Final Report
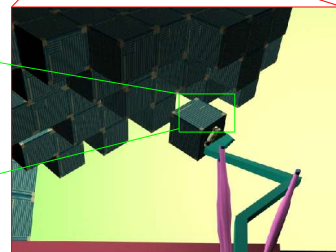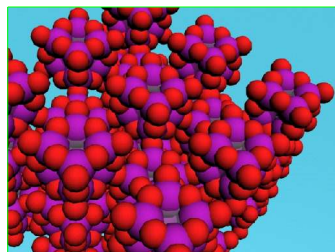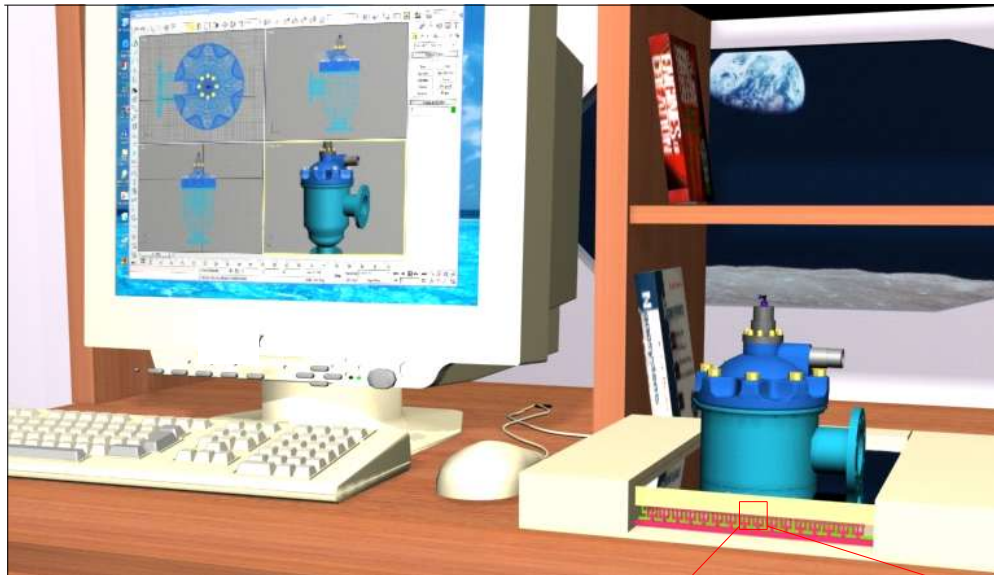
Large-Product General -Purpose Design and Manufacturing Using Nanoscale Modules

NASA Institute for Advanced Concepts

CP-04-01
Phase I Advanced Aeronautical/Space Concept Studies

Principal Investigator: Chris Phoenix
Co-Investigator: Tihamer Toth-Fejel   **May 2, 2005**

1

Cover page: The computer-controlled NIAC desktop nanofactory (top figure) uses nanoscale machinery (lower right) to manufacture a molecularly precise 3-D product—a high performance water filter—out of nanoblocks made using bottom-up techniques, in this case synthesized from 2-layer silsesquioxane deriviatives (lower left).

2

Table of Contents

Summary

The goal of molecular manufacturing is to build engineerable high-performance products of all sizes, rapidly and inexpensively, with nanoscale features and atomic precision. Molecular manufacturing is the only branch of nanotechnology that intends to combine kilogram-scale products, atomic precision, and engineered programmable structure at all scales. It is no coincidence that molecular manufacturing has gone far beyond other branches of nanotechnology in investigating productive nanosystems, because high-performance nanoscale manufacturing systems are the only way that these goals can be achieved.

Building such a product appears to require direct computer control of very small operations. In other words, it needs programmable manufacturing systems capable of acting at the nanoscale. For scalability, the systems themselves must be nanoscale, and many of them must be combined. The need for many systems implies that they must be able to build more of themselves. For convenience, the systems should be combined, and should be able to join their output together into large-scale products. Any system that implements all this functionality can be called a **nanofactory**. The ultimate goal of molecular manufacturing is a high-performance nanofactory and accompanying design capability to make a wide range of high-performance products.

Nanoscale tools promise performance improvements of many orders of magnitude over larger tools due to scaling laws; one of the areas of improvement is manufacturing throughput. Atomically precise fabrication will provide several advantages including natural maintenance of precision. Molecular manufacturing combines these approaches, using nanoscale tools to build precise high-performance products in large amounts and even in large sizes. Productive nanosystems are expected to build exponentially more and rapidly improved manufacturing systems, as well as a wide range of products. The use of productive nanosystems is expected to result in a massive leap in manufacturing capability and product performance.

Architectures for molecular manufacturing have been progressively simplified since the approach was first proposed several decades ago. This report presents further simplifications which point the way to present-day applied development of primitive manufacturing systems. Although the nanoscale is far from fully understood, current knowledge is sufficient to design certain relatively simple structures and machines. These include bearings, motors and actuators, a variety of sensors, several kinds of molecular fasteners, and several implementations of digital logic. Although not all of the designs can be built at this point, it appears possible with present-day capabilities to create a programmable manufacturing system which can build improved systems. Once this is accomplished, the main barrier to advancement will be engineering design.

This report surveys and analyzes several capabilities and architectures of nanoscale systems. In combination, these point the way to manufacturing systems simple enough to be assembled by a scanning probe microscope, but sufficiently functional to build parallel and improved versions. The proposed system could be improved incrementally toward an advanced molecular manufacturing system capable of fabricating integrated products of any size with engineered nanoscale features and atomic precision from small-molecule

5

feedstock. Such products would have extremely high performance due to precision and scaling, and could be quite inexpensive due to the ready availability of manufacturing systems implied by manufacturing systems able to build duplicate systems—exponential manufacturing.

The report is arranged in three parts, "Background," "Proposal," and "Goal," plus four appendices. Because molecular manufacturing is a relatively new and developing field, this report devotes considerable space to providing a context for the new work it presents. This context is recommended reading for anyone who is not already familiar with molecular manufacturing, even if they are familiar with other subfields of nanotechnology. The bulk of the new research done for the NIAC project is presented in Chapters 4, 5, 6, 7, and 9. The compilations, surveys, and overviews in the other chapters and appendices were created for this report.

Part 1 provides background information on molecular manufacturing. Chapter 1, "Incentives," describes the physical reasons that the molecular manufacturing approach will lead to exceptionally powerful manufacturing systems and products. Many of these reasons have been known for many years, but they have not previously been collected and summarized. Chapter 2, "Historical Overview of Molecular Manufacturing," provides a description of the changing approach to molecular manufacturing since its proposal by Feynman. A major theme is the increasing simplicity and power of the proposals. Chapter 3, "Nanoscale Component Fabrication and Assembly," surveys and describes the various forms of chemistry and assembly that can be used to manufacture components of molecular manufacturing systems. It also clarifies important distinctions between various fabrication methodologies associated with molecular manufacturing.

Part 2 presents a proposal developed during the current Phase 1 project for development of an initial programmable nanoscale manufacturing system. Chapter 4, "Molecular Building Blocks," describes several options and methods for creating molecular building blocks—molecular structures a few nanometers in size that can be assembled by simple operations to create functional nanoscale machine systems. Chapter 5, "Early Manufacturing Architectures," proposes an architecture for an extremely primitive nanoscale system that can assemble molecular building blocks into duplicate and improved nanoscale systems, as well as a limited but promising range of products. The system is intended to be simple enough to be created by a combination of synthetic chemistry, self-assembly, and scanning probe microscope operations. But the system is also intended to be complex enough to build parallel systems with higher throughput and improved functionality. Chapter 6, "Early Products," describes three products that might be made by reasonably advanced versions of the system proposed in chapter 5.

Part 3 describes the ultimate goal, and the steps required to get there once the current proposal has been implemented. Chapter 7, "Incremental Improvements," briefly surveys the stages of advancement required to develop the primitive fabrication grid, which must be immersed in a solution of prefabricated feedstock and repositioned by an external actuator, into an advanced nanofactory. Chapter 8, "Covalent Solid Nanosystems," describes some of the structures and systems that might be employed by an advanced nanofactory using advanced materials to build engineered mechanisms. Chapter 9,

6

"Advanced Nanofactory Architecture and Operation," presents a relatively simple and fully scalable architecture for a highly advanced nanofactory, drawing on earlier work by Drexler, Burch, and Phoenix. Chapter 10, "Advanced Product Design and Performance," discusses design effort and methods for nanofactory products, their likely performance and cost, and some possible fields of application including aerospace.

Appendix A, "Nanoscale Properties and Phenomena," describes a variety of nanoscale physical phenomena and presents relatively straightforward nano-mechanical designs that can access or create the phenomena. Appendix B, "Related Concepts, Tools, and Techniques," is a wide-ranging survey of the approaches, tools, concepts, and style of molecular manufacturing. These approaches do not define the field, but collectively they give it shape and direction. Appendix C, "Issues and Constraints," examines the most significant issues that may limit or complicate implementation of molecular manufacturing, as well as several issues that are frequently raised although they do not pose serious problems. Appendix D, "Further Reading," lists several foundational publications in molecular manufacturing, and provides a URL where updated versions of this document will be maintained.

The core of this project is planar assembly: the construction of products by deposition of functional blocks one layer at a time. Planar assembly forms the basis for the primitive manufacturing system proposed in Part 2, as well as the advanced system described in Part 3. The fact that the same architectural approach can be used for systems from the most primitive to the most advanced argues that bootstrapping may be easier than it has been thought. Planar assembly is a new development—less than a year old—in molecular manufacturing theory. It is based on the realization that sub-micron nano-featured blocks are quite convenient for product design as well as manipulation within the nanofactory construction components, and can be deposited quite quickly due to favorable scaling laws. The development of planar assembly theory, combined with recent advances in molecular fabrication and synthesis, indicate that it may be time to start a targeted program to develop molecular manufacturing.

1. Incentives

This section surveys several fundamental reasons why atomically precise nanoscale machines, productive nanosystems, and products made with them, will have far higher performance than today's manufactured systems. Scaling laws indicate that reducing the size of a system will correspondingly increase its performance. The granularity of atoms presents an opportunity for manufacturing systems that create products with no loss of dimensional precision. These factors indicate a definite and substantial advantage of nanoscale machines over larger machines. In addition to strength, the use of advanced materials and sealed dry enclosures will allow manufactured machines to be more efficient than biological machines. Manufacturing of large amounts (grams to tons) of product can be accomplished with large numbers of productive nanosystems; two approaches to scaling up production are presented and contrasted. Large integrated products made with nanoscale machines can preserve most of the performance benefits of the nanoscale.

1.1. Scaling laws

Several important measures of performance improve as machines shrink.[1] Machines can be characterized by simple measures and ratios; for example, a manufacturing system can handle its own mass of materials in a certain number of seconds, and a motor will handle a certain amount of power per unit volume. Broadly speaking, these numbers vary in predictable ways according to the size of the system. These relationships are called "scaling laws."

The speed of linear motion stays roughly constant regardless of scale; this means that as a machine shrinks, its frequency of operation will increase proportionally. Motions cross less distance, taking less time. This in turn implies that the relative throughput of a manufacturing system—the fraction of its own mass that it can process per second—will also increase proportionally to the decrease in size, assuming the components it handles scale along with the machine. Each motion transfers a part proportional to the size of the device, but the motions occur at a higher frequency.

Mass decreases as the cube of the size. This means that, while the relative throughput of a machine shrunk by a factor of 10 will increase by a factor of 10, its absolute throughput will decrease by a factor of 100. To maintain throughput as machines are shrunk to the nanoscale, vast numbers will need to be operated in parallel. If the machines are arranged in a sheet one layer thick, and the area of the sheet does not change as the machines shrink, then the number of machines will increase by the square of the shrinkage; the mass of the sheet will shrink in proportion with its thickness; and the throughput will remain unchanged. Thus, due to the higher relative throughput, the total mass of nanoscale machines can be orders of magnitude less for the same absolute throughput. The rapid decrease in mass with size also means that gravity will usually be unnoticeable

---

[1] Scaling laws are explored in detail in Chapter 2 of *Nanosystems*, available online at http://www.foresight.org/Nanosystems/ch2/chapter2_1.html

in nanoscale machinery, momentum and inertia will be extremely small, and acceleration will be correspondingly high.

The linear relationship between size and relative throughput assumes that the machine handles components scaled to the size of the machine. If the component size is held invariant (e.g. small molecules) as the machine scales, then the dependence of relative throughput on machine size is far stronger. A 10-cm machine such as a scanning probe microscope would take on the rough order of 10^{18} years to manipulate its own mass of individual molecules. But the cube scaling of mass, combined with the linear scaling of operation frequency, implies that a 100-nm machine could manipulate its own mass of molecules in 30 seconds.

Power density varies inversely with machine size. This is because forces are proportional to area and decrease as the square of the size, while volume is proportional to the cube of the size. Speed is constant; power is force times speed; power density is power divided by volume. This implies that nanoscale motors could have power densities on the order of terawatts per cubic meter. (Obviously, it would be difficult to cool large aggregates of such motors.)

In systems that are subject to wear, the lifetime decreases proportionally with the size. This has been a serious problem for MEMS. However, due to atomic granularity and the properties of certain interfaces, atomically precise molecular manufacturing systems need not be subject to incremental wear. Just as replacing an analog system with a digital one replaces analog noise with digital error probabilities, incremental wear is replaced by speed-dependent damage probabilities which typically drop off exponentially with stress on a part.

For a given material, stiffness decreases in proportion to size. This will pose some interesting design challenges for mechanical systems, but the use of stiff covalent solids will help. (Diamond has a Young's modulus of about 1000 GPa.)

1.2. Molecular Fabrication Advantages

There are several benefits of using atomically precise nanomachines to build nanosystems including the machines themselves. Atomic-level design allows much more complex and intricate products to be built. The products can be more precise, which allows higher performance from improved bearing designs. In fact, because covalent bonding is digital, manufacturing precision can be retained through multiple generations of manufacturing without requiring specialized techniques, post-processing, or even sensing (though some sensing is needed in some architectures to prevent accumulation of errors).

9

1.2.1. Product Intricacy

Intricacy of engineered products is a major contributor to their value. But any product can only be as intricate as the information used to build it. This information must be incorporated in the product somehow, for example through machining operations, assembly trajectories, sequences of chemical operations, etc. The use of actuated nanoscale machinery in manufacture allows almost unlimited amounts of information to be incorporated into the product. (See below for further discussion.)

1.2.2. Product Precision

The precision of products has a large effect on their performance. Even at macro scales, more precise surfaces have less wear and friction, so can be driven harder. Precision allows information-processing circuits to be built smaller, using less energy. But these gains are minor compared to the projected gains if products could be made atomically precise.

Atomic precision would allow new ways of thinking about manufacturing variability. Covalent bonds are binary: the bond is either there, or not. Instead of a range of sizes making up a tolerance, parts would be specified chemically, and any deviation would be an error. This means that all non-broken parts of a certain design would be exactly identical. (Thermal noise would not change the average shape of the parts. Isotopes are usually irrelevant, and especially for light elements could be sorted out e.g. with a teragee centrifuge.)

Because of the binary nature of bonds, wear would not be a problem. Friction between hard, smooth, covalent, well-engineered surfaces is also expected to be extremely low (superlubricity has been demonstrated in graphite).

With atomically precise products, nanometer-scale machine components and sub-micron machines become plausible. This presents new opportunities to take advantage of scaling laws. For example, a 50-nm electrostatic motor may be expected to have power density of 10^{15} W/m^3.

In most large-scale manufacturing operations, the product is somewhat less precise than the equipment that produced it. Specialized operations using controlled wear or symmetrical cutting have been developed to increase precision. But in manufacturing at the atomic level, precision is inherent in the binary nature of the bonds. A machine can make a product that is as precise as itself. This makes it easier for machines to build adequate copies; with no precision being lost at each step, an indefinitely long string of copies can be built.

For the purposes of nanoscale machinery, atoms are indivisible, and all atoms of an isotope are identical; different isotopes of the same element are identical for most purposes, and advanced nanosystems will be able to sort isotopes by mass as required. Two components built of atoms covalently bonded in the same arrangement will have identical shapes. (The shapes will be transiently distorted by thermal motion, but the average shapes will be identical.) Construction operations that are sufficiently precise to

cause atoms or molecules to attach in predictable configurations can make perfectly precise products.

The inherent precision of covalent products indicates that manufacturing operations need not involve post-processing steps (analogous perhaps to polishing or lapping) to improve precision. This is equally true regardless of whether the precision in the manufacturing system is achieved by open-loop or closed-loop control. Machines of adequate reliability and repeatability will be able to make perfect products without any observation or feedback. Of course, any system will have a non-zero error rate, but the error will be a random failure, not be the result of imprecision or accumulation of wear. A molecular manufacturing system might make several products with perfect precision, and then a product that is broken due to incorrect molecular structure; it will not make products that accumulate imprecision incrementally. If the products include manufacturing systems, then multiple generations can be made with no loss of precision.

Products built near atomic scale will not be subject to wear in the ordinary sense, because it is impossible to remove a fraction of an atom, and removing an entire atom would constitute breakage, not wear. The forces exerted by nanoscale machinery will typically be too small and distributed to break inter-atomic bonds. Although the granularity of atoms makes perfectly smooth surfaces impossible, smoothly moving bearings can still be implemented; see "Bearings" in the "High Performance Nano and Micro Systems" section.

One problematic consequence of atomic granularity is that machines cannot be designed with the dimensional precision common in macro-scale machining, where a centimeter-scale feature may have a tolerance of a few microns. A nanometer-scale feature in a regular molecular lattice can only have its size specified within a few tenths of a nanometer, though carefully designed modifications of the molecular structure can improve this. The fact that atoms are soft and have smooth interactions reduces the impact of this limitation; what would cause lockup or chatter in a metal machine will simply be a more or less tight fit in an atomic-scale interface.

1.3. Exponential manufacturing

No matter how rapidly a single nanoscale manufacturing system operates, it will not be able to make a gram of product. In order to produce useful quantities of product, vast numbers of machines must be run in parallel. There are several different ways to do this, but all methods involve manufacturing systems that build other manufacturing systems. When the available machinery can double in constant time, throughput increases quite rapidly; only about 60 doublings are required for one 100-nm machine to produce a kilogram of machines.

1.3.1. Scaling up production

Because a billion atoms don't amount to useful quanities of product, any nanoscale fabrication system must be able to build more fabrication systems (possibly via a network of cooperating systems), and the systems must be able to work under automated control. This raises the question of throughput. In today's factories, the question is throughput vs.

cost: will the factory take so much time to produce each piece of product that it adds too much to the product's cost? In nanoscale fabrication, the question is throughput vs. size.

The time for a fabrication system to process its own mass will be a function of two factors: the size of the system, and the rate of processing individual molecules. Carbonic anhydrase is an enzyme that can process 600,000 molecules per second. For machinery, scaling laws indicate that frequency of operation increases linearly with decreasing size, and the operation speed of a 100-nm mechanism should be about 1,000,000 (1E6) operations per second. A 200-nm cube contains approximately 1E9 atoms, so could in theory be built in less than an hour by a single fabrication mechanism. Since ribosomes are only 20-30 nm in diameter, and since engineered molecular features can be on the order of 1 nm in some chemistries, 8,000,000 cubic nanometers (200x200x200 nm) will probably be ample for a set of engineered fabrication machines that could implement the production functionality required for a desktop nanomanufacturing applicance (with external control).

Every factor of 1000 increase in system mass and system throughput requires about ten doublings. It will require approximately 50 doublings to get from sub-micron to gram scale. If each doubling takes an hour, then gram scale fabrication can be achieved in a few days. Megagram (ton) scale fabrication would require only one more day.

1.3.2. Architectures for exponential manufacturing

One approach is for small machines to produce many small products. In this way, the mass of machines increases, but the machine and product size remain at the nanoscale. Large numbers of nanoscale products can be useful as networked computers, as pharmaceuticals, or as components in larger products. Modular robotics may be used to make large products from aggregates of nanoscale robots; proposed architectures include J. Hall's "Utility Fog,"[2] T. Toth-Fejel's "Kinematic Cellular Automata Self-Replicating System,"[3] and NASA's "Autonomous NanoTechnology Swarm" (ANTS)[4]. Toth-Fejel's paper "Legos to the Stars" summarizes earlier work.[5]

Another approach is to build integrated systems containing numerous manufacturing subsystems attached to a framework. This would allow nanoscale building blocks to be joined together, forming a large integrated product. An attractive architecture for a large, advanced, integrated factory is a planar arrangement that produces blocks of approximately micron size from each subsystem, attaching them to a growing product. Scaling analysis indicates that block deposition rate (mass per second, or thickness per second) does not depend on the size of the blocks, since the smaller volume of smaller blocks is compensated by the greater number of producers that can be packed into a plane and the increased operating frequency of each device. A planar architecture is also good for primitive manufacturing systems, since it allows each subsystem to deposit its portion of the product on adjacent sites of a nearby planar substrate; this provides the proper form

---

[2]   http://discuss.foresight.org/~josh/Ufog.html
[3]   http://www.niac.usra.edu/files/studies/final_report/883Toth-Fejel.pdf
[4]   http://ants.gsfc.nasa.gov/
[5]   *The Assembler*, September 1996, http://www.islandone.org/MMSG/

12

for a planar manufacturing system to build a duplicate manufacturing system. As more advanced versions of the system are developed, the planar organization can be retained.

Small manufacturing systems require less internal architecture than integrated systems, but may be difficult to interface with external controllers. It would be difficult to supply high-bandwidth information and power efficiently to free-floating machines. Also, although small products can be useful, large integrated products have a much broader range of applications. Accordingly, this report focuses on integrated manufacturing systems capable of building duplicate or improved systems.

The ribosome implements some of these functions. Its actuation is quite intricate, but its control is rather simple: it is controlled entirely by a stream of instructions in the form of a strand of messenger RNA. The only sensing and processing it does is to detect when a reaction has happened and allow the process to step forward. Even this function is more in the form of a mechanical interlock than an explicit computation. A ribosome includes RNA as well as protein, and so cannot build a full ribosome.

1.4. Information delivery rate

Information delivery to nanoscale products is lacking in today's processes. In bulk processes such as chemistry or nanoparticle production, the ingredients, operation sequences, and a few industrial parameters can be varied. In "top-down" processes like dip-pen nanolithography, an expensive and (comparatively) extremely large machine is used to do just a few operations. Two partial exceptions are biopolymer synthesis and mask lithography. In biopolymer synthesis, monomer addition operations can be repeated indefinitely to make chemicals of (theoretically) unlimited length and arbitrary sequence. This allows as much information as desired to be injected into the process. However, the process is slow, and of limited utility. In mask lithography, information is delivered in the form of an array of light intensities. This allows products with hundreds of millions of sub-micron features to be built for just a few hundred dollars apiece. However, the process has limitations, and the equipment—including the masks—is extremely expensive.

In order to build intricate, precise, heterogeneous nanostructures with individually specified features, enormous amounts of information must be delivered from the computer, through whatever tools are used, to the nanoscale. Existing technologies and proposed nanomanufacturing technologies cannot accomplish this. The few technologies that have a high data rate are not atomically precise because they use beams that cannot be focused tightly enough.

A limitation of self-assembly is that the information must be embodied in the component molecules. For example, DNA structures have been built that contain thousands of bases, but the synthesis process requires many thousands of seconds, corresponding to an information delivery rate of at most a few bytes per second. The manufacture of large unique molecules is expensive, even when stepwise polymerization techniques are used. Mixing multiple prefabricated molecules might help with these problems, but would increase the error rate and the time required for diffusion.

13

Scanning probe microscopes can be atomically precise, but because their actuation systems are large, they move relatively slowly. Semiconductor lithography masks require a long time to manufacture, so although they can deliver information quickly once they are created, the overall data transfer rate from computer to nanoscale is low, and the products are not precise. Electron and ion beams may perform operations corresponding to kilobytes or even megabytes per second of information, but they are not atomically precise tools.

The scaling of operation speed indicates that to embody information in the manufactured product via rapid physical manipulation, it will be necessary to use small actuators. Inkjet printers represent a step in this direction; their print head actuators are a few microns in size, and they can deliver megabytes per second. Furthermore, an inkjet printer can print its weight in ink in about a day. IBM's Millipede, a MEMS-based highly parallel scanning probe microscope array, can modify a substrate rapidly enough to be a serious candidate for computer data storage. Both of these technologies produce only two-dimensional "product," but inkjet technology has been adapted to form three-dimensional products, and scanning probe arrays have been used for dip-pen nanolithography (DPN). Nanoscale actuators, being smaller, will be able to operate faster and handle higher data rates.

To route a high-rate information stream to multiple actuators, to efficiently handle errors locally, and to interpret efficient data formats, not only actuators but digital logic must be shrunk to the nanoscale. Researchers are working on molecular electronics that can perform digital logic from a wide variety of approaches, including carbon buckytube transistors at IBM[6], single atom cobalt-based transistors at Cornell[7], HP's recently announced "crossbar,"[8] and the rotaxane switches at UCLA[9]; this work indicates that transistors can be built from individual molecules and that logic circuits can be built from supramolecular structures.

1.5. Manufacturing Cost

If a nanoscale manufacturing system can build a physical copy of itself, then the cost of developing the first one can be spread over all subsequent copies. If the copying can be done quickly and without human labor, the cost of the systems can approach the cost of the inputs (feedstock and energy). This implies the possibility of a radical decrease in cost per feature, and probably in cost per structure. Even using a crude, primitive,

---

[6] Shalom Wind, Joerg Appenzeller, Richard Martel, Vincent Derycke and Phaedon Avouris, "Vertical Scaling of Carbon Nanotube Field-Effect Transistors Using Top Gate Electrodes", May 20 2002, *Applied Physics Letters*,
http://domino.research.ibm.com/comm/pr.nsf/pages/news.20020520_nanotubes.html

[7] "Vibration-assisted electron tunneling in C140 single-molecule transistors", A. N. Pasupathy, J. Park, C. Chang, A. V. Soldatov, S. Lebedkin, R. C. Bialczak, J. E. Grose, L. A. K. Donev, J. P. Sethna, D. C. Ralph, and Paul. L. McEuen, Nano Lett.; 2005; 5(2) pp 203-207,
http://www.lassp.cornell.edu/lassp_data/mceuen/homepage/pubs.html

[8] http://www.hp.com/hpinfo/newsroom/press/2005/050201a.html

[9] Amar H. Flood, J. Fraser Stoddart, David W. Steuerman, and James R. Heath, "Whence Molecular Electronics?" *Science*, Dec 17 2004: 2055-2056, http://newsroom.ucla.edu/page.asp?RelNum=5718

energy-intensive mechanosynthetic technology, the high strength of defect-free carbon lattice makes it probably competitive with aluminum.[10]

The cost of computation has been dropping dramatically for many decades, and even a primitive molecular manufacturing system can be expected to reduce it still further. The precision and lack of wear in molecular devices implies that complete computer control, with no need for human intervention either during processing or for maintenance, could be applied to all aspects of the manufacturing process.

With today's manufacturing infrastructure, manufacturing capacity is limited by cost and time, and the scarcity of it contributes to product price. But general-purpose manufacturing, in which the same manufacturing system can make either a duplicate system or a range of other products, greatly reduces this source of cost. If scarcity develops to the point that additional manufacturing systems would have higher value than their least expensive product, then existing systems can be used to produce new systems at that cost until the scarcity is relieved. This cost will then be spread over all subsequent products manufactured by that machine—which could easily add up to tens of thousands of times the machine's mass, since scaling laws indicate that a molecular manufacturing system could produce its mass every hour. Thus, the cost of manufacturing capital will become a trivial fraction of the product price.

1.6. New high-performance products

Nanoscale machines can be ganged together to produce large forces or high-throughput operations. The high performance of nanoscale machines implies that a very small volume will usually be sufficient. For example, a large shaft could be driven by a thin layer of motors surrounding it; the driving interface could also function collectively as a bearing for the shaft. (See Chapter 8 on Covalent Solid Nanosystems.) This type of direct interface between sub-micron and large-scale systems implies that the scaling-law performance gains of nanoscale machines can be preserved in large products.

1.6.1. Functional density

The performance of large products will usually be limited by either heat or, more frequently, by design. For example, with motor power density of a megawatt per cubic millimeter, a single layer of motors 100 nm thick would handle 100 watts per square millimeter or 10,000 watts per square centimeter. At reasonable speeds, the motors would be over 99% efficient, producing less than 100 watts per square centimeter of waste heat—comparable to a modern CPU, difficult but not impossible to cool.

Products built with atomic precision, nanoscale features, and high information content can implement capabilities not possible in today's systems. For example, sub-micron medical robots will become possible. High-throughput exponential manufacturing combined with assembly of nano-featured sub-micron blocks into large products implies the ability to build kilogram-scale nanomanufacturing systems capable of turning out kilogram-scale products with near-molar quantities of features and functions.

---

[10]   C. Phoenix, "Design of a Primitive Nanofactory,"
       http://www.jetpress.org/volume13/Nanofactory.htm#s8.6

It goes without saying that using huge numbers of features in an integrated product is not a problem that has been much addressed in today's mechanical engineering. However, this only means that the first products to be built will be simpler and lower performance than they could be. Straightforward simplifications, such as ganging large numbers of high-performance nanoscale motors together to make a large high-performance motor, will allow nanoscale features to be used in engineerable products. There is no immediate need to implement biological levels of complexity; intricate but simple systems can be used. Fault tolerance in massively parallel systems can be handled by simple redundancy at multiple levels.

1.6.2. Advanced materials and structures

Today's manufacturing and machining technologies have made extensive use of metals for structural purposes because they are reasonably strong and easy to machine with crude techniques. However, much of the workability of metals is a result of defects which migrate easily through the material; this allows it to change shape without being destroyed, but also limits its strength. Biological systems are largely dependent for structure on polymer molecules assembled by relatively weak inter-molecular forces.

In addition to being the strongest class of material, covalent solids also have the advantage of stiffness. In sliding interfaces, stiff materials will be able to equilibrate more rapidly than non-stiff materials. This implies lower drag at a given speed. For more information see Chapter 8 on Covalent Solid Nanosystems.

Large structures made of stiff solids tend to be substantially weakened by defects. A small defect causes greatly increased strain in the surrounding material, causing it to fail and propagate the defect. However, anisotropic construction (e.g. fibers) can reduce the size of each component sufficiently that each component will have far less than one defect on average, and the components that do fail can have their load spread over many parallel structures.

Resisting compressive stress over long columns can be hard to do efficiently, because structures may fail by buckling instead of crushing. Adding extra material to resist buckling is inefficient. Trusses can be a lot more efficient of material, but may not be economical if each added component increases the cost. With fully automated bottom-up manufacturing, additional structural complexity will not increase the manufacturing cost, permitting more efficient structures to be used.

1.6.3. Access to the nanoscale

A general-purpose manufacturing system which can make nanoscale structures and machines would greatly facilitate the manufacture of several new types of products, including medical devices, several advanced types of computers, and products using physical phenomena that are currently difficult to take advantage of.

Cells contain molecular and micron-scale components that are currently accessible only with drugs or painstaking manipulation under a microscope. Nanometer-scale devices could be built into sub-micron systems capable of manipulating subcellular structures and interacting with molecules. This would speed medical research as well as making some

16

forms of treatment easier. Energy-efficient designs would be able to intervene in each cell of the body, inserting artificial chromosomes, removing protein tangles, or whatever else was needed for health. Tiny full-featured surgical robots could be the next step from laparoscopy, making surgery as minimally invasive as a hypodermic injection. Tiny sensor arrays could be implanted in the body with less discomfort and fewer complications for long-term monitoring. Computer/neuron interaction could be implemented on a large scale in vivo. These are only a few of the medical applications that could be enabled by even moderately advanced molecular manufacturing.

New types of computers on the drawing board include plasmonic circuits and quantum computers. Both of these require extreme precision, difficult or impossible to achieve with today's micro-fabrication technology. Building engineered, stiff, precise structures from the bottom up would facilitate construction of the required components. Also, as in many other applications, the ability to build a new design rapidly would reduce the cost and delay required for research.

