AI for Scientific Progress Workshop

Sept. 30-Oct. 2, 2016. Palo Alto, California

Nanotechnology & Artificial Intelligence

Concept map developed during the workshop, 2016 Oct 27
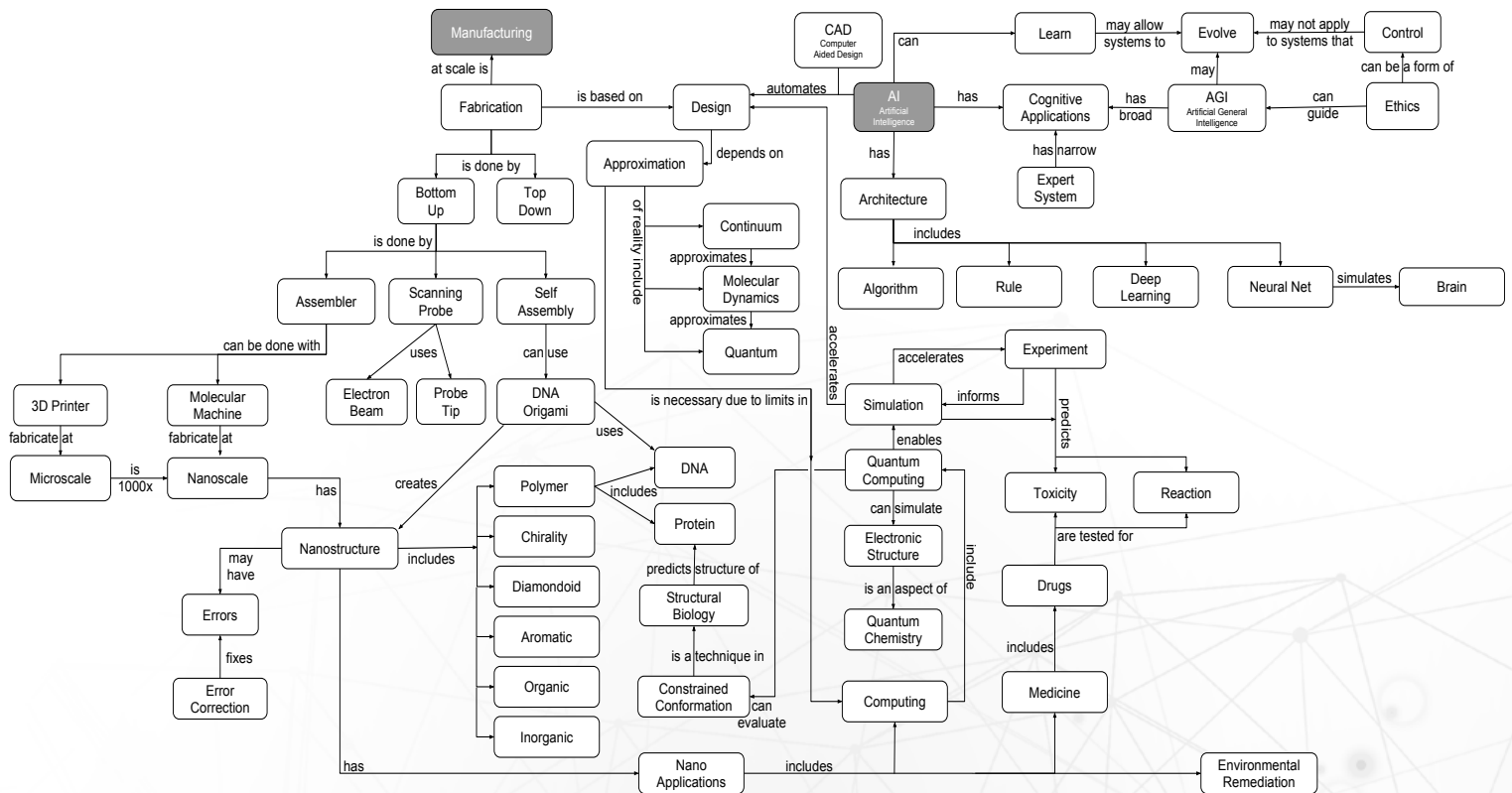
Miguel F. Aznar aznar@foresight.org

Table of Contents

About Foresight Institute

Foresight Institute steers emerging and world-shaping technologies for beneficial purposes and has done so for 30 years. It is our mission to spark innovation across multidisciplinary fields such as synthetic biology, artificial intelligence, and especially nanotechnology. We serve as a nexus for innovation to catalyze research, reward excellence, restrain recklessness, and create community aimed at the long-term flourishing of humanity and the biosphere.

WORKSHOP TEAM

Christine Peterson, Co-Founder
Julia Bossmann, President
Steve Burgess, COO
Allison Duettmann, MS
Maya Lockwood, Director of Engagement
Miguel Aznar, Director of Education
Marcia Seidler, Event Coordinator
Jim Lewis, Technical Editor

foresight.org - foresight@foresight.org

FORESIGHT
INSTITUTE

Participants

Dion Whitehead Amago, Autodesk Research, Bio/Nano/Programmable Matter

Lauren Barghout, Burning Eyedeas, University of California. Berkeley Institute for Soft Computing

Jeremy Barton, Nanofactory Corporation

Angela Morente Cheng, ExO Works

Paul Cherukuri, Institute of Biosciences and Bioengineering, Rice University.

John Denning, Universal Medical Services & Health Management Co.

Martin Edelstein, Covalent / AguaVia

Yad Faeq, 4th n Town Machine Learning R&D

Rob Freitas, Co-Founder/COO of Nanofactory Corporation

Steve Fowkes, Nanopolymer Systems

Franz J. Giessibl, Universität Regensburg, Institute of Expt. and Applied Physics

Don Gilmore, Software Architect

James Gimzewski, University of California, Department of Chemistry and Biochemistry

Dan Girshovich, Zee.Aero, The Embassy Network

William A Goddard, III, Caltech, Department. of Chemistry, Materials Science, and Applied Physics

Ben Goertzel, Hanson Robotics, Aidyia Limited, AGI Society and OpenCog Foundation

Bartosz A. Grzybowski, Ulsan National Institute of Science and Technology, S. Korea

Si-Ping Han, Caltech, Department. of Chemistry

Cosmo Harrigan, Channel Agility

Sergei V. Kalinin, Oak Ridge National Lab, Center for Nanophase Materials Sciences

Kent Kemmish, Stealth Biotech

Rick Lewis, GlaxoSmithKlein

James Lewis, Foresight Institute

Ruiting Lian, Computational Linguistics, Xiamen University, China

Moshe Looks, Google, Machine Intelligence

Joseph W. Lyding, UIUC, Department of Electrical and Computer Engineering

Ralph Merkle, Co-founder/President of Nanofactory Corporation

Neal McBurnett, Data Science

Farhan Malik, Architecture of Things

Nicola Marzari, École Polytechnique Fédérale de Lausanne; MIT

Conrad Pfeiffer, Temple University, Department of Chemistry

Bogdana Rakova, Samsung Research America

Patrick Riley, Google, Machine Intelligence

Natalia Díaz Rodríguez, Data Science

Bruce Smith, Software Developer

Mikhail Soutchanski, Ryerson University, Dept. of Computer Science

Reto Stamm, Two Pore Guys, Inc.

Deepak Srivastava, Taechyon Robotics Inc.

Sara Szymkuc, DARPA Make-It program at GSI, Ltd

E. Dean Tribble, Synchrona, LLC

Aaron Virshup, Autodesk, Bio/Nano Group Scientist

Anatole von Lilienfeld, University of Basel, Department of Chemistry

Peter Voss, AGI Innovations Inc.

Zhiyong Zhang, Stanford University, Department of Computational Science

Gayle Pergamit, AguaVia

WORKSHOP SUPPORT

Jazear Brooks, Technical Product Manager, Gigster

Roby Behrens, Lucid Productions

Justin Kelsey, Lucid Productions

Lori Luttrell, Molecular Biology

Richard Nesbitt, Bio Q Pharma

Chris Sapyta, Menlo School

Max Sims, President Consulting Group

FORESIGHT INSTITUTE

Foreword

At Foresight Institute, we believe that many of the most promising breakthroughs in science happen at the intersection of different fields, both in the long-run and in the short-term.

In the long-run artificial intelligence and molecular nanotechnology are poised to shape our world like no technology that has come before. In narrow ways, artificial intelligence has already achieved superhuman capabilities. Machines beat humans at computing prime numbers, charting itineraries, and steering vehicles, for example. While all of these achievements are impressive in isolation, the race in artificial general intelligence is less about breaking more and more records with carefully coded algorithms, and more about creating systems that can generalize to more tasks and to challenges that require qualities which we consider innately "human", such as creativity and intuition. The victory of the AlphaGo program against Lee Sedol, the best human in the game of Go, for instance, was considered significant precisely because Go, when played at the highest professional level, requires a deep intuition for the game as well as creativity. The level of complexity and the number of possible paths is so high that it cannot be solved algorithmically in the traditional sense. AI technologies such as Deep Learning, together with recent developments in computing hardware, offer a way to tackle this complexity. Ultimately, the arrival of artificial general intelligence, while holding promises and perils, is likely to be at least as revolutionary

in its capacity to shape its environment as was previously the dawn of humanity. Even likelier, it will exceed our world-shaping capacity by orders of magnitude.

