be combined with simple and inexpensive algorithms for finding initial topological structures. Such a combination would open the door for equilibrating new structures that have not been accessible up to now by computer simulation. For this reason, we aim at a *representative sample* of local minimisers of the chosen objective function, and in the next two sections we concentrate on efficient algorithms for the construction of such samples.

6. DISTANCE GEOMETRY ALGORITHMS WITH SMOOTHING OBJECTIVE FUNCTIONS

To ensure stability, most existing energy optimisation codes require a starting model that is adequately close to the real topological structure. Of course, unstable energy-minimisation processes can be controlled by “damping” the refinement (i.e., by decreasing the atomic shift). However, if the energy minimisation is unstable, even with damping, then it is necessary to go back to the starting model and to improve its topological representation. This process is trivial, at least from the theoretical point of view, for structures that can be determined by crystallographic or spectroscopic means (i.e., metals). For complex polymer and composite systems, where the structure may be unknown *a priori* with sufficient accuracy, the construction of a starting model⁴ becomes a nontrivial task of critical importance.

Although distance geometry algorithms have been extensively developed to provide better initial trial coordinates as well as better error refinement, it would be fair to say that such algorithms have been applied predominantly to small molecules. In application to large macromolecules and complex structures such as polymeric systems and composites, the choice of the objective function in the solution of problem (5.6),(5.7) and the procedure for tracing local minimisers of this function deserve much more attention than they were accounted in the past.

Typical choices of the objective function try to make each term responsible for a specific constraint zero if the constraint is satisfied or monotonically increasingly positive as the violation of the constraint increases. This can be achieved for functions taken in the form

$$\tilde{f}_2(\mathbf{x}) = \sum_{(i,j) \in \mathcal{S}} \epsilon_{ij}^k, \quad \text{or} \quad \tilde{f}_2(\mathbf{x}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \epsilon_{ij}^k, \quad \text{where } \epsilon_{ij}^k : \mathbb{R}^n \rightarrow \mathbb{R}. \quad (6.1)$$

Particular examples include

$$\epsilon_{ij}^1 = \max^2 [0, (d_{ij}^2 - u_{ij}^2)] + \max^2 [0, (l_{ij}^2 - d_{ij}^2)], \quad k = 1, \quad (6.2)$$

$$\epsilon_{ij}^2 = \max^2 \left[0, \left(\frac{d_{ij}^2}{u_{ij}^2} - 1 \right) \right] + \max^2 \left[0, \left(\frac{l_{ij}^2}{d_{ij}^2} - 1 \right) \right], \quad k = 2, \quad (6.3)$$

⁴That is, a set of coordinates that defines the approximate geometry of the conformation and configuration of the molecule system of interest, and is input to the minimisation program.

$$\epsilon_{ij}^3 = \max^2 \left[0, \left(\frac{d_{ij}^2}{u_{ij}^2 - 1} \right) \right] + \max^2 \left[0, \left(\frac{2l_{ij}^2}{l_{ij}^2 + d_{ij}^2} - 1 \right) \right], \quad k = 3. \quad (6.4)$$

Such functions are, as a rule, nondifferentiable, and lead to too many minimisers for any realistic structures. It is natural, therefore, to try to transform the original objective function to a smoother function that has a smaller number of local minimisers, and then to trace back those minimisers to the original objective function. In the Wu-Moré method [34] the objective function is also taken in the standard form (6.1) with

$$\epsilon_{ik}^4 = \min^2 \left[\left(\frac{d_{ij}^2}{l_{ij}^2} - 1 \right), 0 \right] + \max^2 \left[\left(\frac{d_{ij}^2}{u_{ij}^2} - 1 \right), 0 \right], \quad k = 4. \quad (6.5)$$

This is reduced to (5.3) for the case where $l_{ij} = u_{ij} = \delta_{ij}$ and $\omega_{ij} = 1$. In this case, the coordinates $\mathbf{x}^1, \dots, \mathbf{x}^N$ from the representative sample we mentioned above should satisfy the following inequalities:

$$\| \mathbf{x}^i - \mathbf{x}^j \| - \delta_{ij} \leq \epsilon, \quad (i, j) \in \mathcal{S}. \quad (6.6)$$

For fixed $\epsilon > 0$ this leads to the problem of finding ϵ -optimal solutions to our distance geometry problem, which is known to be an NP-hard problem [34,38].