As nanoscale technologies develop, new and useful phenomena are being discovered. New applications of photonics and electronics will be enabled by new kinds of nanoscale structures. Although these are often material-dependent, parallel nanoscale robotic systems will be useful in many cases to help organize the materials into structures with long-range precision.

17

2. Historical Overview of Molecular Manufacturing

The idea of using nanoscale machines to build more nanoscale machines as well as products was proposed more than four decades ago. Early proposals were, not surprisingly, incomplete and problematic. Beginning with Eric Drexler's work in the 1980's, manufacturing system architectures began to be investigated in more detail. Drexler tied together the concepts of 1) atomically precise manufacturing via nanoscale machinery; 2) nanoscale fabrication building additional nanoscale manufacturing systems with exponential growth; and 3) large heterogeneous products being made from the combined output of these systems. Several different architectures have been proposed to build large products. Since then, the major trend has been one of simplification: proposed architectures have become steadily more elegant. Accompanying the simplification of the manufacturing process has come a series of increasingly simple and practical suggestions for making large products by using many nanoscale manufacturing systems working in parallel.

2.1. Early proposals

The first proposal for molecular manufacturing is often credited to Richard Feynman. In his 1959 talk, "There's Plenty of Room At the Bottom,"[11] Feynman proposed that structures and machines could be built by machines small enough to manipulate individual atoms. He recognized the benefit of precision due to atomic granularity and the ability to control many nanoscale manufacturing systems in parallel, but did not discuss scaling law advantages.

Feynman proposed using large machines to build smaller machines through familiar machining operations, repeating the process until nanoscale machines were produced. This method is problematic for several reasons. New designs and materials would have to be used at each of several smaller scales, as stiffness changed due to scaling. This would have required the development of new manufacturing techniques at each of several scales. Some techniques that depend on heat could not be used. The source of feedstock at smaller scales is not immediately obvious; a large-scale machine shop starts with blanks of appropriate shapes, but there is no source for micron-scale blanks.

In 1981, Eric Drexler published a paper in the Proceedings of the National Academy of Sciences (PNAS).[12] He proposed that nanoscale machines be built out of engineered proteins and related molecules. This proposal had the practical advantage that proteins could be built with existing biological systems, and the conceptual advantage that biology provides an existence proof of the feasibility of building protein-based scalable manufacturing systems.

---

[11] http://www.zyvex.com/nanotech/feynman.html

[12] K. Eric Drexler (1981), "Molecular Engineering: An Approach to the Development of General Capabilities for Molecular Manipulation." PNAS 78: 5275-5278. http://www.imm.org/PNAS.html

In 1986, Drexler published the popular book *Engines of Creation*,[13] which introduced the term "nanotechnology" to the public. He proposed that large, high-performance products could be built by aggregations of "assemblers" floating in a tank of feedstock. The assemblers would link with each other and with a communication point, and then each build their piece of the product in parallel. Algorithms for control, communication, and navigation were not specified.

2.1.1. Scaffold Filling

In *Engines of Creation*, Drexler proposed that small, self-contained fabrication systems ("assemblers"), with onboard computers and the ability to move and navigate, would join to form a large-scale scaffold. Once in position, they would each build a small piece of the product. During construction, the product would be intepenetrated by the scaffold of assemblers. When finished, the assemblers would exit the product. A feedstock-rich fluid would provide power and materials for the assemblers. Instructions would be transmitted between them once they linked up into the scaffold.

This method uses needlessly complex and inefficient devices. It would be difficult to fit all the required functionality into sufficiently small assemblers; if the assemblers are too big, then each one has to build too much of the product, and overall productivity drops.

2.1.2. Diamondoid and nanofactories

In 1992, Drexler published the technical book *Nanosystems*,[14] an extension of his MIT Ph.D thesis. In this work, he moved away from biologically-inspired molecules, focusing instead on highly crosslinked carbon-backbone covalent solids ("diamondoid"). Instead of using enzymes in solvent to make the molecules, he proposed a new kind of chemistry called machine-phase or vacuum-phase chemistry, in which the position of all reactive molecules would be mechanically controlled. See Chapter 3 on Nanoscale Component Fabrication and Assembly.

In the same book, Drexler introduced the idea of an integrated kilogram-scale manufacturing system in which the nanoscale fabrication machines would be fastened to a framework instead of floating in a solvent. Instead of adding their contribution directly to the adjacent product structure, the nanoscale fabricators would produce molecular components which would be transported to another location for assembly into larger sub-products, which would again be transported to a larger assembly area. This "convergent assembly" approach set the direction for manufacturing system design for the next twelve years.

2.1.3. Modular robotics

An alternative approach to building a single integrated large product is to build a large number of small similar robots that are capable of adjusting their relative position in order to form programmable shapes and structures. Several people in the 1990's proposed

---

[13]   http://foresight.org/EOC/
[14]   Contents and sample chapters at http://www.foresight.org/Nanosystems/index.html and brief summary at http://crnano.org/5min.htm

that modular nanoscale robots could work together to implement larger structures and behaviors.[15] From a manufacturing point of view, the advantage of modular robotics is that they can be made individually by rather small manufacturing systems. If a small system can produce multiple small systems as suggested by Merkle[16] and further explored by Merkle and Freitas,[17] but designs do not yet exist for building larger nanofactories, then modular robotics may be an early way to make large general-purpose functional structures.

J. Hall calculated the requirements for "Utility Fog,"[18] a truss in which each node would be a separate robot with numerous extensible legs. Large shapes could deform and/or flow to reach new configurations. The mechanical performance of the system is not spectacular—the strength per volume and per density is comparable to wood or soft plastic.

Bishop[19] and Michael independently developed the concept of cubes that would fasten at their faces and slide past each other to form large-scale shapes. However, Michael's cubes were designed in multiple sizes, and were intended for top-down construction rather than nano-fabricator construction.
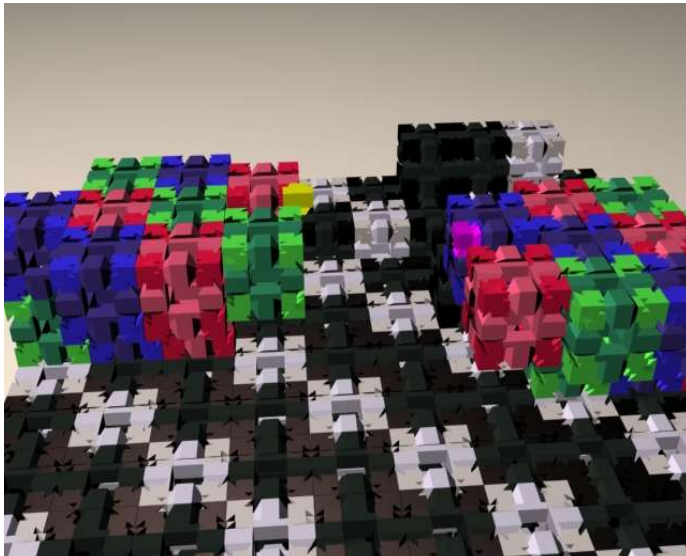
Figure 2.1 *A view of Toth-Fejel's Kinematic Cellular Automata Self-Replicating System.*

T. Toth-Fejel's paper "Legos to the Stars" summarizes earlier work.[20] Toth-Fejel's "Kinematic Cellular Automata Self-Replicating System"[21] (See Fig. 2.1) and NASA's "Autonomous NanoTechnology Swarm" (ANTS)[22] represent more recent work. The NIAC-funded work on Kinematic Cellular Automata by Toth-Fejel examined the

---

[15] See also the extensive bibliography at http://www.foresight.org/Nanomedicine/Swarm.html
[16] http://www.zyvex.com/nanotech/casing.html
[17] http://molecularassembler.com/
[18]   http://discuss.foresight.org/~josh/Ufog.html
[19] http://www.iase.cc/html/active_cells.htm
[20]   *The Assembler*, September 1996, http://www.islandone.org/MMSG/
[21]   http://www.niac.usra.edu/files/studies/final_report/883Toth-Fejel.pdf
[22] *The Assembler*, September 1996, http://www.islandone.org/MMSG/

application of this paradigm for self-replication and made some compelling discoveries, most notably that the system could be less complex than a Pentium.

Mark Yim and others have built physical instantiations of modular robots with varying degrees of success. The biggest problem is that despite their apparent simplicity, they are still so complex that the MTBF is small.

2.1.4. Convergent Assembly overview

In Nanosystems,[23] Drexler published a suggestion of Merkle's: that small parts and blocks could be fastened together more or less permanently within a factory. In convergent assembly the manufacturing system surrounds the product, just as in today's factories. (In contrast, in the scaffold method, the product grows around the manufacturing system.)

Nanosystems 9.7 described ways that nanoscale parts could be attached, including the suggestion that parts could be fastened together using alignment pegs and adhesive surfaces. Even passivated surfaces will attach due to surface forces, and unterminated surfaces will bond; it was suggested that two unterminated diamond (110) surfaces might be made to bond seamlessly to each other. A variety of mechanical fastening strategies could also be used. *Nanosystems* provided projections of factory mass and performance for handling parts at the whole range of scales from nanometer to centimeter, and concluded that a 1-kg factory could manufacture 1 kg per hour of product.

Merkle wrote several papers in the 1990's further developing the convergent assembly concept and simplifying the factory layout.[24] Phoenix published a lengthy paper in 2003 developing a detailed architecture for a convergent-assembly system: physical layout, performance, reliability, control, power, and mechanical and functional fastening of blocks.[25]

2.1.5. Planar assembly overview

In 2004, a collaboration between John Burch and Eric Drexler resulted in a new architecture for a nanofactory.[26] The factory is arranged as a thin plane, with feedstock supplied to one side and product produced at the other side. Instead of multiple stages of assembly, creating parts at multiple scales that need to be handled and joined into larger systems, the planar assembly architecture applies sub-micron building blocks directly to a planar face of the product, building it layer by layer. Because the time required to manipulate and deposit a block scales with the size of the machinery performing the operation, the rate of deposition of the product is not dependent on the size of the block as long as the deposition machinery can scale with the block. It should be possible to deposit products at a rate of several meters per hour.

[23] K. E. Drexler (1992). *Nanosystems: Molecular Machinery, Manufacturing, and Computation.* New York: John Wiley. See http://www.foresight.org/Nanosystems/toc.html

[24] R. Merkle (1997). "Convergent Assembly." *Nanotechnology* 8, No. 1, pp. 18-22. See http://www.zyvex.com/nanotech/convergent.html

[25] C. Phoenix, "Design of a Primitive Nanofactory," http://www.jetpress.org/volume13/Nanofactory.htm#s8.6

[26] Not yet published academically. See http://lizardfire.com/nanofactorySS/index.htm for illustrations and animations of the architecture.

The planar assembly architecture is preferable to convergent assembly for several reasons. It is simpler, not requiring multiple scales of machinery for part placement. It is smaller and scales better; doubling the capacity simply requires doubling the area, and the factory scales to any size. Most importantly, the planar layout lends itself to high-performance systems making advanced products, but it is also suitable for simple machinery doing simple operations.

The following two sections discuss convergent assembly and planar assembly in more detail, contrasting the advantages of each approach.

2.2. Convergent Assembly

Convergent assembly is basically an extension of the assembly-line concept. Small parts are manufactured, then joined together to make larger parts. But whereas a modern factory might have four levels of assembly line from washers to washing machines, a nanofactory might have twenty levels—all the levels would need to be completely automated.

2.2.1. Method

There is a range of approaches to part design. At one extreme, parts are made in unique shapes and proportions, and then fastened together. At the other extreme, modules are all made in the form of cubes or blocks with uniform shape and size, each composed of smaller cubical modules down to a level where a block can be built efficiently by a single fabrication system. In between are functional modules with reduced parts type count, and parameterized parts.

2.2.2. Advantages

If the fastening system does not require much manipulation, if scaling holds, and if part manipulation at the macro level doesn't take a long time, then convergent assembly can be quite fast. Each part may have to travel only a meter or two through the factory, and twenty stages of joining are enough to go from sub-micron to meter-scale. If the parts can simply be pushed together in order to join them (using ridge joints, a strong mechanical zero-insertion-force fastening system proposed by Phoenix; see Chapter 8), then the entire convergent assembly process may require only a few seconds.

Conversely, if robotics are provided to perform intricate manipulations, then a wider range of parts can be joined; this should still be fast, in most reasonable cases.

22
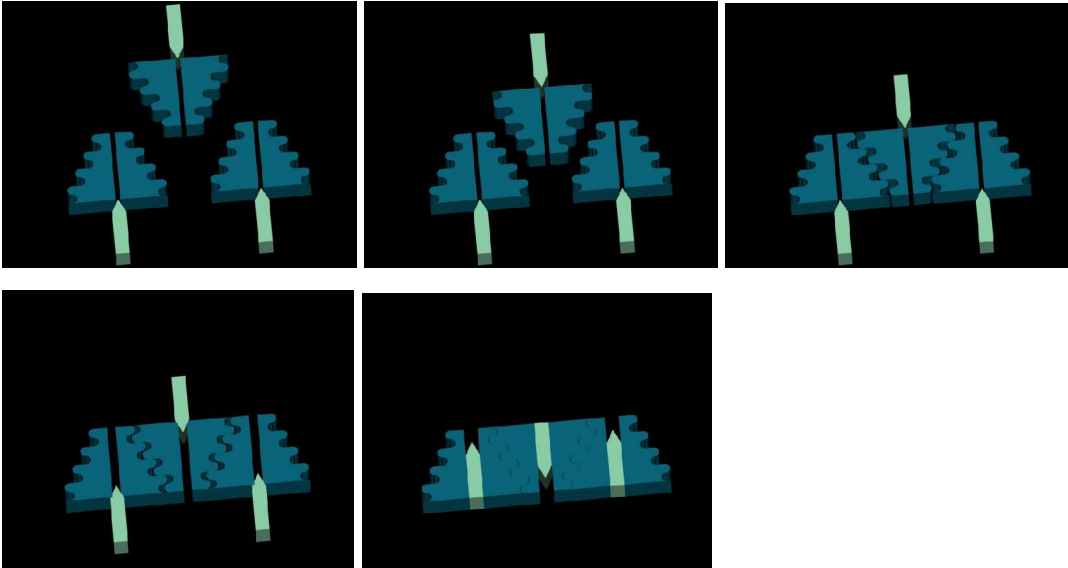
Figure 2.2 *Ridge joint actuation sequence*

2.2.3. Constraints and Problems

Convergent assembly requires robotics in a wide range of scales. It also needs a large volume of space for the growing parts to move through. For example, in the Phoenix nanofactory architecture, the fabricators and computers occupy about 0.5% of the total volume, and over 99% is empty space.

Not only the robots, but their actions, must be designed at multiple scales. If the action consists only of pushing cubes together, this does not appear too difficult. But if more complicated operations are needed, such as joining odd-shaped pieces using small manipulated fasteners, then the problem may quickly become many times more difficult than designing a macro-scale fully automated ("lights out") factory. Each product—and each part of each product—would require its own sequence of operations, delaying the development of truly general-purpose manufacturing.

In the modular or block approach, the assembly robotics are simple, but the design of the cubes is constrained. There are two problems. The first is that every large-scale component, mechanism, or device must be decomposable into cubes. If the component must be placed at a certain location relative to block boundaries, that places significant constraints on the product design. So unless its position within the product is constrained to fit neatly on a multiscale grid, the boundaries of large-scale cubes may cut the component at any position; this would mean that the design must be insensitive to being cleaved at any point.

The problem gets worse if large-scale blocks require large-scale fasteners, because these fasteners would also appear at semi-random places in the component volume. Another problem is that the faces of the blocks must be fairly stiff and reasonably complete (self-supporting) so that all blocks can be handled and joined into larger blocks. This implies that products, as manufactured, will be quite dense, unless voids can be filled with

23

collapsible fine-grained scaffold. Dense products are not a good fit for the high strength promised by some chemistries (it would be wasteful of material); sparse products are not a good fit for a system where the factory encloses the product. In order to fulfill this constraint, products would probably need to inflate or unfold after manufacture. This requires additional design.

2.3. Planar assembly

A few paragraphs of Nanosystems (Sec. 14.3.1a) suggested adding small modules to a planar working face. This approach was not pursued further until Drexler and Burch began to collaborate on an animated illustration of nanofactory function.[27] This approach has several advantages over convergent assembly, and appears to have few disadvantages.

The idea of planar assembly is to take small modules, all roughly the same size, and attach them to a planar work surface, adding to the working plane of the product under construction. In some ways, this is similar to the concept of 3D inkjet-style prototyping, except that there are billions of inkjets, and instead of ink droplets, each particle is molecularly precise and can be full of intricate machinery. Also, instead of being sprayed, they would be transported to the workpiece in precise and controlled trajectories, and the workpiece (including any subpieces) can be gripped at the growing face.

2.3.1. Method

Small modules supplied by any of a variety of fabrication technologies would be delivered to the assembly plane. The modules would all be of a size to be handled by a single scale of robotic placement machinery. This machinery would attach them to the face of a product being extruded from the assembly plane. The newly attached modules would be held in place until yet newer modules were attached. Thus, the entire face under construction serves as a "handle" for the growing product. If blocks are placed face-first, they will form tight parallel-walled holes, making it hard to place additional blocks; but if the blocks are placed corner-first, they will form pyramid-shaped holes for subsequent blocks to be placed into. Depending on fastening method, this may increase tolerance of imprecision and positional variance in placement.

The speed of this method is counterintuitive; one would expect that the speed of extrusion would decrease as the module size decreased. But in fact, the speed remains constant. For every factor of module size decrease, the number of placement mechanisms that can fit in an area increases as the square of that factor, and the operation speed increases by the same factor. These balance the factor-cubed increase in number of modules to be placed. This analysis breaks down if the modules are made small enough that the placement mechanism cannot scale down along with the modules. However, sub-micron kinematic systems are already being built via both MEMS and biochemistry, and robotics built by molecular manufacturing should be better. This indicates that sub-micron modules can be handled by future systems.

---

[27] John Burch and. K. Eric Drexler, "Productive Nanosystems: From Molecules to Superproducts." See http://lizardfire.com/nanofactorySS/index.htm for illustrations and animations of the architecture.

2.3.2. Advantages

This approach requires only one level of modularity from nanosystems to human-scale products, so it is simpler to design. Blocks (modules) built by a single fabrication system can be as complex as that system can be programmed to produce. Whether the feedstock producing system uses direct covalent deposition or guided self-assembly to build the nanoblocks, the programmable feature size will be sub-nanometer to a few nanometers. Since a single fabrication system can produce blocks larger than 100 nanometers, a fair amount of complexity (several motors and linkages, a sensor array, or a small CPU) could be included in a single module.

Programmable, or at least parameterized, (or at worst case, limited-type) modules would then be aggregated into large systems and "smart materials." Because of the molecular precision of the nanoblocks, and because of the inter-nanoblock connection, these large-scale and multi-scale components could be designed without having to worry about large-scale divisions and fasteners, which are a significant issue in the convergent assembly approach (and also in contemporary manufacturing).

Support of large structures will be much easier in planar assembly than in convergent assembly. In simplistic block-based convergent assembly, each structure (or cleaved subpart thereof) must be embedded in a block. This makes it difficult to build a long thin structure that is not supported along each segment of its length, at least by scaffolding.

In planar assembly, such a structure can be extruded and held at the base even if it is not held anywhere else along its length. The only constraint is the strength of the holding mechanism vs. the forces (vibration and gravity) acting on the system; these forces are proportional to the cube of size, and rapidly become negligible at smaller scales. In addition, the part that must be positioned most precisely—the assembly plane—is also the part that is held. Positional variance at the end of floppy structures usually will not matter, since nothing is being done there; in the rare cases where it is a problem, collapsible scaffords or guy wires can be used. (The temporary scaffolds used in 3D prototyping must be removed after manufacture; this is not the best design for a fully automated system.)

The above analysis indicates that large open-work structures can be built with the planar aggregation method. Unfolding becomes much less of an issue when the product is allowed to have major gaps and dangling structures. The only limit on this method is that extrusion speed is not improved by sparse structures, so low-density structures will take longer to build than with convergent assembly.

Planar assembly of sub-micron blocks places a major stage of product assembly in a very convenient realm of physics. The mass of individual blocks is not high enough to make inertia, gravity, or vibration a serious problem. (The mass of a one-micron cube is about a picogram, and under 100 G acceleration would experience a nanoNewton of force. This is comparable to the force required to detach 1 square nanometer of van der Waals adhesion (tensile strength 1 GPa, Nanosystems 9.7.1). Resonant frequencies will be on the order of MHz, which is easy to isolate/damp.) Stiffness, which scales adversely with size, is significantly better than at the nanoscale. Surface forces are also not a problem. They are large enough to be convenient for handling—instead of grippers, just put things in place and they will stick—but small enough that surfaces can easily be separated by machinery.

It should be noted that the problem posed by surface forces in MEMS manipulation is greatly exacerbated by the crudity of surfaces and actuation in current technology. Nanometer-scale actuators can easily modulate or supplement surface forces to allow convenient attachment and release.

Sub-micron blocks are large enough to contain thousands or even millions of features: dozens to thousands of moving parts. But they are small enough to be built directly out of molecules, benefiting from the inherent precision of this approach as well as nanoscale properties including superlubricity. If blocks can be assembled from smaller parts, then block fabrication speed can improve.

Centimeter-scale products can benefit from the ability to directly build large-scale structures, as well as the fine-grained nature of the building blocks (note that a typical human cell is 10,000-20,000 nm wide). For most purposes, the building blocks can be thought of as a continuous smooth material. Blocks can made in partial shapes so that they can be placed to make the surfaces smoother—molecularly smooth, except perhaps for joints and crystal atomic layer steps.
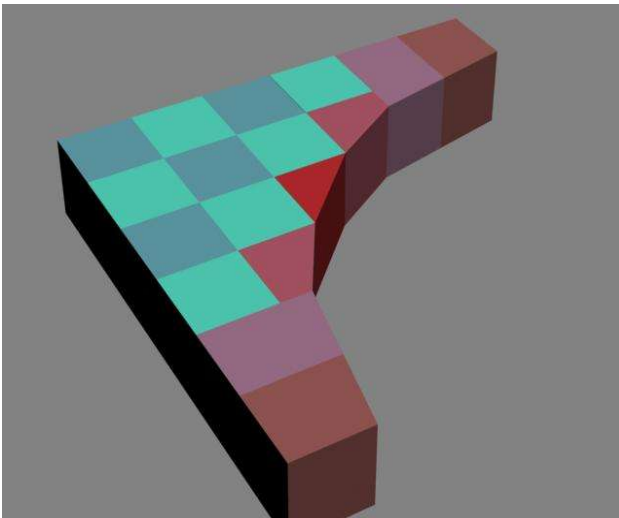

Figure 2.3 *Partial and extended cubes used to create a smooth surface*

2.3.3. Constraints and Problems

2.3.3.1. Extrusion speed

The speed the product can be extruded at is a function of three factors: the time to deliver the block to the surface; the time to fasten the block to the surface; and the fractional area of the working surface where blocks can be installed in parallel.

The block delivery rate can be estimated from the time required to maneuver large blocks in conventional handling operations. A meter-scale palette can be moved by forklift at a rate of multiple centimeters per second, including lifting and placing. Conveyor belts handling uniform products can move at centimeters or even meters per second. Assuming the blocks are delivered by conveyor to a point within ten block widths of the product surface, flexible robotic handling at that point may work at a rate of 1 cm/s or better. Thus 1-cm blocks could be positioned in 10 seconds, 1-micron blocks could be positioned in 1 msec, and so on.

The time to fasten a block to a surface depends on the method of fastening. Any joint that requires only pressure, or is actuated via built-in mechanisms, may fasten in a very small fraction of the block placement time. A joint that requires external robotic manipulation may take many times the block placement time. Since several joints of the former class exist and appear useful at sub-micron scales, it is reasonable to assume that block fastening will not be a significant fraction of assembly time in a good design.

The surface must be held while blocks are being attached. This requires that blocks be maneuvered past the surface-holding mechanism. But the mechanism need not be as wide as a block, so if blocks form a grid pattern, then every other location can be held while blocks are placed into alternate locations. Thus it appears that the overall extrusion rate could be almost half the rate of placing single blocks. For all block sizes down to sub-micron, then, depositing a meter of blocks might take about an hour.

2.3.3.2. Modular design constraints

Although there is room for some variability in the size and shape of blocks, they will be constrained by the need to handle them with single-sized machinery. A multi-micron monolithic subsystem would not be buildable with this manufacturing system: it would have to be built in pieces and assembled by simple manipulation, preferably mere placement. Phoenix's "expanding ridge joint" system appears to work for both strong mechanical joints and a variety of functional joints.

Human-scale product features will be far too large to be bothered by sub-micron grain boundaries. Functions that benefit from miniaturization (due to scaling laws) can be built within a single block. Even at the micron scale, where these constraints may be most troublesome, the remaining design space is a vast improvement over what we can achieve today or through existing technology roadmaps.

Sliding motion over a curved unlubricated surface will not work well if the surface is composed of blocks with 90 degree corners, no matter how small they are. However, there are several approaches that can mitigate this problem. First, there is no requirement

27

that all blocks be complete; the only requirement is that they contain enough surface to be handled by assembly robotics and joined to other blocks. Thus an approximation of a smooth curved surface with no projecting points can be assembled from prismatic partial-cubes, and a better approximation (marred only by joint lines and crystal steps) can be achieved if the fabrication method allows curves to be built. Hydrodynamic or molecular lubrication can be added after assembly; some lubricant molecules might be built into the block faces during fabrication, though this would probably have limited service life. Finally, in clean joints, nanoscale machinery attached to one large surface can serve as a standoff or actuator for another large surface, roughly equivalent to a forest of traction drives.

The grain scale may be large enough to affect some optical systems. In this case, joints like those between blocks can be built at regular intervals within the blocks, decreasing the lattice spacing and rendering it invisible to wave propagation.

2.3.4. Factory architecture

The factory would be in the form of a flat plate composed of several levels. The intake level would take in feedstock chemicals and manipulate them to form molecular parts. This would be in intimate contact with a feedstock/cooling fluid. This will be the most energy-intensive part of the operation, since it manipulates individual molecules; the rest of the layers can probably be cooled by conduction. The parts would then be formed into modular blocks, either in the fabrication layer or after transport. The blocks would be moved to a switching layer where they could be shuffled sideways to the point where they would be delivered to the output layer, which grips the assembly face of the product and places new blocks.

The factory's fabrication speed will be limited by two things: first, the rate of placement of blocks, roughly a meter per hour; second, the rate of fabrication of blocks, in which a fabrication station might make its mass every hour. This means that to extrude a meter per hour of solid product, the fabrication layer of the factory would have to be a meter thick. Several considerations, including cooling and reliability, cast doubt on the workability of this.

However, a square-meter factory producing a kg per hour would require only a mm of fabrication layer containing just a few thousand sub-layers of fabrication stations. As blocks were produced, they would be passed to the location that needed them, so that the kilogram of blocks (>1E16 blocks, a large fraction of a liter) could be placed anywhere within a cubic-meter volume.

Feedstock and cooling fluid would flow through micron-scale channels between banks of fabrication stations. (Exact layout to be added; it's not a difficult engineering constraint.)

The extruded product would be protected by a thin covering or caul that would grow along with the product, to exclude contaminants from the assembly surface and the interior of the nanofactory. A virtually unlimited length of product could be extruded, making it possible for a small factory to make larger factories by growing them edgewise.

28

2.4. Conclusion and Further Work

It appears that planar assembly is a powerful approach to constructing meter-scale products from sub-micron blocks, which can themselves be built by individual fabrication systems implementing molecular manufacturing or directed self-assembly. Planar assembly appears to be competitive with, and in many cases preferable to, all previously explored systems for general-purpose manufacture of large products. It is hard to find an example of a useful device that could not be built with the technique, and the expected meter-per-hour extrusion rate means that even large products could be built in their final configuration (as opposed to folded).

The ability to build large sparse products that do not require unfolding is a significant advantage. Phoenix's 2003 "Nanofactory" architecure required unfolding. His paper handwaved about the ability to unfold a nanofactory, but even that was a comparatively easy case because the interior structure of his nanofactory was almost entirely rectilinear. The ability to build products with more complex internal structure was not well justified, though Phoenix had some preliminary ideas that he hoped could handle curves, elongated shapes, and changes in cross section (density). Planar assembly bypasses the folding problem entirely.

This leaves only the problems of how to build sub-micron placement robotics, how best to utilize large volumes of sub-micron blocks to implement high-performance human-scale products, and how to fabricate molecularly precise sub-micron functional block modules. The first problem—sub-micron placement robotics—should be fairly straightforward once sub-micron functional blocks are available with nanoscale features.

The second problem—building large products out of programmable materials—should be accessible to today's engineering practice. Since the block size is smaller than today's engineering tolerances, and non-jagged surfaces can be built by specifying partial blocks (at least in some chemistries, and in others the blocks may be soft enough to slide past each other), it should be possible to specify parts volumetrically just as they are specified today. Reliability in the face of damage from ambient radiation is a new and substantial problem, but can be dealt with by fairly straightforward redundancy approaches, as explained elsewhere.

The third problem, fabrication of nanoblocks, is beginning to be a focus of research. Blocks built by pure self-assembly may not be sufficiently configurable to make it easy to engineer a nanofactory or complex product. However, even limited tools should make it possible for the self-assembly process to be guided, programmed, and/or restricted so as to make a variety of blocks from one set of inputs. Once configurable functional blocks are available, it appears that manufacture of large products may require only the porting of today's design skills to a new volume-filling (additive or depositional) manufacturing system. The growing number of architectures and demonstrated nanoscale capabilities indicates that a solution may not be far off, at least for simple functions.

3. Nanoscale Component Fabrication and Assembly

The goal of nanotechnology is to build useful nanoscale structures. Molecular manuacturing has a more targeted yet more ambitious goal: to build precise molecular structures, heterogeneous yet precisely organized over long distances, to make engineered products of all sizes. Although molecular manipulation has already been carried out by scanning probe microscopes, large-scale fabrication will require nanoscale tools in order to attain the required throughput. Fortunately, molecular manufacturing is well suited to building such tools. This chapter explores the types of reactions and operations that could be used to accomplish the goal of precise molecular fabrication and nanoscale assembly; many of the reactions are suitable for either scanning probe systems or advanced high-throughput nanosystems, indicating bootstrapping pathways.

3.1. Core concepts

Molecular manufacturing is often associated with a proposal of Drexler's: to do chemistry in vacuum with every molecule attached to a machine-like positioner. However, molecular manufacturing can use a much broader range of techniques. It should be noted that Drexler himself has written for over two decades about molecular manufacturing via solution-phase polymer manufacture (e.g. protein chemistry),[28] and has always advocated solution-phase manufacture as an appropriate starting point.[29]

Mechanosynthesis is the use of mechanical systems to make reactions happen only where and when they are desired. The definition covers several different kinds of fabrication operations and types of control. Mechanosynthesis can be performed either in solution with only some of the molecules controlled, or with all potentially reactive molecules controlled ("machine phase" or "vacuum phase" chemistry). For example, an enzyme that catalyzed a deposition reaction could be attached to a tool tip, with monomers loose in solution. Only when the enzyme was brought near the workpiece would a deposition reaction happen in that location. Another example is the ribosome, which binds amino acids together while removing them from tRNA. The amino acid-tRNA complex floats unconstrained in solution until it binds to the ribosome. Mechanosynthesis can add small molecular fragments to selected positions on a large molecule, add selected monomers to the end of a polymer, or add large molecular building blocks to a larger structure. It can also be used to pull molecules apart or transfer atoms between molecules.

