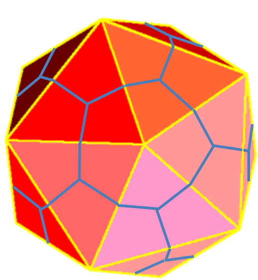
Science in the Workforce

Christoph Hoffmann, Computer Science, Purdue University, November 2007

Over the past decades, the knowledge gathered by science has exploded in volume and has steadily been incorporated into the work place. Increasingly, the problems addressed by industry rely on science and they are expanding in scope. More and more solving them requires interdisciplinary teams. To enable such teams to function effectively, team members must have not only expert knowledge in their respective domains but must also have an operational knowledge and basic understanding of the other disciplines. The role of computing, in particular, is becoming pervasive, necessitating a corresponding change of emphasis in the curriculum.

What methods and algorithms should the science curriculum teach to better prepare students? Given the trend towards large data volumes and complex data interrelationships surfacing in all areas of industry, the repertoire surely should include statistical data analysis and machine learning techniques. While such techniques do not replace the human understanding of the data, the volume of data and its complexity require automated data reduction that distills from a huge base set a sufficiently condensed subset or abstraction that humans can be expected to understand and interpret in a reasonable time span. Because of this ubiquity of large data sets, moreover, emphasis should be placed on the basics of computing with a view towards transferability of techniques to new domains. As an example of such a transfer of knowledge to a new domain, consider how a subject relevant in engineering design could be of relevance to the study of virus assembly and replication.

Today, virtually all artifacts manufactured by industry have been designed first and analyzed by computer. Shape design, in particular, is often parameterized by constraints of size, angle, tangency, and so on. This has the advantage that entire families of parts and assemblies can be designed so that, by varying a few of these constraints, various members of the family are instantiated. To do so, requires constraint solving, a technology in which the geometric configuration and the constraints upon it are translated into a system of nonlinear equations whose solution governs the instantiation of the design. The equation systems easily consist of hundreds of nonlinear equations that must be solved in a few seconds. To do so, many solvers first analyze the constraint system structure to find a decomposition into small subsystems that can be solved separately and whose solutions can be assembled recursively. This is essentially a planning activity, and the analysis phase formulates such a plan. Usually, there are many different viable plans by which to solve the constraint system. In well-constrained problems, all plans ultimately give the same result.



A virus consists of genetic material enclosed in a capsid, a container made from proteins. In 90% of the cases, the capsid has an icosahedral symmetry and the capsid faces are assembled from three essentially identical protein components called monomers. The assembly is rapid and spontaneous. All known viral drugs interfere with the operation of the virus, on entry into a cell, or at replication time of the genetic material and the subsequent protein construction. None interfere with the capsid assembly - yet if the capsid assembly could be blocked, the virus could not reproduce.

It is accepted that the assembly is a geometric process that may begin with two monomers forming an edge, or three monomers forming a face, or five monomers forming a corner. Partial capsid assemblies are known to be rigid, meaning that they are stable attachments with no flexibility. Gluing two cardboard triangles together is not a rigid construction since the triangles can swing relative to each other like a door hinge. Now the planning component of a constraint solver can be used to identify different possibilities of capsid assembly. This is possible because generic rigidity of a partial assembly is analyzed by the planning component of the constraint solver. Then, the possible alternatives can be found, accounting for symmetries, and the energy barriers assessed. In this way, probabilities of particular assembly sequences can be determined.

Acknowledgements: The connection to virus capsid assembly was pointed out by Prof. M. Sitharam, CISE, University of Florida.