With the Gaussian transform $g \in \mathbb{R}^n \rightarrow \mathcal{G}^\lambda(g) \in \mathbb{R}$,

$$\mathcal{G}^\lambda(g) = \frac{1}{\pi^{n/2} \lambda^n} \int_{\mathbb{R}^n} g(\mathbf{y}) \exp \left(-\frac{\|\mathbf{y} - \mathbf{x}\|^2}{\lambda^2} \right) d\mathbf{y}, \quad (6.7)$$

we transform our objective function \tilde{f}_2 defined by (6.1),(6.5) to a smoother function (with controlled degree of smoothing) with fewer local minimisers, then apply an optimisation algorithm to the transformed function, and finally trace the minimisers back to the original function [34]. For functions representable in the form $\tilde{f}_2(\mathbf{x}) = \sum_{(i,j) \in \mathcal{S}} h_{ij}(d_{ij})$ (in our case $h_{ij}(d_{ij}) \equiv \epsilon_{ij}^4$), it was shown in [34] that

$$\mathcal{G}^\lambda(\tilde{f}_2) = \sum_{(i,j) \in \mathcal{S}} \frac{1}{\sqrt{2\pi} d_{ij}} \int_{-\infty}^{+\infty} (d_{ij} + \lambda s) h_{ij}(d_{ij} + \lambda s) \exp \left(-\frac{s^2}{2} \right) ds. \quad (6.8)$$

This integral can be approximated with the Gauss-Hermite approximation

$$\mathcal{G}^\lambda(\tilde{f}_2) \approx \mathcal{G}^{\lambda,q}(\tilde{f}_2) = \sum_{(i,j) \in \mathcal{S}} \frac{1}{d_{ij}} \sum_{k=1}^q \omega_k(d_{ij} + \lambda s_k) h_{ij}(d_{ij} + \lambda s_k), \quad (6.9)$$

where ω_k and s_k , $k = 1, \dots, q$, are weights and nodes, respectively, for the Gaussian quadrature for the integral

$$I(g) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} g(s) \exp \left(-\frac{s^2}{2} \right) ds. \quad (6.10)$$

Then, in order to determine local minimisers of the function (6.9) for different values of λ , $\tilde{f} = \mathcal{G}^{\lambda_k,q}(\tilde{f}_2)$, $k = 0, 1, \dots, p$ ($\lambda_0 > \lambda_1 > \dots > \lambda_p$ and the original function is recovered in the limit $\lambda_p \rightarrow 0^+$), we follow [34] in using the quasi-Newton-type algorithm

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k H_k \nabla \tilde{f}(\mathbf{x}_k) \quad , \quad (6.11)$$

given $\alpha_k > 0$ and the approximation H_k of the inverse Hessian (k is the number of the current iteration). The quality of the result is judged upon the error functions

$$\epsilon_l = \min \left(\frac{d_{ij}}{l_{ij} - 1}, 0 \right), \quad \epsilon_u = \max \left(\frac{d_{ij}}{u_{ij} - 1}, 0 \right), \quad \epsilon_l^u = \max(\epsilon_l, \epsilon_u), \quad (6.12)$$

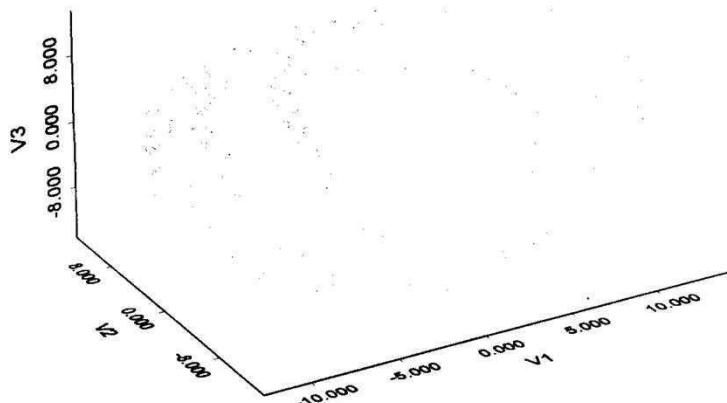


Figure 1. An alkane chain of 250 atoms isolated in space (no periodic boundary conditions; all constraints).

and the corresponding value of the objective function

$$\tilde{f}_2(\mathbf{x}^*) = \sum_{(i,j) \in S} \left\{ \min^2 \left[\left(\frac{(d_{ij}^*)^2}{l_{ij}^2} - 1 \right), 0 \right] + \max^2 \left[\left(\frac{(d_{ij}^*)^2}{u_{ij}^2} - 1 \right), 0 \right] \right\}. \quad (6.13)$$

A typical result of computation with this algorithm, obtained for a short-chain polyethylene structure consisting of 250 backbone carbon atoms (polymer $-\text{CH}_2-\text{CH}_2-$ with three-atom structural unit $-\text{CH}_2-$ [40]), is presented in Figure 1. Such structures correspond to alkane liquids and are important for petrochemicals and lubricants.