Machine-phase chemistry is the kind of mechanosynthesis in which every molecule, or at least every potentially reactive molecule, is mechanically guided and controlled—bonded or strongly adsorbed to a workpiece or positioner. The reaction thus takes place in a vacuum, without the presence of solvent. Some solution chemists have argued that this

---

[28] Drexler, K. E. (1981) "Molecular engineering: An approach to the development of general capabilities for molecular manipulation." *Proc. Natnl. Acad. Sci. U.S.A.*. **78:**5275-5278. http://www.imm.org/PNAS.html

[29] See for example *Nanosystems* Chapter 15 and Section 16.4.

cannot work, but it turns out that some enzymes can indeed work without any solvent.[30] In fact, some enzymes work best without a solvent, as seen in some microdomain structures of hydrophiliehydrophobic block copolymers, which can be a very effective enzymatic reaction medium.[31] Some machine-phase chemistry experiments have been successful using scanning probe microscopes.[32] Machine-phase chemistry is unusual in that reactive species such as radicals and unterminated surfaces can be maintained for long periods of time; this creates many new possibilities for reactions.

Molecular manufacturing can be understood either by its method or by its goal. The method, as explained above, is building precise molecular structures and assembling them into structures under programmable control. The goal is to build large engineered devices and even entire products with nanoscale features and atomic precision.

The original definition of molecular manufacturing required the use of mechanosynthesis.[33] However, a continuum can be drawn from today's methods of solution-phase synthetic chemistry and self-assembly all the way to machine-phase chemistry. The spectrum runs something like this:

1. Make molecules with solution chemistry. Make structures by self-assembly.

2. Make molecules with solution chemistry. Use templates, possibly including active templates (e.g. electrodes), to guide the molecules into position: templated self-assembly.

3. Use primitive nanoscale kinematic machines to form increasingly configurable templates and perhaps to adjust the position of the templates.

4. Use increasingly strong chemical bonds to fasten the molecules together, starting with van der Waals forces or hydrogen bonding and ending with covalent bonds.

5. Use multiple stages of molecular fabrication, turning small molecules into large molecules and/or small devices that are then assembled into product.

6. Retain mechanical control of the molecules at intermediate stages.

7. Use less reactive solvent and more aggressive and mechanically-controlled chemistry. The ultimate extent of this step would use liquid xenon and bind all reactive molecules (which is to say, all molecules that are used) to machinery.

8. Stop using solvent altogether.

Thus, although molecular manufacturing requires the joining of molecular fragments or molecular building blocks to make large programmable structures, it certainly does not require machine phase chemistry, and does not even necessarily require that chemical reactions be guided; the building blocks can be synthesized in bulk, and (at least in

---

[30] An example of argument and counterargument can be found at
http://www.kurzweilai.net/meme/frame.html?main=/articles/art0604.html?

[31]   A. Gupte, R. Nagarajan,SP, and A. Kilara, "Block Copolymer Microdomains: A Novel Medium for Enzymatic Reactions", *Biotechnol. Prog.*, 7, 340-354, 1991.

[32] Some examples are listed at http://www.foresight.org/stage2/mechsynthbib.html

[33] In the glossary of *Nanosystems*, Drexler defines molecular manufacturing as "The production of complex structures via nonbiological mechanosynthesis (and subsequent assembly operations)."

primitive systems) a molecular manufacturing system could guide an assembly process that relied on surface forces rather than stronger chemical bonds.

3.2. Structural materials

3.2.1. Covalent solid crystals

Covalent solid mineral crystals of interest include diamond, alumina (sapphire), and possibly silica, silicon, silicon carbide, and boron nitride. The pure bulk material tends not to be very interesting electronically (e.g. electrical insulators, except silicon carbide). Functionalized or strained surfaces or small volumes can have interesting electronic properties. For example, CVD diamond has been used as a thermionic emitter, and small crystals of silicon fluoresce at size-dependent wavelengths.

The high density of bonds in these materials makes them strong and stiff. The stiffness means that in tensile applications, a flaw could propagate through a cross-section; this implies that large solid fibers will not be as strong as their bond density would suggest. Narrow fibers in a well engineered structure, with much less than one defect per fiber, should approach theoretical strength.

Many ways are known to form diamond that do not require intense heat and pressure. Chemical vapor deposition works because a hydrogen-rich atmosphere keeps the surface from reconstructing into something graphite-like (which would have lower energy). Diamond can also be formed by electron irradiation of bucky-onions[34] and crushing of buckyballs. Preliminary calculations show that diamond will probably be buildable by mechanically guided dimer deposition in vacuum using reactive rechargeable tooltips.[35]

If a set of mechanosynthetic tools can be developed for diamond, then graphene and fullerene structures might be buildable with many of the same tools. This would increase the available structures and electrical properties; fullerenes can be conductors or semiconductors as well as insulators.

Silica can be built in water at ambient temperature. An enzyme, silicatein, has been found that mediates deposition, raising the possibility of manipulating that enzyme to allow growth only in desired directions. Diatoms build complex micron-scale structures out of silica. It appears that silicatein retains its activity when bound to a surface, implying that it could be attached to a mechanically controlled tool tip.[36] Researchers have already used a synthetic variant of a silicatein active site "to create intricate silica patterns at the nanoscale level."[37]

Silicon is well studied due to its use in electronics. In 1994 Aono reported the automated pick-and-place of single silicon atoms from one location to another on a crystal surface.[38] However, this used an STM, and apparently his group was not able to build in three

---

[34] See http://www.nedo.go.jp/itd/grant-e/list/industrial/041t4-e.html
[35] See work by Freitas at http://www.MolecularAssembler.com/Papers/PathDiamMolMfg.htm and http://www.molecularassembler.com/Papers/DMSToolbuildProvPat.htm
[36] http://www.rsc.org/CFmuscat/intermediate_abstract.cfm?FURL=/ej/CC/2004/b410283e.PDF
[37] http://www.sciencedaily.com/releases/2001/09/010920071847.htm
[38] http://www.jst.go.jp/erato/project/agsh_P/agsh_P.html

dimensions, at least before their grant ran out. In 2001 another group reported removal and replacement of a single atom using an AFM.[39]

Alumina is very strong, and even more heat-resistant than graphite. Nanoscale pores can be formed in it. It appears that no one has yet studied the mechanosynthetic or enzymatic creation of alumina.

Boron nitride has not yet been studied as a candidate for mechanosynthesis. The fact that it consists of alternating atoms of very different electronegativity may reduce the number of undesired pathways encountered by deposition operations, making it easier to build by piecemeal mechanosynthesis. Its mechanical properties appear useful:[40] high thermal conductivity, low thermal expansion, good thermal shock resistance, high electrical resistance, low dielectric constant and loss tangent, non toxic, easily machined — non abrasive and lubricious, and chemically inert.

3.2.2. Graphene and fullerene

Graphene and fullerene have bond strength and density comparable with diamond. With five and seven ring insertions, a graphene surface can be made to curve; this could form useful structures. When a sheet of graphite is rolled into a tube (buckytube), the result can be either insulating, semiconducting, or conducting. Metallic buckytubes conduct very efficiently via ballistic electron transport; they are also excellent thermal conductors.

If a set of mechanosynthetic tools can be developed for diamond, graphene and fullerene structures might be buildable with many of the same tools. Graphene and fullerene have also been built by solution chemistry. Designed graphene sheets of 222 atoms have been built,[41] and synthesis of a "buckybowl" (half a buckyball) has also been accomplished.[42]

3.2.3. Polymers

A polymer can be built from monomers, with additions happening at controlled locations and/or in controlled sequence. Examples include protein and nucleic acid synthesis. Branched polymers can directly implement two- or three-dimensional structure, but even linear polymers can fold to form 3D structures.

Proteins, at least natural proteins, are not very stiff. Artificial amino acids, secondary backbones added to nucleic acids, and peptide nucleic acids are all being researched to develop new varieties of polymer. The folding of natural proteins is complex, but artificial proteins can be made somewhat more predictable.

---

[39] Noriaki Oyabu, Óscar Custance, Insook Yi, Yasuhiro Sugawara, and Seizo Morita. (2003). "Mechanical vertical manipulation of selected single atoms by soft nanoindentation using a near contact atomic force microscope" Phys. Rev. Lett. 90, 176102. See http://link.aps.org/abstract/PRL/v90/e176102 and http://focus.aps.org/story/v11/st19

[40] http://www.accuratus.com/boron.html (compressed powder)

[41] Simpson CD, Brand JD, Berresheim AJ, Przybilla L, Rader HJ, Mullen K. (2002) "Synthesis of a giant 222 carbon graphite sheet." *Chemistry* Mar 15;8(6):1424-9. Abstract at http://www3.interscience.wiley.com/cgi-bin/abstract/91016409/ABSTRACT

[42] See http://www.chem.iastate.edu/faculty/Peter_Rabideau/publication.html

Recently, Schafmeister has developed a semi-rigid linear polymer that has two bonds (a spiro link) between each monomer.[43] Depending on the sequence of monomers, curves and relatively straight sections can be built in. The shape is easily predictable from the sequence. This may improve the ability to design structure.

Folded polymers, held together by hydrophilic/phobic interactions and hydrogen bonds, are remarkably floppy compared with any familiar machine part; even the details of their structure are inconsistent. However, they manage to carry out operations in which they keep track of a sequence of states, as in a motor stepping forward on a tubule (rather than going backward or "marching in place"). A better understanding of this may show how floppy, low-energy molecular systems can be used as digital/predictable devices. It appears that the configuration space can include surprisingly steep "walls" that keep the molecule from reconfiguring in a way that loses its state.

3.2.4. Branched polymers

Branched polymers may be useful targets for mechanosynthesis and molecular manufacturing, if they are stiff enough to make well-defined structures and machines. Branched polymer growth and structure would be hard to control by unguided solution chemistry, but might be controlled by mechanically guided protection or deprotection of the numerous polymerization sites. (Dendrimers are branched polymers with a well-defined chemical structure that are grown in solution, but their mechanical structure is floppy and ill-defined due to lack of crosslinking.)

3.2.5. Molecular building blocks

Molecular building blocks are described in detail in Chapter 4, since they are a key piece of the work researched for this project.

3.3. Building molecular structures

An atomically precise structure can be built by any one of, or a combination of, several mechanically guided and unguided processes. Collections of atoms can be joined into larger structures by forces ranging in strength from van der Waals to covalent bonds. Structures can be built in stages: molecules can be fastened into larger molecules, then optionally folded or assembled, then fastened again—or simply held in place by the force that caused the folding. These steps can be repeated as many times as needed, incrementing or multiplying the size of the structure, until a useful component is made. Each of these steps may be carried out under direct mechanical control, or indirect control in which the conditions for the desired operation were set up by earlier manipulations. All that is important for molecular manufacturing is that the end result be programmable.

3.3.1. Mechanosynthesis

There is a wide variety of ways to make programmed parts from molecular feedstock. At one extreme, reactive species weakly bonded to mechanically manipulated "tool tip" molecules can be deposited on a reactive surface in vacuum (or noble gas); this is

---

[43]   http://www.foresight.org/Conferences/AdvNano2004/Abstracts/Schafmeister/

machine-phase chemistry. The ability to use unprotected radicals and other reactive species increases the chemical options, but requires extreme cleanliness.

In solvent, molecules can be forced to bond to selected locations on a surface or molecule. The use of terminated molecules, common (even necessary) in solution chemistry, somewhat reduces the cleanliness requirements. Solvent can be used to deliver feedstock to the tool tip by Brownian motion. Feedstock can either be covalently bonded to the tip via weak or strained bonds, or form a weakly-bound complex. Simply holding the molecule in the right location can increase its effective concentration by many orders of magnitude. Holding it in the right orientation, or applying pressure or electric fields, can further speed bonding to the surface, allowing the use of high-barrier reactions to minimize unwanted deposition.

Deposition reactions due to ordinary unguided chemistry can be used, if protection or deprotection is programmatically controlled. In this case, protection can be either chemical or steric (taking advantage of the spatial arrangement of atoms in a molecule).

In order to make a reaction happen, reactants must be brought together in the right position and orientation. In some cases an energy barrier must be overcome to make the reaction happen. If the barrier is within a certain range, then thermal motion can supply the energy, and the mechanical system can accelerate the reaction simply by holding the reactants near each other, increasing the effective concentration by several orders of magnitude. If the barrier is too high for simple positioning to work, there are several ways to lower it and speed the reaction; these apply to both solution phase and machine phase chemistry. The barrier can often be lowered by pushing the molecules together. (The conventional form of this, applying hydrostatic pressure to affect reaction rates, is called piezochemistry.) Electric fields can adjust the reaction. Extra energy to activate molecules and overcome barriers can be supplied via electric currents or light. Light cannot be focused with nanometer precision, and even with near-field techniques some photons will be created which may be absorbed by distant molecules. (Plasmons might be useful to deliver optical energies more precisely.) However, if the mechanosynthesis technique can guarantee that only the target molecules are in a position in which photochemical excitation will cause bonding, then even diffuse light can be used, and will activate bonding in only the desired locations.

In machine-phase chemistry, the lack of chemically active solvent does not necessarily limit the richness of available synthetic reactions.   It may appear that solution chemistry is more flexible than vacuum mechanosynthesis (machine-phase chemistry). Reactions in which solvent plays a role might in theory be difficult to accomplish in vacuum. Water, a very complex medium, was thought by some to be necessary for complex chemistry. However, the discovery by Klibanov in 1983 that many enzymes work fine in inorganic solvents and even in no solvent whatsoever indicates that solvent plays a less important role than might be assumed, at least in some reactions.[44] A diverse set of machine-phase reactions has been accomplished using scanning probe microscopes.

---

[44]   A. Klibanov describes his discovery that enzymes can work without water or indeed any solvent in this article: http://www.winstonbrill.com/bril001/html/article_index/articles/1-50/article32_body.html

Solvent-based reactions have their own problems. For example, a polymerization reaction may need to remove atoms from the joint and form them into a low-energy molecule that can float away, as well as joining the remaining molecules together—all in one reaction complex. Mechanosynthesis, especially in machine phase, can do reactions in multiple stages without needing to passivate or protect every molecule at every step.

Mechanosynthesis can reduce the rate of unwanted side reactions by preventing the reactants from contacting each other in ways that would allow those reactions to happen, or can enhance reaction rates at desired sites by orders of magnitude. This allows a particular deposition site to be selected from among many chemically similar sites. Many engineered heterogeneous products with large-scale intricacy that would be nearly impossible to build by self-assembly or common solution chemistry would be buildable with mechanosynthetic techniques.

In addition to improved site selection, mechanosynthesis—especially in machine phase— can make use of a broader palette of reactants. In vacuum, long-lived reactive species, radicals, and even unterminated surfaces can be used in ways that would not be possible with most solvents. The ability to create radicals in one process and then bring them together in a chosen orientation in a second, separate process should open new territories of chemistry.

Surfaces unterminated over a large area are likely to rearrange or reconstruct, but removing termination atoms only in the region where new atoms are to be added should avoid much of this problem. Stiffness scales adversely with size, while radicals must be guided very precisely to avoid unwanted reactions. However, calculation indicates that well-designed manipulators built on the 100-nm scale from well-chosen materials can be stiff enough even at room temperature. Machine-phase mechanosynthesis seems especially well suited to the creation of highly crosslinked materials, such as covalent solids.

3.3.2. Folding and assembly

Molecules can gain structure after they are fabricated by folding on themselves or assembling with other molecules. In proteins, folding corresponds to secondary and tertiary structure, and assembling corresponds to quaternary structure. Once the structure is attained, it may be locked in place by additional reactions, either mechanosynthetic or spontaneous. For example, a protein may form sulfur bonds or metal complexes as it folds. Polymers may also be crosslinked by heat or chemical reagants. The cruciform structure of tRNA is another example of polymer folding.

Molecules can join with each other through matching non-bonded interactions: electrostatic or van der Waals attraction or hydrogen bonding. Examples of this are immune complexes, protein quaternary structure, and Langmuir-Blodgett films.

A chain polymer can be built which then folds up into a desired shape. (Post-folding bonding can be used with some chemistries to increase strength and stiffness.) In a system that uses folding to form structure, the selection of which monomer to attach next is the only aspect of operation that needs to be programmable.

36

3.3.3. Bootstrapping

Even if stiff molecular machines are well suited to building additional copies of stiff molecular machines by machine phase mechanosynthesis, it is still necessary to find a way to build the first set of machines. There are several possible approaches. One is to use scanning probe microscopes as mechanosynthetic tools to directly build nanoscale machinery out of high-performance covalent solids. Robert Freitas and Ralph Merkle are the major proponents of this approach.[45]

Another approach is to develop a solution-phase molecular manufacturing capability first, and use it to build machines that can work in vacuum phase. At least two high-performance materials, graphene/fullerene and silica, are buildable in solution phase and will presumably be buildable in machine phase as well. That should allow a set of machines to be built in solution, then dried out, then used to build very similar machines in the new environment. This type of incremental advancement will reduce the difficulty and uncertainty of engineering. Another incremental direction would be to develop a molecular building block system in which the building blocks can be assembled and can function either in solution or in gas/vacuum. Xenon and carbon dioxide are solvents which can be used in liquid, supercritical, or gas phase, leading to vacuum.

3.3.4. Simulation of mechanosynthesis and folding

For several reasons, simulations of mechanosynthetic reactions may have more predictive value than simulations of ordinary solvated reactions. Mechanosynthetic processes can physically constrain molecules in a way that avoids many unwanted reactions entirely. Applying modest pressure to reactants can significantly alter the energetics of the reaction and thus shift reaction rates and equilibria in desired directions. These advantages hold for both solvated and machine-phase mechanosynthesis. The ability to use extremely reactive species in machine-phase mechanosynthesis allows reactions in which the change in energy is far larger than in solution-phase chemistry. If the difference in energy between desired and undesired states is significantly larger than the inaccuracy of the simulator, then computational chemistry tools that must be used with caution for typical solution chemistry reactions may be used with more confidence to evaluate mechanosynthetic reactions.

Natural proteins appear to have evolved not to fold too strongly. Tightly folded proteins may be difficult to metabolize, and in fact indigestible tangles of protein are responsible for several diseases. Proteins designed for folding are easier to predict.

3.4. Joining large molecules

Fabricated parts can be joined chemically, by leaving unterminated radicals (in vacuum) or mutually reactive side chains (in solution). They can also be joined by surface forces between surfaces of complementary shape, complementary patterns of hydrogen bonds or charge between surfaces or dangling strands, or interlocking mechanisms. In systems

---

[45] Examples of scanning probe mechanosynthesis: http://www.foresight.org/stage2/mechsynthbib.html
Proposal for theoretical work: http://www.foresight.org/stage2/project1A.html
Proposal for experimental work: http://www.MolecularAssembler.com/Papers/PathDiamMolMfg.htm

where a general-purpose manipulator does the fabrication, the same manipulator may be used to position the finished parts; alternatively, Brownian assembly might be used, though this is slower for larger parts and is not deterministic.

Several kinds of reaction are available. One is the formation of standard covalent bonds. This can be triggered by light, by electric fields or currents, by mechanical pressure[46], or simply by holding reactive molecules near each other until thermal energy overcomes the reaction barrier. Weaker bonds, including hydrogen bonds and sulfur bonds, can link molecules large enough to include several bond sites. In solvent, free-floating ions can play a part in the bonding; for example, zinc coordinates tetrahedrally with cysteine and/or histidine amino acids, forming a fairly strong bond; see Fig. 3.1. In vacuum or moderate-pressure gas, surface forces (also called dispersion forces and van der Waals forces) can hold things together quite tightly—up to 1 GPa.[47] Binding by surface forces is not actually a chemical reaction, but surface forces form the weak end of a continuum, and the molecular manufacturing design approach described here can apply to systems based on non-chemical fastening—as long as the parts being fastened are atomically precise, and precisely placed.

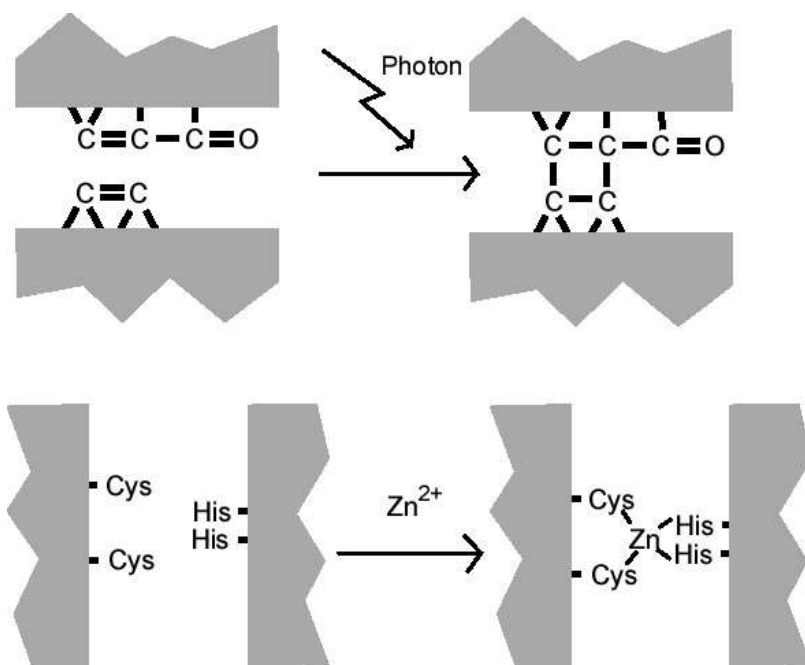
Figure 3.1 Zinc binding and photochemical bonding are proposed for use in primitive molecular manufacturing systems.

[46]   Sullivan, Terena P.; van Poll, Maaike L.; Dankers, Patricia Y. W.; Huck, Wilhelm T. S., "Forced peptide synthesis in nanoscale confinement under elastomeric stamps," Angewandte Chemie, International Edition (2004), 43(32), 4190-4193
[47]   *Nanosystems* section 9.7.1

4. Molecular Building Blocks

Deposition of atoms or dimers on the (terminated or unterminated) surface of a covalent solid is a difficult prospect. It is expected to work, but many of the necessary reactions have not yet been simulated, much less tried experimentally. However, it may be easier to place nanometer-scale molecules into nanometer-scale vacancies in a product. This approach would not require the molecular manufacturing system to perform general-purpose synthesis of molecules. Instead, the system would place a few key types of molecules in programmable patterns. The molecules would be fabricated by ordinary solution chemistry, perhaps with the assistance of self-assembly.

This "molecular building block" (MBB) approach was suggested by Ralph Merkle in 1999[48] and 2000.[49] In the latter paper, Merkle listed six advantages of a molecular building block approach, suggested several ways to join blocks to each other, and discussed the advantages of positional assembly over self-assembly. However, recent advances in the synthesis of molecular structures, as well as a new and simpler architecture for a manufacturing system that can place MBB's into a product, suggest that it may now be possible to develop a primitive but improvable molecular manufacturing system based on MBB's.

4.1. Block composition

Many different options exist for the chemical structure of molecular building blocks. Several likely candidates are listed here. Some possibilities, such as dendrimers, are not listed because they are likely to be insufficiently stiff or difficult to functionalize in specific patterns.

4.1.1. Linear Polymers

Biology contains several examples of polymers that fold to form structures. Protein can fold into very intricate structures implementing complex functionality. Multiple protein molecules can self-assemble to form machines such as ATP synthase. However, proteins are not very stiff—at least this is true for biological proteins; artificial variants are often more stable. Furthermore, protein folding is complex and not fully understood.

Nucleic acids fold predictably. Although a double strand of DNA is not very stiff, variants with four interwoven strands can maintain their structure over multiple nanometers. Another advantage of nucleic acids is that they pair up: a single strand attached to each of two objects can bind the objects together, if and only if the strands have complementary sequences. Nucleic acids can implement some of the same functions as proteins, as in "RNA enzymes."

---

[48] R. Merkle, "Biotechnology as a route to nanotechnology", *Trends in Biotechnology*, July 1999, Vol 17 No 7, pages 271-274, http://www.merkle.com/papers/bionano.html
[49] R. Merkle, "Molecular building blocks and development strategies for molecular nanotechnology", *Nanotechnology* 11 (2000) pages 89-99, http://www.zyvex.com/nanotech/mbb/mbb.html

Recently, Schafmeister has developed a semi-rigid linear polymer that has two bonds (a spiro link) between each monomer.[50] Depending on the sequence of monomers, curves and relatively straight sections can be built in. Because this can form a stiff three-dimensional shape without folding, and because the shape is easily predictable, this family of molecules may be a very convenient backbone for MBB's.

4.1.2. Carbon-cage derivatives

Adamantane is essentially a unit cell of diamond. Merkle suggested that adamantane, either functionalized or with some of its carbon atoms replaced, might make a useful MBB. Inclusion of boron and nitrogen would allow dipolar bonds to be formed between blocks; this has the advantage that a bond can form without releasing any small molecules, making it suitable for machine-phase chemistry.

Merkle also suggested that functionalized fullerenes could make useful MBB's.

4.1.3. Polyhedral oligomeric silsesquioxanes

One ready-made molecular building block that shows significant promise is the fundamental cellular unit of silica, the cubic silsesquioxane $[RSiO_{1.5}]_8$ molecule sometimes known as POSS for Polyhedral Oligomeric Silsesquioxanes (the name is trademarked by Hybrid Plastics).

At this point, silsesquioxane cubes and related molecules are our best bet for a number of reasons. First, it is already possible to obtain semiconductor, conducting, and insulating forms. Second, the silica core is very stable, degrading only at 500 degrees C. Third, the silicon-oxygen bonds are very stiff, therefore silsesquioxane should be almost as strong as carbon nanotubes. Fourth, the DoD has funded research into this class of materials for about a decade, so it's chemistry is understood well enough for two start-ups to provide different types of silsesquioxane in kilogram quantities. Finally, there is lots of high surface area silica in lunar and Martian regolith from which cubic silsesquioxanes could be synthesized.

Figure 4.1. Silsesquioxane cube

Laine's group at University of Michigan has developed methods of making novel, high density functionalized, robust and easily purified cubic silsesquioxane molecules with chemistries that complement those used to form and process dendrimers and hyperbranched materials. For example, it is now possible to selectively synthesize the Br8, Br16 and I8 functionalized OPS (octaphenylsilsesquioxane) cubes in Figure 4.2. High purity has been attained in Laine's lab at in many silsesquioxane products, so it is reasonable to believe that this task is realistic.

---

[50]   http://www.foresight.org/Conferences/AdvNano2004/Abstracts/Schafmeister/

A   **B**

Figure 4.2 Cubic silsesquioxanes. A. Q8 (Q= SiO4) R= H, vinyl, epoxy, 3-hydroxypropyl, amino-propyl, glycidylepoxy, ethylcyclohexylepoxy, methacrylate, etc. B. PhT8/OPS (T = carbon-SiO1.5) R = Br, NH2, alkyl, alkene, acetylene, acyl, azo, etc. R = same or mixed.

Cubic silsesquioxane $[RSiO_{1.5}]_8$ can be used as highly functionalized nanoscale building blocks. They are molecules consisting of rigid, single crystal silica cores with eight vertices (opposing vertex body diagonal are 0.53 nm) to which are appended organic functional groups creating organic/inorganic molecules 1-2 nm in diameter with volumes <2 nm^3. Each group lies in a separate octant in Cartesian space. The positioning, variety, and size of the functional groups, provide many opportunities for using cubes to build nanocomposite/hybrid materials in one, two or three dimensions, one nanometer at a time.

Cubic polyhedral silsesquioxanes have been successfully used to model catalytic surfaces, develop new catalyst and novel porous media, serve as NMR standards, act as novel encapsulants, and as building blocks for nanocomposite materials. Furthermore, the Q8 system is unique and useful because it can be made from any source of high surface area silica, both organic forms like rice hull ash, and inorganic forms such as those found on the Moon or Mars.

A. Core Unit Cube   *B. Shell One*   *C. Shell Two*

Pd-catalyzed Heck coupling   Pt-catalyzed hydrosilylation

Figure 4.3 Synthetic route to a multilayer cube system. Yields are < 30 %.

41

Figure 4.4 *Scheme 1: Synthesis of arylated phenyl octasilsesquioxanes (Ar$_x$OPS). Note that F$_x$OPS is highly photoluminscent with a quantum efficiency near 100 % at 380 nm.*

Figure 4.5 *Scheme 2: Synthesis of Octaaminophenylsilsesquioxane (OAPS) by nitration of OPS followed by reduction.*

Compound	Mn	Mw	PDI	Extent of Reaction (# of silsesquioxane cores added per shell)	
Br$_{5.7}$OPS	921	928	1.01		
Vinyl$_8$Q$_8$	1060	1047	0.99		
H$_8$Q$_8$	947	951	1	GPC	^{1}H NMR
Shell One	8,920	15,340	1.72	6.3	4.3
Shell Two	24,270	37,970	1.57	--	$\approx$ 15

Table 4.1 GPC (gel permeation chromatography) and NMR (nuclear magnetic resonance) analysis of the components and products of core-shell systems.

The orthogonal functionality permits the assembly of cubes using simple chemistries that allow larger core shell materials to be made, as shown in Figure 4.3 (see also Table 4.1). The process uses simple chemistries to combine the cubic silsesquioxanes from Figures 4.2A and 4.2B to yield rigid molecularly precise structures. The method is similar to the softer dendrimers, except that that dendrimers are typically constructed using highly flexible and therefore thermally less stable, organic arms. In silsesquioxane systems, the rigidity of the 3-D structure can be controlled by choosing the bridging tether, combined with the Q8 system of the OPS system. The latter is easily modified using a number of simple catalytic coupling chemistries that extend the cage structure with rigidity or flexibility. The coupling moieties could also act as semiconductors depending on the length of the polyaromatic and its ability to interact with other aromatics on other cubes. One approach to assembly would be to develop donor acceptor (D-A) systems. By printing one half of the couple on a substrate the second half could be added simply by dipping into a solution of the other half of the D-A couple. The octanitro compound shown in Figure 4.5 could provide the acceptor behavior desired (or the hexadecanitroOPS) and the OAPS or a methylated version might be the donor.

Still another way to construct some of the systems desired for a nanofab laboratory could involve the polymerization of the aniline groups with free aniline to form a star polyaniline which, depending on the chain length of the oligo-anilines along each arm, could define the size of a rigid 3-D semiconductor. Also, note that polyanilines are known to offer electrochromic properties making them possible nanoindicators or even sensors in addition to their semiconducting behavior.

Note that in Scheme 1 (Figure 4.4), the Laine group has also made the original octathiophene. It is important to note that the OAPS is primarily meta substituted, the Br$_x$OPS is primarily para substituted, but they made the I$_8$OPS which is entirely para-substituted making it the perfect nanobuilding block. Also, it is much easier to functionalize than the Br$_x$OPS.

Dendrimers are similar to silsesquioxanes in that their chemistry can be tightly controlled, and were also considered. But were rejected because they are too soft and too easily degraded. There is a possibility that a structure of cube-like symmetry could be creating by reacting eight R7POSS(-NCO) with a G1 PAMAM dendrimer carrying eight amino groups. A cursory examination of that option, however, did not lead us to believe that the dendrimer seed could solve any problems that a silsesquioxane cube could not.

4.2. Handling and joining blocks

MBB's will be joined by some kind of chemical bond. A wide range of bonds can be used; some are more reversible than others. In order to place MBB's in the right position, they must be attached reversibly to handling machinery. One relatively simple solution to block placement and handling is to use the same reversible bonding method for both handling and placement, with the bond to the handler being strong enough to prevent the block leaving the handler prematurely, but enough weaker than the bond to the product that the block will transfer reliably from the handler to the product. Of course separate mechanisms for handling and binding are also possible; photochemistry makes a stronger (irreversible) bond with smaller molecular components, so may be preferred for higher-performance systems.

Metallic complexes form bonds that are reversible, but strong enough to make solid structures. An example of this is the zinc binding described in Chapter 3. Distorting the alignment of the amino acids should weaken the bond. If the blocks are covered with appropriate amino acids in a specific pattern, then a binding "socket" could be given a similar pattern, but with fewer or misplaced acids.

4.2.1. Keeping blocks apart

The feedstock of a primitive "wet" molecular manufacturing system will be prefabricated MBB's, a few nanometers in size, dissolved or suspended in a solvent. The function of the manufacturing mechanism is simple: to take individual blocks from solution, move them to a specified location in the product, and attach them strongly enough that they do not unbind. However, the design of the system must ensure that blocks in solution very seldom aggregate with each other or attach to the product where they are not desired, but once attached and bonded to the product they very seldom detach.