The long-term prospects of nanotechnology are on a similarly impactful scale: In 1986 Foresight was founded on a vision of the emerging field of nanotechnology in which current capabilities in nanotechnology lead eventually to fabrication of complex products with atom-by-atom control of the manufacturing process. This ultimate development of nanotechnology, sometimes termed molecular manufacturing and now often termed APM (atomically precise manufacturing), was first described by Richard Feynman in 1959 in his visionary talk "There's Plenty of Room at the Bottom." The ability to manufacture with atomic precision promises a revolution in manufacturing, leading to a world of abundance. Medical nanodevices that cure diseases, can improve health and longevity, nano-enabled photovoltaics that allow for abundant solar energy can heal the environment, and new materials a factor of 100 times stronger than steel can enable space exploration.

For an overview of AI, nanotechnology, and possible roads to artificial general intelligence and molecular nanotechnology, please see the background reading at the end of the white paper.

On a less speculative and more immediate note, the workshop on AI for Scientific Progress arose from a need that became apparent during our workshop series on nanotechnology. Many of the bottlenecks that participants repeatedly stressed during previous technical workshops appeared to be solvable with the help of AI. The design of molecular nanotechnology is a challenge with an unfathomably high level of complexity. Just like a generic computer cannot go through all possible steps in Go in order to beat the game, it cannot model all possible combinations of atoms to form useful molecules. This is where emergent AI technologies give us hope. Qua their computing power and fine-tuned algorithms, AI-tools are uniquely equipped to learn from the abundance of information at the nanoscale, in ways humans are unable to do. Conversely, nanotechnology research can lead to progress in hardware that can tremendously advance computing power and boost progress in AI - accelerating the mutually beneficial feedback loop between the two technologies. The value of this meta-workshop, AI for Scientific Progress, was to propose and discuss the most promising routes for immediate collaboration between AI and nanotechnology experts. From the 60+ possible proposals discussed at this workshop, the most promising ideas were developed into technical research proposals, which comprise the main corpus of this whitepaper. These proposals are deeply interdisciplinary, ranging from machine-learning based simulation of nano-processes to data standards of collaborative AI and Nanotechnology research. The projects have the

potential to advance the state of research in atomic precision and to lead to a virtuous feedback loop that simultaneously advances research in AI significantly. Some of these proposals are now open for collaboration and funding.

The AI tools discussed apply not only to nanotechnology, but can also advance scientific research on a more general level. In a time when borders between different scientific disciplines are broken down, these projects are important signposts to new mutually beneficial research avenues. Opening up and embarking on these avenues requires action. If you want to see a bright future realized, please consider supporting us so we can scale up our efforts. Consider giving to Foresight, get in touch to volunteer, advocate for the beneficial development of technologies, and spread the word about our work.

Here are testimonials from workshop participants:

- "These were some of the most stimulating days of my career."
- "I'm delighted by how many fascinating conversations I was able to have. I was glad to just be there, to meet great people and soak up all the astonishing information concerning what's going on in the current world of AI-Nano synergy."
- A meta-workshop was needed to develop artificial intelligence-assisted tools that could break through the bottlenecks faced by scientific researchers and have the potential to significantly advance scientific research on a more general level

We at Foresight thank all participants for their open-mindedness and ingenuity. A special thanks goes to our volunteers and Lucid Productions who make these workshops into what they are.

Julia Bossmann - Foresight Institute

Allison Duettmann - Foresight Institute

Artificial Intelligence for Scientific Progress

Bringing Digital Control to Physical Matter.

Report Authors:
Jim Lewis, PhD
Allison Duettmann, MS

Foresight Institute,
foresight@foresight.org

FORESIGHT INSTITUTE

Rationale for the workshop

05.

The long-run:
Synergistic world-changing technologies

Foresight Institute's mission is to support the development of world-shaping technologies that lead toward an abundant future. Central to this vision, in the 30 years since its inception, is the concept of the ultimate manufacturing technology, able to inexpensively produce essentially any configuration of atoms compatible with the laws of physics, thus bringing digital control to physical matter. The goal is to create a technological revolution even more important than the revolution that brought digital control to the world of information over the past century. We live in a world composed of atoms; this new technology has the potential to meet all of our material needs inexpensively, rearrange atoms in cells to cure disease and heal injuries, and maintain youthful vigor indefinitely.

This concept has its intellectual origin in a proposal by physicist Richard Feynman in 1959, and was originally named nanotechnology by Foresight's co-founder in 1986. The name has been modified several times to distinguish this advanced, still unrealized technology from other, less precise technological manipulation of matter at the nanometer scale. The key idea is atomic precision, which means specifying the

FORESIGHT
INSTITUTE

> *"Foresight Institute's founding vision includes a mutually beneficial synergy between atomically precise nanotechnology and artificial intelligence."*

number, position, and bonding for each atom in a device or system to achieve the designed function of the device or system. The ultimate manufacturing technology is thus general purpose, high-throughput, Atomically Precise Manufacturing (APM).

Foresight's founding vision anticipates atomically precise manufacturing developing concurrently with advances in other technologies. In particular, the field of Artificial Intelligence (AI), which also had its origins in the 1950s, is anticipated to evolve to enable automated scientific inquiry and automated engineering design. To benefit from the ability to arrange atoms as desired, one needs to know what arrangements will achieve a desired purpose. Conversely, atomically precise manufacturing is anticipated to greatly accelerate the development of more powerful computers to provide the infrastructure for more advanced AI, i.e. Artificial General Intelligence (AGI). Thus, the speed at which either one of these technologies advances can substantially impact the speed at which the other technology develops. This interplay is nicely reviewed by Sacha and Varona in 2013 (1).

The short-term:
Overcoming existing roadblocks in research

> *"Workshops are based on a fast-paced 'Design Shop' group achievement brainstorming process."*

In September of 2014, Foresight began a series of small, fast-paced, interactive, 2.5-day workshops focused on advancing revolutionary technologies by gathering a carefully curated group of wide-ranging thinkers from various disciplines. The goals of these workshops are to assist various researchers in overcoming roadblocks to technology development, facilitate collegiality among researchers from diverse backgrounds while promoting an interdisciplinary approach to identifying and solving research problems, and initiating novel, fundable, collaborative projects. Results are achieved through a fast-paced process based on 'Design Shop' principles that helps participants collectively let down their guard, push the envelope of the possible, and generate lots of ideas quickly to seed entirely new research directions.

This series of technical workshops began with the "2014 Workshop on Directed/Programmable Matter for Energy", followed by the "2015 Workshop on Atomic Precision for Medical Applications", and, most recently, "Breakthrough Technologies for Energy" in May 2016. Over the course of the workshop series, it became clear that the main bottlenecks that repeatedly held participants back from advancing their research could be solvable with the assistance of AI:

"Bottlenecks faced at Foresight workshops that seem treatable via AI: design and simulation, big and small data-analysis, data standards for interdisciplinary research."

- At the Breakthrough Technologies for Energy Workshop, researchers called for the development of software to produce a programming environment that helps design molecular machines, analogous to existing CAD software (see Whitepaper, p. 37).

- At the Atomic Precision for Medicine Workshop, researchers acknowledged that the development of fine-tuned sensors measuring many health values in patients would pose a significant challenge for data analysis and called for software for evaluating large data sets (see Whitepaper, p. 11).

Another opportunity for AI became clear from the interdisciplinary nature of innovation more generally: Foresight believes that innovation is sparked at the intersection of existing fields. However, a significant obstacle to progress with interdisciplinary projects such as the ones developed at the workshops is the difficulty of sharing information that is relevant for the researchers across disciplines in a format and language that is readily accessible and understandable. Much time at workshops is lost in translation of terminology and jargon specific to certain disciplines, but unfamiliar to others.

Workshop Results

"Formerly invitation-only, this was Foresight's first workshop to allow applications from hackers, PhD's, self-taught students, and others."

Forty-four researchers were invited to Palo Alto from Sept. 30 to Oct. 2, 2016 to apply AI, narrowly defined as advanced computer science, to progress nanotechnology and the scientific field generally. Historically, Foresight's workshops were invitation-only. This AI workshop was the first to open attendance to individuals who might have fallen under our radar, such as hackers, PhD's, students, and self-taught individuals with an interest in AI, presenting them the opportunity to bring their diverse skillsets to the challenge. The AI expertise represented included knowledge representation, expert systems, deep neural networks, machine learning (especially deep learning), natural language processing, and artificial general intelligence (AGI). The nanotechnology expertise represented included materials science, computational chemistry, organic synthesis, surface physics, biochemistry, molecular biology, and engineering.

On the nano-side, the workshop focused on the use of AI to speed pathways to the construction of atomically precise 3D objects. These include objects and devices made of DNA, RNA, or proteins, constructed using organic or inorganic synthesis, or using scanning probes to build from the bottom up. AI can speed the design and fabrication of increasingly complex configurations of atoms. Machine learning allows meaning extraction from immense datasets, which is critical as various types of 3D atomically precise sensors require real-time analysis of rapidly increasing amounts

of data and as images of biological structures become more detailed.

From the pool of 60+ research ideas at the intersection of AI and nanotechnology that were generated at the workshop, the most feasible and desirable eight resulted in the creation of technical project proposals. This section gives an overview of the proposals and links to the presentation videos. Several of the projects are now being pursued for further research and are open to funding and collaboration. Please see the participant list for the credentials and affiliations of participants, or contact Foresight Institute at foresight@foresight.org.