In experimenting with this distance geometry algorithm, no bound-smoothing procedures have been implemented. Such smoothing procedures, used in classical distance geometry algorithms, are usually implemented to lower some (unrealistic) default upper bounds in the constraints with the triangle inequality and to raise some (unrealistic) default lower bounds with the inverse triangle inequality. This step is aimed at achieving a better sampling of the conformational space at a high computational cost due to the increased number of effective constraints and the decrease in the regularity properties of the potential function. Despite the elimination of the bound-smoothing step, our experiments with larger chains and increased number of atoms showed that the algorithm described in this section is still computationally expensive. Although very reliable in calculating individual chains isolated in space, the algorithm needs further improvements in application to more complex structures.

Since in materials context the bulk properties are the basis of the economic value of the material, we have to be able to deal with bulk materials simulated by a large collection of molecules. In the next section we describe an efficient distance geometry algorithm capable of dealing with such materials.

7. PERIODIC BOUNDARY CONDITIONS AND MODELLING BULK MATERIALS

The methodology described in the previous section is well suited for the analysis of the behaviour of a single chain isolated in space. However, in the case of bulk polymers and composites where the effect of entangled networks is markedly pronounced (see Section 4), the efficiency of this methodology decreases. We recall that, in order to produce near-equilibrium structures for dense bulk materials, we can apply [41]

- a lattice construction technique,
- the Van der Vegt compression box technique, or
- an experimental density simulation-box technique.

The lattice construction technique is based on

- (a) meshing the 3D space of the experimental polymer volume with regular tetrahedrons,
- (b) placing tetravalent atoms (such as carbon or silicon) in the centre of a regular tetrahedron where their four bonds are perpendicular to the faces of this tetrahedron, and
- (c) constructing the polymer chains by generating random walks following adjacent tetrahedrons.

Although this technique is claimed to scale linearly with the number of atoms [41], its application is limited to several classes of materials only.

The Van der Vegt compression box technique starts with a dilute model polymer and compresses it slowly until the target experimental density is achieved. In fact, the algorithm described in the previous section is relatively easily amenable to this technique, although such amendments come at a high computational cost. In dealing with bulk materials, we can constrain a representative portion of the bulk molecular system to a given shape and then replicate it periodically in the 3D space. The constraint procedure can be built into one of the error functions (6.1),(6.5) by the definition, a new, so-called "shape-function"

$$\tilde{h}(\mathbf{x}) = \begin{cases} \tilde{h} = 0, & \text{if } \mathbf{x} \text{ is on the boundary surface,} \\ \tilde{h} > 0, & \text{if } \mathbf{x} \text{ is outside of boundary surface,} \\ \tilde{h} < 0, & \text{if } \mathbf{x} \text{ is inside the boundary surface.} \end{cases} \quad (7.1)$$

Indeed, to constrain a representative portion of the bulk material, we have to add to the error function the extra term

$$\max(0, \tilde{h}(\mathbf{x})), \quad (7.2)$$

and hence, from a mathematical point of view we reduce a constrained optimisation problem to a series of unconstrained problems.

In the remainder of this paper, we deal with the experimental density simulation-box technique consisting, as the name suggests, of packing chains into a simulation box at the experimental density. Instead of picking up distances randomly (between lower and upper bounds) as we did in the previous section, we pick up 3D coordinates within a chosen periodic cell created by periodic boundary conditions. In this case, we do not compress the structure, as in the classical embedding algorithms, due to its projection from $N - 1$ dimensions to three dimensions. To achieve a higher homogeneity of the resulting structure, we prescribe minimum and maximum bounds on the distance between *each* pair of atoms. These bounds are then used to define bond lengths and bond angles between neighbouring atoms, and the minimum separation of nonbonded atoms. In constructing the code, we assumed that atoms in different molecules can be treated as hard spheres and they have a minimum separation corresponding to a nonoverlap condition which leads to the obvious condition for default lower bounds. The default maximum separation can be essentially arbitrary, but in some cases the assumption that all random points are confined to an "amorphous" cube can lead to "natural" default upper bounds (typically taken as the half of a diagonal of the cubic cell). Atoms in the same molecules separated by many bonds (≥ 3) were treated in the same way as atoms in different molecules (i.e., nonbonded) which leads naturally to the situation where in some case nonbonded neighbours were closer than bonded neighbours. By creating a representative cell of the system and replicating it periodically in all directions in space [4], we begin simulations with random atom coordinates within this "central" cell with periodic boundary conditions. The size of the cell, taken as cubic, was determined by the density of material.

