

Molecular Manufacturing: What, Why and How

Chris Phoenix, Director of Research, Center for Responsible Nanotechnology

Abstract

Molecular manufacturing emphasizes the use of precise, engineered, computer-controlled, nanoscale tools to construct vast numbers of improved tools as well as products with vast numbers of precise, engineered nanoscale features. It has not been clear how to design and build the first nanoscale tools to start the process of scaleup and improvement, or how easily the operation of many advanced tools could be coordinated. This paper develops a roadmap from today's capabilities to advanced molecular manufacturing systems. A number of design principles and useful techniques for molecular construction via nanoscale machines are discussed. Two approaches are presented to build the first tools with current technology. Incremental improvement from the first tools toward advanced integrated "nanofactories" is explored. A scalable architecture for an advanced nanofactory is analyzed. The performance of advanced products, and some likely applications, are discussed. Finally, considerations and recommendations for a targeted development program are presented.

Summary:

The paper is organized into eight sections.

1. "Basic theory and observations" covers topics that apply to a broad range of molecular manufacturing systems.
2. "Primitive molecular manufacturing systems" proposes two designs for systems that appear capable of development with today's technologies.
3. "Incremental improvement" discusses how to develop the primitive systems toward the goal of self-contained kilogram-scale manufacturing systems.
4. "High performance nano and micro systems" describes several designs and techniques that can be useful in advanced nanofactories and products.
5. "Advanced nanofactory architecture and operation" provides an overview of the possible architecture and function of an advanced nanofactory.
6. "Advanced product design and performance" describes the performance available to products built out of strong nano-structured materials, and includes a brief discussion of design approaches.
7. "Incentives and applications" explores some reasons why nanofactory products will be inexpensive to produce and quick to develop, and illustrates several ways that plentiful, inexpensive, advanced products might be used, demonstrating the high value of molecular manufacturing.
8. "Targeted development of molecular manufacturing" explores some of the issues an organization will have to consider in deciding whether to launch a program targeted at developing molecular manufacturing, and some of the desiderata for such a program.

Introduction

As access to the nanoscale improves, increasingly complex materials and devices are being constructed. Recent talk of “nanomanufacturing” emphasizes this trend. Nanomanufacturing refers to advances in conventional materials processing, using large processing equipment and nanoscale phenomena to make small, usually simple products, such as nanoparticles. However, achieving the high expectations that people have of nanotechnology will demand more than incremental development of familiar tools and approaches. Molecular manufacturing is a fundamentally different approach, using nanoscale machines to construct engineered, heterogeneous, atomically precise structures by direct manipulation. Its ultimate goal is to build complex products, both small and large, with atomic precision. It differs from nanomanufacturing in its emphasis on the design, construction, and use of precise and highly intricate nanomachines to handle individual molecules for construction; this enables a reduced reliance on self-assembly to form intricate and large-scale structures.

Basic Theory and Observations

This section explores several topics that will be applicable to a broad range of nanosystems, including nanoscale manufacturing systems. Scaling laws indicate that nanosystems will have extremely high performance. The granularity of atoms presents an opportunity for manufacturing systems that create products with no loss of dimensional precision. Nanoscale manufacturing systems will need a high data input rate in order to create highly functional products—so high that the data will need to be processed at the nanoscale. Sources of error are discussed, along with ways of dealing with error. Two approaches to scaling up production are contrasted. Mechanochemistry, the use of mechanical systems to do controlled molecular synthesis reactions, is too large a topic to discuss in any detail, but the range of available options is briefly surveyed. Design rules and observations relating to efficient motion at the nanoscale are presented, including a comparison of efficient design in soft vs. stiff machines. Finally, the per-atom energy budget of a kilogram-scale nanofactory is discussed.

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.”

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

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 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, then the area of the sheet will not change as the machines shrink; 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 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.)

Atomic granularity

At the energy levels at which nanoscale machinery operates, atoms are indivisible, and all atoms of an isotope are identical (different isotopes of the same element are identical for most purposes; 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.

Information delivery rate

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.

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 IBM², single atom cobalt-based transistors at Cornell³, HP's recently announced "crossbar,"⁴ and the rotaxane switches at UCLA⁵; this work indicates that transistors can be built from individual molecules and that logic circuits can be built from supramolecular structures.

Error sources

A non-zero error rate is inevitable, but the rate can be quite low in well-characterized systems. Wear and manufacturing slop will not be sources of dimensional variation in covalent components. The major source of dimensional variation will be thermal motion, which is ubiquitous and significant in nanoscale systems. (Quantum and Heisenberg uncertainty are typically much smaller effects.) Thermal motion will cause positional variance. For most nanoscale machine components, the effect of thermal motion on position can be estimated from the stiffness of the component. 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. 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.

To resist thermal motion driving a system from a set state to an undesired state, an energy barrier is required. Depending on several factors, a suitable barrier will typically be in the range of 30 to 80 times $k_B T$ (120 to 320 zJ⁶ at room temperature). This does not mean that systems must spend 120 zJ for each state change; see the section on "Energy requirements." Note that a particular state will always encompass a range of positions; some designs may tolerate quite large positional variance, as long as the system is not allowed to slip into a state that is functionally different. Energy will be required to "compress" a system into a tightly constrained configuration, but this energy may be recovered if and when the system is allowed to relax (this is equally true for mechanical and entropic springs).

Due to adverse scaling of stiffness with size, operations that require fractional-nanometer precision require careful design to provide adequate stiffness and alignment. Design and analysis indicate that 100-nm-scale machines built of stiff covalent solids will be stiff

² 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

³ "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

⁴ <http://www.hp.com/hpinfo/newsroom/press/2005/050201a.html>

⁵ 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>

⁶ 1 zJ or zeptojoule = 10^{-21} joules.

enough to maintain the position of an articulated component within a fraction of an atomic diameter.

Ionizing radiation is an inescapable source of damage. An ionizing event can disrupt the structure of a molecular machine. A cubic micron volume will have on the order of several percent probability of damage per year unless large amounts of shielding are used. Non-ionizing photons may excite bonds and cause damage, but can be excluded by a fraction of a micron of metal shielding.

Error handling

In a kilogram of nanomachines, no matter how well designed, there will inevitably be a large number (though small percentage) of broken machines. The appropriate method of error handling will be determined by the architecture of the system. This section will consider error handling in several kinds of manufacturing systems; some of these methods will also be useful for various kinds of products.