Project 1:
Accurate and scalable machine learning based atomistic simulation for nano structures and processes

Team members: Jeremy Barton, Ben Goertzel, Sergei Kalinin, Patrick Riley, Kent Kemmish, Anatole Von Lilienfeld, Jeremy Barton, Zhiyong Zhang and others

PROPOSAL

We will train machine learning models using databases of quantum chemistry [1], as well as experimental results. This approach will deliver atomistic potentials with full quantum mechanical accuracy, while boosting the efficiency of simulation times by 6 orders of magnitude. To extend the databases, we will run high-level quantum calculations on millions of molecules [2] to serve as training data for replacing the approximations of cheaper calculations with machine learned models. Further, we will integrate experimental data, for example, from high resolution imaging and spectroscopic tools (ultrafast STEM and X-ray), and develop inferential tools to incorporate these into our theoretical framework. Our easy-to-use integrated platform will be accessible to simulation experts, but more importantly, to experimentalists who can use the platform to generate and validate ideas for experimental processes designs and validation. We will consider immediate near-term applications of improved calculations. Particularly, in the field of molecular drug design, state-of-the art methods fail to yield predictive results [3]. This implies urgent need for methods which can be used to identify useful drugs. A possible target exemplary drug design challenge is to find small-to-medium sized organocatalysts for the specific and efficient remediation of extracellular matrix covalent cross links, and more specifically, the most abundant and problematic advanced glycation end product in human biology, glucosepane [4].

PRESENTATION BY ANATOLE VON LILIENFELD

https://youtu.be/3MNsHVnTD8I

Transcript

We want to develop machine learning models that are so efficient and so accurate that we can use them to predictively screen the behavior of matter at the atomistic level. These machine learning models can be trained; we showed that that works using using quantum chemistry results just as well as experimental results. So we formed a team where we can do experiments and quantum chemistry simulations to very high accuracy and precision, and we also have machine learning experts to train these models.

This is a really crucial step towards this overall goal - to gain digital control of matter. As seen on the Chematica presentation by Grzybowski (who won the 2016 Foresight Feynman Prize) yesterday, the number of possibilities is really combinatorial. If your computer program is incapable of reliably screening these options, then you are in bad shape if you afterwards want to realize these materials in the real world.

One of the possible applications that we find very attractive is to use these tools to screen large spaces of molecules, and to find interesting molecules. One target is shown [in ref. 4 below]. This is important for aging; actually, you all suffer from this. These are two amino acids that occur naturally in proteins. Imagine my arms are the backbones of two different proteins [holding up his arms]. Lysine and arginine condense together with glucose forming something called an advanced glycation end product. This molecule covalently bonds two different proteins in the extracellular matrix. It is one of the seven well-known factors that contribute to aging.

This is a covalent bond, and we would like to break it. To do so, you need a molecule that will bind to this AGE selectively. There are 10^60 molecules, by conservative estimate, that could be candidates to break this bond. Thus, we need rapid methods to screen and score such molecules. Without such methods, there is no hope of finding an active molecule by chance.

We have a fantastic team with the needed specialties and we have the required datasets. Once we have these tools, we can use them in optimization algorithms to explore this space efficiently; of course, we could always scale this effort up with more funding. We want to combine experiments and quantum simulations so that possibly we could extend known laws of physics if we discover regularities that are unaccounted for by conventional quantum mechanical approximations.

Please see the appendix for a volunteer contribution to this project, "A hypothetical AI-backed CAD for nanotechnology" by Jazear Brooks.

References:

[1] "Fast and accurate modeling of molecular atomization energies with machine learning", M. Rupp, A. Tkatchenko, K.-R. Mueller, O. Anatole von Lilienfeld. Phys. Rev. Lett. 108 058301 (2012); http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.108.058301. "Quantum chemistry structures and properties of 134 kilo molecules", R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, Scientific Data 1 140022 (2014); http://www.nature.com/articles/sdata201422.

[2] "High-level multireference methods in the quantum-chemistry program system COLUMBUS: Analytic MR-CISD and MR-AQCC gradients and MR-AQCC-LRT for excited states, GUGA spin–orbit CI and parallel CI density" H Lischka, R Shepard, R M Pitzer, I Shavitt, M Dallos, T Müller, P G Szalay, M Seth, G S Kedziora, S Yabushita, Z Zhang, Physical Chemistry Chemical Physics 3, 664-673 (2001); http://pubs.rsc.org/en/Content/ArticleLanding/2001/CP/b008063m#!divAbstract.

[3] "Assessment of Programs for Ligand Binding Affinity Prediction", RYANGGUK KIM and JEFFREY SKOLNICK J Comput Chem. 2008 Jun; 29(8): 1316–1331. doi: 10.1002/jcc.20893; http://cssb.biology.gatech.edu/skolnick/publications/pdffiles/277.pdf. "Can we trust docking results? Evaluation of seven commonly used programs on PDB-bind database". D Plewczynski D, M Łaźniewski, R Augustyniak, K Ginalski. J Comput Chem. 2011 Mar;32(4):742-55. doi: 10.1002/jcc.21643; http://onlinelibrary.wiley.com/doi/10.1002/jcc.21643/full. "Small Molecule Hydration Free Energies in Explicit Sol-

vent: An Extensive Test of Fixed-Charge Atomistic Simulations" D. L Mobley, C. I. Bayly, M. D. Cooper, M. R. Shirts, and K. A. Dill, J. Chem. Theory Comput., 2009, 5 (2), pp 350–358, DOI: 10.1021/ct800409d; http://pubs.acs.org/doi/abs/10.1021/ct800409d.
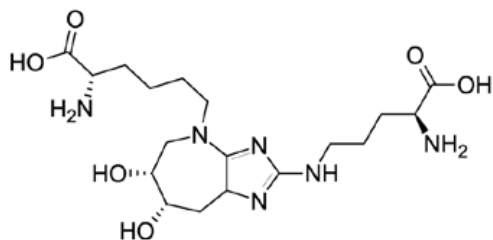
[4] Picture of glucosepane: https://en.wikipedia.org/wiki/Glucosepane

"Quantum Mechanical Design of Enzyme Active Sites". X Zhang, J DeChancie, H Gunaydin, A. B. Chowdry, F. R. Clemente, A. J. T. Smith, T. M. Handel, and K. N. Houk, J. Org. Chemistry 73 (3), pp 889–899 (2008) DOI: 10.1021/jo701974n; http://pubs.acs.org/doi/abs/10.1021/jo701974n

Resources available

Computing: Google cloud, Swiss Supercomputing Center, Stanford Research Computing center, XSED (NSF), Scicore (University of Basel).

Software: AIDA, Games, NWChem, QuantumEspresso, R, WEKA, RDKIT, tensorflow

People needed: Machine learning experts, systems experts, quantum chemists/physicists, experimentalists, robotics

Project 2:
Atomic Precision Metal Clusters

Team members: Steven Fowkes, and others

## PRESENTATION BY STEVEN FOWKES

https://youtu.be/DYY_Eq3NTAw

PROPOSAL

To study and exploit metal catalyst capabilities that are enabled by atomic precision. This would provide new catalysts for relevant organic reaction,s as well as the basic research of metal-metal bonding and redox gradients. The vast number of system parameters will make use of narrow AI to build a profile to develop catalysts for real world applications.

Transcript

This morning, I sat down with Conrad and we started talking about the possibility of blending two different systems of nanostructural self-assembly: Schafmeister's spiral ladder polymers and Nanopolymer System's aromatic polymers. Immediately, we started thinking about a specific example and decided to turn it into a presentation, so what you are seeing here is science and business in 90 minutes.

The idea is the production of catalysts, and Chris Schafmeister has been interested in this for a few years. The idea is to use spiro oligomers to create triangular

CATALYTIC UNIT

SUPRAMOLECULAR SUBSTRATE
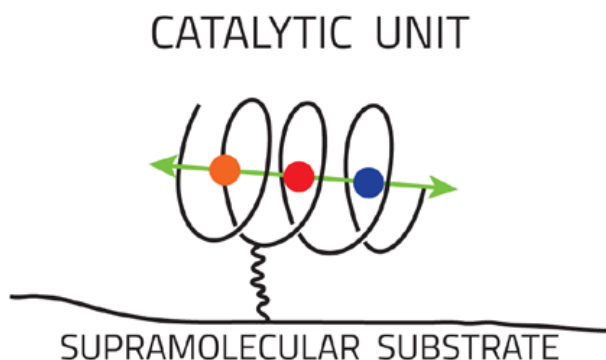
technology could create nanocoils, represented here in black, that would have the ideal bonding interior to coordinate metal atoms and metal ions, and this would provide a way of creating a catalytic unit that could fit into the middle of this cyclic superstructure. So the idea is that the polymer provides the outer structure to prevent the metal atoms that are bonded to each other from undergoing redox reactions and precipitating each other so that you can then have catalytic activity involving direct electron flow from metal atom to metal atom without the need for tunneling.

There are all kinds of great, new theoretical advantages that we can talk about.

- Catalysts for new organic chemistry reactions of the sort that were discussed yesterday with respect to Chematica.

- By attaching photon absorbing groups to the outside of the aromatic polymer, we can input energy into the system to facilitate catalysis. We can have different functional groups at different places on the ring to input photon energies independently into each metal atom, so we can have each one tuned to a different frequency (thank you, Jeremy, for that contribution to our program).

