Integrating Topology and Geometry for Macro-Molecular Simulations (Extended Abstract, Work in Progress, October 18, 2004)

Edward L. F. Moore, Thomas J. Peters **,
David R. Ferguson, and Neil F. Stewart

University of Connecticut, Storrs, CT 06269, USA,
tpeters@cse.uconn.edu,
http://www.cse.uconn.edu

Abstract. Emerging macro-molecular simulations, such as in supercoiling of DNA and in protein unfolding, have an opportunity to profit from two decades of experience with geometric models within computer-aided geometric design (CAGD). For CAGD, static models are often sufficient, while form and function are inextricably related in biochemistry, resulting in greater attention to critical topological characteristics of these dynamic models. The greater emphasis upon dynamic change in macro-molecular simulations imposes increased demands for faithful integration of topology and geometry, as well as much stricter requirements for computational efficiency. This article presents transitions from the CAGD domain to meet the greater fidelity and performance demands for macro-molecular simulations.

1 Introduction and Motivation

Within geometric modeling, the use of the term 'topology' is often informal, sometimes to the point of being misleading. Hence, it is important to decide which topological characteristics of a model should be preserved or changed, as will be expressed here for connectivity and embedding within the domain of molecular modeling and simulations. As context for these considerations, it has been observed [8] that many problems in modern CAGD systems result from poor integration of computational topology and geometry, where fundamental difficulties are often identified only after analyzing unexpected results from a simulation performed on the model.

** Partial funding for all authors was from NSF grant DMS-0138098. Also, E. L. F. Moore and T. J. Peters were partially supported by NSF grant CCR 0226504 and N. F. Stewart was supported in part by NSERC (Canada). All statements in this publication are the responsibility of the authors, *not* of NSF or of NSERC.

Dagstuhl Seminar Proceedings 04351
Spatial Representation: Discrete vs. Continuous Computational Models
http://drops.dagstuhl.de/opus/volltexte/2005/124

As a motivating example, consider connecting six line segments, represented by ordered pairs of their endpoints as, $(A, B), (C, D), \ldots, (K, L)$, where each alphabetic symbol denotes one point of $\mathbb{R}^3$. However, the lengths and orientations of the line segments can be modified, so that the geometric co-ordinates associated with these vertices may change. Figure 1 shows two very different ways to achieve the same connectivity between vertices which result in different embeddings. The distinctions between these two structures are critical to molecular modeling. Indeed, the two structures are homeomorphic but are not isotopic [9, 19].

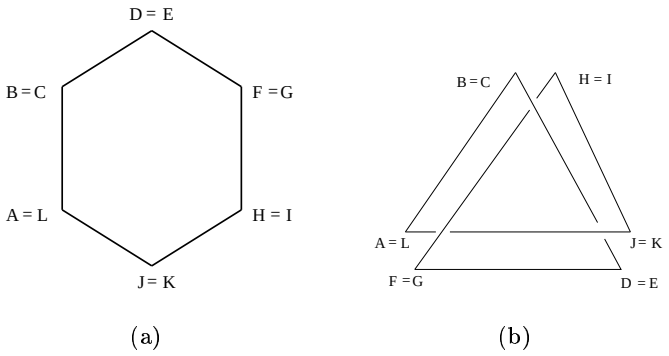


Fig. 1. (a) A planar unknot. (b) A non-planar trefoil.

2 Applications to Molecular Modeling Using 1-complexes

The traditional ball-and-stick model from introductory chemistry serves a valuable role in contemporary molecular modeling, with vertices corresponding to atoms and connecting line segments representing the bonds between atoms. Geometrically, these are, of course, just simplicial 1-complexes [4].

2.1 Preserving topology during molecular dynamics

In some molecular simulations, it is desirable to modify the position of individual atoms, while preserving the simplicial topology of the molecule. The ball-and-stick models and their corresponding 1-complexes will serve as an initial abstraction for considering molecular dynamics. Large molecules may have in excess of ten thousand atoms, so performance considerations are crucial in designing simulation software. Global constraints previously presented [2, 5] may be overly restrictive for such macro-molecules with representations having thousands of vertices. Hence, a new approach which offers more efficient computation

by perturbing only a small number of vertices in a localized region is considered now.