One approach to scaling up production involves many small independent fabricators, each capable of producing more fabricators or small products. If each fabricator is independent of the others (for example, free-floating and driven by broadcast signals⁷), then failure of a fabricator can be ignored, since it will only result in the system producing marginally fewer nanoscale products than requested. (This assumes that failed fabricators will not be able to create generations of flawed-but-functioning fabricators, but this is probably a safe assumption given the digital nature of errors.) As long as the error rate of a fabricator is lower than the number of operations required to build another fabricator or product, it will produce a useful percentage of good product even if no error checking at all is implemented during manufacturing. Of course, a percentage of products will also be non-functional, which may require post-manufacture error detection depending on the application.

The other approach to increasing production uses large integrated manufacturing systems. In such a system, failures must be detected. If each machine is responsible for producing its own fraction of a large product, then the product will have a component failure rate at least as high as the percentage of undetected errors in the system that made it. In multi-generational manufacturing, this would not be sustainable. Even if the resulting manufacturing system could function with large parts of itself non-functional or missing, failure to detect and compensate for the errors would result in accumulation of errors from generation to generation. However, this does not require every operation to be checked for error; all that is necessary is that a broken subsystem not be tasked with producing its full share of the product, and that no critical part of the product be omitted.

Any large nanomachine-based product, including integrated manufacturing systems, must expect and cope with a non-zero component failure rate. In a system that is structured

⁷ For example, see "Casing an Assembler" by Ralph Merkle, *Nanotechnology* **10** (1999) 315-322, <http://www.foresight.org/Conferences/MNT6/Papers/Merkle/>

hierarchically, this can be accomplished by modest degrees of redundancy on multiple levels. In a system organized this way, errors need not be reported outside the node where they are detected. If a sub-node fails, the node reassigns tasks to its remaining working nodes. If too many sub-nodes fail, the node signals failure to its super-node. If the failure rate of sub-nodes is small, then the chance of multiple sub-nodes failing and causing the failure of a node will be quite a lot smaller; a system with multiple levels of redundancy can have rapidly decreasing probability of system failure even as the size of the system increases.

For example, suppose that product operation depends on 2^{57} ($\sim 1.4 \times 10^{17}$) components, and the components are arranged in a hierarchical 8-way tree 19 levels deep with no functional sharing between substages or components (they all need to work). Adding one additional redundant component at the lowest stage, and one additional substage at each of the next three stages (9-for-8 redundancy at each of the four lowest levels), will increase the overall mass by 60%. But if component failure rate is 0.0024 (approximately a ten-year radiation dose for a 200-micron cube), this level of redundancy will ensure that the chance of even a single one of the 2^{45} non-redundant fifth-level substages becoming unavailable is 0.00001.⁸ Additional design effort will greatly reduce the mass penalty of redundancy, but simple redundancy at multiple levels is a useful starting point.

Space or high-altitude applications face an additional complication. Cosmic rays not only increase the dose of radiation, they also create a different pattern of damage. A cosmic ray is a very heavy charged particle—the nucleus of an atom—moving at very high speed. It does not create randomly spaced points of damage; instead it creates a dense track. Where cosmic rays are present, error handling designs must cope with multiple simultaneous failures in adjacent machinery; this means that any critical functionality must be distributed at several separate points.

Scaling up production: 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.

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

⁸ This calculation is from Appendix A of “Design of a Primitive Nanofactory”; see also Sec. 8.5.

J. Hall's "Utility Fog,"⁹ T. Toth-Fejel's "Kinematic Cellular Automata Self-Replicating System,"¹⁰ and NASA's "Autonomous NanoTechnology Swarm" (ANTS)¹¹. Toth-Fejel's paper "Legos to the Stars" summarizes earlier work.¹²

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 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.

Mechanosynthesis

Mechanosynthesis is the use of mechanical systems to control individual molecular reactions with atomic precision in order to build precise structures. 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). 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.

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

⁹ <http://discuss.foresight.org/~josh/Ufog.html>

¹⁰ http://www.niac.usra.edu/files/studies/final_report/883Toth-Fejel.pdf

¹¹ <http://ants.gsfc.nasa.gov/>

¹² *The Assembler*, September 1996, <http://www.islandone.org/MMSG/>

reactants near each other, increasing the effective concentration by several orders of magnitude. If the barrier is too high for simple positioning to work, it can often be lowered by pushing the molecules together. (The conventional form of this, applying hydrostatic pressure to affect reaction rates, is called piezochemistry.) Light can also supply energy to activate molecules and overcome barriers. 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 to activate bonding in only the desired locations.

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. This allows a particular deposition site to be selected from among many chemically similar sites. Engineered heterogeneous products can be built by mechanosynthesis that would be nearly impossible to build by self-assembly or common solution chemistry.

In machine-phase chemistry, the lack of chemically active solvent simplifies computational chemistry simulations. However, it does not necessarily limit the richness of available synthetic reactions. Water, a very complex medium, was thought by some to be necessary for complex chemistry; however, practical experience shows that natural enzymes and antibodies can work as well, if not better, without water—sometimes, without any solvent at all.¹³ A diverse set of machine-phase reactions has been accomplished using scanning probe microscopes.

Lack of solvent allows higher-energy molecules to be used—for example, radicals that would not be stable in solution. 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. In particular, it seems well suited to the creation of highly crosslinked materials, such as covalent solids, since the surface can be locally depassivated in preparation for a precise deposition reaction. The use of higher-energy (extremely reactive) molecules can increase reliability of both the actual reaction and simulations of it, because the energy difference between desired and undesired outcomes can be much larger than is common in solution chemistry.

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 pressure¹⁴, 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

¹³ A. Klibanov describes his 1983 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

¹⁴ Sullivan, Terena P.; van Poll, Maaïke 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

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. 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 5% of the strength of diamond. 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.

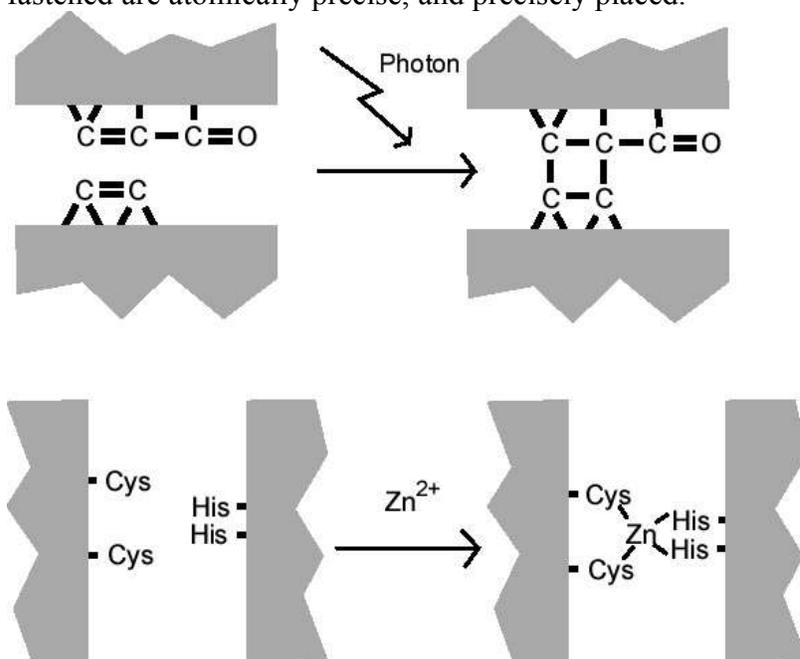


Fig. 1. Zinc binding and photochemical binding are proposed for use in primitive molecular manufacturing systems.