- We can do metallic and subvalent ions so that we can control the redox potential of the construct.

- We can study basic science of metal-metal bonding in a linear system, which has been very difficult to do. We can study the profile of complex redox gradients, and whether or not we can create gating effects for electrons.

- Proof of concept for de novo enzyme design, looking at this as an expert system that will allow us to learn how all of these systems work, and put them together to accomplish specific, predetermined catalytic goals.

We can also take this from linear into to other kinds of geometries in the future, once we have explored the value of this idea.

This is a narrow AI for the system parameters of the catalyst. We will want to pay attention to the metal sequence of the elements, the redox state of each one, the photon inputs, how long are the clusters, the structure of the redox gradient (do we go up and then down, or do we go up in steps), what kinds of catalytic properties can be obtained.

We are projecting a budget of $1 million for phase 1, which is just the initial building of the first prototype. We think it will take five people to do this. Once that is done and we move to phase two, where we do a survey of all the first row transition elements, that is going to involve an AI team, and neither of us is qualified to predict what that is going to entail. Future work would involve collaborations with organic chemists, with industry. Phase 3 would be building super-catalysts for specific applications, like fixing nitrogen or CO2 conversion.

Products would include enzymes, catalysts, energy capture systems (which might lead to an electron-gating element), and catalytic flow reactors having sequences of enzymes arranged in a line.

An obvious source of funding would be government grants, then industry contracts, and then "private sector" to encompass everything else. We are not too proud to take money from anybody. The features that will promote this project include better throughput for enzymes, more economical, more programmable, tunable and adaptable.

References:

"Maximizing the Stereochemical Diversity of Spiro-Ladder Oligomers" Christopher G. Levins, Zachary Z. Brown, and Christian E. Schafmeister. Org. Lett., 8, 2807–2810 (May 23, 2006)
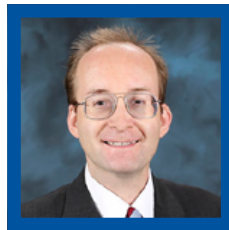DOI: 10.1021/ol060902q   http://pubs.acs.org/doi/abs/10.1021/ol060902q

Nanopolymer Systems: Nanostructural self-assembly: the key to unlocking the power of nanotechnology   http://www.nanopolymersystems.com/home

Project 3:
Atomic Forge

Team member: Sergei Kalinin

PRESENTATION BY SERGEI KALININ

PROPOSAL

Assembling matter atom by atom, precisely and controllably, is the ultimate goal of nanotechnology. The reigning paradigms to enable this goal are scanning probe microscopy (SPM) and synthesis. SPM assembly dates back to seminal experiments by Don Eigler, who demonstrated single atom manipulation and writing. Yet stability and throughput remain issues, and only in the last decade synergy of STM and surface chemistry was used to make several-qubit devices. The molecular machines approach harnesses the power of modeling and synthetic chemistry to build individual functional blocks, yet strategies for structural assembly remain uncertain.

We propose a third paradigm — the use of atomically focused beam of scanning transmission electron microscope to control and direct matter on atomic scales. Traditionally, scanning transmission electron microscopies (STEMs) are perceived only as imaging tools, and any beam induced modifications are undesirable beam damage. Yet in the last five years, our team and several groups worldwide demonstrated that beam induced modifications can be more precise. We have demonstrated ordering of oxygen vacancies, single defect formation in 2D materials, and beam induced migration of single interstitials in diamond like lattices. What is remarkable is that these changes often involve one atom or a small group of atoms, and can be

monitored real time with atomic resolution. This fulfills two out of three requirements for atomic fabrication.

In this project, we seek to transition from effect and observation to control. We seek to harness these beam induced high energy phenomena to actively control matter. As a proof, we have implemented beam control in STEM to create single digit nanometer structures that can be formed and imaged with atomic resolution. We believe that this approach can be extended to create 3D structures in the bulk with atomic precision.

Note that AI will be a central element of this development. First of all, STEM already use machine learning to form the beam (64 tunable elements of aberration correction). In order to implement the Atomic Forge, we will need to:

- create the library of structures and beam induced transformations (much like cause and probability of effect look up table),

- create rapid image analytic tools to identify observed atomic structures from ptychography and local imaging,

- rapid decision making (given observed structures and known cause-effect relationship, what do we need to do to make atoms move where we want).

Speaking of feasibility:

- There is a large fleet of STEM platforms that can be repurposed for these applications (so scalability is taken care of).

- What we need in year 1 is a rapidly built up knowledge base. The proposed cost for contracts, postdoctoral effort, and theory development is approximately $1 million in year one.

- The expected deliverable is to make an atom move in selected lattice site in the bulk/form sub 5 nm crystalline line/create given sparse vacancy

- Year 2-3 will require rapid cost ramp up to demonstrate assembly and establish theory base.

Ask me how, why our team, why now.

Transcript

This morning, I want to give you the atomic forge, so let me tell you what it is. Most of this meeting basically talked about how we can make and direct matter on the atomic level. There are two paradigms to do that. One is just to start from scratch. Let's pick atoms on the surface; let's manipulate them. This is how nanotechnology started. This is what we have been able to achieve in 25 years. It works. It costs a lot of money; it requires integration between STM and surface chemistry, but we can make multiple-qubit devices now. The problem is that it is surfaces, it is slow, and it is a limited range of materials.

The second paradigm, which has attracted attention since Eric Drexler's book, is the molecular machine. We use chemistry, we synthesize them, we hope that at some point they will assemble, much like the Terminator. The truth is, however, we probably have to somehow manipulate them. So what we do is assemble them bottom-up, but we still need a way to assemble these machines into functional contraptions. So what I want to bring to your attention -- what we think is the third way of doing it.

Many things are started by a random observation. Several years ago, it was noted that if you stick material covered by an amorphous version of the material into the electron microscope, the electron beam can induce the crystallization of the material. While it does it, we can actually observe it with atomic resolution. Notice, when we talk about atomic fabrication we come up with the requirement that at the atomic level there should be feedback that should be correctable. In this case, unlike molecular machines, and much easier than with STM, we get atomic resolution for free. That is what electron microscopes do.

Three years ago, we used an atomic force microscope (AFM) to draw things on the surface; we wrote the logo of the Department of Energy, and also the logo of (the German band) Rammstein, and my colleagues can still not forgive me for this because they think it is not music. What we can do is take the AFM controller electronics, and connect them to the electron microscope, which is a really expensive piece of equipment, which is really not designed for being controlled. But we have done it, and it works.

This is an example of the structure, which is written in the amorphous material in letters that are basically ten atomic forms around. If you have done it once, it makes sense to do it twice, so we incorporated helium atom and liquid sources so we can draw from gases and liquid phase.

The question is, what can be done in the future? We have observed -- both our group and our colleagues worldwide -- that actually an electron beam can produce a tremendous amount of controlled modification of a surface. Three years ago, we were able to burn holes in two-dimensional materials and observe how the materials reconstruct into a different crystalline phase. Also two years ago, we were able to observe how a single interstitial atom is basically kicked by the electron beam inside the lattice so we can move it where it wants to go. We were able to observe how the oxygen vacancies inside the solid order to form the grain boundary planes.

The question that I bring to your attention is, can we take the electron beam control and do the same thing that Don Eigler did 25 years ago, but do the same thing in the bulk, not on the surface? This is the whole idea of the atomic forge -- use the electron microscope not as an observational control, but as a tool to control matter. I've shown you that we can do this with small modifications. Now we just need to learn to control it. This is where I believe that AI is absolutely essential, for very simple reasons. First of all, when you form the subatomic beam, you use machine learning in order to use aberration correction. The aberration corrected microscope appeared ten years ago because computers allowed to tune the 64 elements mathematically, and not do it by hand. That is the only reason why it was not done 30 years ago. It generates large volumes of data, in some cases as much data as the large hadron supercollider. This data controls the information of the solid.

An important question is, if this one machine is sustainable, can we scale it up? There are hundreds of thousands of STEM platforms over the world that are not really used, so if it works, we can scale up really fast.

The AI needs here are really straightforward. First of all, we need to do the libraries of geometric structures of atomic level radiation damage. So we need to build the libraries of cause and effects for specific structures. Ideally, we want to do it on the fly. We need to have the AI controls of position and intensity, and what we will get is atomic fabrication.

References:

"Flexible metallic nanowires with self-adaptive contacts to semiconducting transition-metal dichalcogenide monolayers" J Lin et al. Nature Nanotechnology 9, 436–442 (28 April 2014) doi:10.1038/nnano.2014.81.
http://www.nature.com/nnano/journal/v9/n6/full/nnano.2014.81.html

FORESIGHT
INSTITUTE

"Atomic-Level Sculpting of Crystalline Oxides: Toward Bulk Nanofabrication with Single Atomic Plane Precision" S Jesse et al. Small 11, 5895–5900 (19 October 2015) DOI: 10.1002/smll.201502048
http://onlinelibrary.wiley.com/doi/10.1002/smll.201502048/abstract

"Paving the way to nanoionics: atomic origin of barriers for ionic transport through interfaces" MA Frechero et al. Scientific Reports 5, 17229 (17 December 2015) doi:10.1038/srep17229.:

"Directing Matter: Towards Atomic Scale 3D Nanofabrication" S Jesse et al. ACS Nano, 10, 5600–5618 (May 16, 2016).   DOI: 10.1021/acsnano.6b02489

"Dynamic scan control in STEM: spiral scans" Sang, X., Lupini, A.R., Unocic, R.R. et al. Advanced Structural and Chemical Imaging 2: 6 pp 1-8 (13 June 2016) DOI: 10.1186/s40679-016-0020-3

Project 4:

Data standards for Collaborative Nanoscience and AI

Team members: Aaron Virshup, Neal McBurnett, Bruce Smith, Jim Lewis, Lauren Barghout, Ruiting Lian, and others

PRESENTATION BY AARON VIRSHUP

https://youtu.be/BFeRxWxlIz0

PROPOSAL

To collaborate on nanoscience, humans and artificial intelligences need a common language. We will develop a suite of data formats that can relate molecular representations to experimental data, physical properties, synthesis pathways, human cell assay data, and new data types yet to be dreamt of.