Before proceeding to our distance geometry algorithm, a few additional comments on the size of the cell should be made. First, we recall that when periodic boundary conditions are applied in molecular simulations, it is customary [42] to choose the Lennard-Jones cutoff distance smaller than half the edge of the cubic cell. This will preclude interactions between an atom and its

“image”. Second, we note that following a procedure described, for example, in [42] one can construct a larger cell, when appropriate. The basic idea lies with the fact that the nature of nonbond interactions in MD simulations can be easily linked to the cutoff distance, typically ranging from the Gaussian “phantom” chain limit (where the cutoff is zero⁵) to the Lennard-Jones limit. If one needs to increase the cutoff distance beyond the Lennard-Jones limit, the following procedure could be applied. If, for example, we deal with three molecules, we can use the final configurations corresponding to each of them inside of four cells previously considered for the construction of a new cell consisting of these 12 molecules all together, and then we can use the MSI amorphous cell methodology in order to arrange these molecules at the same temperature and the same density (see further details and references in [42]).

The basic idea of our algorithm is similar to classical procedures [16,30,43] in that it consists of

- (a) minimising the error function by generating random conformers and then
- (b) refining the obtained coordinates against an error (objective) function chosen here as

$$\tilde{f}_2(\mathbf{x}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \epsilon_{ij}^5, \quad \text{where } \epsilon_{ij}^5 = \max(l_{ij} - d_{ij}, d_{ij} - u_{ij}). \quad (7.3)$$

The main difference of the proposed algorithm compared to classical DG procedures is in a number of steps that were taken to achieve savings in computational time for producing better starting conformations, in the sense that subsequent minimisation of the error function more often succeeds in reaching a value close to zero. Compared to classical distance geometry algorithms such as the metric matrix DG, our algorithm is considerably simpler. In fact, we do not base our constructions on computing (three) largest positive eigenvalues and corresponding eigenvectors of the trial matrices which correspond to some conformations in \mathbb{R}^n and converting such matrices into a set of trial coordinates in \mathbb{R}^3 .

The main steps of our algorithm are as follows:

- we select an atom from the given list at random (let it be, say, atom i);
- we calculate the distance between this atom and *each* other atom;
- if the calculated distance falls outside of the prescribed bounds, we calculate the difference, i.e., the bounds error computed according to (7.3), and identify the largest of such bounds error (attained, say, for atom j);
- then we move atom i along the path $i - j$ in order to satisfy the prescribed bounds by sampling the new distance randomly from a quadratic distribution between the minimum and maximum bounds (i.e., using a uniform distribution in \mathbb{R}^3);
- we repeat the above procedure for all atoms until all bounds are satisfied.

The algorithm starts by choosing a random vector $\bar{\mathbf{x}}_0 \in \mathbb{R}^{3N}$ where N is the number of atoms in the structure. In most cases for generating $\bar{\mathbf{x}}_0 = (\bar{\mathbf{x}}_0^1, \dots, \bar{\mathbf{x}}_0^N)$ we apply the following rule:

$$\bar{\mathbf{x}}_0^j = l * a_i * \text{Rand}(), \quad \bar{\mathbf{x}}_0^j \in \mathbb{R}^3, \quad j = 1, \dots, N, \quad (7.4)$$

where a_i is the length of the corresponding edge of the cell ($i = 1, 2, 3$), l is the scaling factor, and `Rand()` is a pseudorandom number generator. To ensure that we do not have a strong dependency of our results on the choice of $\bar{\mathbf{x}}_0$, we experimented with various random number generators (using different computational platforms), including uniform pseudorandom generators of mixed congruent type, the L'Ecuyer random number generators with Bays-Durham shuffle, and added safeguards, etc. (see also [34,44] and references therein).

This algorithm has been tested for a number of bcc and fcc lattices and allows us to quickly obtain initial topologies for structures many times larger than those reported in the previous

⁵Which formally corresponds to excluding the nonbond energy term from (3.3).

section. Typical computational results for long linear alkanes, confined to a periodic cell, i.e., bulk polyethylene (solid or liquid), are presented in Figure 2 for a chain of 1000 backbone atoms. This simple, widely used material seems to be most appropriate for our study. Indeed, this material has the markedly pronounced non-Gaussian character [45,46], and hence, emphasises the significance of the effect of topological constraints discussed in Section 4 of the paper. The link between the atomistic structure of this material and its macroscopic properties is under continuing scrutiny of the researchers [45,47]. By considering long polymer chains, we also address well-known difficulties in simulation procedures dealt with so far by very few researchers [10]. Although in many cases direct calculations of such important characteristics of entangled polymers as bulk and interfacial viscoelastic properties remain well beyond the capability of even the most powerful computers, it is well understood that a reliable methodology for the prediction of such properties requires a good understanding of molecular structure and dynamics over a wide spectrum of length scales, from the level of individual segments to the level of entire chains [10]. Understanding the role of such structures in determining the properties of the material is complicated because of a very large range of conformational states accessible to such chains. Moreover, if the structure contains flexible side groups, one should deal not only with the lack of regularity in the main chain conformation, but also with the different local packing conditions of the groups which under certain temperature conditions behave differently (this is the case for many pendant groups, e.g., methyl groups, dynamics of which could be accessible sometimes with MD simulations by counting the number of state transitions occurring during a fixed simulation time at each temperature [42]). Of course, the algorithm proposed here in its own right cannot resolve all these difficulties of molecular modelling. However, in addition to its simplicity and computational cost effectiveness, it has an appealing feature of the distance geometry approach allowing us to describe any (macro)molecular system as a set of interatomic distances between all pairs of atoms in the (macro)molecule, whose complete conformational space is contained within this set. By proposing an efficient new algorithm on how to create a *sample* of conformers that are consistent with those distances, we enhance further existing computational procedures of the MD simulation leading to the possibility of equilibrating new complex structures.