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 far greater than the inaccuracy of the program, then computational chemistry tools that must be used with caution for unconstrained solution chemistry may be used with more confidence to evaluate mechanosynthetic reactions.

Energy requirements

The large number of operations necessary to build large objects by manipulation of individual molecules is grounds for concern about the energy budget of such operations. An early target is 100 kWh/kg, which would cost about \$10/kg at today's energy rates. More advanced nanofactories that implemented recovery of mechanosynthetic reaction energy might achieve 10 kWh/kg, suitable for a desktop device. This corresponds to about 720 zJ (7.2×10^{-17} J) per atom. (Some products, such as nano-structured computer chips, would be worth many millions of dollars per kilogram in today's market. However, the corresponding power densities would make small kilogram-per-hour nanofactories impossible to cool.)

Digital computation can use a lot of energy. An early design for a 32-bit 1 GHz CPU was calculated to use 74,000 zJ per clock cycle. This design was mechanical, but reasonably efficient; it used reversible logic, and split its energy between overcoming friction and erasing bits. (For thermodynamic reasons, clearing a byte of information dissipates about 23 zJ at room temperature: $\ln(2)k_B T$ per bit.) This indicates that custom circuits (or even mechanically fixed trajectories) rather than general-purpose computation should be used for low-level control of machines handling small numbers of atoms per operation.

A covalent reaction may dissipate around 500 zJ, although this energy can be recovered if the reactants are held by mechanisms in a sufficiently stiff manner to avoid “snapping” from one position to another.

It is often useful for a subsystem to maintain a certain state until pushed to the next state by another subsystem. This is useful for at least two reasons: it simplifies the control architecture, and it saves the energy that would be required to “compress” a system from an unknown state into a known state. In a macro-scale system, keeping a component in a definite state might be accomplished with a detent or ratchet that imposes an energy barrier to movement. However, the size of the barrier does not scale to the nanoscale; no matter how small the machine, the barrier must be at least 120 zJ to resist thermal motion (at room temperature). Simply pushing past the barrier, without a mechanism to recover the 120 zJ as the system snaps into the new position, would be quite wasteful. An efficient design will attach to a system in a known state, remove the barrier, move the system to the new state, replace the barrier with the system now on the other side, and then release the attachment. The only losses in such a mechanism will be frictional. In Fig. 2, the function of the barrier is fulfilled by the latching pin; the external latch is engaged by the driving pin; the bar actuator moves the bar to the new state after the latching pin is withdrawn; and then the driving pin is removed after the latching pin is re-engaged. No energy is wasted in abrupt “downhill” transitions.

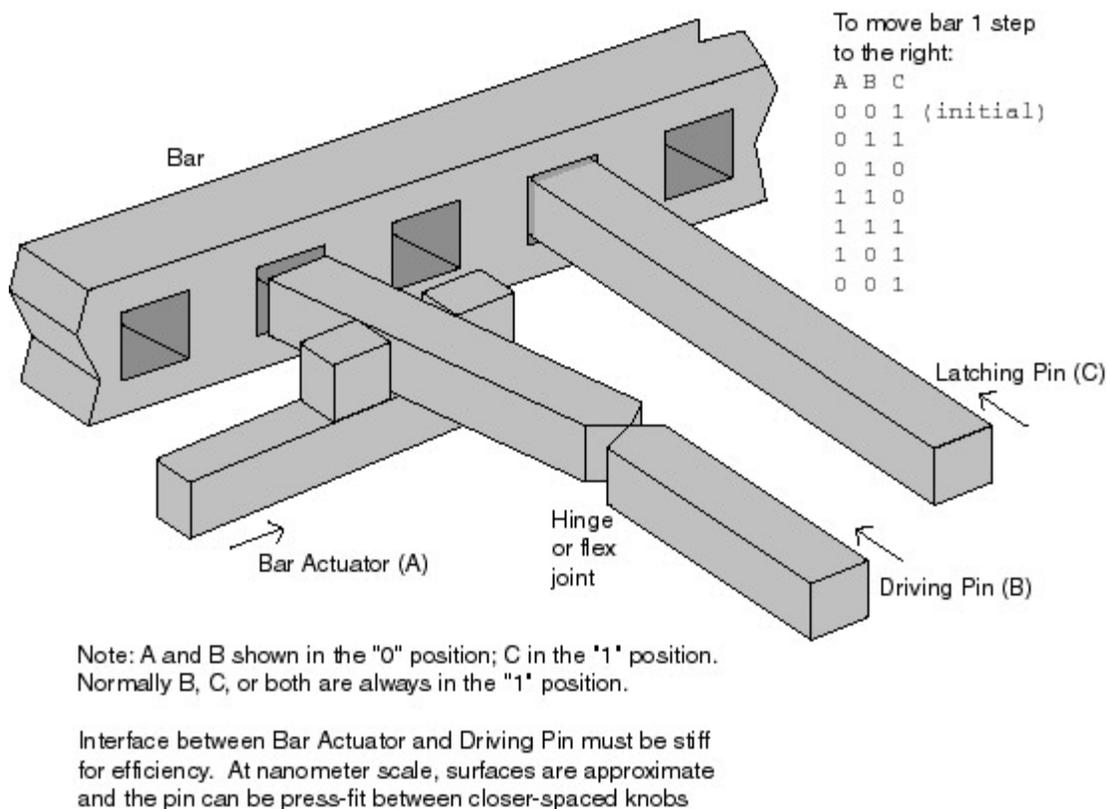


Fig. 2. A schematic of a thermodynamically efficient repositioning system.

Primitive Molecular Manufacturing Systems

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 "nanoblocks" 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 nanoblocks and the functionality of nanoblocks and nanosystems. Finally it briefly analyzes the productivity of the general nanoblock-placement approach, and discusses how productivity could be scaled up.

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.¹⁵ 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.¹⁶ 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 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.¹⁷

To form a solid product, blocks must fit together. See Fig. 3. 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.

¹⁵ 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>. See also http://wise-nano.org/w/Doing_MM for updates to this paper.