A charged object in solution will be surrounded by counterions of the opposite charge. In polar solvents like water, the solvent molecules will align to nearby charges and partially screen them; in nonpolar solvents, charges can affect each other over longer distances and are apt to bind to the object. If two objects of the same charge approach each other, they will repel. Less polar solvents will extend the repulsion zone, requiring more energy to force the objects together.

For the purpose of joining blocks, the reaction should happen only when the blocks are pushed together, not when they are floating in the feedstock solution. If zinc binding is used, then excluding zinc from the feedstock solution and excluding stray blocks from the product area can prevent the feedstock blocks from aggregating in solution. The "Silkscreen" approach (see Chapter 5) was designed to maintain a product environment of different composition than the feedstock solution. If photochemistry is used, then stray blocks must be prevented from contact while the photons are being delivered. Putting a charge on the blocks can keep them separated in solution, preventing unwanted aggregation even if no membrane is used to separate feedstock from product environment, as in the "Tattoo" approach.

A simple approach to block fastening is to cover the block with zinc-binding amino acids forming half-binding sites (two acids forming half of a tetrahedral site). Design the block with a negative charge equal to the number of sites. Zinc ions, with two positive charges apiece, will bind to each site, giving the block a net positive charge; this will keep it separated from other blocks in solution. If two blocks are pushed together strongly enough to overcome their repulsion, then half of the zinc will be squeezed out, leaving exactly enough zinc to neutralize the charge on the two blocks, and fastening the blocks strongly together. If the blocks are reasonably rigid, then it will be energetically unfavorable for extra zinc (along with its associated counterions) to squeeze in and allow them to separate, because multiple zinc ions would have to enter the tight inter-block space simultaneously. Thus the feedstock solution, with loose blocks, zinc, and counterions, could be allowed to contact the product without dissolving it or accreting to it. This would be ideal for the "Tattoo" approach.

4.2.2. Other fastening options

Attachments formed by photochemistry or electricity should work with either the Tattoo or the Silkscreen approach. Although it is somewhat more complicated, requiring delivery of light or electricity as well as some way to cancel the charge on the blocks as they are added to the product, it has the advantage that it will work equally well in more advanced solvent-free systems. Another complication is the need to keep reactive molecules (such as oxygen) away from the photochemical sites.

If the zinc binding approach and the photochemical approach conflict with some aspect of block or system design (for example, if the blocks cannot be prevented from colliding and accreting too frequently due to Brownian motion), it should be possible to use a pair of molecules that forms a bond via a reaction that is energetically favorable but has a high energy barrier. Such a reaction will happen very seldom between blocks in solution, because both the reaction's energy barrier and the block's repulsion must be overcome. But once blocks are confined and pressed together, the block's repulsion will no longer impose an energy barrier, and the effective concentration of the reactants will increase by several orders of magnitude; proper alignment of the reactants may also help. Together, these factors should make the reaction happen many orders of magnitude faster, allowing a fast assembly rate to coexist with a low error rate. Many such reactions will work without solvent.

In order to manipulate blocks mechanically, they must be attracted from solution and attached to a manipulator. This will happen if a binding site ("socket") in the manipulator is made attractive to the blocks, for example by giving its interior a charge opposite to the charge on the blocks. This is an application of self-assembly. For some blocks, it will be important to orient them correctly. In these cases, patterns of charge, asymmetrical shape, short complementary strands of DNA, and weak bonds such as hydrogen bonds can be used to cause the block to favor a particular orientation and can make a socket specific for a particular block type.

4.2.3. Using multiple block types

The manufacturing system will need to be able to place more than one kind of block. There are two ways to accomplish this. One possibility is that the block types will be mixed in the feedstock solution, in which case the sockets must be block-specific (meaning that they must be reconfigurable, or there must be multiple sockets). The other possibility is to flush through one feedstock solution at a time, with each solution containing a different block type. When a solution is flushed out, blocks will remain in sockets, but can be dealt with simply by depositing them onto the product in an appropriate location.

If the system uses zinc binding, then putting a few zinc binding sites in the binding surface of the socket can be used to bind to the block strongly enough to hold it reliably, but weakly enough to let it go without damage when the actuator is retracted from the product. Slight misalignment of the binding sites can reduce the binding force, and adding more sites can increase the force.

4.3. Molecular electromechanical actuators

A machine is not much good unless something moves it. Actuation is probably the first feature after structure that will have to be included in primitive molecular manufacturing systems. (Power and information delivery may be done optically in early machines, not requiring structures.) In more advanced systems, electrostatic actuators can be built from only conductors and insulators, but these will require intricate structure as well as not using a solvent that causes problems through charge screening. For early designs, actuators that occupy a single MBB will be desirable. This section surveys several likely candidates.

Electromechanical actuators convert electrical energy into mechanical work. Some "molecular muscles" currently being researched in many labs are electromechanical actuators fabricated from redox-active polymers. The actuation mechanism of some of these materials depends on volume changes caused by the uptake or expulsion of counter-ions during a redox (reduction/oxidation) cycle, as in the case of IMPC (Ionic Metal-Polymer Composites). Unfortunately, this electromechanical actuation in redox-active polymers is a bulk property that is not necessarily intrinsic to individual molecules.

Piezoelectric materials deform because the spacing between charged atoms in the crystal varies slightly under an electric field. Although they require high voltages at millimeter scale, sub-micron thicknesses should be activated by sub-volt fields. However, they have very low strain (at most a few percent) so may be difficult to use.

Some electrically deformable polymers work by ion intercalation/expulsion, which may release unwanted ions into solution and may not work in single molecules at all. Redox reactions can cause changes in a molecule's electron distribution, which can cause large changes in its shape or in how it fits together with other molecules. The mechanism of known redox actuators involves protonation, and this may not work in some environments; search is ongoing (in other research efforts) for molecular actuators that use only electron exchange.

46

Annulene-based actuators can deliver strains of almost 6%.[51] Poly(calix[4]arene-bithiophene) has been calculated to produce 78% strain, and another thiophene, poly (quaterthiophene) (poly(QT)), has exhibited reversible strains of 20% at 5 volts[52]. Such actuators would seem to provide the required performance. These redox powered actuators could presumably be protonated by battery-type reactions that are driven by varying electric currents from nearby nano-wires.

These molecular devices are single-molecule actuators based on redox-active reactions. However, there are two difficulties with single-molecule electromechanical actuation in a redox-active polymer:

1. Few molecules will satisfy the design requirements for a molecular electromechanical actuator (in terms of strength, speed, stability, and most important, actuation distance).

2. Traditional analytical techniques can't differentiate between bulk mechanism of ion intercalation/expulsion (which we don't want) and intrinsic single-molecule mechanisms (which we do).

In addition to giant piezoelectric materials, annulenes, and polymer/platinum nanofilms, viral protein linear (VPL) motors, and interlocking rotaxane dimers may be useful.[53]

Bio-motors probably will not be as useful for the planned systems. The F0-F1 ATPase, the kinesin, myosin, and dynein superfamily of protein molecular machines, and bacteria flagellar motors all depend on either ATP or ion gradients for energy. Motors that use ion gradients, ATP, or DNA binding are inferior to electromechanical actuators for a number of reasons:

1. Electrons move faster than protons, and much faster than larger molecules.

2. Purely electrical processes will almost always require less material input.

3. It will probably be architecturally difficult to supply ATP or maintain ion gradients, either from external sources or from local molecular machines.

4.3.1. Giant piezoelectric materials

Piezoelectric materials can change dimensions (strain) slightly under an applied voltage (and vice versa). The new materials PZN-PT (solid solution of PbZr1/3Nb2/3O3), PMN-PT (solid solution of PbMn1/3Nb2/3O3) and PbTiO3 have strains as high as 2% (40 times greater than in conventional piezoelectric materials). Fu and Cohen showed (from first principles) that large strains can be produced via polarization rotation.[54]

---

[51]   Michael J. Marsella, Rodney J. Reid, Samia Estassi, and Li-Sheng Wang, "Tetra[2,3-thienylene]: A Building Block for Single-Molecule Electromechanical Actuators", J. AM. CHEM. SOC. 2002, 124, 12507-12510.

[52]   Patrick A. Anquetil, Hsiao-hua Yub, John D. Madden, Peter G. Madden, Timothy M. Swager, and Ian W. Hunter, "Thiophene-based conducting polymer molecular actuators", Smart Structures and Materials 2002: Electroactive Polymer Actuators and Devices (EAPAD), Yoseph Bar-Cohen, Editor, Proceedings of SPIE Vol. 4695 (2002)

[53]   See http://www.bionano.neu.edu/AR220-BE06-10_001-033_.pdf for descriptions of several types of molecular motors.

[54]   H. Fu and R.E. Cohen, "Polarization rotation mechanism for ultrahigh electromechanical response in single-crystal piezoelectrics", Nature (403), 281-283, January 20, 2000. See

47

Piezoelectric materials are characterized by separation between the centers of positive and negative ions. When an electric field is applied, a tiny shift occurs in the ions, causing contraction in the crystal as a whole. With lead zicronate titanate (PZT), which displays the largest conventional piezoelectric effect, this change is only approximately 0.01% in a 100 V/mm field.

Utilizing the principle of nano-symmetry of point defects, Xiaobing Ren and his colleagues controlled the symmetry of point defects to coincide with the crystal symmetry. When a field is applied to material in this state, domain switching occurs (simultaneously generating huge electro-strain).

To make this electro-strain effect reversible, certain doping elements needed to be added, and the material subjected to an aging treatment. A BaTiO3 sample containing a trace amount of Fe had a huge reversible electro-strain of 0.75% in a low field (comparatively speaking, for piezoelectric materials) of 200 V/mm.[55]

4.3.2. Annulenes

Annulenes are cyclic molecules that undergo a redox-induced transformation between a planar shape and a "tub" shape.
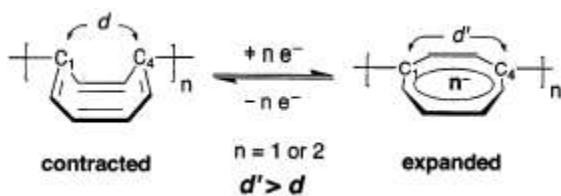

Figure 4.6 *Annulene reconfiguration from Marsella et al. Used by permission.*

Tetra(2,3-thienylene) is a thiophene-fused [8]annulene capable of undergoing redox-induced dimensional changes of 3.54-5.92% (depending how the basic cyclic molecule is modified, for example by taking two of them and making a dimmer). Unfortunately, it cannot achieve a planar conformation.[56] However, thiophene-fused didehydro[12]-annulene seems a better candidate for actuation, exhibiting a redox-induced change of 19%.[57]

4.3.3. Polymer/Platinum NanoFilms

Polymer/metal cluster nanocomposite thin films have been synthesized with a layer-by-

---

http://carnegieinstitution.org/YearBook_99_00/scientists/Cohen/Cohen_eYearbook/cohen.html

[55] SeeXiaobing Ren, "Giant Field-Induced Strain Effect Based on Novel Principle", http://www.nims.go.jp/eng/news/nimsnow/Vol2/No7/p2.html; J.S. Forrester, R.O. Piltz, E.H. Kisi and G.J. McIntyre, "Temperature induced phase transitions in the giant piezoelectric material PZN-4.5% PT", J. Phys: Condens. Matter (2001) 13 825-833; http://www.ansto.gov.au/ainse/prorep2001/R_01_086%20revised.pdf

[56] See Michael J. Marsella, Rodney J. Reid, Samia Estassi, and Li-Sheng Wang, "Tetra[2,3-thienylene]: A Building Block for Single-Molecule Electromechanical Actuators", Journal of the American Chemical Society, 2002, V124, pp12507-12510.

[57] See Marsella, M. J.; Piao, G.; Tham, F. S., "Tetra[2,3-thienylene]: A Building Block for Single-Molecule Electromechanical Actuators," Synthesis 2002, 1133-1135.

48

layer electrostatic self-assembly process. Platinum-nanocluster colloids were protected by cationic polyelectrolyte molecules of poly(diallyldimethylammonium chloride) (PDDA), resulting in Pt:PDDA nanoclusters ~2.1 nm in size with a positive surface charge. These were anchored in a monolayer to a substrate, followed by a layer of poly-S-119, each molecule of which was oriented by the electrical attraction of its negatively charged –SO3 functional group. An 11-bilayer poly-S-119/Pt nanocluster film, with a 3.4 weight percentage of platinum and a thickness of approximately 30.2 nm (~3 nm per bilayer), exhibited an actuation response similar to that of conventional ionic polymer–metal composite films for low applied voltage in air. The film has a uniform nanostructure, with the Pt nanoclusters dispersed within the multiple molecular layers. Because of its unusual laminated structure, it seems to have different actuation mechanisms from those of traditional piezoelectric materials, and from other EAP materials.[58]

4.3.4. Thiophene-based Polymers

Another EAP consists of thiophene based conducting polymer molecular actuators. These actuators operate by conformational rearrangement of the polymer backbone at the molecular level, resulting from pi-pi stacking of thiophene oligomers upon oxidation, producing a reversible molecular displacement.

Conducting polymer materials such as those based on polypyrrole or polyanilines generate forces per cross-sectional area that are up to two orders of magnitude greater than human muscle (40 MPa), with equal power to mass ratios (39 W/kg). Unfortunately, while these actuators require low drive voltages, they typically offer limited active strain (2%) and limited active strain rate(1%/s). Even worse, they are based on electrochemical oxidation and a resulting diffusion and intercalation of ionic species into the polymer bulk film.
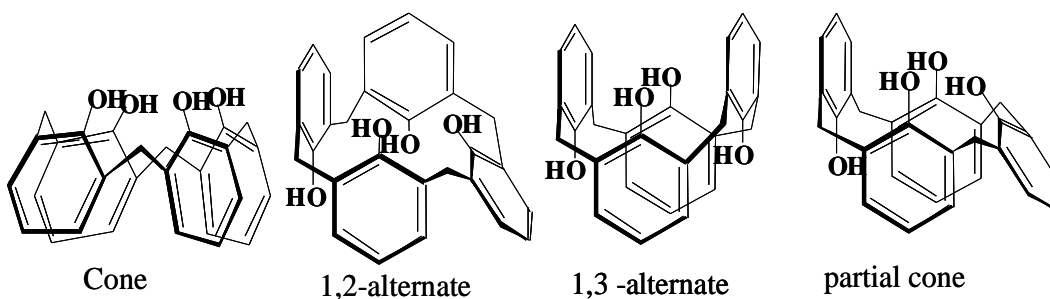
Figure 4.7 Different conformations of calix[4]-arene. Figs. 4.7-4.9 used by permission of P. Anquetil.

Thiophene-based actuators depend on calix[4]-arene (figure 2) to actually provide the motive force, while the thiophene (a ring of four carbons and one sulfur) is a rigid rod. The calix[4]arene molecule has the ability to exist in a number of different conformations: the cone, 1,2-alternate, 1,3-alternate, and partial cone.

---

[58]   T Zeng, R Claus, F Zhang, W Du1 and K L Cooper, "Ultrathin film actuators fabricated by layer-by-layer molecular self-assembly", 2001 Smart Mater. Struct. 10 780-785

One thiophene-based molecular actuator candidate, calix[4]arene-bithiophene, shown in Figure 1, employs hinge molecules of calix[4]arene (figure 2) interconnected by rigid rods of quaterthiophene (figure 4.8). The rods attract one another in the oxidized state, contracting the material. This actuation characteristic results from pi stacking of thiophene oligomers upon oxidation, producing a reversible molecular displacement.

Theoretical calculations show that poly(calix[4]arene-bithiophene) can achieve up to 78% strain, and (with lower strain) achieve a maximal stress of 321 MPa, with electromechanical coupling efficiency as high as 66.7%.
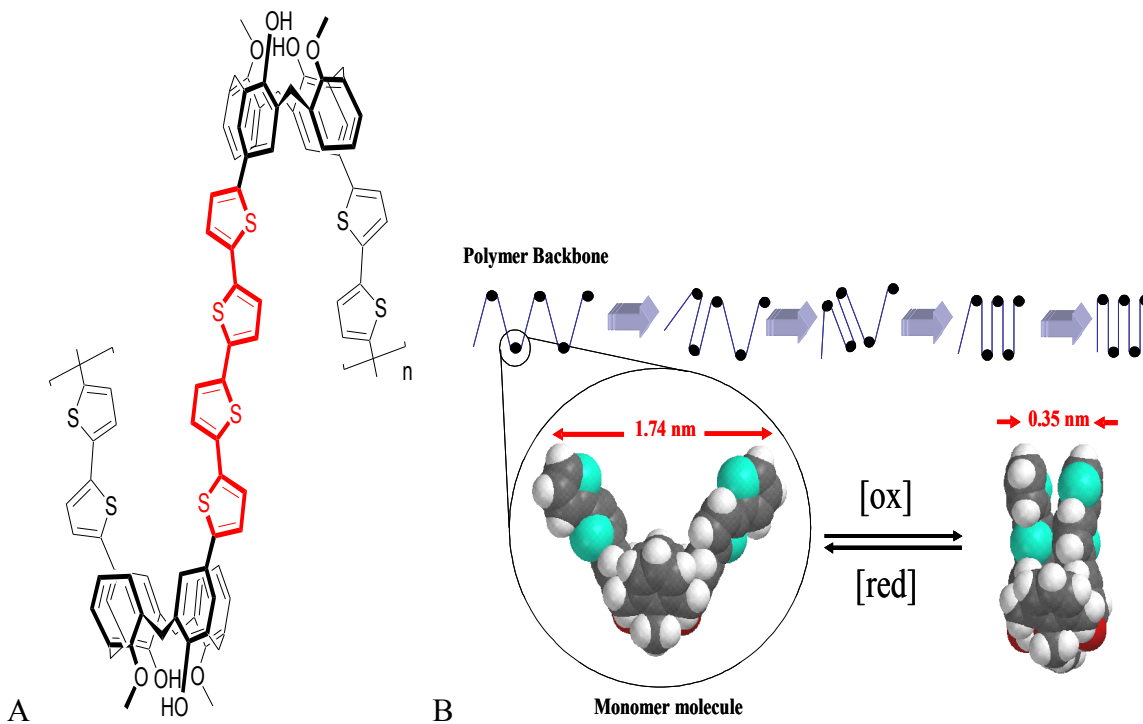

A   B

Figure 4.8 Poly(calix[4]arene-bithiophene) : A one monomer; B details and overview

The second candidate actuator is poly(quaterthiophene) (poly(QT)), which can be used to study the stacking effect of the oxidized thiophene oligomers upon oxidation. In other words, this latter candidate is a non-hinged version of the poly(calix[4]arene-bithiophene) molecule.
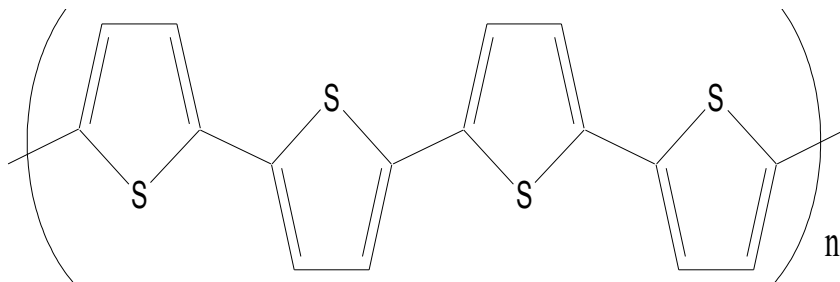
Figure 4.9 *Quaterthiophene*

Experimental poly(QT) films have exhibited reversible strains of $20 \pm 3\%$ at 5 volts.[59]

4.3.5. Viral Protein Linear (VPL) motors

The Influenza envelope glycoprotein hemagglutinin (HA) and bromelain cleaved hemagglutinin (BHA) undergo similar pH-dependant conformational changes. HA consists of two polypeptide chain subunits (HA1 and HA2) linked by a disulfide bond. HA1 contains sialic acid binding sites, which respond to the cell surface receptors of the target cells and, hence, help the virus to recognize a cell. According to the widely-accepted spring-loaded conformational change theory, there is a specific region (sequence) in HA2 that tends to form a coiled coil. Upon activation by acidic conditions (a pH around 5), this 36 amino-acid residue region makes a dramatic conformational change from a loop to a triple-stranded extended coiled coil along with some residues of a short a-helix that precedes it. This process relocates the hydrophobic fusion peptide (and the N-terminal of the peptide) by about 10 nm. Mavroidis and his group are engineering a peptide identical to the 36-residue long peptide mentioned above, which they call loop36. Cutting out the loop36 from the VPL motor, they obtain a peptide that has a closed length of about 4 nm and an extended length of about 6 nm, giving it an extension by two-thirds of its length. Once characterized, the peptide will be subjected to conditions similar to what a virus experiences in the proximity of a cell, that is, a reduced pH.[60] Because pH changes involve the donating of a proton or the acceptance of an electron pair, it seems possible that VPL motors could be adapted to electrical actuation. It is not clear yet what causes the conformational change.

4.3.6. Interlocking rotaxane dimers

When a ring is threaded onto a rod, it can either rotate around the axle or undergo a translation movement. When two rings are attached to two rods, the movement is limited to linear movement.

---

[59] Patrick A. Anquetil, Hsiao-hua Yub, John D. Madden, Peter G. Madden, Timothy M. Swager, and Ian W. Hunter, "Thiophene-based conducting polymer molecular actuators", Smart Structures and Materials 2002: Electroactive Polymer Actuators and Devices (EAPAD), Yoseph Bar-Cohen, Editor, Proceedings of SPIE Vol. 4695 (2002).

[60] A. Dubey, G. Sharma, C. Mavroidis, S. M. Tomassone, K. Nikitczuk, and M. L. Yarmushc, "Computational Studies of Viral Protein Nano-Actuators", Journal of Computational and Theoretical Nanoscience, Vol. 1. 1–11, 2004, http://www.bionano.neu.edu/JCTN0101_003.pdf ; see also C. Mavroidis, A. Dubey, and M.L. Yarmush, "Molecular Machines", Annu. Rev. Biomed. Eng. 2004. 6:10.1–10.33, http://www.bionano.neu.edu/AR220-BE06-10_001-033_.pdf
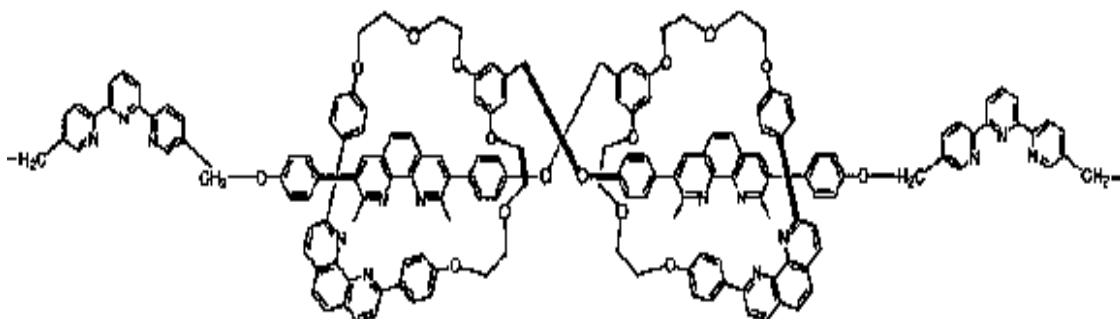
Figure 4.10 *Two interlocking rotaxane dimers*

The dimer in figure 12 is based on transition metal complexes that can be activated by set either an electrochemical or a photochemical signal. By sending a chemical signal to the molecule, one can readily convert the contracted compound (copper(I) complex) to the extended situation (zinc complex) and vice versa.[61] Unfortunately, these molecules are very difficult to synthesize.

---

[61] "Towards molecular machines and motors based on transition metal complexes" , Benoît Xavier Colasson, Christiane Dietrich-Buchecker, Maria Consuelo Jimenez-Molero, Jean-Pierre Sauvage, *Journal of Physical Organic Chemistry*, Volume 15, Issue 8 , Pages 476 – 483 Special Issue: Organic Reactivity Special Issue . Issue Edited by Mirjana Eckert-Maksi, Marie-Françoise Ruasse. See also "Chemically Induced Contraction and Stretching of a Linear Rotaxane Dimer", Maria Consuelo Jimenez-Molero, Dr., Christiane Dietrich-Buchecker, Dr., Jean-Pierre Sauvage, *Chemistry - A European Journal*, V8N6 , Pages 1456 – 1466, 15 Mar 2002, http://www3.interscience.wiley.com/cgi-bin/fulltext/91016413/PDFSTART

5. Early Manufacturing Architectures

Once a general-purpose manufacturing system capable of building duplicate and improved systems is developed, further progress and improvement can be rapid. However, it has not been clear what the first such system would look like, or how it could be built. This section proposes two architectures for primitive systems that could be developed with today's technology. Both use relatively uncomplicated nanomechanical systems, attached to larger actuators, to bind prefabricated molecular building blocks (MBB's) from solution and fasten them to chosen locations on the product. Each design is intended to be capable of manufacturing an array of independently controlled duplicate systems.

This section provides general background, then describes the physical layout and general functionality of each approach. Then it focuses on handling and joining MBB's and the functionality of MBB's and nanosystems. Finally it briefly analyzes the productivity of the general MBB-placement approach, and discusses how productivity could be scaled up.

5.1. Background

It is too early to tell whether the first molecular manufacturing systems will be based on solvent-immersed mechanisms assembling prefabricated molecular building blocks or on scanning probe systems doing machine-phase mechanosynthesis to build covalent solids. As nanoscale technologies and molecular manufacturing theory have developed, it has at times appeared that the most effective development approach would be to develop a scanning-probe system that could do machine-phase mechanosynthesis, and use it to construct, slowly but directly, a nanoscale system capable of using machine-phase mechanosynthesis to build duplicate nanoscale systems more rapidly.[62] Recent progress in molecular building blocks, along with more detailed understanding of how a primitive "wet" system could be improved incrementally to an advanced "dry" system, suggests that a "wet" context is also a good place to start. (It is worth noting that Drexler, who is often associated with "dry," highly advanced diamond systems, has always recommended starting with "wet" systems.)

Several different classes of molecules can implement an engineered structure. DNA forms predictable folded structures, and a variety of shapes have been made with it. Although double-stranded DNA is too compliant to make multi-nanometer structural components, stiffer four-stranded variants have been used. Nadrian Seeman has made crystals up to a millimeter in size, ordered to 1 nm resolution.[63] Protein design (which is easier than predicting how natural proteins will fold) has resulted in novel folds, and can be used to produce small complex shapes. Perhaps the most promising molecular backbone is being developed by Christian Schafmeister: a polymer that is relatively stiff

---

[62]   Robert Freitas and Ralph Merkle are proponents of this approach. See
         http://www.foresight.org/stage2/mechsynthbib.html and http://www.foresight.org/stage2/project1A.html
[63]   http://www.math.uwaterloo.ca/~bingalls/BIRS/abstracts/seeman.html

even in single strands, using spiro linkages between cyclic monomers. A library of monomers allows the design of a wide variety of twists and shapes.[64]

To form a solid product, blocks must fit together. See Fig. 5.1. To help with alignment and insertion, a completed layer of blocks should form shallow pits between the blocks into which the next layer of blocks will fit. A cube standing on one corner would have an appropriate shape and orientation to form a solid structure, but many shapes will work. In effect, each completed layer forms a regular grid of vacancies for the next layer of blocks to be placed into.
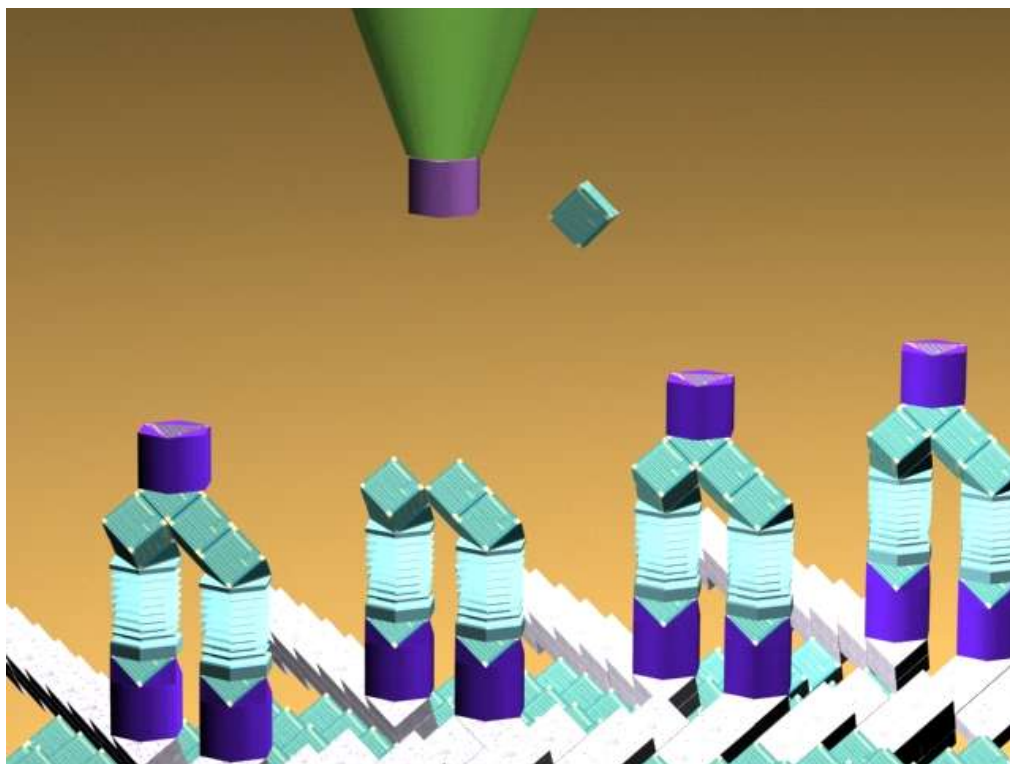
Fig. 5.1. ***Tattooing a Row:*** *Depositing MBB's via scanning probe to build an array of "Tattoo" block-placement machines. Note the orientation of the deposited blocks. The free-floating MBB is about the enter the corner-shaped cavity in the probe tip. The tip will then push the block into the nearly completed nanoprobe directly underneath it. When in operation, each nanoprobe in the row can be extended by establishing a voltage difference between the appropriate "x-wire" (the row of white conducting blocks that runs almost horizontally near the bottom) and "y-wire" (the conducting rows that extend toward the upper left).*

A 5-nm molecular building block could contain thousands of atoms. This is small enough to build via advanced synthetic chemistry (perhaps aided by self-assembly of multiple molecules), and to maintain a well-defined structure even if the framework were somewhat sparse, but large enough to include several molecular actuators, a molecular electronics component, a fullerene-based structure, or a fluorescent nanoparticle or

---

[64]   http://www.foresight.org/Conferences/AdvNano2004/Abstracts/Schafmeister/

molecule. The surface of the block would have room for several dozen amino acids or photochemical sites, the position and orientation of which would determine its bonding behavior.

5.2. "Tattoo" architecture

The Tattoo architecture for programmable heterogeneous assembly of MBB's is based on the possibility of making MBB's that will not aggregate or bond in solution, but will bond when pushed together by mechanical force. See below in "Handling and joining molecular building blocks" for discussion of several ways in which this might be accomplished.

Given a way to reversibly bind a MBB to a weakly binding site ("socket") attached to a scanning probe microscope tip, and a surface that (like the surfaces of blocks) will bond to a block that is pushed against it but will not accrete blocks randomly, the scanning probe tip can be used to deposit multiple blocks at selected locations on the surface. The tip with bound block might be used to image the surface, though fluid immersion (for non-contact or tapping-mode AFM), block fragility or unwanted bonding (for contact-mode AFM), and block conductivity (for STM) may present practical problems. Once a layer of blocks is completed, a second layer can be deposited on it, and so on. Solutions of different MBB's can be flushed in, allowing each layer to be made of as many block types as desired.