Let v be a vertex of interest from a 1-complex, denoted as M. It is essential to ensure that no new self-intersections are introduced in M during the geometric perturbations of v in order to preserve the isotopy class of M [2, 5]. Prior to perturbing v, find $Star(v)$, which is defined as the union of all segments that have v as one endpoint. This is trivially done by use of the connectivity information given for v. Once $Star(v)$ is determined, the next step is to perturb v to its new position v^* and then find the bounding box that contains the union of $Star(v)$ and $Star(v^*)$. If this bounding box, denoted B, is disjoint from all edges of M that were *not* in $Star(v)$, then the topological equivalence has been preserved. (This follows since the perturbation of v to v^* has compact support that is disjoint from $M - Convex\_Hull(Star(v) \cup Star(v^*))$, as can be shown from a slight extension of a previously presented argument [2], where $Convex\_Hull(Star(v) \cup Star(v^*))$ denotes the convex hull of $Star(v) \cup Star(v^*)$.) If the intersection of B with $M - Star(v)$ is not disjoint, then a more detailed analysis is needed, considering each segment of M that intersects B. Fortunately, identification of all these segments is done efficiently within standard clipping techniques from computer graphics [3]. The above description is captured in the following pseudo-code.

Pseudo-code:
For each $v \in M$ to be moved to a new position v^* :
  Create $Star(v)$; // Uses connectivity data for v //
  Create $Star(v^*)$; // Uses connectivity data for v //
  Create $B = Bounding\_Box(Star(v) \cup Star(v^*))$;
    if $B \cap (M - Star(v)) \neq \emptyset$, return **false**,
      else, return **true**.

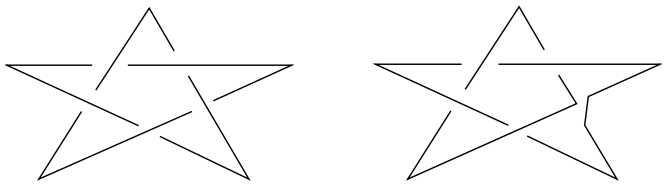
A return of **false** for any vertex means that a more detailed collision detection algorithm must be performed. The more detailed geometric analysis would then be completed by testing whether any of the identified segments are intersected during any part of the perturbation being performed on $Star(v)$. If *all* vertices return **true**, then the selected perturbations are permissible.

2.2 Changes of topology

There are also cases in molecular dynamics where it is critical to change the isotopy equivalence to reflect the chemical properties of a particular molecule. This is a distinct problem area requiring attention to self-intersections produced by perturbations.

The chemical importance of changing both connective and embedding characteristics is shown in Figure 2, as reported elsewhere [21]. Figure 2(a) shows a knotted molecular model. Under the chemical action known as a 'strand switch' [21], the result would be Figure 2(b). Both connectivity and embedding changes

have occurred. In Figure 2(a), five vertices are cyclically connected to form the knot. After the strand switch, three new vertices have been introduced. Of the original vertices, the top-most and the bottom left-most have been newly connected to an added vertex to form a triangle. The other two new vertices have been introduced to avoid intersection with this triangle. Furthermore, even though the triangle and the five sided figure are disjoint, they form interlocking objects that cannot be separated from one another by any action that leaves both of them intact and also does not cause any intermediate intersection between them. This example was chosen to illustrate the critical role for the identification of crossings, where the general problem of identifying crossings is computationally difficult [18].



(a) (b)

Fig. 2. (a) A knotted molecule. (b) A resulting catenane.

3 Extending to 3D Molecular Representations

While the 1-complexes for ball-and-stick models are useful initial abstractions, more interesting and useful images can be created from models of the boundary surfaces of 3D molecules. As is typical in graphics, a static model might be created and then triangulated. Then, for animations of such a triangulated model, the algorithm outlined in the Section 2.1 is extensible to 3 dimensions. This is similar to work already published [17], where the advantage of the algorithm of Section 2.1 is in its performance improvements for local, rather than global, perturbations. These local considerations are expected to be useful in simulating properties of macro-molecules having in excess of ten thousand atoms. In particular, earlier work [2, 17] relied upon the expensive computation of all distances between disjoint vertices, edges and faces. It was specifically noted [2] that these pairwise distance computations could be done prior to any single perturbation. However, when many perturbations are being executed, the overhead associated with repeating all these distance computations for each iteration may be prohibitive, particularly for the dynamic updates expected in simulations. The more localized methods presented here afford significant performance advantages when much of the geometry can be culled by use of neighborhoods of compact

support. This is consonant with the spirit of many algorithms in graphics and animation, where acceptable performance is achieved by significant culling of geometric objects that are irrelevant to a particular operation.