¹⁶ <http://www.math.uwaterloo.ca/~bingalls/BIRS/abstracts/seeman.html>

¹⁷ <http://www.foresight.org/Conferences/AdvNano2004/Abstracts/Schafmeister/>

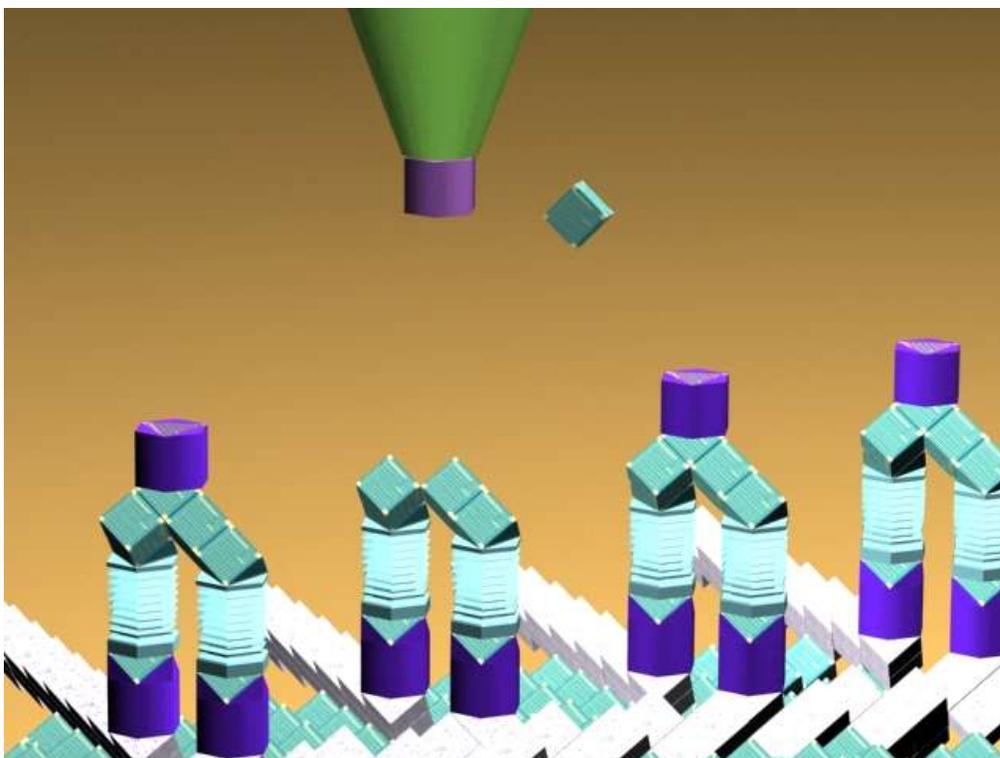


Fig. 3. Tattooing a Row: Depositing nanoblocks via scanning probe to build an array of “Tattoo” block-placement machines. Note the orientation of the deposited blocks. The free-floating nanoblock is about to enter the corner-shaped cavity in the probe tip. The tip will then push the block into the nearly 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), 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 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.

“Tattoo” architecture

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

Given a way to reversibly bind a nanoblock 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 nanoblocks 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 is established, the next step is to build a "tattoo needle." 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 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. 4A. Alternatively, a special prefabricated socket structure could be attached to a distinctive mounting point by self-assembly. See Fig. 4B.

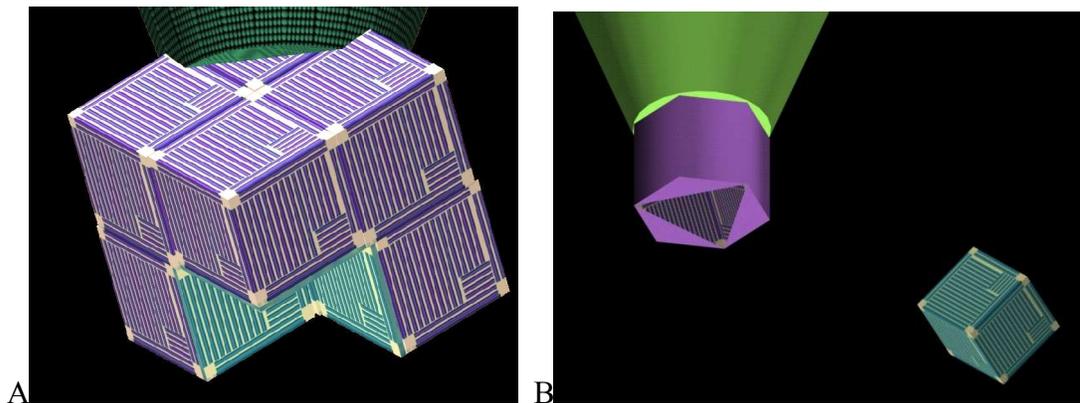


Fig. 4. Two ways to implement a socket: A) Place several blocks to form a cavity, perhaps followed by post-processing to modify the inner faces. B) Use a special-purpose molecular building block.

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 or array of machines. These machines could then be used in parallel for higher-throughput manufacturing of larger arrays.

If the machines 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 nanoblock-specific sockets can be built on different machines in the grid, then multiple nanoblocks can be mixed in one feedstock solution and each machine activated only when its selected type of nanoblock 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 nanoblocks 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 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.

“Silkscreen” architecture

The Silkscreen architecture is based on the idea of separating a solution containing nanoblocks 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.) Like the Tattoo approach, the Silkscreen approach repositions the placement machine relative to the product using a 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.