Once the ability to build patterns of blocks using a macro-sized scanning probe microscope is established, the next step is to build a "tattoo needle" made up of MBB's. This is a socket attached to a nanoscale actuator which can be individually activated—via electrical connection, if the actuator is electrical (optical actuators will not need physical connection). The actuator needs only one degree of freedom. Its purpose is to push the socket away from the surface with a displacement on the order of a nanometer. (The "needles" will be positioned near the product by a large positioner after blocks have bound to their sockets, and selected actuators will be activated to drive the blocks the final distance into the product.)

Construction of the socket will probably require special design. Several stackable blocks placed in a triangle will make a block-shaped cavity. In operation, the inner cavity must attract blocks but not bond to them; this can be accomplished by a charge on the inner faces opposite to the normal block charge. However, in solution, the cavity blocks must not aggregate with the product or with each other. Because multiple block types can be used, blocks with a single special face can be placed in a cluster to form the socket. A chemical post-processing step may be required to modify the special surfaces of the blocks. See Fig. 5.2A. Alternatively, a special prefabricated socket structure could be attached to a distinctive mounting point by self-assembly. See Fig. 5.2B.
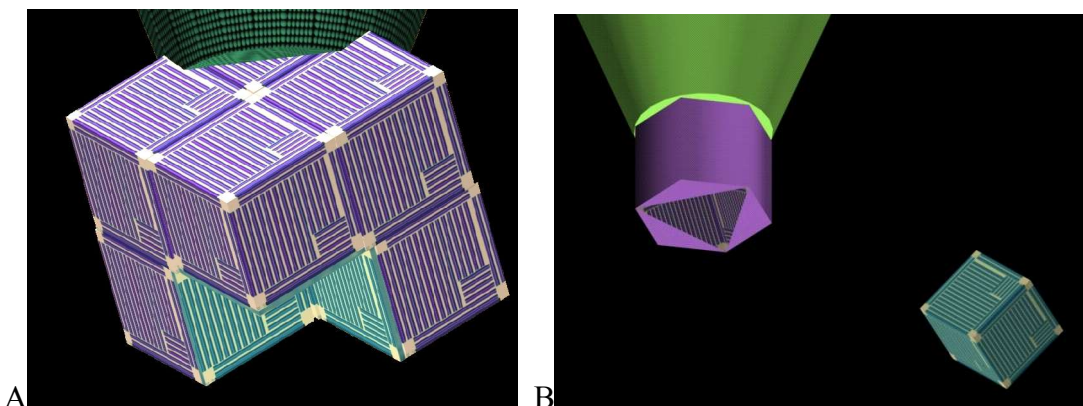
Fig. 5.2*. Two ways to implement a socket: A) Place several blocks to form a cavity, perhaps followed by post-processing to modify the inner faces. B) Use a special-purpose molecular building block.*

Once the actuator-and-socket machine is built, it can be used to deposit blocks on a surface held in front of it. If that target surface is attached to a three degree of freedom manipulator, which is used to reposition the surface relative to the deposition machine, then blocks can be deposited in selected locations and in layers. (If a scanning probe microscope is used as the manipulator for the target surface, then scanning the surface relative to the machine's socket might be used to image the surface, though there are several practical problems with this.) The goal is that the machine should be able to build a second, identical machine on the surface that is presented to it by the manipulator; and not only one, but a row of machines. This row of machines could be used in parallel to build an array, and these machines could then be used in parallel for higher-throughput manufacturing of larger arrays.

If the "tattoo nano-needles" can be independently controlled so that only some of them are made to deposit blocks at any given placement, then they can be used to build heterogeneous structures or regular structures with different periods than the machine spacing. Independent control can also be used for error correction; if a machine is known to be non-functional, another machine can be translated over its position to place the required blocks.

If different types of MBB-specific sockets can be built on different machines in the grid, then multiple MBB's can be mixed in one feedstock solution and each machine activated only when its selected type of MBB is wanted. This would remove the need to flush multiple solutions past the machines, eliminating the corresponding penalty in time, materials, and possibly errors resulting from contamination with the wrong type of block.

An even simpler variant of this approach might be useful to test the block-binding surface functionality before complicated MBB's and nanoscale machines are developed. A tower or needle built without an actuator but with a socket at the tip could be used to deposit passive chromophore-containing blocks by moving the surface it is attached to in a way that presses the socket against an opposing surface. The light from a simple heap of blocks would be visible with an optical microscope; several heaps spaced several microns

56

apart in an identifiable pattern would confirm success. A more ambitious goal would be to use the tower to construct more socket-tipped towers, and then test their functionality. With sufficient precision, an exponentially growing number of towers could be created.

5.3. "Silkscreen" architecture

The Silkscreen architecture is based on the idea of separating a solution containing MBB's from the substance or condition that would cause the blocks to bond together. Instead of the array of "needles" in the Tattoo architecture, the Silkscreen is a membrane with an array of holes. The membrane serves several purposes. It separates the feedstock blocks from the product, and can maintain distinct conditions (such as concentrations of zinc) on each side. Its primary purpose is to control the position and timing of block passage through the membrane to the product.

Each hole in the membrane contains an actuator which can reversibly bind to a single block, transport it through the membrane, block the hole to prevent mixing of solutions, and present the block to the product. (Something shaped like a cutaway wheel or disc, with a socket in its rim and mounted on a torsion hinge, can perform all these functions simultaneously with only one degree of freedom and no bearings. See Fig. 5.3.) Like the Tattoo approach, the Silkscreen approach repositions the nanoblock-deposition machine relative to the product using a macro-scale three degree of freedom manipulator.

The membrane would be closely fitted to the growing product, and could be aligned to it by local forces. A block passing through a hole in the membrane must be able to reach only one vacancy in the product; the block's motion can be constrained by the manipulator until it is bound to the product.

An initial Silkscreen membrane might be built by any convenient combination of self-assembly, synthetic chemistry, and lithography. The grid of holes could be created either by lithography or by a self-assembled membrane such as DNA tiles. Each hole would be filled by a molecular actuator system. Once constructed, the first system could be used to build larger membranes and improved designs.

The simplest membrane might have only one hole. Its actuator could be activated by light (which would not need to be focused) or by passing current between the product and feedstock side of the membrane. The hole might be constructed by slow lithographic techniques such as ion milling or dip-pen nanolithography (DPN). Even with only one hole, the size of the product would be limited only by the speed and reliability of deposition and by the range of the product positioner.
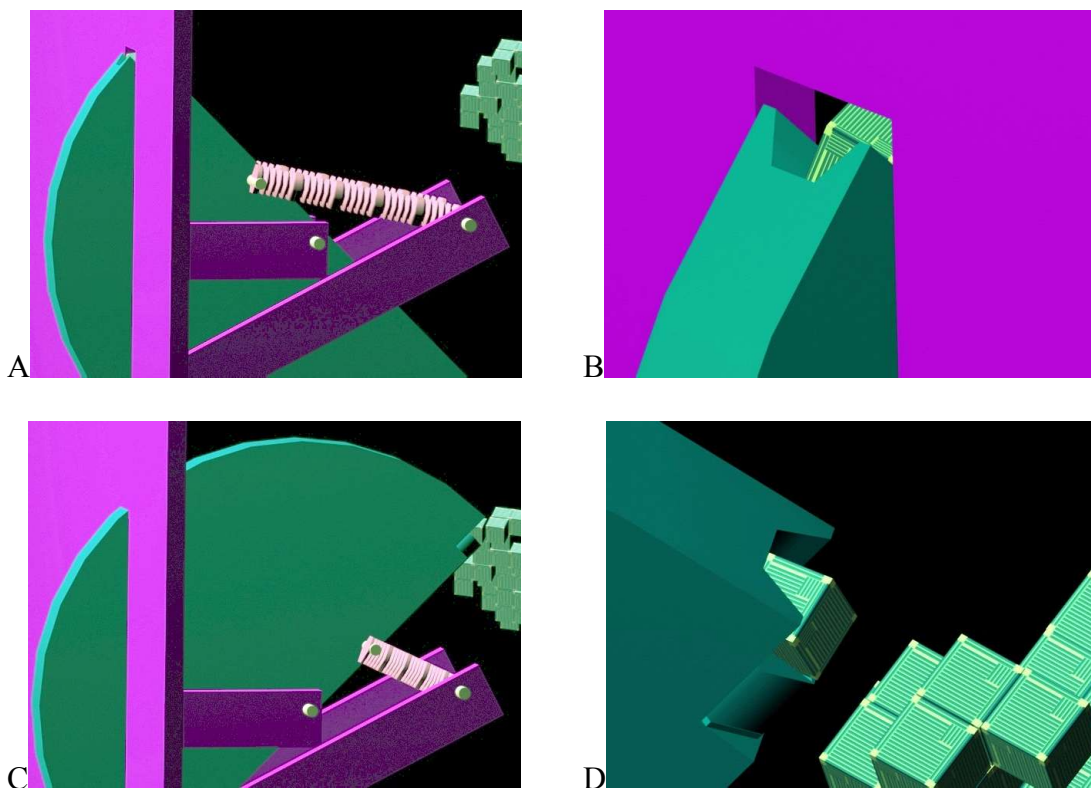
Fig. 5.3. A) The manipulator is in position to gather a block from solution. B) A close-up of the socket's position with a block inserted. C) The manipulator has moved to attach the block to the product. D) A close-up of the block about to be pressed into place.

After manufacture of a membrane, passive gaps and actuators that are jammed open could be sealed shut by (for example) putting half of a binary adhesive on each side of the membrane. A plug will form wherever a gap allows the two components to mix.

5.4. Molecular building blocks and nanosystem functionality

The simplest manufacturing systems only need to extend or retract a bound block; this requires only a linear actuator with one degree of freedom and a small range of motion, and a way to control individual actuators. There are several kinds of molecular actuators that may be suitable, and several kinds that are less suitable for one reason or another; actuators are explored in Chapter 4, "Molecular Building Blocks."

Speed and addressability will be important for any practical nanofactory. DNA binding actuation, though quite flexible, is also quite slow: many seconds are required for the strands to diffuse and link. Molecular precision and small size are important; this may rule out some actuators that depend on bulk effects.

Some molecules are responsive to light; they would have to be placed at least several hundred nanometers apart to be individually addressable,[65] but this may be acceptable in

---

[65]   Unfocused light can be delivered to a sub-wavelength region through proximal near-field effects, but would likely spill over to adjacent actuators.

early designs. Light has the advantage that it requires no physical structure to deliver it; it can be focused in a pattern from a distance. Also, light can be switched with very high bandwidth, though the response time of a slow actuator might negate much of this advantage. A focused pattern of light will have low spatial precision by nanoscale standards; this may be partially overcome if molecules can be found that are sensitive to specific wavelengths of light, so that several different actuators can be used within a single pixel of a multicolored pattern.

Several kinds of electrically driven actuators could be used, as explained in Chapter 4. A separate wire need not be run to every individual actuator. Even without digital logic, an X-Y grid of wires can be used to control an actuation system at each of the points where the wires cross.

Sensing will be important, not just to prevent accumulation of errors from generation to generation, but also for research: to provide early confirmation that new designs are working as intended. The important question will be whether a block has been placed as intended. The block can be detected by contacting it with a physical probe, which would be similar to a block-placement probe without the socket and with a weak actuator. Full extension of the probe would indicate an absent block.

Information must be returned from the nanoscale. Electrical signal return could use a simple mechanical switch, as suggested for digital logic. Information could also be returned optically; fluorescent nanoparticles can be held near quenchers when the probe is retracted, and extension of the probe would cause significant increase in fluorescence. Single-molecule fluorescence has been detected. If probes can be individually activated, then they can be spaced closer than the diffraction limit; probes spaced far enough apart could still be operated in parallel. Alternatively, careful detection of light levels could indicate how many probes within a single pixel had been extended, and this could be compared with the intended result; if an error was found, then single-probe actuation could be used to isolate it.

5.5. Throughput and scaleup of molecular block placement systems

Activating a molecular actuator via photons or electric fields might take on the order of a millisecond. Diffusion of MBB's to fill the sockets might take several milliseconds; this depends on many factors, including block size, the concentration of the blocks, and the viscosity of the solvent. If ten five-nanometer blocks per second can be placed, then a single placement machine would be able to deposit a square micron area of a single layer, five nanometers thick, in a bit over an hour, or a 100 nanometer cube in 20 minutes. The placement machine itself will probably be built on a 100-nanometer scale (20 blocks on a side), implying a very high manufacturing throughput relative to its size.

Diffusion and binding of blocks to sockets is probabilistic, and without the ability to detect when a socket is filled, a relatively long time must be spent waiting until it is very likely to be filled. If this is the limiting factor in deposition speed, the ability to sense when a socket is filled would allow faster deposition. Operations could be scheduled as blocks became available. This might allow an average of 100 or even 1000 blocks to be placed per second per machine. A system containing 10,000 machines placing 1000 five-

nanometer blocks per second apiece would build only a few nanograms of product per hour, a cube several microns on a side, but that corresponds to several billion blocks—enough to build a powerful CPU, for example.

6. Early Products

This section explores some products that might usefully be made by even a primitive version of a planar nanofactory. The ability to manufacture large (millimeter-scale or larger) arrays of strongly joined, nano-structured building blocks would make it possible to build a range of useful electronic and physical components. These could be higher performance and less expensive than components made by current manufacturing processes.

On long-duration or colonizing missions, the use of a single compact manufacturing system to make a range of intricate components would substantially reduce the amount of equipment required for self-sufficiency. Early manufacturing systems would make small, difficult-to-manufacture components of larger products, allowing a reduction in the equipment needed for self-sufficiency. Later systems could manufacture complete products, as shown in the illustration below.

6.1. Water Filter

The Filtomat M102C high performance self-cleaning water filter[66] can clean 110 gallons/minute with a screen area of 90 square inches. The smallest particle size particle that it can screen is 50 microns, though the manufacturer also makes thread filters going down to 3 microns.[67]

Building these filters, and all the different filters in their product line, requires a serious investment in equipment: about a hundred people with a few acres of manufacturing plants. A small-scale manufacturing facility that could build water filters in non-terrestrial locations would require about 3000 square feet and a few tons of heavy equipment.[68] If NASA establishes a significant Lunar or Martian presence, it will require the presence of manufacturing capability equivalent to the Filtomat factory, not just for water filters, but also for the range of products needed for exploration and life support. With an advanced nanofactory, however, such huge manufacturing plants would no longer necessary. The only necessities will be the nanofactory, energy,
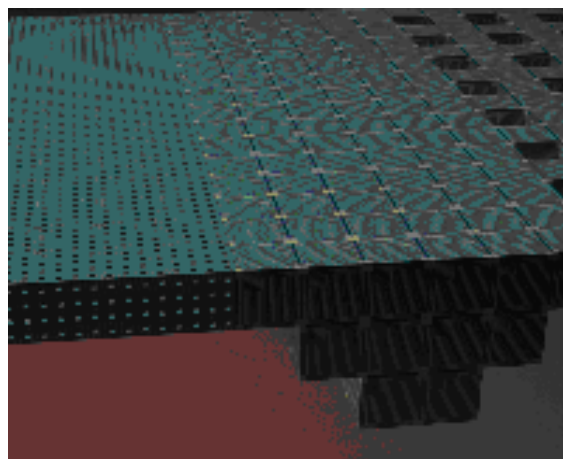
Figure 1. A filter with 0.3 nm pore MBB's; standard solid MBB's make up the central support structure (center) and coarser screens (upper right).

---

66   Filtomat, Inc., has been selected as the sole-source supplier for the Navy's CVN21 (next generation aircraft carrier) program. Product information at http://www.filtomat.com/100-series/100-c-series. Modeled in 3DMax by Anthony Mcfadden http://www.3dcafe.com/asp/industry.asp.

67   Under NSF/ANSI Standard 42 this is a Class VI filter (> 50 microns at 85% efficiency), where class I is 0.5 to < 1 micron and can generally be only be done by carbon/charcoal adsorption.

68   Personal conversation with Dr. Yossi Mor, V.P. of Business Development, Amiad USA, April 6, 2005.

61

and equipment to supply a few different types of molecularly precise MBB's. Although primitive nanofactories may not be able to build large products efficiently, they could build the most intricate and delicate components of a wide range of products, reducing the amount of special-purpose equipment required.

Not only would a desktop nanofactory make it possible to have a much lower overhead for manufacturing support, but the final products will be superior. The screening ability of an M102C water filter, for example, could be improved from 3 microns to .3 nanometers: an improvement of four magnitudes. Since water vapor has a kinetic diameter of .265 nanometers, nothing else would get through (except some light gases and small ions). Even with such extreme filtration, the pressure drop of 2 psi across the current membrane could be lowered to about 0.5 psi.[69]

Using the nanofactory/MBB approach, such an improved filter would be built using two types of input MBB's: solid structural MBB's and MBB's with .3 nanometer pores through them. Since a single layer is desirable, the pore MBB's need to be supported by a fractal structure of the solid MBB's. Figure 2 shows an interface between pore and solid MBB's. Even without specialized pore MBB's, simply omitting blocks would produce a filter with MBB-sized holes. Such a filter would pass larger molecules such as ethanol, benzene and other larger organics, but would exclude all infectious organisms. On the other hand, if the specialized pore blocks could be functionalized to repel other molecules (such as sodium ions, as is done by aquaporins), then the filter could also be used to purify salt water and industrial pollution.

The same technique could be used to build air filters to screen out nanoscale particulates; functionalized pore MBB's would be needed to reliably filter out carbon dioxide or carbon monoxide.

6.2. Artificial Kidney

If the nanofactory had inputs of biocompatible pore MBB's with functionalized pores of different sizes, then a different "blueprint file" could produce a double-filter device that screened blood for particular molecules such as urea; this could produce the core of a portable or even insertable dialysis machine.[70]

In a weight-limited, long duration mission, NASA could not afford to pack a dialysis machine. But in an emergency, it would be very nice to have the capability to build one, or any of a vast array of equally complex machines. Just as a printer can print both novels and blueprints, the flexibility of the nanofactory approach extends far beyond water filters, air filters, and kidney dialysis machines.

---

[69] Personal email from Gayle Pergamit, Biophiltre LLC and Agua Via LLC, April 8, 2005

[70] Gayle Pergamit and Martin Edelstein, "Near-term Medical and Environmental Products Built at the Atomic Scale: Artificial Kidney and Water Purification", 1st Conference on Advanced Nanotechnology: Research, Applications, and Policy, Oct 22-24, 2004, http://www.foresight.org/Conferences/AdvNano2004/Abstracts/Pergamit/

6.3. Laptop Supercomputer

Using a set of four types of MBB's (conductor, insulator, p-type semiconductor, n-type semiconductor), about 300 MBB's could form a NAND gate, as shown in Figure 3. For five nanometer MBB's, the design rules would need to change from those in use for .20 micron photolithography, but perhaps not drastically; since each MBB contains between 130 and 3,500 atoms (up to several nm in size), conventional transistor designs may still work. An advantage is that, IC "chips" would no longer be limited to two dimensions, so circuit densities per volume would increase significantly; the a relatively primitive nanofactory could produce laptops two or three magnitudes more powerful than those available today.

Figure 3. A design for a NAND gate made up of conducing, insulating, p-type, and n-type MBB's.

With a relatively small set of more complex electronic MBB's, much greater performance gains become possible. The research in transistors made of carbon nanotubes[71], single cobalt-atoms[72], and rotaxane[73] indicates that logic circuits can be built from supramolecular structures of under 1000 atoms. If these can be built (either by self-assembly, positional assembly, or some hybrid) then MBB's could make it possible to build a computer system that can run 10^{11} MIPS, nine magnitudes better than today's products.

---

[71] Shalom Wind, Joerg Appenzeller, Richard Martel, Vincent Derycke and Phaedon Avouris, "Vertical Scaling of Carbon Nanotube Field-Effect Transistors Using Top Gate Electrodes", May 20 2002, *Applied Physics Letters*, http://domino.research.ibm.com/comm/pr.nsf/pages/news.20020520_nanotubes.html

[72] "Vibration-assisted electron tunneling in C140 single-molecule transistors", A. N. Pasupathy, J. Park, C. Chang, A. V. Soldatov, S. Lebedkin, R. C. Bialczak, J. E. Grose, L. A. K. Donev, J. P. Sethna, D. C. Ralph, and Paul. L. McEuen, Nano Lett.; 2005; 5(2) pp 203-207, http://www.lassp.cornell.edu/lassp_data/mceuen/homepage/pubs.html

[73] Whence Molecular Electronics? Amar H. Flood, J. Fraser Stoddart, David W. Steuerman, and James R. Heath, *Science*, Dec 17 2004: 2055-2056, http://newsroom.ucla.edu/page.asp?RelNum=5718

7. Incremental Improvements

Although it would have practical applications, the main purpose of a primitive molecular manufacturing system would be to build a more advanced system than itself. The ultimate goal is an easy-to-use nanofactory that can rapidly build a complete duplicate nanofactory, as well as a wide range of high-performance products, from inexpensive feedstock. Although this is vastly beyond the primitive nanofactory, a set of incrementally developed capabilities can form a sequence of transitional improvements ending at the goal. These capabilities could be added one at a time, which would allow each capability to be developed and tested separately. This approach would also produce spinoff manufacturing systems which would be increasingly useful.

The capabilities to be added are largely independent, so could be added in any of several sequences of development. To the extent it is possible to validate future designs, planning several steps ahead may guide design choices and allow faster development. Although detailed designs for each capability would be beyond the scope of this paper, this section explores some of the issues and techniques that might be involved in improving the nanosystems, removing the solvent from the nanofactory interior and product space, manufacturing blocks internally, and improving the mechanical design.

7.1. Improving nanosystems

Digital logic is not necessary for the earliest systems, but will quickly become convenient and then necessary as the system scales up. For example, multiplexers and demultiplexers can greatly reduce the number of wires that must be connected to the nanoscale. Molecular electronics is a rapidly developing field, with transistors being made out of carbon nanotubes and single organic molecules. Simple electrical switches operated by molecular actuators may also be used; the several-nm gap that would be left by removing a conducting nanoblock from a chain is too large for significant tunneling of electrons.

With more efficient and intricate mechanical systems, the manufacturing process can be made more flexible and functional without sacrificing efficiency. Actuators can be improved at any stage. Although bearings are high on the list of priorities, a lot can be accomplished with pantograph-like trusses. Covalent bonds do not fatigue; they can be flexed indefinitely without weakening, and the range of motion is considerably higher than with MEMS. Once stiff smooth surfaces can be constructed, a new type of bearing may be built; see "Bearings" in the "High Performance Nano and Micro Systems" section.

7.2. Removing the solvent

Solvent is convenient for diffusive transport, but it creates drag. Once feedstock blocks are bound to manipulators, there is no need for solvent unless it supports some particular operation, such as supplying ions for bonding or for actuators (though the preferred actuators will not use ion movement but electronic shifts). There are many ways of

64

fastening blocks onto the product that do not require solvent. Pure electrostatic actuators are excellent choices for a dry environment, and should also work in clean non-polar solvent.

A Silkscreen type factory could maintain solvent on the feedstock side of the membrane and low-pressure gas on the other side. In early designs, some solvent would presumably enter with the blocks or otherwise leak through. The solvent, operating temperature, and block material should be chosen to limit adsorption and promote evaporation in the "dry" interior; this implies a pressure difference between the "dry" and "wet" sides, which may limit the size of the membrane. Later designs, with smaller feedstock molecules and better construction methods, should make it possible to operate without any solvent leaks, allowing operation at equal temperatures and pressures on both sides; that would cause the solvent to condense on the dry side, except that no solvent will be present.) Liquid xenon may be a good choice for a solvent.[74] A major advantage of xenon is that it is chemically inert, so it will not combine with any reactive molecules used in mechanosynthesis.

Many reactions require no solvent; the combination of mechanically mediated reactions and lack of solvent is unfamiliar to most chemists, but has been demonstrated in the laboratory for a few reactions, and theory suggests that it should be a very rich chemical domain. Molecules with radicals that could not survive in solvent can be maintained in vacuum, held separate by mechanical control (acting on the non-reactive portions of the molecules) until the radicals are reacted according to plan. This should provide access to chemistries that require the maintenance of conditions far from equilibrium, and should facilitate the step-by-step construction of highly crosslinked covalent solids.

In the absence of high drag and friction, nanoscale structures have resonant frequencies around the gigahertz range. This implies that a dry nanofactory could work much faster than the wet versions, assuming that the actuators were responsive. Electrostatic actuators should be very fast, limited only by the current carrying capacity of the wires.

7.3. Manufacturing blocks

Blocks in the five nanometer size range are too large to be easily synthesized (in fact, they may be expected to be quite expensive), and do not diffuse as rapidly as might be wished. However, blocks made too much smaller might not have enough internal space to hold the desired functionality, and could reduce throughput. The nanofactory will be more efficient, its products will be cheaper, and it will be easier to design and create new block types, if small molecules can be used as feedstock, combined into large blocks internally, and the large blocks placed in the product layer. (A variant of this is to use one system to create and stockpile the blocks, and then another system to assemble them into the product. This requires essentially the same functionality as the integrated nanofactory.)

Small molecules will not have room for elaborate molecular attachment mechanisms, but there are many approaches to joining molecules. One approach is to use reactions that

---

[74] "Xenon is also nearly inert chemically, but because of the xenon atom's large electric polarizability the liquid is an excellent solvent-particularly for aliphatic materials."
http://cfa-www.harvard.edu/Walsworth/Activities/Physical%20Science/liquid-xenon.html

form bonds between structures of a few atoms when pressed together, but do not crosslink rapidly in solution. If the reaction releases a small molecule, then the factory design must allow it to escape, but many reactions (for example, cycloadditions) rearrange bonds to produce a single product molecule. Another approach is to begin with inert molecules, then make them reactive by removing passivating structures, and finally position them so as to bond spontaneously in the desired configuration.

There is a wide variety of ways to make programmed parts from molecular feedstock. Even a preliminary exploration of the available reactions would require many pages. At one extreme, molecular fragments weakly bonded to "tool tip" molecules can be deposited on a depassivated surface in vacuum (or noble gas). This is called "machine-phase chemistry." The ability to use reactive species such as unprotected radicals, carbenes, and silylenes increases the chemical options, but requires extreme cleanliness.

In solvent, molecules can be forced to bond to selected locations on a terminated surface. The use of solvent allows feedstock to be delivered to the tool tip by Brownian motion. Simply holding the molecule in the right location can increase its effective concentration by many orders of magnitude. Holding it in the right orientation, or applying pressure or electric fields, can further speed its bonding to the surface, allowing the use of high-barrier reactions to minimize unwanted random deposition. The use of passivated surfaces, with pressure or other targeted stimulus being required to cause a reaction, would somewhat reduce the need for absolutely clean working conditions.

A linear polymer can be built which then folds up into a desired shape. (Post-folding bonding can be used with some chemistries to increase strength and stiffness.) The programmable control system need only select which monomer to attach next, and force it to bond to the chain. This is the function performed by the ribosome.

Component molecules can be used which hold their position by electrostatic or van der Waals attraction or hydrogen bonding.

Deposition reactions can happen through conventional, non-mechanical chemistry, with the location controlled by mechanically mediated protection or deprotection of deposition sites. In this case, protection can be either chemical or steric (physical blocking of a site).

It may be useful to build blocks out of molecules that are already structural or functional components. Alternatively, very small molecules could be aggregated to build arbitrary structures; this would make the factory extremely flexible, and allow for rapid improvements in product design without changing feedstock or factory design. Providing diverse functionality from a few simple molecules appears feasible, because conductors and insulators are all that are needed to build electrostatic actuators. With actuators and mechanical switches, digital logic and sensors can be built.

7.4. Improving the mechanical design

Directed internal transport of product components within the factory would be useful for several reasons. It would allow broken machinery to be bypassed. When a factory that manufactures blocks internally is forming a product with large voids, internal transport

would allow the transfer of blocks from the entire factory's block-manufacturing machinery to dense regions of the product, alleviating a potential bottleneck.

Specialized block-manufacturing equipment would increase the efficiency and speed of the molecular manipulations. Instead of programmable robotic systems, fixed-path machinery could do common operations to combine small molecules into frequently used structures. A well-designed "mill" would use only one degree of freedom in its operations, making it straightforward to power and synchronize with other machines.

8. Covalent Solid Nanosystems

Diamond, graphene, and fullerene can be produced by a wide range of reactions. New low-temperature records are continually being set for fullerene synthesis, and solution-phase synthesis of C_{60} (buckyballs) is being pursued. The flexibility of "dry" chemistry appears well-suited to building diamond lattice by incremental deposition. Accordingly, this section will assume that a high performance nano-manufacturing technology will be able to use materials of diamond-like strength. More information on high-performance covalent solid nanosystems can be found in *Nanosystems*[75] and *Nanomedicine I*.[76]

This section includes discussion of the convenience of surface forces for manipulating micron-scale components; this was developed in the course of this study. A mechanical fastening system that requires no external manipulation, previously developed by Phoenix,[77] is presented because it may be useful for fastening sub-micron components in a planar assembly process. The rest of the chapter summarizes work from *Nanosystems* on an efficient class of bearing; electrostatic actuators, including a reversible motor/generator; digital logic; and mechanochemical power conversion. Many other mechanical designs, including nanometer-scale gearboxes, teragravity centrifuges, and a wide variety of sensors, are developed in *Nanosystems* and *Nanomedicine I*.

8.1. Van der Waals Robotic Gripping At Micron Scale

The use of micron-scale mechanical systems (MEMS) to manipulate micron-scale objects has been hampered by attractive forces between surfaces which make it difficult to let go of anything. However, with only slightly more sophisticated robotics, these same forces can be the basis of a very convenient and general-purpose method of handling objects. This section explains how such an approach would work, with preliminary engineering calculations.

8.1.1. Van der Waals Forces

Due to electron interactions, an attractive force will develop between almost any two surfaces in close proximity. The force is actually a collection of mechanisms totaling on the order of a nanonewton per square nanometer for surfaces in close contact.[78] For a cubic micron part, massing on the order of 10^{-15} kg and weighing $\sim 10^{-14}$ N, the contact force of a single square nanometer provides 10^4 times the force of gravity. Thus there is no need for any stronger gripping mechanism to manipulate parts: simply touching a plate to a surface will attach the two. Although the force can be screened by liquids, in gas or vacuum it provides a convenient means for handling or attaching micron or sub-micron blocks.

---

[75] http://www.foresight.org/Nanosystems/toc.html
[76] http://nanomedicine.com/NMI.htm
[77] Chris Phoenix, "Design of a Primitive Nanofactory" Sec. 3.2,
       http://www.jetpress.org/volume13/Nanofactory.htm#s3.2
[78] See *Nanosystems* section 3.5.

The force between two surfaces in close contact can approach 5% of the tensile strength of diamond. This indicates that closely fitting surfaces which must be pulled apart must be strongly built. However, the force falls off rapidly with separation. According to calculation, separating two surfaces by 0.2 nm (~ 1 atomic diameter), for example with atomic-scale spacer bumps, will reduce the force by almost an order of magnitude.

According to Nanosystems Fig. 3.10 d, the potential energy per square nanometer of two diamond surfaces touching (0.2 nm apart) vs. separated is about 225 zJ. (The PE at 10 nm separation is 0.1 zJ, well below thermal noise.) If 100 kT energy is taken as the criterion for a reliable joint, then the contact area should be at least 1.8 nm^2. If spacers are used, or materials with lower Hamaker constant, then the area would need to be increased. For example, with 0.2 nm spacers, the energy per nm^2 is about 56 zJ, and 7.4 nm^2 of contact area would be needed.

8.1.2. Grippers

As implied above, a gripper need be no more than a square nm or so of contact area. The end of a structural beam will suffice; there's no need for "hands" or "fingers."

One useful approach could be to make the end of the manipulator hemispherical and provide a matching cavity in the surface to be picked up. This will allow the object to rotate around that point of contact. A tripod of transversely and longitudinally stiff arms, or a hexapod of longitudinally stiff arms, will be sufficient to orient the part in space. In the latter case, if the length of the arms is variable, the result is a Stuart platform.