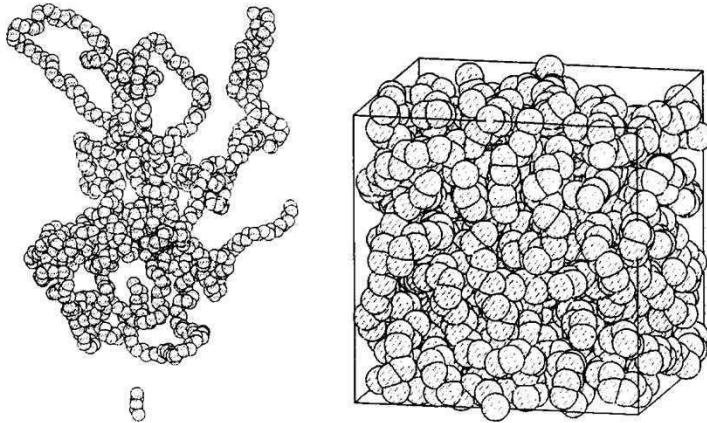


Figure 2. Structure from the representative sample for a 1000-atom alkane chain with periodic boundary conditions (all constraints); unpacked (left) and packed (right).

In this series of experiments, the size of a periodic cell was determined from the prescribed density of polyethylene at $\rho = 0.8 \text{ g/cm}^3$ and temperature 300°K . In particular, in Figure 2 the size of the cubic cell is 32.1430 \AA , while for a 2500 backbone atom structure it was 43.6305 \AA (we computed structures with 2500 backbone carbon atoms and larger systems). The apparent separation of several atoms noticeable in this figure is due to the imposed periodic boundaries in this case. Since for many realistic polymer and composite systems, we have to deal with

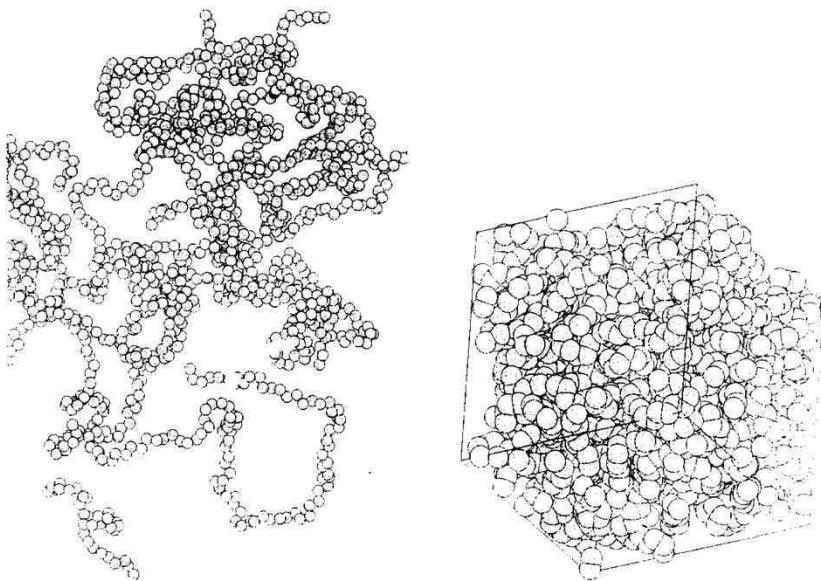


Figure 3. Structure 1 of the representative sample for 10 alkane chains, 150 atoms each (periodic boundary conditions, all constraints); unpacked (left) and packed (right).