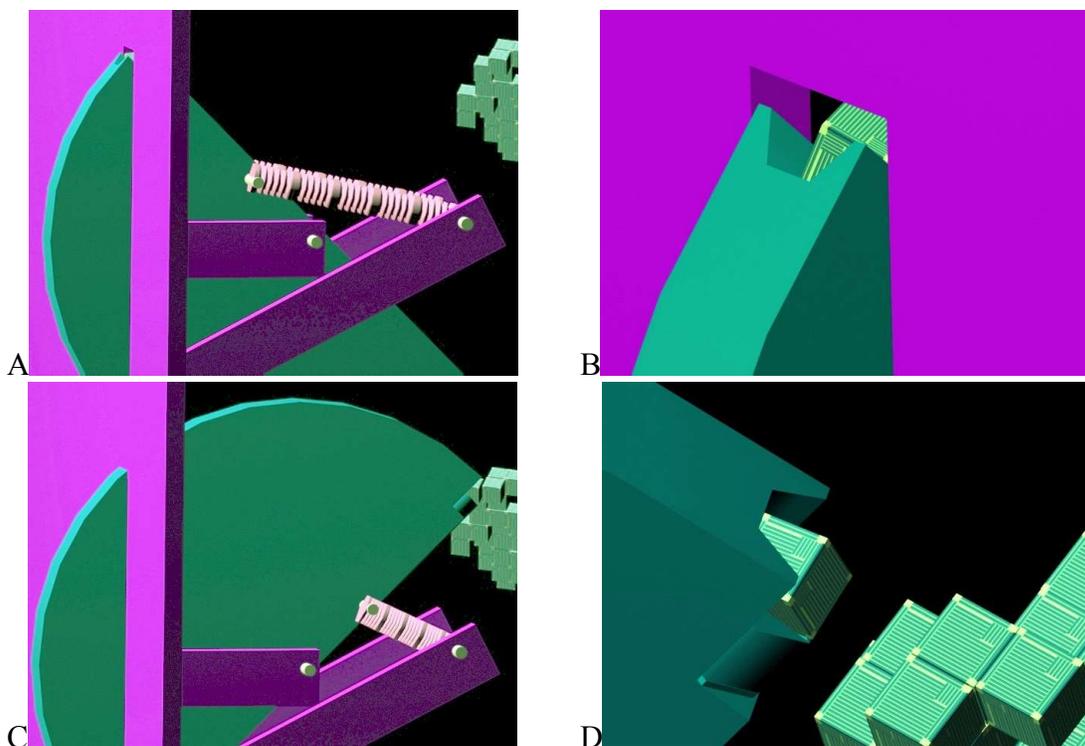


Fig. 5. A) The manipulator is in position to gather a block from solution. B) A close-up of the socket's position with a block inserted. C) The manipulator has moved to attach the block to the product. D) A close-up of the block about to be pressed into place.

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.

Handling and joining nanoblocks

The feedstock of a primitive “wet” molecular manufacturing system will be prefabricated molecular or macromolecular blocks, a few nanometers in size, dissolved or suspended in a solvent. The function of the manufacturing mechanism will be 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. 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.

There are many ways that blocks can be strongly joined—a wide range of chemical reactions or intermolecular forces can be used. Two attractive possibilities are photochemical crosslinking and the binding of zinc or other metal ions to certain amino acids, both described in the Mechanosynthesis subsection.

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 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.

Attachment mediated 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. Patterns of charge, asymmetrical shape, short 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.

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 mis-alignment of the binding sites can reduce the binding force, and adding more sites can increase the force.

Nanoblock 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.

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, but this may be acceptable in 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.

There are several kinds of electrically actuated materials. 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 are probably not suitable. 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 for molecular actuators that use only electron exchange. Annulene-based actuators can deliver strains of almost 6%.¹⁸ Poly(calix[4]arene-bithiophene) has been calculated to produce 78% strain, and another thiophene, poly(quarterthiophene) (poly(QT)), has exhibited reversible strains of 20% at 5 volts¹⁹. Such actuators would seem to fulfill the required function. These redox powered actuators could presumably be protonated by battery-type reactions that are driven by varying electric currents from nearby nano-wires. 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

¹⁸ 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.

¹⁹ 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)

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.

Throughput and scaleup of nanoblock placement systems

Activating a molecular actuator via photons or electric fields might take on the order of a millisecond. Diffusion of nanoblocks 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.

Incremental Improvement

Although it could have practical applications, the main purpose of a primitive molecular manufacturing system would be to build a better system. The ultimate goal is an easy-to-use nanofactory which 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. The 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 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.

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.

Removing the solvent

Solvent is convenient for diffusive transport, but 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 actuators (though the preferred actuators will not use ion movement but electronic changes). There are many ways of fastening blocks into 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. (In later designs, with smaller feedstock molecules and better construction methods, it should be possible to operate without any solvent leaks.) Liquid xenon may be a good choice for a solvent. 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. Radicals that could not survive in solvent can be maintained in vacuum, held separate by mechanical control of their molecules until they 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 far 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.

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 another 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 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, the factory design must allow it to escape, but many reactions (for example, cycloadditions) rearrange bonds to produce a single product structure. Another approach is to begin with inert molecules, then make them reactive by removing passivating structures, and finally position them to bond spontaneously.

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. This somewhat reduces the cleanliness requirements. It also allows the solvent to be used to deliver feedstock 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.

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.) 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.

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.

High Performance Nano and Micro Systems

Basic scaling analysis and theoretical designs using known materials indicate that the ultimate performance limits of nanomachinery may be many orders of magnitude higher

than is achieved by either biology or contemporary machines including computers. This section focuses on high performance nanoscale designs; the next section addresses the integration of these designs into large products.

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 C₆₀ (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 uses materials of diamond-like strength. Much of this section is based on calculations from *Nanosystems*.

This section includes discussion of the convenience of surface forces for manipulating micron-scale components; a mechanical fastening system that requires no external manipulation; an efficient class of bearing; electrostatic actuators, including a reversible motor/generator; digital logic; and mechanochemical power conversion.

Surface forces and component manipulation

Due to electron interactions, an attractive force will develop between almost any objects in close proximity. The force between objects in close contact is on the order of 1 nN/nm², though it decreases rapidly with distance.²⁰ For a cubic micron part, massing on the order of 10⁻¹⁵ kg and weighing ~10⁻¹⁴ N, the contact force of a single square nanometer provides 10⁴ times the force of gravity. 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.

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.

The potential energy per square nanometer of two diamond surfaces touching (0.2 nm apart) vs. separated is about 225 zJ. A contact area of less than 1 nm² should be adequate to hold against thermal vibration. If materials with lower Hamaker constant are used, or spacers are added to reduce material strength requirements, then the area would need to be increased. For example, with 0.2 nm spacers, the energy per square nanometer is about 56 zJ, and 2.2 nm² of contact area would be needed.

These numbers indicate that manipulation of micron-scale blocks does not require mechanical grippers. Contact pads of a few square nanometers area can implement the functionality of grippers. 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

²⁰ Calculations in this section are based on *Nanosystems* Sec. 3.5 and Fig. 3.10 d.

few square nanometers is sufficient to determine which of two opposing "grippers" a part will stick to when the grippers (pads) are pulled apart. Assuming that binding force and energy are proportional to area for all pads, 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 a manipulator, and install contacts of total area $2A$ on any convenient wall or other fixed surface. To stick the part to the wall, move it into contact, and then withdraw the manipulator's contact points one by one. To remove it from the wall, simply retract the manipulator with all three contacts engaged, and the block will come loose from the wall.

More elegant designs can be easily imagined. If the manipulator is capable of bending motions, then rotating its pads on edge will allow them to disengage from the block, while a combination of translation and slow bending will rotate the whole block and disengage it from the wall. Alternatively, a small plunger can be used to push the block away from a large pad. A sphere-and-cup contact can allow the block to rotate; six struts of variable length with sphere-and-cup contacts could be used to manipulate the block like a Stewart platform.