If an arm-and-wrist positioning motif is desired, rather than multiple simple arms, then the gripping area must be able to control rotation as well as position. A single 4-square-nanometer contact point vs. a micron-long cantilever gives a mechanical disadvantage of 1:1000. This indicates that several contact points spaced several nm apart will be desirable. To control shear (slippage) and rotation, the surface cannot be completely smooth. However, atomic surfaces are never completely smooth. The grippers can be corresponding crystals with proper register and spacing. For more positive control, larger surface irregularities can be used.

8.1.3. Transferring and releasing

Because gripping requires no moving parts, any surface can be thought of as a gripper. If well-designed contact points are built onto the surface, it will be possible to deposit and retrieve the part. This section discusses the design of such passive contacts, as well as a simple active contact that allows deposition and retrieval. Passing the part between two active-contact grippers will require no extra design work.

If a block is held at opposite ends between identical grippers which are slowly pulled apart, the block will randomly stick to one or the other. If one gripper is slightly stronger, the block will stick to that one more of the time. There are two different strength parameters. One is the force holding the gripper to the block. The other is the potential energy of the joint. If the geometries or materials differ, then a joint with higher force at contact may have lower total potential energy. Vibration (including thermal vibration) might cause a slight random separation, reducing the force at that joint and causing the

joint release more easily than its force at contact would indicate. In the general case, then, it will be important to calculate both energy and force. But if the grippers are similar, then either can be used.

As calculated above, an area of a few square nanometers is sufficient to hold a block against thermal noise. By extension, a difference in area of a few square nanometers is sufficient to determine which of two opposing grippers a part will follow when the grippers are pulled apart. Assume force and energy are proportional to area for all grippers. Let A be the area required to hold a part against thermal noise. Then a simple approach to setting down a block and picking it up again would be to put three contact points, each of area A, on the end of the robot arm, and two on the wall. To stick the part to the wall, move it into contact, and then withdraw the contact points one by one. To remove it from the wall, simply retract the arm with all three in position, and the block will come loose. A single plate that is flexible or hinged, and is peeled up from one side, can be used instead of three separate contact points.

A little area and a degree of freedom of actuation can be saved by use of a plunger on the robot. Let the plunger have area P, which may be significantly smaller than A. Put a patch on the wall of any size W greater than A+P. Put a plate on the robot of any size greater than W+P. With the plunger retracted and even with the robot's plate, the robot can pick up the part. To deposit the part, the robot extends the plunger (or equivalently retracts the plate). The plunger holds the part to the wall until the plate is withdrawn, then is itself withdrawn. Note that this design gives much freedom to the size W of the wall mounting. The only restriction is that the force on the plunger not be great enough to cause bond rearrangement or cleavage.

Yet another way to transfer parts involves twisting the part to partially break contact. If the robot twists its plate sideways while keeping it at the same spatial position, it will twist away from the part and can then be removed. However, if the robot moves its plate sideways while twisting, then the part will twist away from the wall and will come away with the plate.

8.2. Mechanical fastening: Ridge joints

For some applications, such as fastening micron-scale blocks together without complicated manipulations, it may be desirable to use a strong mechanical joint that requires low insertion force, but activates itself to lock in place. Surface forces can be used to power the mechanism of the joint. One such joint is the "expanding ridge joint."

Each mating surface is covered with small "ridges" that are roughly triangular in cross section. See Fig. 6. All exposed surfaces are non-reactive (e.g. hydrogen-passivated diamond). The ridges on each face interlock with the ridges on the opposing face. As the joint is pressed together, the ridges split and expand sideways. The sloped faces of the ridges are not smooth, but are shaped to grip the opposing ridge, with scallops deep enough to form overhangs when viewed perpendicular to the block face. A scallop is chosen instead of a sawtooth or ratchet profile in order to avoid crack formation at sharp concave angles. Scallops also make assembly motions smoother, and allow the un-powered assembly described below. The expansion of the ridge opens a space in its

70

center, which is then filled by a shim which sits above the almost-closed gap between the two halves of the ridge. Once the shim is in place, the volume of the joint cannot easily be compressed, and the surfaces of the ridges cannot easily slide past each other; pulling apart the joint would require compressing a solid mass of diamond by several percent or breaking at least half of the ridges simultaneously. If the ridges all run in the same direction, the joint may be able to slide freely. Crossed ridges will produce a joint that is quite stiff against shear.
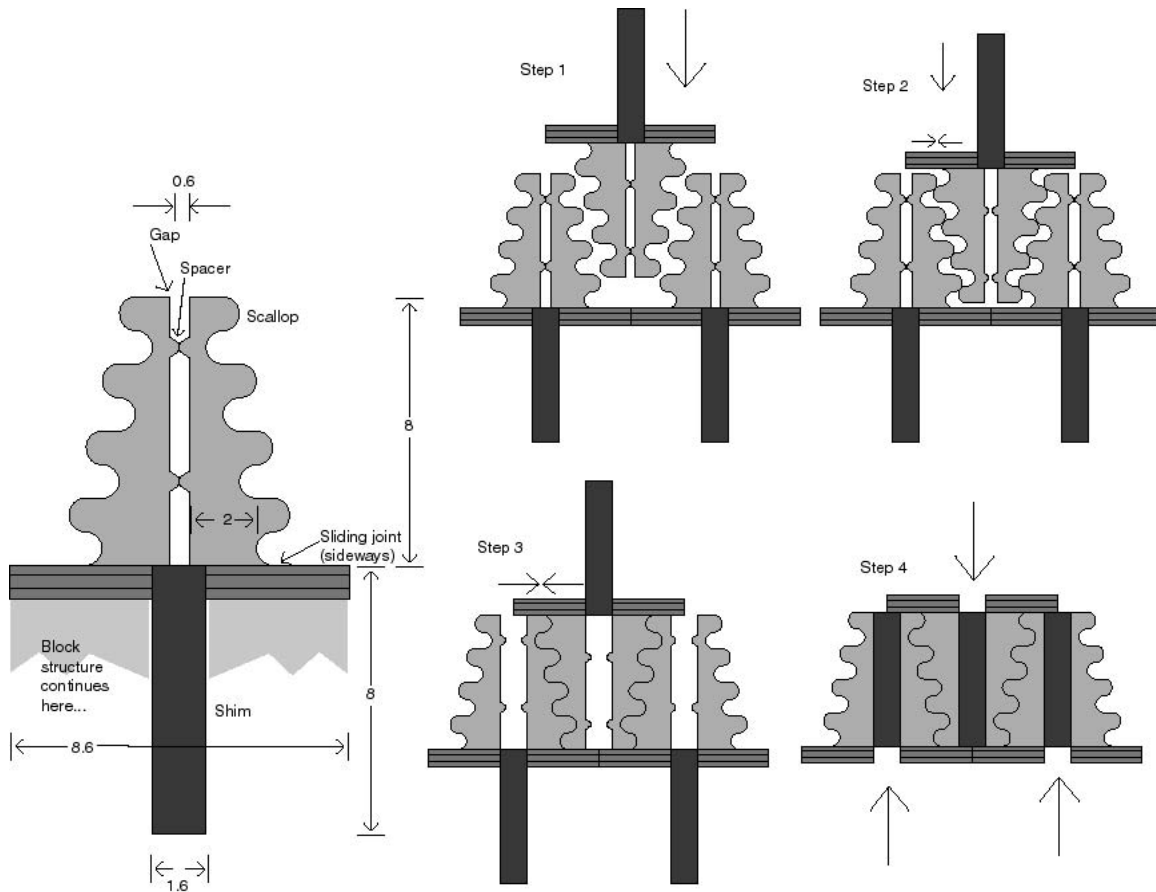
Fig. 6. Operation of a ridge joint. Ridges are pulled apart by surface forces without external actuation or manipulation.

As opposing ridges approach each other, the scallops pass in close proximity. Surface forces between the ridges will tend to pull the ridge apart. When the ridges are fully pulled apart, surface forces can pull the shim into position and hold it against thermal noise. Careful design to balance the surface forces may allow this approach to work with as little as 12 nm^2 of ridge surface. If the shim is retained from entering the gap, then the mechanism will form a weaker and reversible joint, useful for placing blocks in a temporary configuration.[79]

---

[79]   See Sec. 3.2.1 and 3.2.2 of the "Primitive Nanofactory" paper for detailed calculations and further discussion of the applications of ridge joints. http://www.jetpress.org/volume13/Nanofactory.htm#s3.2.1

8.3. Efficient bearings between stiff surfaces

Although the granularity of atoms makes a perfectly smooth surface impossible, low-friction bearings can still be constructed out of stiff materials. Mechanically, atoms are soft, rather than hard-surfaced spheres; bonds are also somewhat compliant. Atoms overlap when they bond, reducing the irregularity of the surface.

If the atoms on two facing surfaces have the same pattern and spacing, the atoms of one surface will fit between the atoms of the other, requiring high force to slide the surfaces. To prevent this, the atoms can be placed out of register, with different spacing or orientation; this can be implemented by using dissimilar materials, by rotating one of the surfaces, or by choosing different crystal planes for the surfaces. In this case, the transverse forces on the atoms will almost completely cancel, leaving a very small net force (and hence very low static friction) at any displacement. Superlubricity is a condition of very low friction between surfaces with atoms in non-corresponding positions. Superlubricity has recently been observed between graphite sheets rotated to out-of-register angles.

Drexler has proposed that nested hoops with different numbers of atoms on the inner and outer surface should show a similar effect, especially if the least common multiple of the numbers is large. (Relatively prime numbers have large LCM, though a small common factor may be desirable to maintain symmetry and reduce oscillation.) Nested carbon nanotubes have been observed sliding and rotating freely, an apparent example of this prediction. This would allow building low-friction rotational bearings.

Because atomically precise surfaces can slide past each other without stiction or wear (and for some surfaces, also with low drag), there is no need for lubricants. This is fortunate because a single atom is larger than the gap between two flat stiff surfaces. (This implies that such surfaces form a sliding impervious seal.) Without lubricants, the perpendicular stiffness of a sliding bearing is high, being a function of surface forces between two stiff surfaces.

Another kind of bearing uses a covalent single bond. This is only suitable for relatively low loads, but may be useful in some small machines. It is expected to have especially low drag.

Efficient nanoscale bearings are expected to have effectively zero static friction. The net force exerted by a bearing surface will usually be far smaller than the forces used to drive the machinery. Low-speed, high-efficiency systems may use driving forces even smaller than the forces exerted by efficient bearings; however, the energy barriers created by bearing forces will be lower than thermal energy. This means that even with a vanishingly small driving force, thermal energy will move the system past the barriers. Systems with no static friction can be run as slowly as desired. Dissipation increases at least linearly with speed, so slowing down a system by a factor of 10 will allow it to dissipate at least 10 times less energy in performing the same operation. Some of the orders of magnitude increase in nanoscale manufacturing throughput and power density can thus be traded for significant improvements in efficiency.

Mechanical parts interacting at high speed lose some energy to vibration. For example, rapid compression of a component will require a certain amount of energy. If the compressed object does not have time to equilibrate during compression, then some of the energy used to compress it will be lost. Equilibration time is limited by the speed of sound in the material, but the speed of sound is high in stiff materials. Softer materials, including many biological structures, will take longer to equilibrate, so cannot move as fast with the same efficiency.

8.4. Electrostatic actuators

At large scales, electrostatic actuators require high voltages and have low power density. However, a potential difference of 5 volts across a gap of 4 nm will produce a force of 7 pN/nm^2, or 1 nN from a 12 nm square plate. This is a usefully large force for nanosystems.

An efficient 10-volt electrostatic motor can be built on a 50-nm scale. With a rim speed of 1000 m/s, which is within diamond breaking strength, the power density of such a motor would be awesome: 10^{15} W/m^3. Even at a 1 m/s rim speed, 10^{12} W/m^3 is several orders of magnitude better than any existing natural or manufactured motor. By running it in reverse, the same device would become a generator; in effect, its design is "reversible" in the sense of reversible logic, since it can run in either direction without converting significant amounts of energy to heat.

Diamond is an excellent insulator. Some carbon nanotubes are excellent conductors. Thus an all-carbon system would not be limited in its ability to handle and use electricity.

8.5. Digital logic

A lower bound for the performance of digital logic can be set by a simple, easily analyzed, purely mechanical design. Nanoscale rods that move to block the motion of other nanoscale rods can implement logic gates. A logic gate built with this approach might occupy a volume of 10 nm^2, switch in 0.1 ns, and dissipate less than 10^{-21} J per operation. Computers could perform 10^{16} instructions per second per watt.[80]

The mechanical approach to logic is relatively insensitive to material choice. Because it does not rely on electronic effects, components can be packed tightly without the limits that would be imposed on electronic logic by electron tunneling. Error rates can be extremely low, because an error would require a logic rod to slip past a physical obstruction. Thus, although mechanical logic is surely not the fastest switching element, it may be approach maximum compactness, perhaps even competing with molecular logic.

8.6. Mechanochemical power conversion

In reactions between small molecules or non-stiff components of large molecules, the bonding force drives the reaction to completion quite rapidly. However, if reactants are stiffly held, the reaction can constrained to move slowly through the intermediate stages.

---

[80]   See Nanosystems Ch. 12.

This will exert a force on whatever is holding the reactants, and energy can be extracted from this force. Because this is not a heat engine, it is not limited by Carnot efficiency; in theory, nearly 100% of the chemical energy can be recovered. Electrochemical conversion, carried out by fuel cells, also is not Carnot limited. Drexler estimates that the feasible power density of mechanochemical energy converters is on the rough order of 10^9 W/m^3. Freitas calculates (*Nanomedicine I* 6.3.4.5) that the power density of a single pore in a proton exchange membrane may be on the order of 10^{11} W/m^3, though of course fluid volume and surrounding structure would reduce that figure considerably.

9. Advanced Nanofactory Architecture and Operation

Although any design for an advanced nanofactory must be tentative, the physical architecture and operating parameters of a reference design can be sketched out. The design presented here modifies the Burch/Drexler planar assembly architecture[81] and extends it with concrete physical suggestions and calculations, using some calculations and formulas from *Nanosystems* and applying earlier architectural and quantitative work done for the Phoenix "primitive nanofactory" convergent assembly system.[82] This chapter discusses transport of blocks within the factory and placement in the product, control of the factory including a possible data file organization, block fabrication machinery, physical layout, power and cooling, maintenance of internal cleanliness, and reliability in the face of random damage.

The planar assembly structure proposed for the primitive molecular manufacturing system is also an effective structure for high-performance nanofactories. The "nanoblocks" that would be attached to the growing product would be somewhat larger than the molecular building blocks used in the bootstrapping design, on the order of 100 nm to 1 micron. (A 200-nm block contains about a billion atoms.)

Blocks would be manufactured from small feedstock molecules, and then transported to the assembly face. Maximum product deposition speed would be on the order of 1 mm/second (block placement frequency increases as block size decreases, so deposition rate is independent of block size), though the rate of deposition would in many cases be limited by the rate at which blocks could be fabricated unless the blocks were prefabricated. Feedstock and cooling fluid would be distributed to the input side of the nanofactory via pipes. Design and analysis indicate that a 1-kilogram factory might manufacture and deposit 1 gram of product per second. In a kilogram of nanomachines, errors are inevitable, so each section of the factory must be designed for error detection and redundancy.

9.1. Block delivery

Product deposition speed will be limited by three factors: block delivery speed, block placement speed, and block fabrication speed (unless blocks are prefabricated). A kilogram of blocks distributed over a square meter would make a 1-mm thick layer. However, the use of high-strength covalent solids would allow the product to have very low density; product structural members may be inflated volumes with walls just a few microns thick, or composed of lightweight truss or foam. One kilogram of well-built covalent solid (e.g. diamond) would be able to replace at least 100 kg of today's materials, with even greater weight savings for active components. Accordingly, it is reasonable to provide a large product deposition area even for a kilogram-per-hour factory, since it will be able to make products that today would weigh hundreds of kilograms.

---

[81] An illustration of this architecture is online at http://lizardfire.com/nanofactorySS/index.htm

[82] C. Phoenix, "Design of a Primitive Nanofactory," Journal of Evolution and Technology #13 (2003) http://www.jetpress.org/volume13/Nanofactory.htm

For efficient operation, the fabrication of blocks can continue during times when the mass deposition rate is lower than the mass fabrication rate; and when high concentrations of mass are to be deposited, blocks should be delivered from the entire fabrication volume. This means they will have to be transported from the area where they are manufactured to the area where they are deposited.

Blocks will be large enough to be handled mechanically, and can be transported by any convenient mechanism. A layer of small "cilia" that pass the block overhead is mechanically simple and allows redundant design. Grippers are unnecessary—contact points for surface forces are sufficient.

Depositing a thin, solid column of blocks at maximum speed requires blocks to be delivered from all over the factory to a small area. This determines the thickness of the routing structure, which is roughly the same as the desired thickness of the deposited part divided by the ratio of delivery speed to deposition speed. For example, to allow deposition of a 1-cm solid block at 1 mm/s, if the blocks can be moved internally at 1 cm/s, then the routing structure must be about 1 mm thick.

9.2. Factory control and data architecture

There are enough different ways to implement digital logic that it seems safe to assume that any system capable of building a nanofactory can also include large numbers of digital computers. General-purpose (microprocessor style) computation is energetically expensive because bits must be erased at an irreducible cost of $\ln(2)kT$ per bit, so it will be infeasible to use many CPU cycles per atom in a high-throughput kilogram-scale nanofactory. But special-purpose logic (state machines) can be used to good effect in controlling repetitive tasks such as feedstock placement, and need not erase bits. To plan the handling of larger blocks (many millions of atoms), general-purpose computation will add minimal cost.

The control system will be deterministic and detailed, making it possible to specify any nanoblock at any position. The blueprint can still be small, since one nanoblock design can be re-used at many locations. Arrangements of nanoblocks can also be copied (tiled) and used to fill solid geometry volumes.[83]

There are only about 10^{15} 100-nm blocks per gram, so with the computing resources that can be included in a kilogram-scale nanofactory, it will be quite possible to plan the path of each individual block from where it is fabricated to where it will be placed.

The blueprint/control file will be broadcast equally to all parts of the factory; this allows a large number of local computers to be accessed with a simple network design. The control file will be sent in two parts. First will be a hierarchical solid geometry description of the product, which describes the block patterns in each volume and sub-volume of the product. This information will instruct each fabricator as to what block pattern it needs to make. The number and type of blocks required for each layer can be

---

[83] The discussion of design in section 5 of the "Primitive Nanofactory" paper applies here, with the additional freedom of specifying empty volumes and not having to worry about large-scale joints between large blocks-of-blocks. http://www.jetpress.org/volume13/Nanofactory.htm#s5

76

calculated in advance. These can then be distributed over the capacity of the fabrication volume to keep all the fabricators busy.

The distribution planning is complicated by the fact that at different times, concentrations of mass in different parts of the product may draw blocks from all over the factory. Pathological cases can be designed in which it is impossible to deposit blocks at full speed. However, preprocessing (generating broad plans for each product) along with tolerance for mild inefficiency will allow block production and delivery to be planned with simple deterministic algorithms. Each fabrication region will be able to compute exactly what point in the product volume its block is destined for, and when the hierarchical geometry/pattern description of the product is delivered, each fabricator will be able to identify exactly which block type to build. (A processor's eye view: "I'm 54,925.2 microns from the edge of the product... that means I'm 2,142.6 microns from the edge of sub-pattern K, whatever that is... Here comes the spec for K... I'm in sub-sub-pattern XB, 152.4 microns in... That means I'm building block type KXB4W.")

Fabrication instructions for individual blocks will be delivered next. Each fabricator will know which part of the instruction stream to pay attention to. In practice, a local computer will likely control multiple fabricators, parsing the instruction stream and sending appropriate instructions to each fabricator it controls. A few redundant fabricators for each computer will allow broken fabricators to be left idle. Blocks will be built sequentially, so local processors will not need to remember the entire instruction stream. They will receive a string of instructions for a given block, and place the next ten thousand atoms while instructions for other block types are delivered. Again, it is possible to design pathological cases where this doesn't work well, such as all block types requiring complex instruction at the same point in each construction sequence, but this approach typically should be reasonably efficient.

This plan assumes that all fabricators will be working in parallel. If the fabrication mechanism allows blocks not to be built in lockstep (see next section), then blocks with simple blueprints (which can be remembered by local computers) can be built out of step. Alternatively, the blueprint file may be sent multiple times in parallel with several different timings. This can be accomplished by having broader communication channels to each local computer, or different fabricators can be hardwired to different communication channels to get blocks ready at different times. Blocks will probably take minutes to hours to build, which is enough time to transmit many gigabytes of data.

9.3. Block fabrication

The most flexible way to build large (million- to billion-atom) blocks from molecules is to use a general-purpose manipulator system to add molecular fragments one at a time to the block. In this scheme, the blocks would all be built in lockstep, and deposition would start after fabrication was finished. This is fast enough to provide high performance: scaling laws indicate that it might take an hour for a single 100-nm manipulator to build a billion-atom (200-nm) block one atom at a time. But the delay before the first block is finished could be several times longer than the time needed to deposit the product.

One way to speed up the process is by building several block components in parallel, either at general-purpose workstations or on special-purpose fixed-path mills, and then combining them to form the block. The sub-parts could be either general-purpose parts, such as lonsdaleite cylinders to be added to a diamond crystal,[84] or special-purpose parts like computational elements. Either way, this could speed up block construction severalfold, allowing deposition to begin sooner.

Of course, the fastest way is not to manufacture the blocks in the factory, but to prefabricate them using any convenient combination of synthetic chemistry, self-assembly, mechanosynthesis, and simple mechanical assembly. A variety of design options is available as to what size the input blocks should be, how they should be delivered (solvated or clean and packed), how they will be fastened, and whether the nanofactory will include an intermediate block-assembly stage before depositing the blocks onto the product. Separating block-fabrication machinery from product-assembly machinery would not necessarily reduce overall efficiency, and could substantially reduce both the time needed to assemble a product and the heat generated in the assembly device.

9.4. Physical architecture

The product is deposited from a block-placement plane which is studded with manipulators to push the blocks onto the product surface. The manipulators may hold the growing product, making it possible to build several disconnected parts that will be fastened together after further deposition. Alternatively, the product may be held externally; this would require disconnected parts to be connected by temporary scaffolding.

Between the manipulators are holes through which blocks are supplied from the routing structure. The routing structure is composed of braided or interwoven block delivery mechanisms (probably cilia). This allows blocks to be shipped crosswise, routed around damaged areas, and so on. Power and control for the delivery mechanisms, as well as for the placement plane mechanisms, run inside the delivery structure.

Below the routing structure is the fabrication volume. This contains most of the mass of the nanofactory. It will be arranged in long, thin fins with 1-micron gaps between them for cooling/feedstock fluid to circulate; cables or tension struts will bridge the coolant gaps to contain the coolant pressure. The interior of each fin will be hollow, providing workspace for fabrication and space for transporting parts and blocks. The walls will be lined on both sides with fabrication systems. Feedstock will flow with the cooling fluid past the first fabricator stage. Molecules will be taken in through semi-porous active membranes (for solvent-filled factory designs) or tight-binding receptors (for vacuum-filled factories). The fin may be 4-6 times the width of a completed block--on the order of a micron--and about 2/3 empty space. In combination with the fluid channel volume, this means that the density of the fabrication volume will be about 0.1 g/cm^3. For a kilogram-scale factory with a square-meter deposition area, the fabrication volume will be about 1 cm thick. This provides about 10,000 square meters of surface area for feedstock intake and cooling; diffusion and heat transfer will not limit the nanofactory speed.

---

[84] See *Nanosystems* 8.6.5b and 8.6.6.

Below the fins are fluid supply and return ducts. Low-viscosity fluid can flow a distance of 3 cm at 1 cm/s through a 1-micron wide channel with a pressure drop of 6 atm, so the fluid can be injected from the ducts, flow along the fins to the top of the fabrication volume, and return, without excessive supply pressure. No fractal plumbing is needed for a factory that manufactures 1 kilogram per hour per square meter.

Because the physical architecture of the nanofactory is planar, with feedstock intake and processing located on the opposite side of the plane from product deposition, there is no need to change any dimension of the factory's nanoscale components in order to increase the manufacturing capacity and deposition area. In effect, multiple square-meter designs can be abutted to make as large a factory as desired.

9.4.1. Power and cooling

Most of the energy used by the factory will be used in the block-fabrication area, since handling the smallest components (feedstock molecules) will require the majority of operations. Fortunately, this is the area that is closest to the cooling/feedstock fluid. A cooling fluid made of a low-viscosity carrier of small encapsulated ice particles can provide very efficient heat removal;[85] a flow of 1 liter/second can cool 100 kW—more than an efficient nanofactory will need. Cooling by phase change also has the advantage of keeping the whole factory at an almost uniform and constant temperature. Because only about 1 gram per second of feedstock is needed, feedstock molecules can be dissolved in the cooling fluid at about 1000 PPM.

Within the nanofactory, power can be distributed very efficiently at all scales by rotating shafts. Electrostatic motor/generators can be used to interface with an external electrical power system.

A nanofactory manufacturing a kilogram per hour and drawing 1.4 kW, probably the upper limit for an air-cooled "desktop" model, would have an energy budget of about 100 zJ (10^{-19} J) per atom. Achieving this would require recovering the chemical energy of bonding operations, since a single bond contains several hundred zJ of energy. Recovering energy from chemistry requires controlling the reactions with machinery that is stiffer than the bond strength, so that the reaction can pull the machinery along smoothly. This is thought to be possible with advanced design. A less advanced design would require an external cooling system.

9.4.2. Clean internal environment

The internal environment of the factory must be kept clean of contaminant molecules that could cause undesired reactions or jam the moving parts. The factory interfaces with the environment at feedstock delivery and at product deposition. The feedstock molecule intake mechanism will deterministically manipulate the desired molecules, which provides an opportunity to exclude other molecules. This will be relatively easy if the molecule is small, like acetylene, or at least compact so that a shaped shell closed around it can exclude all solvent molecules. Small feedstock is preferable for several other reasons as well, including feedstock cost and flexibility of manufacture.

---

[85]   *Nanosystems* 11.5

Environmental contaminants can be kept out of the product deposition mechanism by extruding a thin shell or sealing sheet to cover the product and any unused area of the deposition surface. Before the product is removed, a second covering must be deposited to seal the deposition surface.

9.5. Reliability

Most of the mass of the nanofactory will be used for manufacturing blocks from feedstock. This implies that many fabricators per second will be damaged, and a tiny but non-zero fraction of blocks under construction will not be completed. This can be dealt with by building duplicates of all block types. Since the number of block types must be far smaller than the total number of blocks, this is not too onerous. Excess good blocks can be retained for later use or added to a reserved "dump" volume of the product. Because products must be designed for redundancy, single blocks that fail to function will not typically impair the functionality of the product. This implies that a block need not usually be replaced by an exact duplicate; instead, a block of approximately correct mechanical properties could be substituted. This further decreases the difficulty of coping with blocks that are missing due to unexpected fabricator failure.

Because damage rates per sub-micron module will be less than 1% per year, self-repair seems likely to be a worse choice than redundancy in terms of mass and complexity. This means that damaged machinery will simply be taken off-line permanently. Damaged or incomplete blocks in a damaged fabrication area can be retained there permanently, since that area will not be used again. (In the event that damaged parts need to be moved and stored, the required dump volume would be small.)

Blocks damaged after fabrication while in transit to the product assembly surface need not be replaced. The product's design will need to cope with radiation damage immediately after manufacture, and its lifetime will be far longer than the manufacture time; therefore, its design must be able to deal with a small fraction of damaged blocks. Missing blocks could pose a larger problem, but radiation damage will not significantly change a finished block's physical shape, so the block should still be able to be added to the product.

The transport mechanism will consist of many redundant arms / struts / cilia that work together to move the block along surfaces. The random failure of a small percentage of these will not compromise the ability of the rest to transport the blocks. If a patch of them fails, blocks can be diverted around the area.

There is limited room at the planar assembly surface, and the robotics may be more complex than for block transport. However, the volume of nanoscale radiation-sensitive machinery associated with the surface will be small. A square-meter area of machinery, 100 nm thick, and with machinery occupying 5% of its volume, contains about 5 milligrams of machinery. The entire volume may be hit about 4 times per millisecond. If repair requires a few microseconds--which may be plausible given the small scale and correspondingly high operation frequency--then the entire placement operation could stop during each repair without greatly reducing performance. A simpler solution is to make the placement machinery flexible and large enough that machines can reach to do the

work of a disabled neighbor. This would slow down the deposition slightly, but if most machines in an area were undamaged, then many machines could participate in taking up the slack by each doing some of their neighbor's work to distribute the load of the missing machine.

As discussed in sections 6.1 and 8.5 of the "primitive nanofactory" paper, computers can be made as redundant as necessary.

10. Advanced Product Design and Performance

To benefit from the high performance of nanoscale machines and materials, large products will have to combine large numbers of nanoscale components in efficient structures. Preliminary architectural work indicates that nanoscale performance advantages can be preserved in large-scale integrated systems. Large aggregations of high-powered machinery would need to be cooled, but this would be less of a handicap than in today's products due to more efficient machinery and more effective cooling.

It is hard to imagine the extent of what could be done with advanced nanomanufacturing. Clearly, products could gain radically in performance and efficiency. This chapter argues that in addition to higher performance, the cost and time to develop products could drop significantly, and manufacturing costs could drop precipitously. The final subsection describes a few applications of the technology that are likely to be influential.

10.1. Optimal use of high-performance materials

Strong, stiff solids in tension are often greatly weakened by minor flaws. Strain concentrates around any crack or other flaw, causing local failure which rapidly propagates. Solid carbon lattice (diamond) is no exception. To take advantage of the theoretical strength of carbon-carbon bonds, it will be necessary to prevent crack propagation. This can be done by building the material in long thin fibers attached in a way that does not propagate failure. Similar approaches are used today in advanced polymers and fibers, but molecular manufacturing construction would give more control over structure. This would allow fibers to be perfectly aligned, and crosslinked to each other or attached to other structures with minimal strain. Most important is that most of the fibers would be defect-free.

Strong fibers can also form the basis of energy-absorbing materials. Interlaced fibers with high-friction molecular surfaces could be designed to slip past each other under stress slightly less than that needed to break bonds in the material. The limit to the energy absorbed in such an "unbreakable" material is the heat capacity of the material; when its temperature rises too high, its bonds will be weakened. In theory, this much energy could be thermalized (absorbed) in just a few nanometers of motion; longer fibers would allow the material to absorb repeated impacts after it had cooled.

A solid block, slab, or beam of material typically is not efficient at resisting compressive stress. With no cost penalty for manufacturing more complex objects, it will be possible to make a variety of efficient structures such as honeycombs and fractal trusses. A thin pressurized tank will resist compression at any point, transferring compressive stress to the contents and imposing tensile stress on the walls.

A diamond shaft rotating at high speed can carry power at 10 watts per square micron.[86] This may be the most compact way to transmit power within a product. Stretching a spring made of tough diamond structure can store energy equal to a significant fraction of

---

[86]   See *Nanomedicine* 6.4.3.4, available online at http://nanomedicine.com/NMI/6.4.3.4.htm

the bond energy of the spring's component atoms. Such a storage system could be charged and discharged quite rapidly, and store energy without leakage.

10.2. Performance of advanced products

The performance of a product will depend on its mass, power, and heat budgets. To a large extent, mass can be traded for efficiency, by adding more systems and running them more slowly to obtain reduced drag. Given the extremely high power densities of electrostatic motors, and the smaller but still quite high power density of electrochemical (fuel cell) or mechanochemical processors (perhaps more than 1 GW/m^3)[87], it is safe to assume that power transformation will not be a significant part of the volume of most meter-scale products. If the product expends its power externally, for example on propulsion through a viscous medium, then only a small fraction (probably well under 1%) of the total power handled need be dissipated as internal frictional heating.