Yet another challenge for surface models of molecules has been presented in a review article [20]. There it is observed that contemporary 3D models of DNA that rely upon tubular models of constant circular cross section for computing stress and strain are overly simplistic, despite their significant value in providing first order dynamic approximations at modest computational expense. This convenience arises since the engineering equations for stress and strain exist in closed form and are easy to implement for efficient performance. A more ambitious goal, though, is to create a more realistic geometric model and subject it to a comprehensive finite element analysis.

From CAGD, splines have proven to be valuable as general geometric representations, though recently alternatives have arisen in form of α-shapes [7], subdivision surfaces [6], and point-cloud data [11]. Future work is to determine the best use of the mathematics underlying these various representations to develop optimal geometric representations for simulations of macro-molecules. We note that our past interest in spline modeling offers the perspective that low degree splines are likely to be preferred for simulations, but that implies that any model of a large molecule will need to be composed of many spline surfaces, typically requiring their pairwise intersection to determine common boundaries where they can be joined, as has been typical in CAGD [14]. This raises the algorithmic issue that such intersections are only approximated [13] , and it will be necessary to ensure that appropriate error bounds on those approximations are preserved during simulation in order to maintain the topological integrity of the model.

There are also promising developments from computational topology that may contribute to this problem. It has been shown [1] that it is possible to change the knot equivalence class when creating a PL approximation of a smooth unknot. That example is repeated here. Figure 3 shows two simple homeomorphic space curves, where the PL curve is an approximation of the smooth curve. However, these curves are not isotopic because they depict different knots: the smooth curve is the unknot, whereas the PL approximation is the figure-8 knot.

In the right half of Figure 3, the z coordinates of some vertices are specifically indicated to emphasize the four knot crossings in $\mathbb{R}^3$ (all other end points have $z = 0$). All end points of the line segments in the PL approximation are also points on the original curve. Since many graphic and simulation algorithms are executed upon PL approximations of complex geometric models, it should be clear that having this knotted curve as an approximant to the original unknot would be undesirable in those circumstances.

Similar pathologies can occur when approximating surfaces. For example, consider a surface created by sweeping a circle of constant radius, along the unknot curve shown in Figure 3 to form a pipe surface [12], where the radius of the curve would be constrained so that the resulting pipe surface would be non-self-intersecting [10]. Then it should be clear that an ill-advised tessellation of this

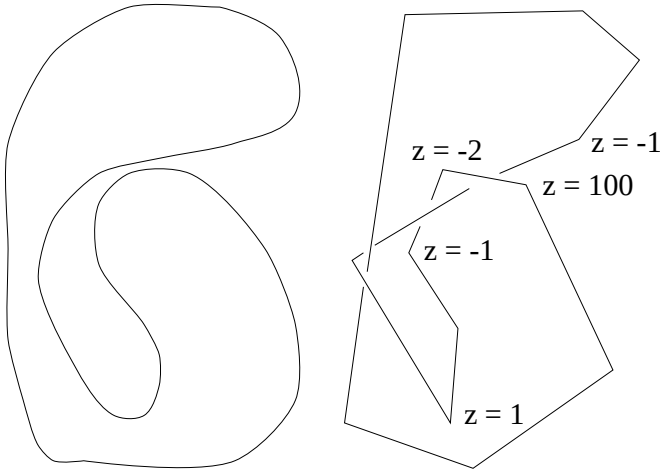


Fig. 3. NONEQUIVALENT KNOTS

swept surface could, itself, become knotted. Recently published constraints [15, 16] establish when approximations of 2-manifolds preserve the original isotopy class. The proofs of these theorems establish a tubular neighborhood in which the approximation is guaranteed to be isotopic to the original surface. For each manifold, its tubular neighborhood will be explored as a constraint within which the isotopy class of dynamically changing molecules can be preserved. When multiple surfaces are combined in a model, the issue of how these constraints should interact is a non-trivial open problem. Such techniques may be useful for macromolecular simulations where preservation of topology is required for dynamically changing molecules.