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 unpowered assembly described below. The expansion of the ridge opens a space in its 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.

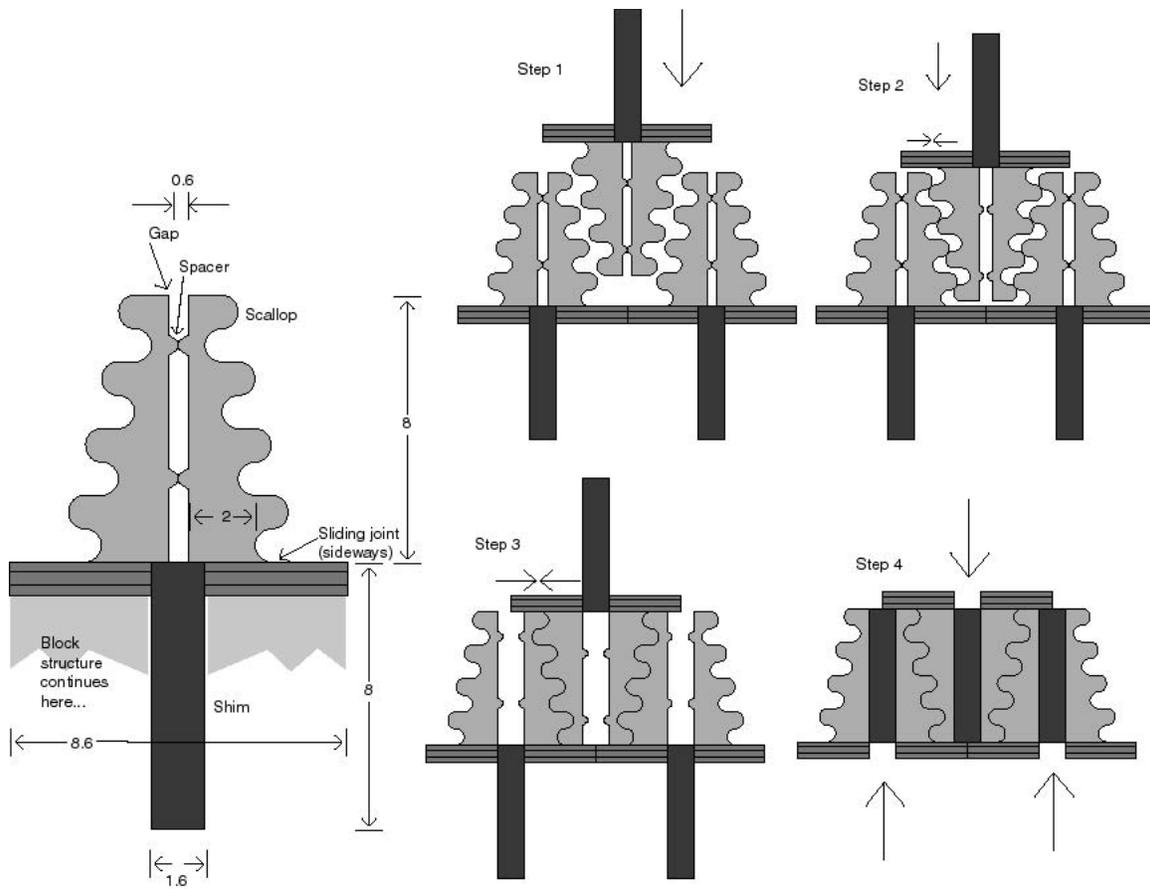


Fig. 6. Operation of a ridge joint. Ridges are pulled apart by surface forces without external actuation or manipulation.

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.²¹

Bearings

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.

²¹ 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>

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. 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. 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 systems may use such a small driving force that this is not the case; 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.

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/nm², 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/m³. Even at a 1 m/s rim speed, 10^{12} W/m³ 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 fact, its design is thermodynamically efficient (“reversible”, in the sense of reversible logic).

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.

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 nm², switch in 0.1 ns, and dissipate less than 10^{-21} J per operation. Computers could perform 10^{16} instructions per second per watt.²²

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 limits imposed by electron tunneling. Error rates can be extremely low, because an error would require a logic rod to slip past a physical obstruction.

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 be constrained to move slowly through the intermediate stages. 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/m³.

Advanced Nanofactory Architecture and Operation

Although any design for an advanced nanofactory must be tentative, the rough outlines of a reference design can be sketched out. The design presented here is derived from the Burch/Drexler planar assembly architecture²³ and the Phoenix primitive nanofactory. This section discusses transport of blocks within the factory and placement in the product, control of the factory including a possible data file organization, block

²² See Nanosystems Ch. 12.

²³ An illustration of this architecture is online at <http://lizardfire.com/nanofactorySS/index.htm>

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 blocks that would be attached to the growing product would be somewhat larger than in the primitive 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 could be 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.

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. Accordingly, it is reasonable to provide a large product deposition area even for a kilogram-per-hour factory.

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 can 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 1 mm thick.

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 for repetitive feedstock-placement tasks, 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.²⁴

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 hierarchical solid geometry descriptions of the product, which describe the block patterns in each 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 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

²⁴ 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.

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, 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 may be sent 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.

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, 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 as described in the "Primitive" section, 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.

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 would 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 resist coolant pressure. The interior of each fin will be hollow, providing workspace and space for transporting parts and blocks. The walls will be lined on both sides with fabrication systems. 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; added to the fluid channel volume, this means that the density of the fabrication volume will be about 0.1 g/cm³. 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.

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 adjacent to 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.

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; a flow of 1 liter/second can cool 100 kW—more than an advanced 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.

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.

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.

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

percentage 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. Damaged or incomplete blocks in a damaged fabrication area can be retained there permanently, since that area will not be used again. (Self-repair seems likely to add more mass and complexity than redundancy. If 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 radiation-sensitive machinery is correspondingly small. A square-meter area, 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.

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 limit performance less than it does in today's products due to more efficient machinery and more effective cooling.

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.²⁵ 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 the bond energy of spring's component atoms. Such a storage system could be charged and discharged quite rapidly, and store energy without leakage.

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 (as much as 1 GW/m³), it is safe to assume

²⁵ See *Nanomedicine* 6.4.3.4, available online at <http://nanomedicine.com/NMI/6.4.3.4.htm>

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,²⁶ 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 many 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.²⁷

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 diamond-like material cost \$100 per kg to produce, nanofactory-built products would generally be competitive with current products.

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, 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.

²⁶ 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. Minimizing the sum of these requires spending about 10% more energy on friction. See <http://www.cise.ufl.edu/research/revcomp/theory/Memo19-tradeoff/Memo19-RodLogic.doc> for Frank's finding.