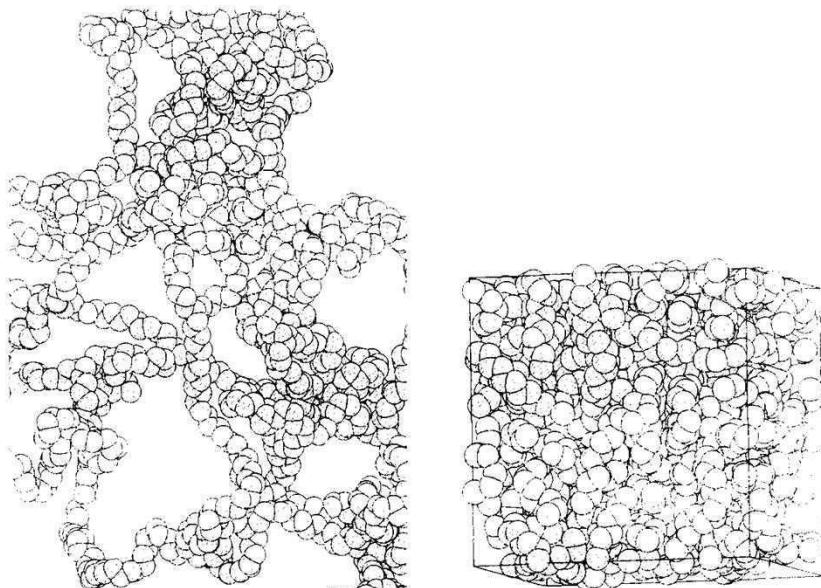


Figure 4. Structure 8 of the representative sample for 10 alkane chains, 150 atoms each (periodic boundary conditions, all constraints); unpacked (left) and packed (right).

many chains confined to one molecular structure, in Figures 3 and 4 we present two typical members of our representative sample for bulk linear polyethylene constructed from 10 chains of 150 backbone carbon atoms each. The cell box in this case was determined as 36.799 Å. It can be seen that all presented structures are sufficiently homogeneous to be relaxed efficiently with molecular dynamics codes, and we shall present the results of the MD refinement of these structures elsewhere.

Finally, in Figure 5 we demonstrate the results of computation for bulk decane (decane molecule is ethylene's derivative with the degree of polymerisation equal to five) $C_{10}H_{22}$ simulating with 30

molecules confined to a periodic cell 21.3331 \AA at density $\rho = 0.73 \text{ g/cm}^3$ and temperature 300°K . Although hydrogen atoms are often omitted from consideration by assuming that their positions can be well approximated and they can be added later by “decorating” the structure (this idea was used in the previous groups of experiments), we took hydrogen atoms into account via packing constraints for this particular case. Since decane can produce a wide spectrum of volatile hydrocarbon products on degradation [48], the effect of inclusion of hydrogen atoms into the model requires further investigation (this might also be useful in the context of polymers with pendant groups [46,49]).

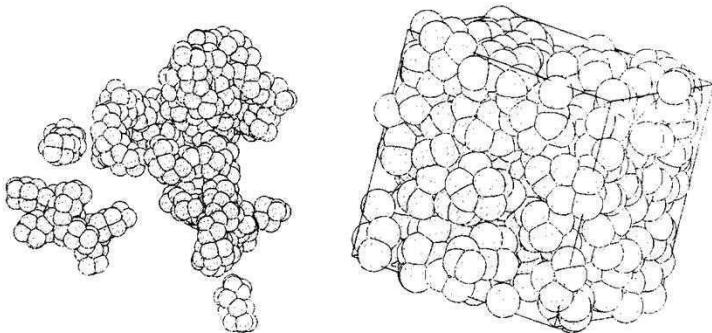


Figure 5. A member of the representative sample for bulk decane simulated by 30 molecules (periodic boundary conditions, all constraints); unpacked (left) and packed (right).

The obtained conformations were analysed using the concepts of pair correlation (i.e., radial distribution functions), dihedral distribution, radius of gyration, cohesive energy, free volume, and Voronoi statistics [50,51]. In particular, the radial distribution function for the conformation shown in Figure 5 confirms the presence of two peaks, one at 1.54 \AA , and the other at 2.52 \AA , as expected.

In modelling complex polymer and composite systems it is important to achieve a high level of vectorisation of computational procedures. Using NEC SX-4, we monitored the level of vectorisation of the algorithm presented in this section in a series of model experiments reported here. In particular, generating a representative sample of 10 structures for a 1500-atom model (10 chains, 150 backbone carbon atoms each) took 1302 sec with the level of vectorisation exceeding 99%. A similar computation for the decane model required 4876 sec with vectorisation 99.5%.

8. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we considered efficient distance geometry algorithms capable of dealing effectively with individual polymer chains isolated in space, as well as with large material structures, including bulk materials. By providing good approximations to topological molecular structures of materials, such algorithms can increase the efficiency of available molecular modelling codes, and therefore, can lead to the possibility of calculating structural and mechanical properties of polymer and composite structures using available methodologies [6]. In this sense, the paper is a contribution to the development of effective computational mathematics methodologies to design, characterise, and optimise polymers and composites before undertaking expensive experimental work.