We are already working with practitioners at the University of California San Diego (UCSD) Memorial Sloan Kettering Cancer Center (MSKCC), Merck, and Autodesk, and using lessons learned from legacy formats such as PDB, mmCIF, and CML.

This project builds the digital infrastructure that allows experts pursuing nanotech, via multiple paths, to bring digital control to physical matter.

Edit: This project to facilitate the sharing of nanotechnology-related data among nanotechnology researchers, machine learning experts and other interested parties was placed on Github on October 27, 2016:

https://github.com/alchemistry/fileformat/blob/master/ai-nanotech-foresight-project.md

Transcript

I think one of the most interesting things we've seen over the past few days is that everyone here speaks slightly different languages. What this project proposes to do is to figure out how we are all going to communicate, both as researchers and with our software.

We can start thinking about all of the different regimes that really go into nano-science — things like small molecules, like proteins, like nanoparticles and surfaces. We have a lot of researchers here who focus on slightly different parts of each one of these problems, and oftentimes have overlaps between these different problems. As we start to look out at the implications of their work, the people they are talking to, and the research they are connecting to in the broader world, we can start looking several overlaps through to much bigger concepts like DNA origami, or how cells work, etc.

The question is, how do we have file formats that let us easily look at all of these systems in the level of representation that a given researcher actually wants to look at, but without losing all the data that the people looking at the smallest scales want to look at, too?What we are really here to do is figure out, how do we make this all accessible to people doing AI and learning, and how do we make it readable by both the AI software and the practitioners?

Our proposal is to build a common language for computational nanoscience. This would involve an ecosystem of well-defined, unambiguous file formats, one that is readable by scientists from different domains, both by themselves and with their software packages, and most importantly, one that is readable by AI software and can be understood and parsed by AI researchers.

I am going to zoom in quickly and look at one of the overlaps that comes up in this field a lot, which is the overlap between small molecules and biological proteins. The problem here is that the data formats we have now are not extensible, are unlinkable, most of them come from ancient Fortran codes, and many of them were defined by a graduate student a number of years ago. We sometimes actually have difficulty interpreting the data unambiguously.

FORESIGHT
INSTITUTE

We are thinking of solving this problem with an already existing collaboration. As an example, running a very quick quantum chemistry calculation (of a benzene molecule) produces data in a few different ways: a 2-dimensional chemical graph, a 3-dimensional representation, a bunch of molecular orbital information, and a large amount of numerical data, such as the overlap matrix, which corresponds to, here, a 70 by 70 array of floating point decimal numbers. How do we store these results in such a way that someone who wants to look at it in 2D can look at it, someone who wants the numbers can look at it, and someone who wants to look at orbitals can look at it? Here are the problems with some of the current file formats:

- Tripos Sybyl MOL2 chemical modeller input file (MOL2): you are going to have to write a parser for that; it doesn't look anything like Extensible Markup Language (XML) or JavaScript Object Notation (JSON). It has a lot of information, but certainly nothing quantum chemical.

- PDB: does not store bond orders and does a fairly lousy job of describing atoms.

- XYZ: good because at least it is unambiguous — list of atoms and their x, y, and z coordinates. However, there is nothing here that stores the data that we need to communicate with each other.

What we are proposing to do is to start looking into modern, hierarchical data storage — things like JSON, that stores all of this: stores the molecular topology, stores its name, stores its provenance (tells us where it came from), and stores all the numerical data that certain researchers might need — things like the calculated electronic wave function, all of the orbitals and their overlap matrix, and all of this in an easy to parse manner. We have an existing collaboration, and this is our goal: file formats that cover this space.

References:

Current related efforts:

      Mosaic: https://github.com/mosaic-data-model

      H5MD: http://nongnu.org/h5md/index.html

      AiiDA: http://www.aiida.net/docs/

An incomplete list of popular formats:

      https://en.wikipedia.org/wiki/Chemical_file_format

Data:

This format should encode data from:

      Structural databases (PDB, Materials Genome)

      Mechanics codes (ReaxFF, OpenMM, LAMMPS)

      Quantum codes (Gaussian, Quantum Espresso)

      Informatics codes (OpenEye, RDKit)

Resources:

Existing informal computational drug discovery collaboration:

http://github.com/alchemistry/fileformat

Project 5:
Miniaturization as a Service (MAAS)

Team members: Don Gilmore, Max Sims, Tad Hogg, Angela Morente Cheng, Si-Ping Han, Farhan Malik, and others

PRESENTATION BY MOSHE LOOKS AND SI-PING HAN

https://youtu.be/auoL2UMFmdA

PROPOSAL

Given a macroscale design specification, e.g., an application-specific integrated circuit (ASIC) for matrix multiplication, use AI to create a plan (sequence of well-defined steps) that constructs a functionally equivalent nanoscale object. Applications include accelerated neural networks, faster memory, and central processing units (CPUs). By working on the nanoscale, we can design and construct lower power and faster functional equivalents that initially complement and eventually replace their macroscale equivalents. On the AI side, we will formulate the problem as translation subject to constraints (minimize cost, number of steps, reversibility, etc.), and solve using program induction, planning, and neural reinforcement learning.

Transcript

Moshe: I am going to talk about automated design of nanosystems and how we are going to provide miniaturization as a service.

FORESIGHT
INSTITUTE

Our mission is to accelerate the transition going from the world of macro scale components big circuits, CPUs, RAM and make them at the nanoscale. The problem is that while there are multiple benefits of doings things at the nanoscale, the design and manufacturing principles are totally different. You can't just shrink a macroscale design. It will not work because the physics are totally different. Even if you could just shrink it and have it magically work, we could not build those things. So we need a way of going from a macroscale design to an actual procedure, a sequence of elementary operations that we know how to do that lead to a functionally equivalent nanoscale object.

The approach that we are going to take is to treat this as a machine learning task. We will use AI to translate a given macroscale design, which includes a full functional input-output specification of the behavior that we want, along with constraints on the final solution, such as the number of steps, the power consumption, the amount of compute time, etc,. to create an object that is functionally equivalent at the nanoscale.

On the AI side, algorithmically, we are going to need to take a hybrid approach. To establish a correspondence between macro- and microscale, we are going to need to start out with some kind of training data: macroscale designs, and then sequences of steps that we can take at the nanoscale which will allow us to build a functional equivalent.

Once we have a few of these, we can use them to learn a sequence-to-sequence model, or a variant thereof, which is the standard framework for translation problems. The utility of having a model like this is that even if the results are not very good, we can run inference on it very quickly, and that will serve as a starting point at which we can apply more computationally expensive methods, such as program induction to take large steps and transform designs in creative ways to exploit the unique physics of the nanoscale to create machines that are functionally equivalent, but perform that function in a radically different way.
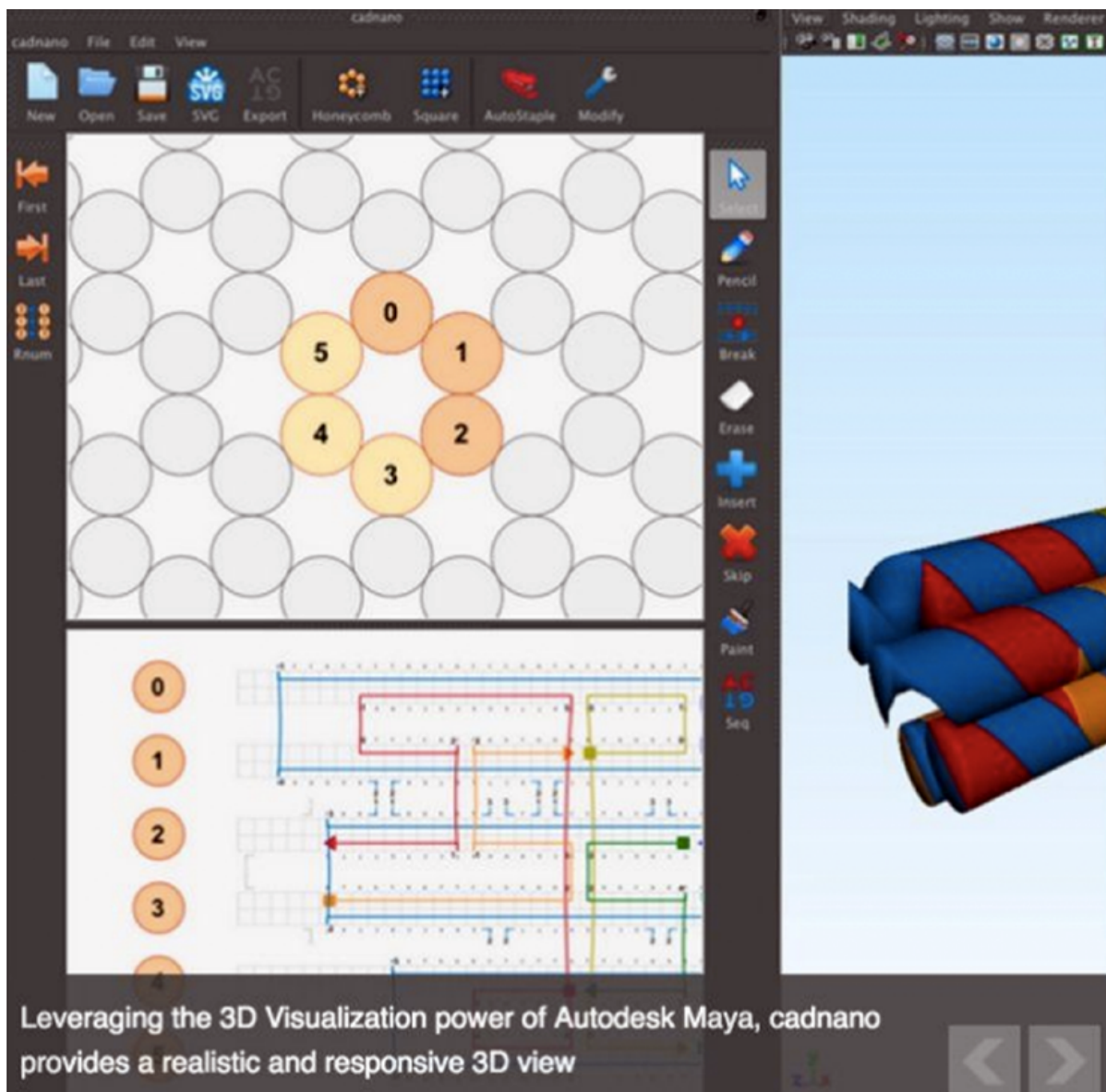
Finally, we are going to include classical planning, based on forward and backward chaining. If there is a design from program induction that is almost correct, you can use planning to exhaustively search small subspaces in order to find the exact sequence of steps that you need to get functional equivalence.