²⁷ See *Nanosystems* 11.5.

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 microblocks that the product is assembled out of. A microblock will be a small fraction of the size of a human cell. This is a convenient size to contain a basic functional unit such as a motor or CPU. A library of such microblock 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.

Incentives and Applications

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 section 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.

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.)

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 energy would not be limited. (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 would 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. 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.

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, better understanding of causes, effects, and feedback loops in the body would 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. See Fig. 7. Even something as simple as widespread use of inexpensive greenhouses could save substantial amounts of water, topsoil, pesticides, fertilizer, and land area.

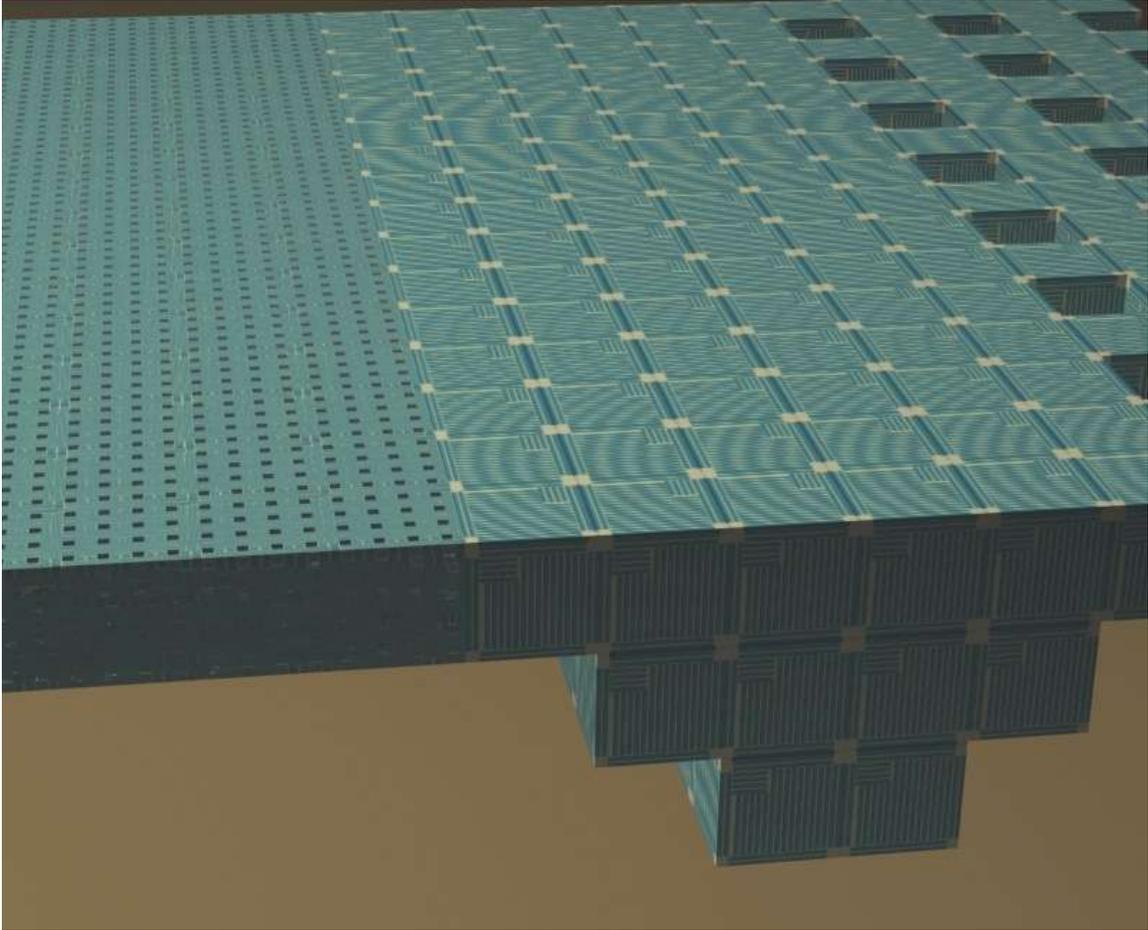


Fig. 7. A water filter with .3 nanometer pores (on left) would clean water down to the atomic level with minimal pressure drop due to drag. Such thin membranes would need to be supported by struts (center). Larger pore sizes are possible (on right). The very smooth surface on top would reduce fouling.

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.

Targeted Development of Molecular Manufacturing

The high performance promised by molecular manufacturing will induce some organizations to consider whether they should work to hasten its development. This section explores some of the issues involved in such a decision, and some of the desiderata for a targeted development program.

Existing vs. Required Capabilities

In theory, all that is needed for an advanced nanofactory-type molecular manufacturing system is the ability to build precise structures of conductors and insulators. The conductors can be used for switches, as well as electrostatic motors and solenoids, and that is sufficient to build digital logic and actuated machinery. Mechanical systems can perform whatever tactile sensing is required to avoid accretion of errors.

In practice, the nanoscale is a rich source of new phenomena. There is no practical or theoretical reason to exclude any phenomenon that is useful. In recent years, non-bleaching chromophores, mechanical programmable DNA-building systems made of DNA, several new kinds of molecular actuators, and many other helpful and relevant technologies have been developed by nanoscale technology researchers. This provides ample reason why a molecular manufacturing program will benefit from and further motivate existing nanoscale technology research.

Currently, most nanoscale technology research is not targeted at the development of nanoscale construction equipment. The synthesis of stiff covalent solids in arbitrary shapes is beyond the capabilities of today's unguided solution chemistry, so such structures have received little attention. However, they are a natural fit for building with nanoscale mechanosynthesis. Likewise, conventional mechanical engineering at the nanoscale would be difficult with the soft and flexible molecules that are available from many of the solution chemistry and biomolecular lines of research. With stiff covalent solids, nanoscale mechanical engineering requires substantial modifications from macro-scale practice, but is not completely alien.

The emphasis on mechanical rather than electronic functionality confers temporary limitations on molecular manufacturing, but also substantial advantages. Nanoscale mechanics depends on atoms, which are far heavier than electrons. This means that many quantum or unfamiliar phenomena can be factored out of the analysis of many machines. There is no need to use the full palette of nanoscale phenomena, because the basic task of molecular manufacturing is quite simple: to join molecules repetitively and reliably, in programmable location or sequence, and then to perform ordinary mechanical operations on the result.

In summary, a targeted molecular manufacturing development program would encourage basic research in a wide range of nanoscale technologies, while working to develop nanoscale substances, tools, and engineering disciplines that are not much emphasized

today. Although a development program may be accelerated by future nanoscale science breakthroughs, and would do well to cultivate them, development does not appear to be dependent on any additional fundamental breakthroughs.