There is a subtle, but significant, point to observe about this unknot example. It is easy to visually inspect the original curve of Figure 3 to determine that it was the unknot by the change of isotopy class discussed previously. However, it has been shown that even *identifying* the unknot is not an easy algorithmic process [18]. Hence, the strategy when working with molecular models should *not rely* upon being able to give a mathematical specification of the knot type. Indeed, such a notion would likely be quite foreign to most practitioners.

4 Concluding Remarks

Our exploratory efforts to date are promising. The formalisms discussed here can lead to software tools to ensure that if the topology of a model is judged to be correct, then subsequent geometric approximations and dynamic perturbations will *not* alter that topology if the approximations and perturbations are appropriately constrained. Although it is clear that approaches to these molecular problems will rely upon dynamic intersection detection, the specialized techniques

described here avoid many of the well-known performance problems associated with the general geometric intersection detection problem. The challenging task remains to formulate broad algorithms that preserve the efficiencies presented for these known molecular cases, but which are sufficiently extensible to cover the range of input data.

References

1. Amenta, N., Peters, T. J., and Russell, A. C., *Computational topology: ambient isotopic approximation of 2-manifolds*, invited paper, Theoretical Computer Science, 305, 3-15, 2003.
2. Andersson, L-E., S. M. Dorney, Peters, T. J., Stewart, N. F., *Polyhedral perturbations that preserve topological form*, CAGD, 12, 785-799, 1995.
3. Angel, E., Interactive Computer Graphics, Second Edition, Addison-Wesley, Reading, MA, 2000.
4. R. H. Bing, *The Geometric Topology of 3-Manifolds*. American Mathematical Society, Providence, RI, 1983.
5. Andersson, L-E., Peters, T. J., Stewart, N. F., *Equivalence of topological form for curvilinear geometric objects*, International Journal of Computational Geometry and Applications, 10 (6), 609-622, 2000.
6. Desbrun, M., Schröder, Barr, A., *Interactive animation of structured deformable objects*, Proceedings of Graphics Interface, 1 – 8, 1999.
7. Edelsbrunner, H. and Mücke, E. P., *Three-dimensional alpha shapes*, ACM Transactions on Graphics **13**, 43 – 72, 1994.
8. Farouki, R. T., *Closing the gap between CAD model and downstream application*, SIAM News, 32(5), June 1999
9. Flappan, E., When Topology Meets Chemistry, Cambridge University Press, 2000.
10. Maekawa T., Patrikalakis N. M., Sakkalis T., Yu G., *Analysis and applications of pipe surfaces*, Computer Aided Geometric Design, 15, 437-458, 1998
11. Mitra, N. J., Nguyen, A., and Guibas, L., *Estimating surface normals in noisy point cloud data*, to appear, special issue of International Journal of Computational Geometry and Applications.
12. Monge, G., Application de l'Analys à la Geometrie. Bachelier, Paris, 1850.
13. Mow, C., Peters, T. J., Stewart, N. F., *Specifying Useful Error Bounds for Geometry Tools: An Intersector Exemplar*, Computer Aided Geometric Design, **20** (5), 2003, 247–251.
14. Peters, T.J., *et al, Computational topology for regular closed sets (within the I-TANGO project)*, Topology Atlas Invited Contributions vol. 9, no. 1 (2004) 12 pp. `http://at.yorku.ca/t/a/i/c/50.htm`
15. Sakkalis, T. and Peters, T. J., *Ambient isotopic approximations for surface reconstruction and interval solids*, Proceedings of the Eighth ACM Symposium on Solid Modeling and Applications, 176–184, June, 2003.
16. Sakkalis, T., Peters, T. J. and Bisceglio, J., *Application of ambient isotopy to surface approximation and interval solids*, invited paper, CAD, Special Issue, Solid Modeling Theory and Applications, G. Elber and V. Shapiro, (ed.), **36** (11), 1089 – 1100, 2004.
17. Peters, T. J., *et al, Propagating topological tolerances for rapid prototyping*, 1996 ASME IMECE Conference Proceedings, MED-Vol. 4, 487 – 498.

18. Pippenger, N., Hass, J. and Lagarias, J. C., *The computational complexity of knot and link problems*, Journal of the ACM, 46 (1999) 185-211.
19. Rolfsen, D., Knots and Links, Publish or Perish Press, Berkeley, 1976.
20. Schlick, T., *Modeling superhelical DNA: recent analytical & dynamic approaches,* Current Opinion in Structural Biology, 1995, **5:** 245 – 262.
21. Seeman, N., `seemanlab4.chem.nyu.edu`