What is interesting about this is that if we can get it to work, it can train itself offline; it can bootstrap. There are many more examples of macroscale objects than

FORESIGHT INSTITUTE

microscale, so if we have a large library of macroscale objects, we can be churning away and searching for translations into the microscale, and use this to improve the model better and better.

On the nanoscale side, we have some suggestive proofs of concept. This a tool, "caDNAno". This screenshot (Illustration 1) shows what the output of the process might look like. This might be a way for people to inspect the models that are generated.

Si-Ping Han: I will talk about examples of actual molecular manufacturing.

Leveraging the 3D Visualization power of Autodesk Maya, cadnano provides a realistic and responsive 3D view

These are some examples of a class of nanostructures called DNA origami. There are 2D examples and 3D examples, and also today we can build larger complexes of DNA origami. The way that these structures are made is that a large DNA strand, an elongated DNA strand, is held into this shape by several hundred small DNA strands that all come together in a salt solution and self-assemble. When we specify the nanoscale designs, the idealized example of the nanoscale design is very different from the actual shape that you get in solution, because at this scale, the structures are actually dominated by the very strong interactions of the molecular scale forces. We can make quite sophisticated types of DNA structures, but it is also true that assembly is error prone. For example, assembling a large 2D array of DNA origami tiles should produce a uniform grid, but instead errors and incomplete assembly can happen at the edges. In any given system, we will have to take all of these effects into account and incorporated in physics-based models, including redundancies and errors and translating them to the nanoscale.

References:

Proof of concept for nanoscale design: "Designer Nanoscale DNA Assemblies Programmed From The Top Down" R Veneziano, S Ratanalert, K Zhang, F Zhang, H Yan, W Chiu, M Bathe. Science 352, 1534 (24 Jun 2016) DOI: 10.1126/science.aaf4388
http://science.sciencemag.org/content/352/6293/1534 –

Standard neural translation model: "Neural Machine Translation By Jointly Learning To Align And Translate" D Bahdanau, K Cho, Y Bengio. Published as a conference paper at ICLR 2015.
https://arxiv.org/pdf/1409.0473.pdf (19 May 2016)

Neural module architectures: "Neural Module Networks" J Andreas, M Rohrbach, T Darrell, D Klein.
http://arxiv.org/pdf/1511.02799v3.pdf (1 June 2016)

Program induction: "Competent Program Evolution" Moshe Looks. metacog.org/main.pdf (December 2006)

Proof of concept application of classic planning to similar domain: "Pattern Decomposition with Complex Combinatorial Constraints: Application to Materials Discovery" S Ermon, R Le Bras, SK Suram, JM Gregoire, CP Gomes, B Selman, RB van Dover.

https://arxiv.org/pdf/1411.7441.pdf (27 Nov. 2014)

"Folding DNA into Twisted and Curved Nanoscale Shapes" H Dietz, SM Douglas, WM Shih. Science 325: 725-730 (7 August 2009). DOI: 10.1126/science.1174251
http://bionano.physik.tu-muenchen.de/Source/curvature_dietz_et_al.pdf

"Self-assembly of a nanoscale DNA box with a controllable lid" ES Andersen, M Dong, MM Nielsen, K Jahn, R Subramani, W Mamdouh, MM Golas, B Sander, H Stark, CLP Oliveira, JS Pedersen, V Birkedal, F Besenbacher, KV Gothelf, J Kjems. Nature 459: 73-76 (7 May 2009) doi:10.1038/nature07971
https://www.tu-braunschweig.de/Medien-DB/pci/_Tinnefeld/nature07971.pdf

"Programmable molecular recognition based on the geometry of DNA nanostructures" Sungwook Woo, PWK Rothemund. Nature Chemistry 3: 620–627 (10 July 2011) DOI: 10.1038/NCHEM.1070
http://www.dna.caltech.edu/Papers/stacking-bonds2011.pdf

"Crystalline Two-Dimensional DNA-Origami Arrays" W Liu, H Zhong, R Wang, N C Seeman. Angew. Chem. Int. Ed. 50, 264–267 (January 2011)   DOI: 10.1002/anie.201005911
http://ww2.chemistry.gatech.edu/~lw26/bCourse_Information/DNA_technology/papers/seeman_orgami.pdf
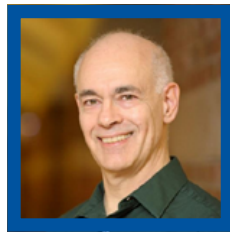
Resources available

https://materialsproject.org/, http://cadnano.org/, http://cando-dna-origami.org/

Project 6:
Nanoscale 3D Printer with error correction

Team members: James Gimzewski, Reto Stamm, Robert Freitas, Paul Cherukuri, Joe Lyding, and others

PRESENTATION BY JOE LYDING

https://youtu.be/QeOM68gZNLY

PROPOSAL

Idea: Design a bottom-up, 3D printer that can programmably construct an atomically precise macroscale object with built in error correction.

Impact: Atomically precise assembly will lead to the fabrication of new nanoscale devices and materials that will have major impact on society.

Approach: Fabricating an atomic scale 3D printer by combining an interferometer stage with a scanning tunnelling microscope (STM) to rapidly assemble individual atoms into layers and then layer-by-layer.

Hardware: 3D Printer Construction
- Ultra High Vacuum STM
- Interferometer guided piezo controlled stage

Software: AI Controlled Error Correction
- Slicing the model into single atom layers
- Deposition/process control

FORESIGHT INSTITUTE

- Image recognition to check single atom stampable layer
- Error correction - controlling the bumping out of place atoms into place with the STM before stamping
- Post-deposition single atom doping
- Actuator control to do the actual stamping

Products
- New transistor technology
- NEMS - Nanoelectric mechanical systems
- Catalytic surfaces
- Mechanically strong low density structures
- Platforms for self-assembling features
- Nano imprint templates

Advantages
- Feasible today
- Reducing cost for atomically precise surfaces and 3D shapes
- Rapid fabrication of nanoscale devices with atomically precise structure

Proposed budget: As a full-time project this project requires $3-5 million for 3 years to

- support 3 people
- 500k for the STM
- 500k for the interferometer stage

Figure 1

Use a moveable type approach to 3D print atomically precise objects.

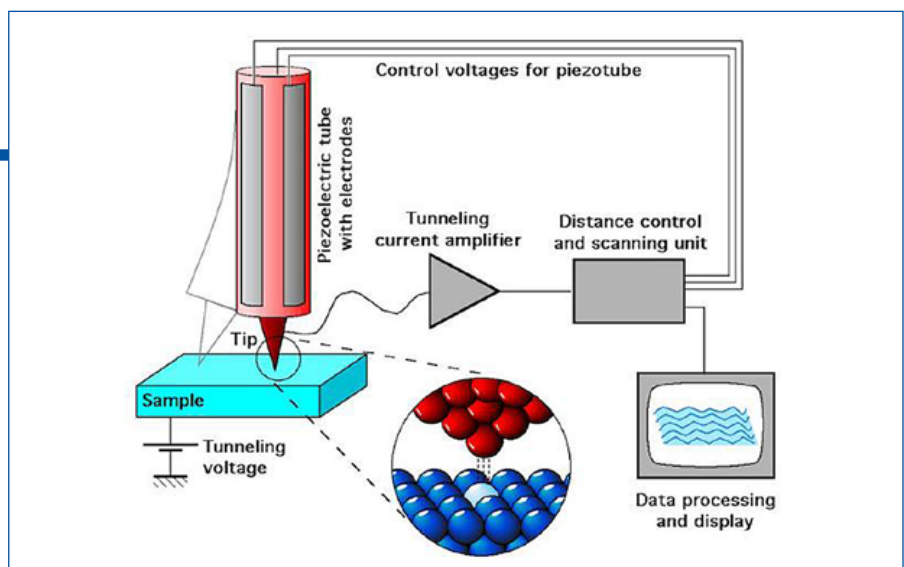Make "molds" for atomically precise layers using an STM.