Evaluating the need for targeted development

Several questions must be asked and answered in order to decide whether it is in an organization's best interest to expend resources on working toward molecular manufacturing in a targeted way. Several different kinds of factors will influence the decision.

The first question is the ultimate capability of molecular manufacturing and its products. A closely related question is how soon that capability might be developed, and how it will compare with competing technologies which will exist at that time. It is worth noting that development of even primitive general-purpose nanoscale manufacturing would probably facilitate rapid development of improved versions. In other words, development of a molecular manufacturing platform that can build products only slightly ahead of the competing technologies could result in the rapid development of more advanced capabilities. The development trajectory of a rapid-prototyped technology would likely advance more quickly than a competing trajectory which did not have the benefits of rapid prototyping. Thus, it is likely that whatever the physical limits of molecular manufactured products are, an accelerated molecular manufacturing development effort will have a good chance of accessing those capabilities significantly in advance of their expected schedule.

The cost/benefit tradeoff of funding or not funding development should be considered. This includes the benefit to the organization of owning the technology, which must be balanced by the chance that the organization will be unable to retain ownership of its work due to competition or regulation. Another factor that should not be overlooked is the cost to the organization if molecular manufacturing is developed elsewhere in a way that locks them out. With such a powerful technology, it is easy—perhaps too easy—to predict winner-takes-all dynamics. More subtle questions include the relative merits of owning the technology vs. encouraging its broader growth for the indirect benefits, and the break-even point(s) for money spent vs. acceleration gained.

Although there is insufficient information to make a projection of development cost, it is likely that the cost is decreasing rapidly. Computers for simulation are becoming exponentially more capable, along with core molecular biology technologies that may be useful in development. Molecular manufacturing theory is continually advancing, finding cheaper and faster development pathways. The number of organizations that could launch a rapid development program will increase as cost decreases.

After considering the benefits and costs of various development scenarios, an organization may conclude that it should promote, ignore, or perhaps retard the development of molecular manufacturing. Separate strategies may have to be considered

for the organization itself, its partners, and its competitors. Guidelines for implementing a strategy to promote development are discussed in the next section.

Guidelines for a targeted development program

A program targeted at rapidly developing advanced molecular manufacturing should choose goals extending several steps ahead. There are at least three possible pathways to develop an advanced system: direct scanning-probe manufacture of stiff machines; engineering of biomolecules to develop an improving series of biomimetic machines leading eventually to high-performance integrated systems; and fabrication of small molecular building blocks (not particularly biological) with which to build simple and improving nanomachine block-handling systems. Looking only one step ahead will provide no basis for choosing between these approaches. Selecting more ambitious goals and planning backward as well as forward will help in selecting an approach. (The other approaches could be pursued in the ongoing basic nanoscience effort, to allow their value to be developed in a less structured context.)

If a goal several steps ahead is known, then simulation may allow tentative designs to be developed in advance of laboratory capabilities. An important benefit of such activity is that, even if the simulation is not completely accurate, the experience will help to train designers for the time when the physical work catches up to the plans. To the extent that the simulations are accurate, development of workable designs can help to inform the next step, and a stockpile of designs verified in simulation can speed development once experiment catches up. Even if the simulations have to be adjusted to conform to experimental reality, the design stockpile will often provide a good starting point for revision rather than complete redesign.

An organizational culture should be developed that promotes creativity and exploration. Participants who do not see the value in pursuing approaches that are not guaranteed to work would hamper exploratory efforts. Many scientists are not comfortable building theories about unexplored areas of applied science. Of course, every idea needs to be critically evaluated before too much effort is spent on it. But critical evaluation should apply different standards in guiding exploratory work than in guiding major commitments to development strategies. Good ideas can be killed by premature demands for “proof,” and new ideas should have a chance to take root unless they are obviously and demonstrably flawed.

Along the same lines, acceptance of a high percentage of dead-end efforts will reduce development time, though there will be a tradeoff in cost. If only one idea at a time can be tried, then time and effort must be spent selecting the best one. If it turns out not to work, not only money but substantial time will have been lost; this will discourage the people involved from admitting that they should try a different idea. Conversely, if multiple ideas can be tried in parallel, then trying a less good idea costs considerably less time. An institutional structure that does not penalize abandoning a moribund approach

will help to minimize the amount of effort spent on attempts to justify previous unfruitful effort.

Modular design, in which the goal is broken down into sub-goals which can be solved in any of several ways, will make it easier to plan multiple parallel efforts to solve pieces of the problem. For example, the method of joining building blocks has some effect on the design of the actuators. However, a well-designed but broadly conceived specification for actuator characteristics will allow a range of compatible block and actuator technologies to be developed in parallel.

Finally, the leaders of the program must remember that nanoscale technology and molecular manufacturing theory will continue to develop rapidly, and they must incorporate new advances in their program, either by refocusing existing efforts when sufficient reason arises, or if resources allow, starting new parallel efforts and allowing old efforts to dwindle naturally if and when people are tempted away to better approaches.

Conclusion

Currently available theory and analysis indicate that molecular manufacturing will lead to the development of extremely high performance nanoscale machines, far exceeding anything available today in engineering or in biology according to simple measures of performance. These machines can be integrated into manufacturing systems of any desired scale, capable of processing their own mass in hours or minutes. Products built with these manufacturing systems would be extremely valuable.

The concepts presented in this paper suggest that practical development efforts could be launched using today's tools and theories. The cost of such efforts, and their effect on the speed of advancement of molecular manufacturing and its many spinoff technologies, is currently unknown. Because the timeline for development will be affected by the timing, resources, and number of targeted development programs, the time cannot be predicted either, although it is likely to be somewhat in advance of the schedule for independent development of capabilities equivalent to molecular manufacturing.

The author wishes to thank Eric Drexler, Tihamer Toth-Fejel, Jeffrey Soreff, and Robert Freitas for helpful comments, and Tihamer Toth-Fejel for the rendered illustrations.

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 analysis of a kilogram-scale integrated molecular manufacturing system. Although the “convergent assembly” design is superceded by the “planar assembly” approach described in the current paper, many concepts and calculations may be adapted for other nanofactory designs. It 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 have been verified by computational chemistry. A slide show and links to the animation are available at <http://lizardfire.com/nanofactorySS/index.htm>

Portions of this paper summarize work done by the author under a Phase 1 grant from the NASA Institute for Advanced Concepts. That work is linked from http://wise-nano.org/w/NIAC_Results

This paper makes many suggestions based on currently available tools and designs. Many of these suggestions will become rapidly obsolete as better tools become available. The latest version of the paper will be available at http://wise-nano.org/w/Doing_MM Note that the website allows interested readers to comment on and even edit the paper.