As a development of the presented work, we are now undertaking a comparison of the performance of several distance geometry algorithms for simulating bulk molecular cross-linked structures using end-to-end distributions, radii of gyration, and radial distribution functions. It is important to get further insight on the role of entanglement when the degree of chemical cross-linking goes from high (where most of entanglement is locked) to relatively small. A related avenue of our current and future research lies with a detailed investigation of the dependence

of distance geometry algorithms on the total numbers of atoms and on chain lengths for bulk materials, including high molecular weight polymers where the concept of entanglement becomes especially important.

Finally, it should be emphasised that not only distance geometry algorithms are complementary to the molecular dynamics methodology, but the molecular dynamics methodology itself is complementary to measurement techniques such as neutron scattering. This makes the field truly interdisciplinary where combined efforts of theoretical and chemical physicists, engineers, molecular biologists, and pure and applied mathematicians are required.

REFERENCES

1. L.E. Nielsen and R.F. Landel, *Mechanical Properties of Polymers and Composites*, Marcel Dekker, New York, (1994).
2. W. Huiszinga *et al.*, From simulation data to conformational ensembles: Structure and dynamics based methods, Preprint SC 98-36, ZIB, (1998).
3. J.D. Honeycutt, A general simulation method for computing conformational properties of single polymer chains, *Computational and Theoretical Polymer Science* **8**, 1-8, (1998).
4. B.R. Gelin, *Molecular Modeling of Polymer Structures and Properties*, Hanser, Munich, (1994).
5. E. Leontidis *et al.*, Monte Carlo algorithms for the atomistic simulation of condensed polymer phases, *J. Chem. Soc. Faraday Trans.* **91**, 2355-2368, (1995).
6. W.A. Goddard III *et al.*, Atomistic simulation of materials, In *Molecular Modeling*, (Edited by M.A. Chaer Nascimento), pp. 65-130, World Scientific, Singapore, (1994).
7. M. Murat and K. Kremer, From many monomers to many polymers, *Journal of Chemical Physics* **108**, 4340-4348, (1998).
8. M. Karplus, Molecular dynamics: Applications to proteins, In *Modelling of Molecular Structures and Properties*, Studies in Physical and Theoretical Chemistry, Vol. 71, (Edited by J.-L. Rivail), pp. 427-461, Elsevier, (1990).
9. P. Comba and T.W. Hambly, *Molecular Modeling of Inorganic Compounds*, VCH, Weinheim, (1995).
10. V.G. Mavrantzas *et al.*, End-bringing Monte Carlo: A fast algorithm for atomistic simulation of condensed phases of long polymer chains, *Macromolecules* **32**, 5072-5096, (1999).
11. W.R. Newell *et al.*, Construction of genetic maps using distance geometry, *Genomics* **30**, 59-70, (1995).
12. D.C. Spellmeyer *et al.*, Conformational analysis using distance geometry methods, *Journal of Molecular Graphics and Modeling* **15**, 18-36, (1997).
13. N. Karasawa and W.A. Goddard III, Force fields, structures, and properties of poly(vinylidene) crystals. *Macromolecules* **25**, 7268-7281, (1992).
14. L.M. Blumenthal, *The Theory and Applications of Distance Geometry*, Chelsea, (1970).
15. T.F. Havel, I.D. Kuntz and G.M. Crippen, The theory and practice of distance geometry, *Bulletin of Mathematical Biology* **45**, 665-720, (1983).
16. G.M. Crippen and T.F. Havel, *Distance Geometry and Molecular Conformation*, John Wiley & Sons, New York, (1988).
17. R.V.N. Melnik, On consistent regularities of control and value functions, *Numer. Funct. Anal. and Optimiz.* **18**, 401-426, (1997).
18. M.R. Nyden and J.W. Gilman, Molecular dynamics simulations of the thermal degradation of nano-confined polypropylene, *Computational and Theoretical Polymer Science* **7**, 191-198, (1997).
19. F. Bornemann, *Homogenization in Time of Singularly Perturbed Mechanical Systems*, Springer-Verlag, Berlin, (1998).
20. P.S. Shenkin and D.Q. McDonald, Cluster analysis of molecular conformations, *Journal of Computational Chemistry* **15**, 899-916, (1994).
21. E.R. Duering, K. Kremer and G.S. Grest, Relaxation of randomly cross-linked polymer melts, *Physical Review Letters* **67**, 3531-3534, (1991).
22. E.R. Duering, K. Kremer and G.S. Grest, Structure and relaxation of end-linked polymer networks, *J. Chem. Phys.* **101**, 8169-8192, (1994).
23. R.F.T. Stepto, Editor, *Polymer Networks: Principles of Their Formation, Structure and Properties*, Blackie Academic & Professional, London, (1998).
24. R.W. Ogden, Elastic deformation of rubberlike solids, In *Mechanics of Solids: The Rodney Hill 60th Anniversary Volume*, (Edited by H.G. Hopkins and M.J. Sewell), pp. 499-537, Pergamon Press, Oxford, (1982).
25. M. Kroger and H. Voigt, On a quantity describing the degree of chain entanglement in linear polymer system, *Macromol. Theory Simul.* **3**, 639-647, (1994).
26. K. Kremer, Numerical studies of polymer networks and gels, *Computational Materials Science* **10**, 168-174, (1998).
27. M. Doi and S.F. Edwards, *The Theory of Polymer Dynamics*, Clarendon Press, Oxford, (1995).
28. R. Everaers and K. Kremer, Test of the foundations of classical rubber elasticity, *Macromolecules* **28**, 7291-7294, (1995).