Because signaling over even modest distances slows a computer system significantly, and because erasing bits has an irreducible thermodynamic cost, massively parallel computer systems are likely to be concentrated sources of heat. Reversible logic can be used to reduce the number of bits that need to be erased, but this comes at the cost of increasing the total amount of logic and thus the frictional losses. Increasing the size of the computer also increases the signal path length, requiring slower operation. In general, a computer using reversible logic will be most efficient when it spends about half its energy on erasing bits and half on friction,[88] which means that Drexler's reference calculation with its modest use of reversible logic was actually close to ideal.

By today's standards, computers will draw very little power: a billion 32-bit 1-GHz processors would use about 10 watts. For most applications, new algorithms would be required to make use of such a highly parallel system. Cooling a cubic-centimeter volume of computers (which would produce about 10^5 watts of heat) can be accomplished via branched tubing and a low-viscosity coolant fluid using suspended encapsulated ice particles.[89]

Using the full strength of diamond, handling compressive stress efficiently, and using much more compact motors, computers, and sensors (with less mass required to mount them), products could be built with a small fraction—usually less than 1%—of the structural mass required today. This would often be inconveniently light, but water ballast could be added after manufacture. This shows that nanofactories will be competitive for more than just high-tech products. Even if highly structured products composed of diamond-like material cost $100 per kg to produce, nanofactory-built products would generally be competitive with current products.

---

[87]   See *Nanomedicine I* Sec. 6.3.4.5 http://www.nanomedicine.com/NMI/6.3.4.5.htm
[88]   Mike Frank has found that the tradeoff between entropy savings and hardware blowup is approximately a power law: bit-erase energy saved equals hardware penalty (which is approximately friction) to the ~0.64 power. See http://www.cise.ufl.edu/research/revcomp/theory/Memo19-tradeoff/Memo19-RodLogic.doc. Minimizing the sum of these requires spending about 10% more energy on friction.
[89]   See *Nanosystems* 11.5.

10.3. Design of advanced products

To date, product complexity has been limited by the need to manufacture it via relatively simple and crude processes, and minimize the number of operations to reduce manufacturing cost. A nanofactory will impose essentially zero penalty for each additional feature[90], and will provide several design choices per cubic nanometer. Although a simple product such as a cube filled with inert matter would be easy to specify, designers will want to use nanoscale features in heterogeneous structures without being overwhelmed by complexity.

Perhaps the simplest design task will be to re-implement existing products in the new technology. In today's products, a volume of the product usually implements a single function: a motor, a computer, a structural beam. Many of these components will be able to be replaced by a higher-performance component without changing the product's functionality. In some cases, inert mass will have to be added to preserve kinematic characteristics and avoid excessive redesign; for example, replacement of electromagnetic motors by nanoscale electrostatic motors may require a flywheel to be added to replace the missing rotor mass.

For drop-in replacement of today's components, the key ingredients are well-characterized nanomachines and design libraries that combine them into larger structures. Product designers should not need to worry about nanoscale phenomena in nano-motors, nor about designing a multiply branching distribution structure or converging gear train, nor about implementing fault tolerance.

It seems likely that design of nanofactory-built products will follow a method similar to software engineering: build high-level designs on top of many levels of predictable, useful capabilities encapsulated in simple interfaces that allow the low-level capabilities to be parameterized. One simple but useful design technique will be to design a structure that can be repeated to fill a volume, and then specify the desired volume. (The structure will likely be part of a pre-supplied library.) That volume, full of whatever functionality was designed into its structure, can then be used as a component of a larger, higher-level structure.

The lowest level of structure that designers of large products will usually be concerned with is the individual nanoblocks that the product is assembled out of. Nanoblocks as small as 100 nm can contain basic functional units such as a motor or CPU, whereas even a 1-micron nanoblock would be a small fraction of the size of a human cell. A library of such nanoblock designs will be available for combining into larger functional units.

At the highest level, designers who are merely trying to recreate today's level of product performance will find it easy to fit the required functionality into the product volume.

---

[90] Because the same number of molecular construction operations would be required to fabricate the same mass regardless of design, the only cost would be the time and energy required to process a slightly larger specification file.

84

10.4. Rapid R&D and deployment

Manufacture of prototypes is currently a costly process. Rapid prototyping machines can help, but so far they can only make passive components, not integrated products, and the machines themselves are costly. Assembly is still required. Manufacture of a prototype product also takes substantial time.

Today, high volume manufacturing may require overhead including expensive molds, training of workers, and procurement of supplies. The design of products intended for high-volume manufacturing must take this into account; the product must not only be useful, but also manufacturable. Designing the manufacturing process is a significant part of the total design cost, in addition to the constraints it places on product design.

A nanofactory would be equally well suited to rapid prototyping and high volume production. A prototype could be produced in an hour for a few dollars per kilogram. This would allow rapid testing of new product ideas and designs, more like compiling software than like today's prototyping process.

Once a design was approved, it could immediately be put into production. Depending on how the nanofactories were deployed, production could be at point of sale or even point of use. Warehousing and shipping would not be required, substantially reducing the expense and delay of product deployment. Less reliance on economies of scale would allow efficient test marketing, potentially reducing initial marketing costs. Lower costs for R&D, and far lower costs for initial deployment, would allow greater tolerance of failure and thus more innovation.

Nanoscale and microscale designs could be developed by a rapid genetic algorithm-type process with physical evaluation of the fitness function. An array of varied designs could be built and tested, and the results used to refine the design parameters and specify the next test array. Rapid construction and testing of millions of components would allow physical implementation of genetic algorithm design methods. (Note that the products would not build copies of themselves—there would be no self-replication. The nanofactory would build each successive generation of products.)

10.5. Low cost of manufacture

As a general rule, prices are driven by a balance between demand and scarcity. A nanofactory capable of building a duplicate in an hour would be able to support a rapid increase in the number of nanofactories to any desired level. There would be no scarcity of manufacturing capacity; if nanofactories ever became more valuable per kilogram than their cheapest products, they would be used to manufacture more nanofactories. Even if the feedstock is relatively expensive, this argument implies that products with high value per gram such as computers and pharmaceuticals would not lack for manufacturing capacity. Of course, this argument ignores sources of artificial scarcity such as patents.

Although early nanofactories might require expensive feedstock and consume large amounts of power, a combination of deliberate design processes and genetic algorithm approaches could produce rapid improvements in nanofactory component design, allowing the use of simpler feedstock. Similar rapid design effort might be implemented

to develop nanofactory-built chemical processing plants, reducing the cost of feedstock. Because small designs could be built in large arrays, the processing system could use any convenient combination of mechanosynthesis, microfluidics, and industrial chemistry.

If production capacity became non-scarce (at least to the patent holders) then it is not obvious what resource would be scarce. Lightweight solar collectors could produce a plentiful supply of energy. If the products include solar collectors, then a few days of manufacture could supply as much energy as required. (The active component of a solar collector can be quite thin, and the structure can be low-mass; a structure massing a kilogram per square meter would recover the energy cost of its manufacture in a day or so. Thermionic solar cells have been built out of CVD diamond.) Feedstock could be small carbon-rich molecules; carbon is readily available anywhere on Earth, and nanofactory-built equipment might be used to process cheap carbon sources into feedstock. A similar argument applies to silica- based feedstock; silica is also plentiful on the Moon and Mars. Nanofactories would not require special skill to operate, and would not require working conditions that would be expensive to maintain.

With near-zero labor costs, low environment and capital costs, and moderate energy and feedstock costs, there is no apparent reason why the per-item cost of production should be more than a few dollars per kilogram regardless of device complexity. (Of course, the cost of design, including the amortized design of the initial nanofactory, will be far from trivial.) This contrasts sharply with the value of the products to consumers—by today's standards, nano-manufactured computer components would be worth billions of dollars per gram. There would be a huge incentive for profit-taking; it is not at all obvious how soon consumers would see the full benefits of the new manufacturing technology, even if its military implications did not cause it to be restricted.

10.6. Applications

A general-purpose manufacturing system capable of making complete high-performance products will have many applications. These include massively parallel sensors and computers, military force multiplication, wholesale infrastructure replacement, ecological footprint reduction, and aerospace.

The ability to build kilograms of fully engineered nanoscale products means that vast numbers of sensors could be produced at near-zero cost. These could be integrated into one structure for optical or medical data-gathering, or could be incorporated in small distributed sensor platforms. More compact functionality, easier fabrication, and more efficient use of power would give nano-built sensor platforms a significant advantage over MEMS technologies.

The ability to gather large amounts of physiological data (e.g. chemical concentrations or electrical potentials) in real time using a sensor array small enough to be inserted into the body without damage would be a huge help to medical research. Early and accurate detection of health conditions would help in the mitigation of many diseases. Early detection of adverse reactions to treatments would allow doctors to design more aggressive and experimental treatments without compromising safety; it might even be possible to bypass clinical trials entirely. At the other extreme, implanted sensor arrays

could provide better understanding of causes, effects, and feedback loops in the body; this could allow more subtle and less invasive treatment. Cell-sized machinery raises the possibility of cell-by-cell interventions, even genetic interventions, with more control and flexibility than current therapies. Interfacing to neurons, for both sensing and controlling and controlling neural signals, could be done more delicately and on a far larger scale than with today's electrode technology.

Massively parallel computers would be one of the first products of a nanofactory development program. This would have applications in simulation, sensor array data processing, data-mining, pattern recognition, symbol manipulation, and neural networks. Precise construction could also be expected to be useful in building quantum computers.

Portable high-volume general-purpose manufacturing of advanced products would greatly increase the flexibility and power of military activities. Sensor and computer improvements would greatly improve situational awareness. Manufacture of products at the time and place they were required would improve logistics and transportation. The ability to build high-performance computers and actuators into any (or every) cubic millimeter of products would allow new kinds of weapons to be developed. The ability to inexpensively and rapidly build, test, and deploy new weapons would accelerate progress in weapons technology.

Low cost, high throughput manufacturing could be used to build very large quantities of product. Inefficient products, and even networks of inefficient products, could be replaced with little manufacturing effort. Of course, replacement depends on many other factors including design effort, installation effort, political resistance, compatibility, and consumer acceptance. But infrastructure is largely invisible to consumers, is often administered centrally, and is often aging or inadequate. Infrastructure may provide many suitable targets for widespread deployment of nanofactory-built products.

As the efficiency of infrastructure is improved, and the ability to monitor the environment increases, it will be increasingly possible to reduce humanity's ecological footprint. Accidental or unnoticed damage could be reduced, and the consequences of deliberate activities could be better known. Mineral extraction and processing activities, including fossil fuel consumption, could be reduced. Water could be used and re-used more efficiently. Even something as simple as widespread use of inexpensive greenhouses could save substantial amounts of water, topsoil, pesticides, fertilizer, and land area.

10.7. Aerospace applications

Aerospace hardware depends on high performance and light weight; this is especially true for spacecraft. Orders of magnitude improvement in material strength, power density of actuators, and computer performance, make it reasonable to think of designing rockets with hardware mass a small fraction of payload mass. Lightweight inexpensive hardware makes it easier to design combination airplane/rocket systems. It might even be worthwhile to include an efficient gas processing system and fill a collapsible liquid oxygen tank after launch. Combination airplane/rocket systems capable of reaching orbit could be much smaller than rocket-only systems. Today, orbital access is expensive due to minimum rocket size, high construction cost, and the additional work required to avoid

expensive failures. If spacecraft were smaller and failures were substantially less expensive, then R&D could proceed with less deliberation and more testing of advanced chemical and non-chemical designs.

Greatly increased device density and near-zero cost of adding additional devices would make it possible to add large numbers of sensors to any product. This would allow better characterization both during testing and during operational use, showing where mass could be saved and warning of impending failures.

Once nanofactories are developed to the point that they can be taken on board spacecraft, they can serve as general-purpose manufacturing systems for flight hardware. This will be especially useful in long-duration missions. Any hardware that could be replaced in flight would have an extra margin of safety against premature failure or design flaws discovered after takeoff. New hardware designs could be incorporated after launch. Missions to unknown environments would have an added level of flexibility, which could increase safety. Medical support equipment could be built on an as-needed basis, both saving weight and allowing more targeted response to developing conditions. In large-scale missions, addition of a product recycling system would allow wholesale reconfiguration of the spacecraft and equipment at different phases of the mission. Because the nanofactory, by design, uses only a few different kinds of feedstock substance, recycling would be simpler than in today's industrial systems.

In extremely long-range missions, the use of covalent solids made from light atoms implies the ability to gather new materials from lightweight space objects such as comets and carbon asteroids, without having to deal with the gravity wells and mining operations that would be required to extract heavy minerals from larger bodies.

A. Nanoscale Properties and Phenomena

A.1. Nanoscale Properties

The line between properties and phenomena is blurry, especially at the nano scale. Basically, phenomena emerge from properties, but these phenomena can then be considered as properties for another level of emergent physics. A property might be electron path length; electrical resistance could then be considered a phenomenon.

The most basic properties are the electron quantum equations interacting with atomic nucleus positioning and passing fields and photons; these form the basis of almost everything we'll be considering here. Chemistry itself emerges from this level; the ability to form chemical bonds is a phenomenon of electron density, and bond strength (including angle bending and torsion) is a phenomenon of bonds.

The possibly-useful phenomena are placed on a size scale at approximately the size where they begin to be important. Some of the effects, such as surface forces and diffusion, are important or useful over several orders of magnitude. Others, such as chemical bonds, have a very narrow or characteristic length. Question marks indicate that I'm not sure where on the list to put a given property.

Any given nanomachine will only use a fraction of these effects. As implied in Nanosystems, entire kilogram-scale products can be built using only a few of them (physical/mechanical properties; electrostatics, conductivity, and tunneling; molecular binding/recognition; and bond formation). Control of a small subset, then, will be enough to implement a general-purpose manufacturing technology.

A.1.1. Sub-Angstrom

High-energy radiation cross-section

Electron excitation, atomic data storage, isotopes

NMR (may be influenced by molecular structure)

A.1.2. 1 Angstrom

Atomic magnets

Atomic bonds: Covalent, Hydrogen, Sulfur, Ionic, etc

High-energy radiation damage

X-ray diffraction

Electrochemical effects

Spectroscopy

89

Triboluminescence

Thermoluminescence, optically stimulated luminescence

A.1.3. 1 Nanometer

Strength

Validity of continuum mechanics for some types of structure

Surface forces

Photoelectric effects

Electrostatic interatomic interactions

Superlubricity

Heat (thermal bath)

Brownian motion/docking

Electron tunneling

Photoconductivity

Thermionic emission

Peltier effect

Mechanical interlocks

Binding

Adsorption

Solvation

Semi-permeable membrane pores

Electron migration (conductivity)

Electromechanical effects

Electromigration

Piezoelectricity

IPMC

A.1.4. 10 Nanometers

Thermoelectric

Magnetic domains

Entropic springs

Diffusion

Osmosis

Plasmons

Superconductivity

Concentration gradients

Electrostatic actuators

Magnetostriction

A.1.5. 100 Nanometers

Electrostatic motors

Robots

Holograph features, photonics

A.1.6. 1 Micron

Electromagnets

Weight

Probabilistic radiation damage becomes significant

Diffraction-limited optics

A.2. Functions and Functional Nanostructures

This is a partial list of classes of functions that may be useful in building engineered systems. Each class may include several functions, and each function may be implemented in various ways at various scales by various technologies.

Inert structures will be useful for only the simplest of operations, such as particle filtration. But as will be seen below, there are many ways to implement the various functions. Not every property need be used to do this. For example, a stiff but non-conductive substrate could be used to implement mechanical information storage and computation, while a conductive but floppy substrate could be used to build molecular electronics for the same purpose.

For many purposes, it will be unnecessary to implement functions at scales much smaller than a micron. This means that some functions can be implemented using engineered nanosystems rather than molecular electronic properties or quantum effects. Paragraphs below that are labeled *Nanomachine* describe ways to implement a function that do not depend on molecular electronics, and so could be implemented in a variety of materials.

A.2.1. Photon Detection

Photons are generally detected by their effects on electrons. Near and above optical frequencies, their energy is enough to cause a change in electron distribution within a molecule. Photons can also cause electrons to be emitted from a charged surface. (see Electron Manipulation)

Photons may supply enough energy to kick metastable atoms back to a lower-energy state.

Photons with energy below ambient heat (e.g. radio) can be detected electrically with an antenna. This works well with continuous waves, but spread-spectrum shows that even discontinuous waves below noise level can be detected: "the system can work at negative SNR in the RF bandwidth" http://www.sss-mag.com/ss.html

Photons may also be "detected" by their effect on other photons in nonlinear optical media.

Nanomachine: Nanoscale rectennas might be used to collect energy from optical photons.

A.2.2. Photon Production and Manipulation

Photons can be produced from a wide range of energy sources, including stored strain in crystals (thermoluminescence), electron/hole recombination as in LEDs, and of course heat.

Photons can be manipulated or modulated through a wide range of types and configurations of material, including organic dyes, nanometer-scale (strained) silicon crystals, and sub-micron optically active arrays. Photons can be converted to excitons and vice versa, slipping through sub-wavelength holes in metals, and being guided (even around curves) by 50-nm channels (see http://nanophotonics.ece.cornell.edu/).

Nanomachine: Arrays of radiating antennas might be used to produce photons up to optical frequencies, including control of phase (implies hologram and laser generation). Rotated or shaken dipoles can produce microwave or possibly far-IR photons.

A.2.3. Motion Detection

For the present, motion can be defined as a coordinated displacement of nuclei. This is related to conventional mechanical ideas of force, leverage, etc. But nuclei moving relative to each other can cause a variety of useful changes in their electron clouds. These can be observed as piezo____ or tribo____ effects.

Motion can be detected by simple mechanical interference, in which one part blocks another part from moving.

Relative motion of separated objects can still be detected if they are close enough for their energy states to interact. For example, nanoscale "diffraction gratings" can be used to detect atomic-scale motion. Also, energy transfer between photon absorbers and emitters can be used to measure their separation. From one point of view, this can be considered a type of spectroscopy (since it deals with the interpretation of interactions between matter and radiation).

92

Capacitive or tunneling-current sensors can detect displacement with atomic-scale resolution.

A.2.4. Motion Production

In general, piezo effects can be reversed to produce motion. Electrostatics can produce motion. Thermal expansion can be used for actuation down to MEMS scales, but this may not be relevant at nanometer scale. Some molecules will flex/reconfigure in response to energy input; for example, bacteriorhodopsin has a piece that swings when it's hit by a photon. All these effects do not depend on solution chemistry.

In solution, motion can happen through binding and unbinding of dissolved species. Enzymatic action on those species can fuel a motor (linear or rotational). Molecular binding, as demonstrated in "DNA tweezers" and other examples, can be used to constrain motion, having the effect of actuation if this compresses a spring. Also, molecular turbines can be driven by concentration gradients, as in the bacterial flagellum. The electronic configuration of a solvated molecule can be changed by the pH of the solution. This can displace other molecules, as in rotaxane actuators.

Actuation systems that depend on pH-mediated or other electronic changes can presumably also be driven by electrodes. This implies that several rich chemistries can probably be adapted for actuation.

Nanomachine: Clutches, detents, shafts, and gears appear to be buildable on nanometer or even sub-nanometer scales. Although thermal noise would cause substantial slop relative to familiar metallic mechanical systems, this slop would not be damaging unless it was so extreme as to jam moving components in a way that caused direct mechanical force transfer. This can be seen by the fact that a thermal process which added energy to a mechanical system, or frequently concentrated energy in quantities greater than several k_bT (Boltzmann's constant times the temperature) on a particular point, could be used as the basis for a perpetual motion machine extracting work from thermal noise. In a well-engineered covalent system, energy less than k_bT will not cause bond rearrangement, and therefore thermal slop will not cause wear as it would in a large metal-based system. Mechanical constraints, such as adding housings around gears, can be used to prevent unwanted slippage. This means that mechanical motion can be transported and switched mechanically.

Nanomachine: Electrostatic actuators should work on nanometer scales for any system that can implement conductors and insulators. Power density scales favorably with decreasing size, implying nanoscale electrostatic motors could have power densities of 10^{15} W/m^3.

A.2.5. Chemical Detection

Molecules can be detected if they change the electronic states of adjacent molecules, as in gases adsorbed on a conductive buckytube. To some extent this can be used to analyze the electronic properties of the molecules.

Molecules can fill a binding site and be detected by a variety of means. This should be able to achieve the specificity of antibodies, which have been developed to distinguish between armchair and zig-zag buckytubes. Detection methods for bound molecules (complexes) include direct measurement of mass (cantilevers), the formation of three-part complexes (for example, linking a dye and a magnet, or a chromophore pair), and steric hindrance (mechanical blocking), as in cellular channels.

Molecules can be detected at a distance through spectroscopy or NMR.

Nanomachine: Functioning binding sites have been built out of polymer molded around molecules; this implies that almost any material should be able to make binding sites. An enclosed pocket, or an open but strongly binding site, can be mechanically probed to see whether or not it contains a solid molecule.

A.2.6. Chemical Manufacture

Bulk chemistry, even with catalysts, is rather imprecise, with yields of 99% being rare even for fairly straightforward reactions. This can be improved by limiting the available side reactions and adjusting reaction conditions so as to improve the energetics.

An example of programmable chemistry is the cellular manufacture of nucleic acid and protein sequences. This is done by very specialized enzymes that can select from only a small pool of possible monomers. Even so, without error correction, the enzymes working alone frequently have error rates worse than 10^{-6}.

Nanomachine: Mechanosynthesis is the use of mechanical motion to initiate chemical reactions in programmable locations. Lab demos to date have often added electric currents and/or fields to the mix. Electric current approaches may not be applicable to highly parallel manufacturing, but fields can be adjusted by positioning semi-charged atoms near the reaction site. Simulations appear to show that carbon dimers can be deposited on at least one unterminated diamond surface without undesired reconstruction, even at room temperature. Mechanosynthesis is attractive because in theory it allows a few deposition reactions to be used to build an extremely wide range of structures.

A.2.7. Structure Formation

Structures can be formed by covalent bonds and by non-covalent binding. Examples of the latter include antibody-antigen complexes, protein folding, and nucleic acid pairing. Structures formed by reversible binding can be used for detection or transport of molecules.

Polymers with a single bond between the monomers are generally too floppy to be considered as structure. But braiding several single-chain polymers together can form stiffer structures.

Molecules with two- and three-dimensional bond arrangements can be much stiffer, and single molecules can form useful structures. Examples are buckytubes, diamond, and crosslinked polymers.

Nanomachine: Surface forces will work at any sub-micron scale to bind stiff parts with matching surfaces into larger structures. Mechanical interlocks, such as knobs snapping into bulb-shaped holes, should work down to nanometer scales. Phoenix has described a mechanical "expanding ridge joint" which should be able to preserve >50% of the strength of solid material.

A.2.8. Electron Manipulation

Quantum dots can store electrons. Four dots arranged in a square, with two electrons in opposite corners, can be used as a component of a "wire" and/or a logic element.

"Virtual atoms" can be formed by corralling electrons within multi-atomic structures. These can be expected to be more different from normal atoms than germanium is from carbon; it's not clear that predictions of high-strength materials or programmable physical properties are justified. One useful feature of virtual atoms is that they can interact strongly with magnetic fields.

Physical rearrangement of atoms can cause large-scale displacements of electrons, as in piezoelectricity. Heat or electric fields can cause the emission of electrons from surfaces. Sufficiently energetic photons can cause separation of charge, as in ionization or electron/hole pair creation.

Electrochemistry is a huge field. In fuel cells and batteries, atoms need to become ions in order to move through an electrolyte; their lost electrons can be sent through an electric circuit to extract power. (Conversely, chemistry can be driven "uphill" by electricity.)

Manipulation of electrons based on their spin ("spintronics") may become useful for implementing digital logic.

Nanomachine: Electrostatic generators can be implemented by moving an isolated conductor through an electric field. A 3x10 nm conductor is about the right size to hold one electron. This is more or less the principle behind a Van de Graaff generator. The electric field can be generated by any charged conductor, or by using atoms of varying electronegativity.

A.2.9. Electronic Detection

Detecting local electronic states or fields may be useful for chemical detection, digital logic, position sensing, etc. Electron states can influence conduction, as in field effect transistors. A shift in electron state may also cause a molecule to reconfigure, as in piezoelectricity. The spectroscopy can also be expected to change.

Nanomachine: There's a wide range of ways to convert electron state to force, including simple electrostatic attraction or repulsion; even insulators are attracted toward electrostatic fields. Force can be converted to displacement and measured.

A.2.10. Information Storage

Information can be stored in any measurable configuration of a system which is resistant to random erasure. This means that there must be an energy barrier of $\sim 100 \ k_b T$ between the set of configurations which signify the desired value and any configuration signifying

a different value. However, this does not mean that this much energy must be consumed to store the information. For example, consider a mechanical system which locks a marker in one of two slots. If the slots are separated by a detent, moving the marker may be quite inefficient. If the marker is weakly held in place while a separator is inserted, thermal noise may cause the marker to jump to the wrong side at the wrong moment, causing an error. But slow compression of a powerful spring will be almost completely reversible, and a sufficiently strong spring will keep the marker in its proper position while the separator is inserted. For a more detailed exploration of this, see Nanosystems Sec. 12.4. Electronic designs can also be made efficient.

There is an inescapable cost to storing information. The forgetting of a piece of information involves the state of its marker becoming less constrained. To continue the mechanical example, removal of the separator will allow the marker to wander freely between the slots. This is a waste of energy corresponding to $\ln(2)k_bT$. Storing a new bit of information will require the use of that much energy to compress the marker (which implements an entropic spring) into a single slot. Attempts to get around this by measuring the state of the marker will not help; the original marker may be released efficiently if its state is known, but the release mechanism will then know its state—and need to forget it, with accompanying waste of energy.

A.2.11. Quantum Entanglement and Logic

This will probably be very useful in the future, though few current devices use it. Precise nanoscale construction of large-scale devices will make it easier to build whatever structures are required by quantum computer designs.

A.2.12. Digital Logic

Digital logic does not really belong in this list. Almost any interaction that can be driven (amplified) can be used to implement digital computation. So digital logic is an engineered function, not a physical effect or phenomenon: if interactions are set up correctly, and the results are interpreted properly, the result can be treated as a digital computation. It should be noted that many effects can be used to implement reversible logic, which becomes more energy-efficient (but less space- and time-efficient) without limit as the driving force and speed are reduced. In general, a reversible computer will be most efficient overall when it is spending about half its energy on bit erasure and half to overcome friction.[91]

Nanomachine: An architecture for digital logic circuits sufficient to implement a CPU is given in *Nanosystems* Ch. 12.

---

[91] Mike Frank has found that the tradeoff between entropy savings and hardware blowup is approximately a power law: bit-erase energy saved equals hardware penalty (which is approximately friction) to the ~0.64 power. See http://www.cise.ufl.edu/research/revcomp/theory/Memo19-tradeoff/Memo19-RodLogic.doc. Minimizing the sum of these requires spending about 10% more energy on friction.

B. Related Concepts, Tools, and Techniques

The concepts and approaches discussed in this Appendix are significant aspects of molecular manufacturing. They do not define it and are not all required in every implementation, but collectively they give it shape and direction.

The goals of molecular manufacturing are high performance automated manufacturing of high performance products. This requires programmability, precision, and nanoscale tools (which benefit from scaling laws).

The use of effective engineering techniques is both a goal and a method. Easier engineering enables faster development of more advanced products. Although many aspects of the nanoscale are complex, poorly understood, or otherwise difficult to engineer, there are enough straightforward and easily characterized phenomena to build computers, actuators, sensors, and structures—the basic ingredients of mechanical engineering. The vast number of nanoscale parts in any large product is both a challenge and an opportunity, since it allows design at multiple levels of abstraction.

Several tools and techniques useful for nanoscale engineering are listed and briefly described.

The final sections of this Appendix discuss incremental improvement toward advanced molecular manufacturing capabilities, and compare and contrast molecular manufacturing approaches with biological designs.

B.1. Goals

A major purpose of the goals listed here is to enable a manufacturing system to build a complete duplicate system, possibly with improvements. Multiple generations of such duplication can result in exponential growth of manufacturing capability, leading from a single nanoscale fabricator to many grams per hour of production. This requires that precision be maintained from generation to generation, and that the system be fully automated, programmable, and flexible.

B.1.1. Precision

Mechanically guided chemistry has the potential to restore precision: to make a product that is more precise than the manufacturing operations involved in its fabrication. This is because covalent bonds are non-linear: in general, atoms are either bonded or they are held separate by a substantial repulsive force.

A wide variety of reactions and techniques is available for fabrication of molecular structures. Some of these techniques involve transferring single atoms or small groups from a "tool" molecule to a growing surface.[92] Other approaches, including the "Tattoo"

---

[92] David J. Mann, Jingping Peng, Robert A. Freitas Jr., Ralph C. Merkle, "Theoretical Analysis of Diamond Mechanosynthesis. Part II. C2 Mediated Growth of Diamond C(110) Surface via Si/Ge-Triadamantane Dimer Placement Tools," *J. Comput. Theor. Nanosci.* 1(March 2004):71-80. http://www.molecularassembler.com/JCTNMannMar04.pdf.

design described in this report, fasten together larger molecules by means of functional groups placed on their surfaces.

The various approaches vary equally widely in the precision they require for placement. Some advanced deposition reactions that have been investigated involve depositing a dimer on an unterminated surface with many reactive points and many possible final states. Such a reaction requires a placement precision of a fraction of the width of an atom. In contrast, approaches which place large molecular building blocks into vacancies between previously placed blocks may only require precision on the order of a nanometer.

Because molecular manufacturing uses a limited set of incremental construction operations in programmable sequence, simply varying the sequence can produce a different output. The ability to build programmable shapes implies that a wide range of machines with widely varying functionality can be built. The advantages of atomic precision are further explored in Chapter 1, "Incentives."

B.1.2. Programmability

In order to build a wide range of intricate products, a manufacturing system needs to be controlled externally. Since human control is incompatible with the vast numbers of operations required for nanoscale construction, the system needs to be programmable.

In order to automate a system, it must be well characterized. This encourages a minimal set of operations. One way to minimize the operation set is to use operations that can be done many times apiece while varying simple parameters. A number of chemical synthesis operations fit this requirement. For example, adding monomers to a polymer chain in programmed sequence, as ribosomes and DNA synthesis enzymes do, can build extremely large and complex molecules with only a few operations.

When building crosslinked covalent solids (e.g. diamond), the deposition of small molecular fragments at programmably selectable points on the surface can allow a vast range of structures to be built with, again, only a few operations. Selecting between chemically similar sites is very difficult in solution chemistry. But if the deposition reaction is mechanically guided, then the control of position can be separate from the chemistry, and the similarity of the chemical operation becomes an advantage.

Even modern computers would be incapable of controlling the huge numbers of operations required by an advanced molecular manufacturing system. The energy cost of today's transistors, combined with the energy, time, and design complexity required to drive a large number of signal pathways to and from the nanoscale, imply that the manufacturing system must include nanoscale computers capable of processing control data that is efficiently coded (compressed).

B.1.3. Performance

As explained in the chapter on Incentives, scaling laws indicate that nanoscale machines can have performance that is many magnitudes higher than today's machines. This is a major reason for using nanoscale manufacturing machines (productive nanosystems) rather than larger, slower systems. To make the most of this performance, it will be

98

important to minimize all sources of drag. One source of drag is immersion in fluid, implying that the machine should be vacuum-filled. A medium of vacuum also eliminates the need to provide extremely pure solvent.

Without a fluid medium, moving parts will be in direct contact. However, superlubricity (see below) should eliminate static friction, and drag can be minimized by minimizing the area of contact. Drag can also be minimized by using mechanical components that are stiff and equilibrate quickly. (Proteins, due to their low stiffness—especially with regard to side chains, which may function as energy-balancing entropic springs and which will detune if deformed too quickly—may be expected to be inefficient at even moderate speeds.)

B.2. Engineering approaches

B.2.1. Predictable and deterministic systems

Systems are easiest to engineer when they can be easily and accurately modeled—when they follow simple rules and always do the same thing in the same situation. If they incorporate mechanisms to restore precision, to correct small errors before they become large enough to affect operation, then the systems can carry out extremely intricate operations in extremely large numbers with high reliability. The best example of this is computers. A desktop computer can carry out Avogadro's number of transistor operations in just a few days, executing software far more complex than any physical object ever designed, without a single error. Digital logic and "digital" mechanical operations will go a long way toward enabling nanofactories and their products to perform similar feats in the mechanical realm.