Figure 2
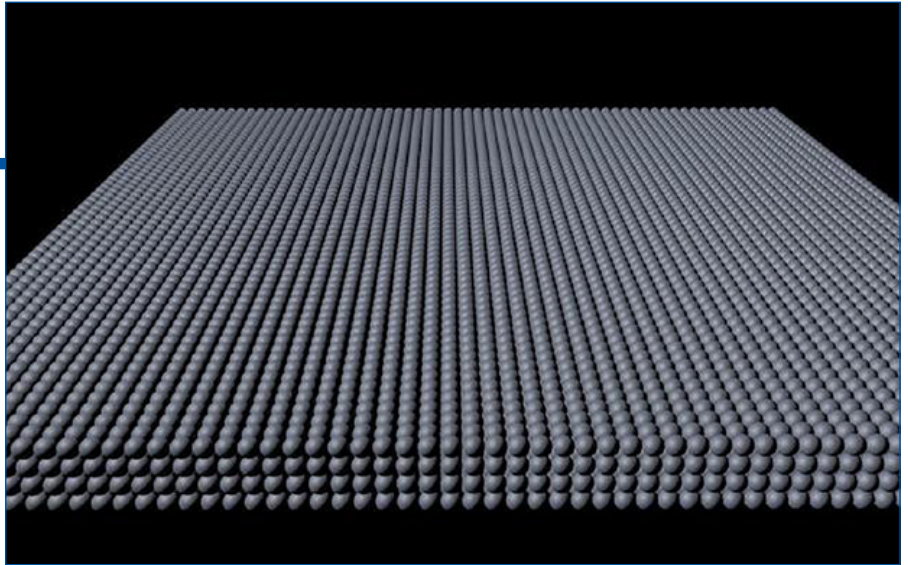
Use STM manipulation to make an atomi-
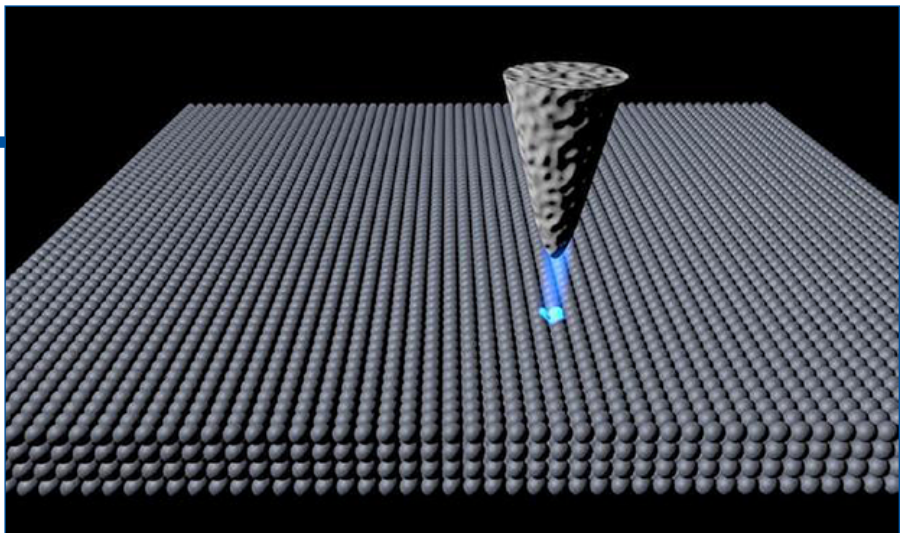cally precise template.

Figure 3

Remove an atom.

Figure 4

Remove more atoms to create an atomi-
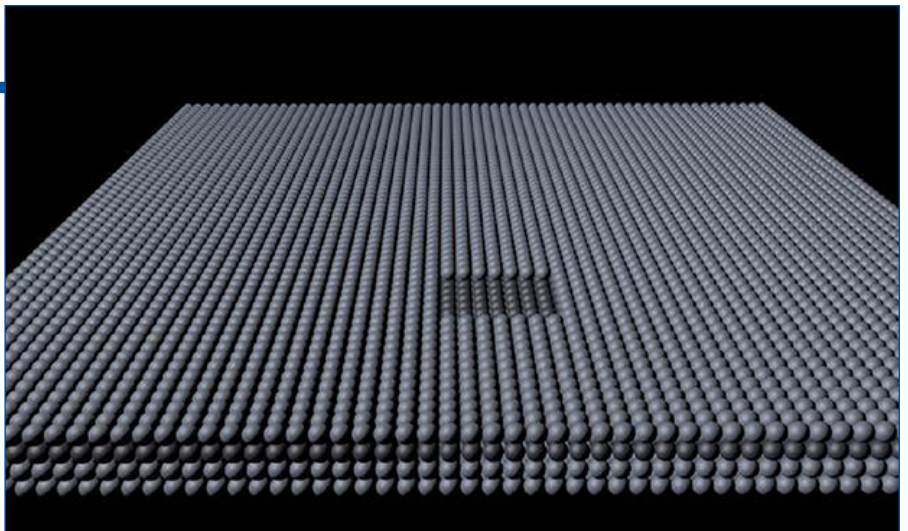cally precise mold.

Figure 5

A lower resolution cartoon showing a
second layer of atoms removed to create
an atomically precise mold

Figure 6

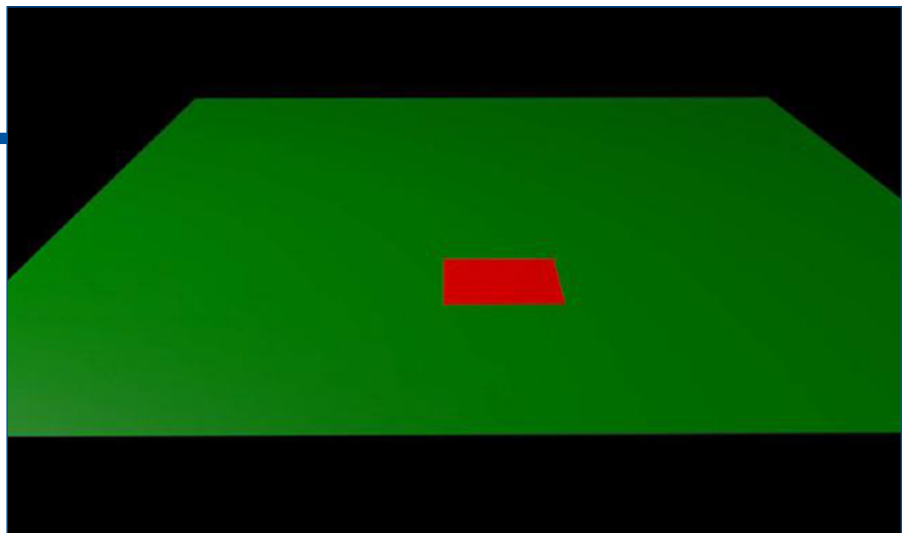Material to be transferred self-assembles
into mold to atomic precision.

Figure 7

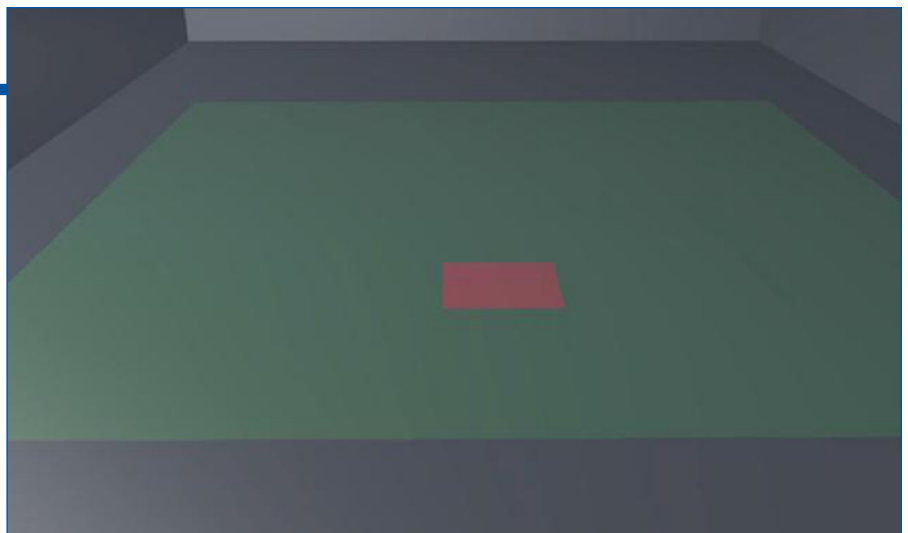Approaching with a stamp to remove the
material from the mold.