29. R. Everaers and K. Kremer, Topological interactions in model polymer networks, *Physical Review E* **53**, R37–R40, (1996).
30. J.M. Blaney and J.S. Dixon, Distance geometry in molecular modeling, In *Review in Computational Chemistry*, Vol. V, (Edited by K.B. Lipkowitz and D.B. Boyd), pp. 299–335, VCH Publishers, New York, (1994).
31. M.E. Hodsdon, J.W. Ponder and D.P. Cistola, The NMR solution structure of intestinal fatty acid-binding protein complexed with palmitate: Application of a novel distance geometry algorithm, *Journal of Molecular Biology* **264**, 585–602, (1996).
32. D.N. Theodorou and U.W. Suter, Detailed molecular structure of a vinyl polymer glass, *Macromolecules* **18**, 1467–1478, (1985).
33. A.Y. Alfakih, A. Khandani and H. Wolkowicz, Solving Euclidean distance matrix completion problems via semidefinite programming, *Computational Optimization and Applications* **12**, 13–30, (1999).
34. J.J. Moré and Z. Wu, Global continuation for distance geometry problems, *SIAM J. Optim.* **7**, 814–836, (1997).
35. S. Al-Homidan and R. Fletcher, Hybrid methods for finding the nearest Euclidean distance matrix, In *Recent Advances in Nonsmooth Optimization*, (Edited by D.-Z. Du, L. Qi and R.S. Womersley), pp. 1–17, World Scientific, (1995).
36. R. Bhatia, *Matrix Analysis*, Springer, New York, (1997).
37. S. Barnett, *Matrices: Methods and Applications*, Clarendon Press, Oxford, (1992).
38. B. Hendrickson, The molecule problem: Exploiting structure in global optimisation, *SIAM J. Optimization* **5**, 835–857, (1995).
39. A. Uhlherr, MC conformational sampling of the internal degrees of freedom of chain molecules, *Macromolecules* **33**, 1351–1360, (2000).
40. R.J. Young and P.A. Lovell, *Introduction to Polymers*, Chapman & Hall, London, (1991).
41. J.S. Tokarski *et al.*, Molecular modelling of polymers 17. Simulation and QSPR analyses of transport behavior in amorphous polymeric materials, *Computational and Theoretical Polymer Science* **7**, 199–214, (1997).
42. F. Alvarez *et al.*, Origin of the distribution of potential barriers for methyl group dynamics in glassy polymers: A molecular dynamics simulation in polyisoprene, *Macromolecules* **33**, 8077–8084, (2000).
43. G.M. Crippen, A.S. Smellie and W.W. Richardson, Conformational sampling by a general linearised embedding algorithm, *Journal of Computational Chemistry* **13**, 1262–1274, (1992).
44. J.I. Gail *et al.*, Computer simulation studies of molecular orientation in polyethylene networks, *Macromolecules* **33**, 4966–4971, (2000).
45. D.J.R. Taylor *et al.*, Computer simulation studies of molecular orientation and the stress-optical properties of polyethylene networks, *Macromolecules* **32**, 1978–1989, (1999).
46. R.A. Jones *et al.*, Atomistic modeling of the formation of radial pairs in irradiated amorphous poly(propylene), *Macromolecules* **32**, 8350–8355, (1999).
47. T.M. Madkour and A.M. Barakat, Computer simulation of polymers, *Comput. Theoret. Polymer Science* **7**, 35–47, (1997).
48. N.S. Allen and M. Edge, *Fundamentals of Polymer Degradation and Stabilisation*, Elsevier, London, (1992).
49. F. Alvarez, A. Arbe and J. Colmenero, Methyl group dynamics above the glass transition temperature: A molecular dynamics simulation in polyisoprene, *Chemical Physics* **261**, 47–59, (2000).
50. *Polymer: User Guide*, Parts 1–3, Biosym/MSI, San Diego, (1995).
51. K.S. Schweizer and J.G. Curro, PRISM theory of the structure, thermodynamics, and phase transitions of polymer liquids and alloys, *Advances in Polymer Science* **116**, 319–377, (1994).