A machine will be deterministic if it has high barriers between the state it is supposed to be in and any error state. As explained below, the use of high barriers does not imply that energy equivalent to the barrier height must be wasted in moving the system from state to state.

B.2.2. Classical physics and mechanics

It is often said that quantum effects replace classical physics at the nanoscale. It is true that quantum effects are easier to observe, and can be significant. However, *Nanosystems* Chapter 5 compares classical and quantum analyses of mechanics and concludes, "At ordinary temperatures in all but very light, stiff systems (e.g., single atomic masses and bond-stretching displacements), the classical value for the positional variance given by Eq. (5.4) is accurate enough for all but the most precise work." In other words, when analyzing the mechanics of even a single-nanometer component, quantum effects and uncertainties can be ignored.

Biological molecules are often "squishy." This appears to be adaptive—artificially modified versions of proteins are often more stable than the natural version, implying that the natural version could be stiffer if that were advantageous in the context of biological systems. It may be that the softness of natural proteins makes them easier to metabolize

—an important consideration in a system employing self-repair—or allows them to use entropic springs for fine-tuning of energy efficiency.

In any case, the softness of natural proteins should not be taken as an indication of the desirable, much less the maximum, stiffness of engineered nanoscale machines. Nanoscale components built of crosslinked covalent solids can easily have a stiffness in excess of 10 N/m. With this stiffness, thermal noise will vary the position of the surface by less than an atomic diameter; the component will have a definite shape. This does not mean that thermal noise can be ignored, but that it can be coped with, just as vibration is an important but far from crippling factor in large-scale engineering.

B.2.3. Hierarchical design

In designing systems with millions of components and functions, it is necessary to be able to treat groups of components collectively. Grouping components into systems that have simpler function than the sum of the parts is one way to accomplish this. "Libraries" of well-characterized and easy-to-specify systems can be recombined and re-used to make quite complicated systems. This approach is utterly necessary in computer design, which has many levels of abstraction between the user and the silicon, allowing human programmers to produce programs with millions of instructions that guide trillions of computations. When designing large products composed of nanoscale machines, the physical hierarchy of size scales suggests a good fit with a hierarchical design approach. A single component at the millimeter scale, easily specified and comprehended, may be composed of thousands of micron-scale functionally simple components, which may in turn be composed of thousands of nanometer-scale components.

The use of easily specified systems suggests another design technique: the use of design rules. This means that components are expected to work correctly within a certain range of conditions, and will only be used within that range. This implies that the components need not be completely characterized over all conditions—only over the conditions of intended use. Unfamiliar nanoscale phenomena can be ignored as long as they fall outside the planned scope of the design rules.

Of course, these approaches cannot access the full power of nanoscale design. Hierarchical design suppresses emergent properties, and design rules encourage the ignoring of potentially useful phenomena. If an engineering discipline arises that can deal with these complexities—genetic algorithms, for example—then nothing prevents the use of that discipline to augment the range of products that can be designed. However, as in computers, the available power and intricacy of even a restricted mechanistic design space will keep designers busy for many years. And even mechanistic designs will far outperform today's products.

B.3. Parallel and reliable operation

Because nanoscale machines have higher performance than larger machines (due to scaling laws as well as other factors), it will be advantageous for large products to use many nanoscale machines in parallel to perform large-scale functions. For example, instead of using one large motor, using 100 motors 1/10 as large will deliver the same

performance with 1/10 the volume and mass. Because the nanoscale is about a million times smaller than today's machinery, using 10^{12} nanoscale motors will provide the same power in one-millionth the volume. If ten times that many motors are used, then they can be run ten times slower, which will reduce drag by an order of magnitude—and will still be 10,000 times smaller than the motor they're replacing.

To run many devices in parallel requires predictable operation. This can be attained by conservative engineering. However, it also requires that failure be dealt with, because some small fraction of those trillions of components will inevitably fail (though the failure rate may be dominated by background radiation).

To deal with failure, simple redundancy at multiple hierarchical levels is sufficient. For example, if a system has a 1% chance of failing, then a non-redundant collection of 10 systems has almost a 10% chance of failing. But a collection of 11 systems, of which only 10 have to work, has only a 0.5% chance of failing. Calculations done for one kilogram-scale nano-structured hierarchical machine show that a 60% increase in mass will produce a reliable system out of 10^{17} sub-micron components with realistic failure rates of 0.024% per year per 200-nm cube.[93]

B.4. Efficiency, reversibility, and far-from-equilibrium machines

A system will be efficient when it does not convert high-quality energy to heat. This means that the most efficient systems will maintain a nearly constant energy balance: they may convert energy from one form to another, but the total of all non-thermal energy will stay the same throughout the machine's operation.

A machine that maintains a constant energy balance is called *reversible*. Because it uses no energy in going forward, it requires no energy to push it backward. Of course, in reality, any action will dissipate some energy through drag. But in the absence of static friction (see below), the smallest force can move a reversible machine in the desired direction, albeit at low speed.

A machine with varying energy levels can be part of a larger system with constant energy levels. For example, a mechanical process that experiences different forces at different points in its trajectory can be coupled to a cam-following spring that provides an equal and opposite force. (At low energy levels comparable to thermal noise, it becomes easier to analyze the behavior of such systems in terms of energy rather than force, because the effectiveness of a barrier to motion depends not on the force but rather on the energy; thermal noise can push past a high but narrow barrier. But in the absence of high forces, either analysis will work.)

Digital, reliable, predictable operation is compatible with reversibility. Broadly speaking, a system is digital when it has high barriers between the desired state (more precisely, the collection of desired states) and any undesired state. A ratchet is such a system; a spring must be compressed in order to move the system from one state to another. But the release of the spring wastes the energy of compression. An efficient system can be moved from one known state to another known state, and locked in place in the new state,

---

[93] "Design of a Primitive Nanofactory," http://www.jetpress.org/volume13/Nanofactory.htm#s8.5

without any waste of energy. This requires additional mechanisms to lock and unlock the system at the right time, but for nanoscale operations conducted frequently, it is worth doing.[94] Fig. B.1 illustrates such a design.
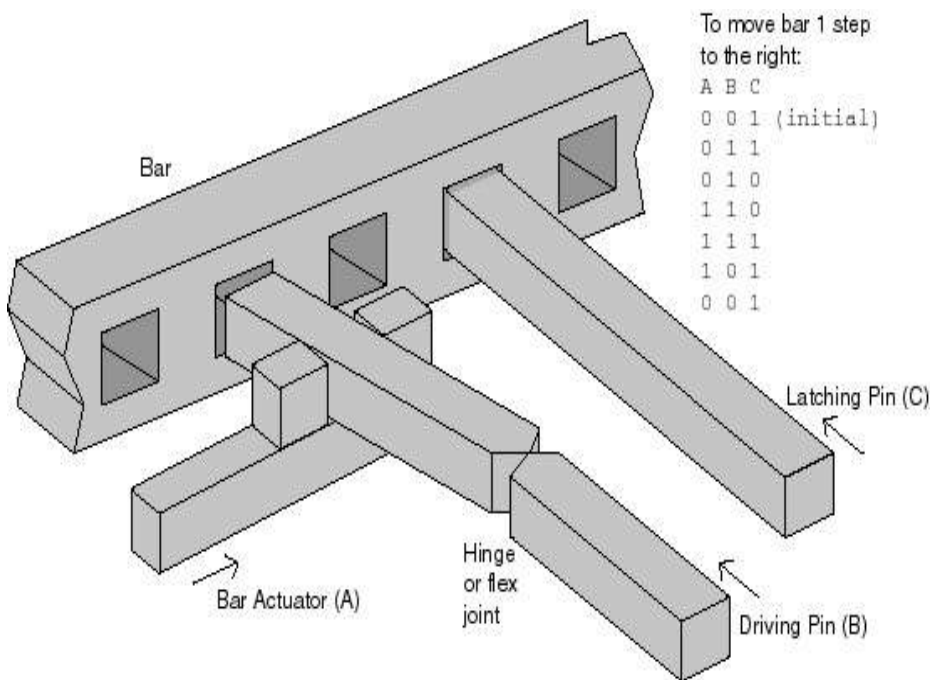
Note: A and B shown in the "0" position; C in the '1' position.
Normally B, C, or both are always in the "1" position.

Interface between Bar Actuator and Driving Pin must be stiff for efficiency. At nanometer scale, surfaces are approximate and the pin can be press-fit between closer-spaced knobs

Figure B.1 *A thermodynamically efficient stepping drive.*

A system that maintains a state far from equilibrium need not expend any energy to do so, just as a table expends no energy in supporting an object. A system can transition between a set of states at the same out-of-equilibrium level, just as a ball rolling on a table does not gain or lose energy. And a system with one component that can store or release energy as needed to keep itself in balance can make other components transition smoothly between different energy levels, again with no energy lost. Although calculating a system's total energy is important in traditional chemistry and in some branches of physics, it is irrelevant to the operation of nanosystems. What is important is the relative change in energy over the system's trajectory.

---

94   See "A thermodynamically efficient stepping drive,"
      http://www.jetpress.org/volume13/Nanofactory.htm#s3.1

B.5. Transport

Transport of small objects such as molecules poses an interesting conceptual challenge. It seems wasteful of both mass and energy to dedicate a mechanism to individually transporting each molecule. However, intuition is incorrect in this case, and mechanical handling of molecules can in fact be the best design choice.

Diffusive transport, in which a suspended particle moves from a source to a destination via Brownian motion, is often assumed to be efficient—probably because nothing directly touches the molecule. However, diffusion must be driven by a concentration gradient, which must be maintained at some energy cost. Diffusion requires mass and volume for the solvent. Also, the speed of diffusion decreases rapidly as the distance increases; it is only adequate at the smallest scales. Finally, once the particle arrives at its destination, its arrival must be sensed, and this requires energy. In fact, moving a particle at a certain speed via diffusion, and then detecting its arrival, requires as much energy as directly dragging it through the fluid.

If diffusive transport is no more efficient than direct dragging through fluid, then direct dragging without fluid will be more efficient. In other words, keeping hold of the particle and moving it through low-density gas or vacuum will use less energy than diffusion—as long as the friction in the mechanism is less than the drag that the particle would have experienced. But the mechanism's drag depends on its design, and can be reduced substantially. To take a simple example, a large wheel mounted on a small bearing can transport a particle at its rim for a long distance with only a slow motion over a small area at the bearing surface.

Diffusive transport may still be useful for design convenience, or for delivering molecules from bulk storage in solvent to a mechanical handling mechanism at the edge of the tank. But it cannot be supported on the grounds of efficiency.

B.6. Efficient sliding interfaces

Chapter 10 of *Nanosystems* analyzes several types of sliding interfaces between stiff covalent solids. It points out that interatomic forces in different contact regions between incommensurate surfaces can cancel each other, leaving a very low net force and a very low energy barrier at every point in the desired motion trajectory. This has special relevance for sleeve bearings, which can be designed so that the least common multiple of the period of the inner and outer surfaces is quite high, resulting in a force (or energy) pattern that has very high frequency and very low amplitude—in other words, the bumps are smoothed out.

Chapter 7 of *Nanosystems* analyzes the known mechanisms of energy dissipation. The basic conclusion is that energy dissipation is usually proportional to speed, and for reasonable design parameters (e.g. a speed of one meter per second) the dissipation will usually be small compared to k_BT.

These conclusions may seem surprising in light of the significant difficulties that friction has caused for MEMS. But it must be remembered that MEMS fabrication technology is incapable of producing precise smooth surfaces. Superlubricity, an orders-of-magnitude

103

reduction in friction between surfaces that are out of alignment, has been observed in graphite[95] and probably also in carbon nanotubes by Zettl.[96]

B.7. Other useful techniques

Electrical currents are used hardly at all in biological systems, perhaps because they are immersed in water. Ion currents are used for signaling, but only a few animals use them for delivering power. In constructed nanosystems, electricity may be used for a variety of purposes. Simple electrostatic actuators can deliver nanoNewtons of force in a few cubic nanometers; nanoscale electrostatic motors can deliver many microwatts per cubic micron. Electricity is probably not as dense or efficient a means of delivering power as rotating shafts, but may still be useful.

Digital control is a good fit for predictable, reliable, automated machines. Although general-purpose computation with standard CPU/memory architecture can be energy-intensive due to the need to erase bits (which has an irreducible thermodynamic cost), special-purpose digital logic that merely has to track a machine's state can be reversible.

Advanced materials can take advantage of nearly the full strength of covalent bonds. Although an imprecise structure can be weakened by small defects, anisotropic structures can result in components small enough that most components will be defect-free.

As explained in Chapter 9, micron-scale components can be very conveniently manipulated using surface (van der Waals) forces. This reduces the complexity of robotic handling of components. Instead of complicated grippers and attachment mechanisms, simple contact points will be sufficient.

Cooling will not be a problem for isolated nanosystems, even systems with high power density. However, large aggregates may produce thousands of watts of heat per cubic centimeter. *Nanosystems* section 11.5 describes a branched plumbing structure and a composite phase-change cooling fluid (using suspended encapsulated ice particles) that can remove 10,000 watts per cubic centimeter from a centimeter-thick slab.

Nanosystems Chapter 13 describes a variety of mechanism to import molecules from disordered (e.g. solvated) environments and process them. *Nanosystems* describes a variety of other useful designs, including gears, seals, motors, and sensors.[97] Robert Freitas's book *Nanomedicine I*[98] describes a number of additional devices and capabilities. These devices are implemented in a largely mechanical approach, so should be adaptable to a range of materials and construction techniques. Although it will surely be desirable to use other nanoscale phenomena, the ability to do so much with basic mechanics provides reassuring evidence of the utility and applicability of molecular manufacturing even in the absence of specific designs.

---

[95]   Enrico Gnecco , Anisoara Socoliuc, Ernst Meyer, Alexis Baratoff, Roland Bennewitz,Martin Dienwiebel and Joost Frenken, "Slippery Nanoworld," Europhysics News (2005) Vol. 36 No. 1, http://www.europhysicsnews.com/full/31/article2/article2.html
[96]   http://www.physics.berkeley.edu/research/zettl/projects/LinearBearing.html
[97]   see http://www.foresight.org/Nanosystems/toc.html
[98]   see http://nanomedicine.com/NMI.htm. The entire book is available online.

B.8. Incremental approaches to advanced functionality

The nanosystem concepts presented in Drexler's *Nanosystems*, which have become identified with the field of molecular manufacturing, are sufficiently different from present nanotechnology research directions that it may be difficult to see how such an approach can be developed. Although some researchers, notably Merkle and Freitas, advocate a direct approach using scanning probe microscopes to fabricate diamond-lattice machines, an incremental approach may be easier to fund and to accept the plausibility of. Fortunately, each aspect of design can be varied incrementally from current practice to advanced nanosystems.

The most familiar medium for chemistry is water; all known life uses water, it is a powerful solvent, and its polar nature makes it useful for electrochemistry and ion transport. However, water is not as necessary to chemistry as has been believed. Klibanov has found that some enzymes do not require water, and may not require any solvent at all. Other, less universal solvents may be useful for improved cleanliness and reduced reactivity. Perhaps the most chemically inert solvent is liquid xenon, which is nonetheless a good solvent for some kinds of molecules. Although some molecular machine mechanisms depend on solvent, such as some kinds of sliding interfaces and actuator molecules, they can be replaced when appropriate by others that do not—but that will still work in solvent.

Advanced molecular manufacturing aims to fabricate large and intricate molecular structures out of small and inexpensive molecules such as acetylene. However, this is not a requirement of molecular manufacturing systems. System design can progress from attaching large molecular building blocks in programmable locations via surface forces or weak bonds, to attaching smaller molecules via stronger bonds (either in programmable locations on a growing 3D structure, or in programmable sequence to lengthen a polymer chain). Eventually, as the sophistication, speed, and working conditions (e.g. lack of solvent) of the nanoscale machinery improve, more reactions can be done under mechanical control.

Self-assembly is a convenient way to attach relatively complex molecules into relatively simple structures. Pure self-assembly is limited in that the structure can contain no more information than is embodied in the molecules themselves. However, self-assembly can be made less "pure" by degrees. Potential self-assembly sites can be protected and deprotected either mechanically or chemically. Stronger bonding mechanisms can be added to secure molecules which have self-assembled in the right position. Molecules can be steered by external fields. Intermediate binding structures, which may be positionable, may confine molecules in a way that greatly speeds their reaction rate in chosen positions. Other conditions such as pressure can be added to invoke reactions in desired locations. Once reactants are being moved mechanically rather than by diffusion, they can be modified in several stages including the use of very reactive intermediates that are simply not allowed to contact anything that would spoil them.

Because primitive tools exacerbate such practical difficulties as thermal noise, it is currently difficult to track exactly what happens at the nanoscale. With improved nanoscale structures, it will be possible to implement deterministic nanoscale machinery

even at room temperature, and basic theory shows that deterministic (mechanically constrained) operation need not be expensive in terms of energy. Deterministic machines will probably be desirable for easier engineering. But until they are developed, modular unsynchronized machines can be separated by buffers and combined into useful systems, as they are in biology. As the means become available to synchronize adjacent machines in a chain of operation, the chain can gradually be linked together.

Digital delivery of information to nanoscale manufacturing systems will be necessary to turn terabytes of efficiently coded blueprints into advanced nano-products. In early nano-manufacturing systems, such high-bandwidth channels and receivers may not yet exist. In this case, most of the product's information can be built into molecular building blocks, with only basic positional information being delivered explicitly. A very early system might control only a three-axis positioner, a one-axis block-placer, and a microfluidic system to flush appropriate building blocks past the block placement system in the correct sequence. Later systems could use more programmable block placement systems and dispense with the microfluidics. Blocks could be made smaller in exchange for more placement information. Eventually, blocks would be replaced by feedstock, and full fabrication information would be delivered.

B.9. Molecular manufacturing contrasted with biological function

It is instructive to compare and contrast molecular manufacturing with biological systems. Biology is currently the most capable nanoscale manufacturing system in existence, and approaches to engineered nanoscale fabrication would do well to look for what can be emulated and what can be improved on.

B.9.1. Differences

Biological life must include several sophisticated functions that will not be needed in molecular manufacturing systems. Biological systems need to grow—to modify their structure over time. They need to reproduce—to form small viable packages that can carry on their genes. They must evolve; although evolution is usually not a requirement for any given organism at any given point in time, an organism that lost its capacity to evolve usually would not last long. They must deal efficiently with changing life stages and environmental conditions. With all the functionality required for growth and homeostasis, self-repair is a natural capability to possess. Although organisms cannot always completely correct accumulated damage, even at the cellular level—Alzheimer's, CJD, and aging itself are examples of aggregating problems—their self-repair capacity is still impressive. The ability to self-repair efficiently requires metabolizing and re-using their own substance; this parallels the ability to metabolize complex mixtures of complex chemicals obtained from the environment.

By contrast, machines—including nanomachines—have a much simpler task. They do not have to metabolize complex chemical mixtures from the environment—at least if they follow current engineering trends, any required substances will be delivered already more or less purified. Machines do not have to reproduce, grow, or evolve—in fact, it is far preferable that they do not. This means that machines do not have to incorporate tiny

integrated self-contained manufacturing centers. Even nanofactories, which are designed for manufacture of diverse products including duplicate nanofactories, do not need to include self-contained manufacturing units. Power conversion, operational control, environmental shielding, feedstock delivery, component fabrication, and component assembly can all take place in separate regions of the factory. This can simplify design considerably.

Although nanomachines will fail from time to time, it seems likely that it will be simpler and more mass- and time-efficient to include redundant units instead of self-repair or self-construction capability. A modest increase in mass and a minimal increase in functional complexity can allow a system to lose capacity gradually over a period of decades or even centuries—by which point a more advanced nanofactory would long ago have replaced it with a more advanced product.

B.9.2. Similarities

Despite the differences in function and purpose between biology and engineering, many of the same techniques can be used. Biology uses non-equilibrium chemical states to drive a diversity of reactions and processes; so will molecular manufacturing. Both use covalent structures. Both use thermal noise as a lubricant to smooth over small energy barriers in operation. Both use digital information storage; though biology has developed several impressive compression techniques, molecular manufacturing will be able to use more information and transmit it more efficiently. Both use macro-scale aggregates of molecular mechanisms working in coordination. Both use digital control to convert simple inputs into intricate outputs.

Any "trick" that is used by biology can be used also by either basic or advanced molecular manufacturing systems—the basic systems by incorporating biological materials and structures, the advanced systems by building the required structures. Stiff molecular structures can be built with hinges where reconfiguration is useful. Complexity can be built into engineered systems where it is useful; though usually the engineers prefer to avoid this, new methods or principles of design may be developed (perhaps through study of biology) that allow complex systems to be created. In any case, complexity, like digital logic, can be implemented in almost any physical substrate; construction by molecular manufacturing techniques will not prevent its use.

107

C. Issues and Constraints

Several practical issues will add to the difficulty of molecular manufacturing. Although none is an insurmountable problem, they will have to be taken into account in detailed designs.

C.1. Small size of molecules

Most molecules are too small to be conveniently sensed or manipulated with today's tools. Nanoscale tools will have more precise features and be more sensitive. Physical binding pockets, whether designed or evolved, will be able to bind molecules; even crude polymer molding techniques can make effective binding sites for molecules. Once physically bound, a molecule can be sensed by physical probing, locked in place by reconfiguring the binding site, and manipulated at will.

Another issue raised by the small size of molecules is the low throughput of manufacturing devices that handle molecules. However, scaling laws come to the rescue here. A molecule-handling machine made of a billion atoms, built on a 100-nm scale, would be able to move very quickly, performing perhaps billions of operations per hour. That would allow it to build its own mass in about an hour—a very good throughput for any manufacturing system. The problem then shifts from low throughput to the need to build and control vast numbers of manufacturing devices in parallel. But this is conceptually straightforward, and recent architectures such as planar assembly have shown that physical implementations can also be straightforward.

C.2. Thermal noise

Thermal noise causes a background of random motion that is pervasive and inescapable. (Quantum and Heisenberg uncertainty typically have much smaller effects.) For most nanoscale machine components, the effect of thermal motion on position can be estimated from the stiffness of the component. For nanoscale components at room temperature, the random displacements can be significant, especially in systems that are not stiff. Thermal motion is random, but high energy motions are rare; in general, thermal motion will not provide enough energy to break a single bond, and a single bond provides sufficient stiffness to limit motion to a fraction of a nanometer. In stiff systems, such as can be built on a 10-nm scale out of diamond, thermal displacements can be reduced to less than an angstrom. Such precision is usually not required; for mechanically interlocked systems, all that is needed is that the displacements do not cause parts to slip past each other, and it is not hard to engineer systems with displacements smaller than nanometer-scale features. This implies that even very small interlocked machine parts, on the scale of a single nanometer, can be designed not to slip past each other due to thermal perturbation.

One task that requires high precision is mechanosynthesis. The precision required depends on the type of synthetic operation. Depositing a molecular building block or adding to a polymer may require only nanometer precision, because there may be no

available side reactions to cause problems. However, operating at a chosen site on a covalent lattice surface may require sub-angstrom precision. This may be achievable at room temperature; if not, cooling to liquid nitrogen temperature should be sufficient. Selected single atoms have been removed from covalent lattice surfaces at room temperature.[99]

Thermal motion can supply enough energy to break chemical bonds, but only unusually weak or strained bonds will break at room temperature.[100]

C.3. Other error sources

Background radiation can cause errors; it requires a lot of mass to reduce and cannot be entirely eliminated. It may cause a failure rate on the order of 3% per cubic micron per year. Depending on the design, that cubic micron may contain many nanomachines in parallel, in which case the damage will only cause a reduction in capacity, not a complete loss of function. (A cubic micron has room for many millions of nanometer-scale features; most nanomachines will be smaller.) Light can supply enough energy to break bonds, but can be shielded by a sub-micron layer of metal.

Entropy is often cited as a source of error. Entropy has several different meanings. Thermal noise has already been considered. The word is often used by non-physicists to mean the tendency of things to fall apart or be imperfect, with a vague reference to thermodynamics. However, a more reliable rule of thumb is that "Diamonds are forever." Entropy will not cause a single error in a ton of diamond in a million years. Entropy is also cited as preventing well-organized systems from being built. However, this would only be a problem if power could not be supplied. Remarkably small amounts of power are required to overcome entropy and create organization; it must be accounted for in detailed molecular-level models, but is merely a correction factor in the operation of most machines.

Design errors can cause machines not to work as designed. These can be caused by bad logic or inaccurate models. Once an error is discovered, through a machine failing to work as expected, the error can be researched and corrected.

Some designs may allow failures to propagate. A failing component can overload adjacent components, causing them to fail as well. This can be seen at scales as large as power grids and as small as crack propagation. Obviously, this pattern of failure is not unique to nanomachines. Because radiation-induced failure is inevitable, all large nanosystems will have to be built with redundancy, and the frequency of failures will provide many tests of the systems.

C.4. Heat removal

Single nanoscale machines will be easily cooled by conduction. Aggregates of machines may need cooling due to high power density. However, efficiency can be increased at the

---

[99] C. R. Kinser, M. J. Schmitz, and M. C. Hersam, "Conductive atomic force microscope nanopatterning of hydrogen passivated silicon in inert organic solvents," *Nano Lett.*, **5**, 91 (2005).
[100] *Nanosystems* 6.4

expense of speed, so adding more machines and running them all more slowly will reduce the total heat generated. For example, a one-megawatt bank of electric motors might occupy a cubic millimeter and generate a kilowatt of waste heat—rather difficult to cool. But a cubic centimeter containing a thousand times as many motors could convert a megawatt of power while generating only one watt of heat. In general, systems expected to exceed the power densities of today's systems by only a few orders of magnitude will not have a problem with cooling.

C.5. Complex and poorly understood phenomena

Several phenomena at the nanoscale are not well understood today. The folding of proteins is becoming somewhat predictable, and several proteins have been designed including a novel fold. However, proteins need not be used; nucleic acids fold quite predictably, and Schafmeister's polymers are stiff and do not need to fold in order to take their designed structure.

Mechanisms of dissipation are not fully understood, but superlubricity has been demonstrated in graphene systems.

Semiconductor electronics become problematic at the nanoscale because of quantum effects. Although quantum effects can usually be ignored when dealing with the mechanics of molecular structures, electrons are smaller and more strongly affected. However, complete nanosystems including computers can be built without the use of electronics. Electrical currents and electrostatic actuators with multi-nm separation to reduce tunneling will work just fine.

Chemistry is not fully understood. However, only a few reactions are needed for mechanosynthesis, since a programmable mechanosynthetic system can re-use reactions in different sequences or positions. Computational chemistry can provide fairly good indications of which reaction trajectories will produce the desired results, especially since mechanical confinement of the reactants imposes a very high energy cost on unwanted trajectories. When likely reactions are identified, the models can be fine-tuned by targeted experiments.

C.6. Need for extremely reliable reactions

The yield of wet-chemistry reactions is rarely above 99%. However, mechanosynthesis requires far higher yields, because many operations in sequence are required to make a product.   A primitive manufacturing apparatus made out of relatively large molecular nanoblocks may require only a few dozen blocks. However, 50% yield after only 50 steps requires 98.6% yield at each step. As systems become larger and include more components, the reliability of each step must improve to "six nines" or even "nine nines" for a billion-atom machine constructed one or two atoms at a time.

Reliability of a reaction can be improved by several means, including mechanically blocking unwanted reaction trajectories, purifying the feedstock (reactants) with carefully designed binding/sorting mechanisms, reducing the size of the feedstock to make it easier to sort and purify, and lowering the temperature to reduce the effects of thermal noise.

110

Note that lowering the temperature does not necessarily reduce the reaction rate, since mechanical pressure can be used to reduce or even eliminate reaction energy barriers.

C.7. Need for manufacturing closure with limited palette

In order for molecular manufacturing to work, small manufacturing systems must make many duplicate manufacturing systems under automated control. Although complete blueprints do not exist for an engineered nanoscale collection of sensors, actuators, and structure that will be capable of using its few operations to build a complete duplicate of itself using only simple feedstock, energy, and information, there is every reason for optimism.

There are several polymer systems in which intricate and useful three-dimensional structures can be specified by the sequence in which monomers are added to the chain. These include DNA, RNA, proteins, and Schafmeister's stiff polymers. This is one pathway by which a limited set of operations can make a virtually unlimited set of structures. Another pathway is to control the position of deposition operations on the surface of a growing object in order to build arbitrary shapes. Freitas has written that the number of tools necessary to build diamond lattice may be as few as six.[101]

Given the ability to build arbitrary structure, the next requirement is to implement actuation, control, and sensing. Control and computation can be implemented with sliding mechanical interlocks. Computers based on this approach could actually be far in advance of today's technology in terms of size and computations per watt, though the clock speed might be limited to ~1 GHz. Tactile sensors, including mechanisms to detect whether a molecule is occupying a binding site, could be implemented mechanically. Actuators could be implemented either with stepping drives controlled directly by mechanical digital logic, or if electrical power is available, actuators could be electrostatic and controlled by mechanical switches.

Mechanosynthetic systems can be implemented with these basic ingredients of robotics. A stepping drive can be used to control one leg of a Stewart platform; two drives in parallel with slightly different step sizes can implement Vernier control. If a small set of molecule-manipulating "tool tips" cannot be found that is capable of synthesizing the whole set, then small molecules can be pre-built and imported into the product. As long as the number of tool tips is fairly small, this is not onerous even if the synthesis is expensive; the mass fraction of the tool tips will be a tiny fraction of one percent in nanofactories, and will drop to zero in nanofactory products that do not need to do mechanosynthesis.

Although a broader palette of functionality would of course be desirable, and additional functions will be able to be implemented in almost any construction system, the above list of basic functions—structure, sensing, control, actuation, and mechanosynthesis—form a solid foundation for manufacturing closure.

---

[101] http://www.foresight.org/stage2/project1A.html

D. Further Reading

Nanosystems by K. Eric Drexler (Wiley Interscience, 1992) remains the foundational book on molecular manufacturing and the design of stiff precise nanoscale machines. A very brief summary of the book is at http://crnano.org/5min.htm. A detailed table of contents, with links to Chapters 1 and 2 and the Glossary, is at http://www.foresight.org/Nanosystems/toc.html

Nanomedicine Volume I: Basic Capabilities by Robert Freitas (Landes Bioscience, 1999) explores the application of high-performance nanosystems to medicine. Much of the book is not specific to medicine, but rather a generally applicable exploration of product design and performance. The entire book is available online at http://nanomedicine.com/NMI.htm

"Design of a Primitive Nanofactory" by Chris Phoenix (*Journal of Evolution and Technology*, Oct. 2003) is a system-level analysis of a kilogram-scale integrated molecular manufacturing system. Although the "convergent assembly" design is superceded by the "planar assembly" approach described in the present document, many concepts and calculations may be adapted for other nanofactory designs. The paper covers mass, physical layout, throughput, power, computation, cooling, error handling, product design, and a strong nanometer-scale mechanical fastener useful for joining micron-scale or larger parts. It is available at http://www.jetpress.org/volume13/Nanofactory.htm

An animation by John Burch and Eric Drexler shows how an advanced nanofactory might be organized and how it might function. The molecular transfer tools illustrated in the animation have been verified by computational chemistry. Illustrations and links to the animation are available at http://lizardfire.com/nanofactorySS/index.htm

Portions of this document were adapted for a paper presented to the Society for Manufacturing Engineers conference, "Molecular Nanotechnology and Manufacturing: The Enabling Tools and Applications." That paper may be more readable for a general audience. A version is available online at http://wise-nano.org/w/Doing_MM

This document includes many suggestions that are based on currently available tools and designs. Some of these suggestions will become rapidly obsolete as better tools and theories are developed. Updated versions of this document will be published at: http://wise-nano.org/w/NIAC_Results
The website allows interested readers to comment on the paper, and an errata list will be maintained.