FORESIGHT
INSTITUTE

Figure 8

Transferring the two atomic layers from the mold to the stamp.

Transcript

Our idea is a 3D printer with error correction for atomically precise macro objects. We want to leverage a couple of techniques: one is scanning probe microscopy for atomic-scale fabrication, but also some of the semiconductor industry techniques. With the interferometry driven servo stages for chip manufacture, you can actually process 10 chips per second with x, y, z precision to 1.6 nm. That is one of the things we will be utilizing.

We want to create the atomically precise template once and then use it repeatedly. To do that, we will use an STM. I will not go through the details of that and how the patterning works, but here is a cartoon where we would take a surface, say an atomically clean silicon surface. We come in with an STM tip and pluck out an atom. There are publications where that has been done. We then repeat that process to clear out an area.

We might hydrogen-passivate the surface, and then use selective chemistry to put the material that we want to transfer into that hole. It would self-assemble into that hole with atomic precision. Then we would come in with a stamp, and transfer to that stamp. We could repeat that process, transferring the next layer to a stamp, etc. You could have differently shaped pieces that you are going to transfer. They are all one atomic layer thin, so you could transfer them sequentially and build objects like a 3D printer, where each layer determines how the structure evolves. Another thing you could is, if you had a second type of material, a different material, you

FORESIGHT INSTITUTE

WORKSHOP RESULTS

could say, "I am going to build that into the structure", and then you might be able to use that as a release layer; you might be able to subsequently remove it. So now what you've done is to open up the opportunity to make a nanoelectromechanical systems (NEMS) type of actuator.

You might have the scanning probe a millimeter away from where the transfer is going, so you go back and forth between them because you will need to do some error correction. You will need to verify, "Is the stamp still accurate? Do I need to put an atom back or take an atom out because of an error that has occurred during the processing?"

We would use off-the-shelf technology in terms of the UHV STM and interferometry-driven servo stages. As I said, the chip manufacturers can move a centimeter with 1.6 nanometer x, y, z accuracy 10 times a second. That is including the exposure step, so it is actually faster. We envision that if we are moving a millimeter, we might be able to do this whole process 100 times a second. The molecules or atoms that are going to be transferred would be introduced in the vapor phase, would rapidly self-assemble into that little pit that we made into the surface, and then the stamp would come in and take it up.

There are a lot of interesting opportunities for AI control and error correction: building a model, deposition, etc. It is a system, like a self-driving car, where you have to account for lots of things going on at the same time. There is a lot of image recognition error correction. For example, we analyze the stamp to see if it is still atomically precise. If it is not, we have to fix it.

What can you make with this thing? You could make new transistor technology, like atomically precise FinFETs. If you use release layers, you could make objects that could be moved. With atomic precision, you could have electromechanical actuators for sensors, or mechanical computing -- a radiation-hard computer would just have mechanical switches, for example. Small switches operating at 100 gigahertz are pretty interesting. Catalytic surfaces -- you could design surfaces to taste for catalytic processes. You could make a porous type of structure that is mechanically strong but doesn't weigh much. Nanoimprint templates -- you could make a template that can then be used to emboss very rapidly copies of itself. The one thing that is not on this list, interestingly, is that the semiconductor industry wants atomically precise templates so that they can calibrate their lithography systems. You could make those.

39   "AI FOR SCIENTIFIC PROGRESS WORKSHOP"

Project 7:
NLP and General AI

Team members: Rick Lewis, Peter Voss

PRESENTATION BY RICK LEWIS

https://youtu.be/aZRSs9WTSvM
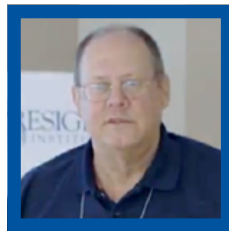
PROPOSAL

This proposal concerns a very specific target and can hardly be considered as general AI (whatever that may be). It should, however, be considered as a small step in the direction of creating a general AI application.

The medical literature is vast and the accumulation of potentially relevant information is accelerating. The information is vital to the successful development of safe and effective pharmaceuticals. In spite of general criticism in the media concerning overpriced medications and excessive profit, the success of the pharmaceutical industry can be considered positive and beneficial. Human life expectancy has significantly improved during the 20th century and a number of diseases which manifested considerable morbidity or were uniformly fatal can now be either cured or significantly ameliorated. It is anticipated that this trend will most likely accelerate.

The medical literature is important in several ways:

1. It serves as a repository of data which may reveal new emerging safety signals.
2. It serves as a means of benchmarking anticipated efficacy.
3. It serves as a means of identifying potential adverse events which are commonly associated with the underlying disease state.
4. It serves as a means of hypothesis generation concerning possible pathophysiologic pathways for toxicity or new areas for possible efficacious use.

The pharmaceutical industry is required by regulations to review literature for potential clinical safety signals. In addition, during the development of the clinical protocols, scientific staffs are often faced with questions during protocol design which require additional investigations of the literature.

The ascertainment of worthwhile information from the literature can be a time-consuming and laborious process which, unfortunately, is often repeated in large part because searches are not saved and results are not shared between teams. Most companies do not have a centralized repository of literature which is annotated and rapidly available to staff through simple "natural language" searches.

The emergence of natural language processing and, in particular, the advances in artificial intelligence and neural networks provide opportunities to improve this process. It is anticipated that early progress will be made by making the acquisition and indexing of literature more efficient. Eventually, it may become possible to train intelligent machines to provide additional value-added functions.

The most important resources that any knowledge worker has are training and time. The repetitive and time-consuming tasks should be made as brief as possible. The attached is an example of an exploratory approach to consider for the improvement this valuable process. It time a general AI tool would progress beyond simple assistance in identifying and curating the references.

Data & Resources

For this exercise, a seemingly straightforward question was asked concerning a possible adverse event occurring as the result of a medicinal product. The product in question is a monoclonal antibody which is an antagonist of interleukin-6. The concern rests with exposing patients who have been infected with hepatitis B to this medication which is an immunosuppressive. It is known that severe and often fatal (50% mortality) reactivation of the hepatitis can occur as a result of treatment with

some immunosuppressive agents (e.g., rituximab).

The questions is: I am hepatitis B core antibody positive but HBsAg negative. I will receive tocilizumab for rheumatoid arthritis, should I take antiviral prophylaxis?

To answer this question, the Embase database was searched for the following terms: tocilizumab, hepatitis B, and prophylaxis. Twenty unique references were captured.

The review was then undertaken by two different methods:

Adobe Acrobat

1. The identified Embase abstracts were downloaded as a text file and then printed using Adobe Acrobat.

2. Using Adde Acrobat and its redaction function, three key terms were specified and assigned different colors throughout the document. This makes it relatively straight-forward to identify each individual abstract as to its informational content.

    a. Hepatitis B (Yellow)
    b. Tocilizumab (Purple)
    c. Prophylaxis (Red)

3. Each abstract was reviewed and assigned a significance value grading its usefulness in answering the question. The scoring is defined as 0 - 10. For example:

    a. 0 = no value
    b. 4 = Provides valuable information
    c. 10 = Answers the question fully (includes references which concerning prospective trials designed to specifically address the question).

4. The number of hits of each of the terms of interest was counted and recorded.

5. A very simple analysis of the correlation between the "significance value" and the number of hits (total and by individual term) was employed.

Linguamatics I2E.

The same text file which generated the Adde document was imported into I2E

and indexed. It was then queried using the I2E software.

Results are available on request.

Transcript

We want to develop better text mining tools. They really are necessary to search really large datasets. The way that we have done literature in the past just is not very effective. I deal a lot with medical literature; it is very important for signal detection, and you get information concerning efficacy of compounds. You can look for adverse events that occur in the setting of the disease. That last part is very important, because otherwise they may just be background noise. Another thing is that it can help you with hypothesis generation. And we are required by law to do this. It has to be done in a methodical manner. The basic reality is that time is precious. I don't have enough of it. The only way that I can improve my productivity is by getting smarter, learning new skills, and using improved tools. Right now, the tools suck!

I want to ask a specific question [of the literature]. I have Hepatitis B core antibody. (Some of you have probably heard some talk about this.) But I am hepatitis B antigen (HBsAg) negative. That means that I have suppressed my infection completely. No symptoms. Zippo! But I am going to get this immunosuppressive agent, called tocilizumab. Do I need to take antiviral prophylaxis? Because when you are on an immunosuppressive, when you have been exposed to Hepatitis B, you can reactivate the virus, and if you reactivate, you have a 50% chance of death. So it is kind of a big deal.

The methods that I use to search the literature: I use Embase [https://www.elsevier.com/solutions/embase-biomedical-research] a lot; it has tons of literature. I created a CVS file, put it into a relational data model, and from that you can create queries, base reports, outputs, and that kind of stuff. I also went into the text files and created what I call the Adobe Process. It is really conceptualization. Then, Linguamatics I2E [https://www.linguamatics.com/] is a Natural Language Processing (NLP) text mining program that I have access to.

The query process and output from Embase shows many hits for individual search terms, like tocilizumab, hepatitis B, and prophylaxis, but searching for combinations of those terms greatly reduces the number of hits. That is really obvious; everyone does that. This is really an explanation of what Embase is, what Paradox does [relational database management system <https://en.wikipedia.org/wiki/Paradox_(database)>].

I use Adobe Acrobat Professional for this because it has a redaction function so you can go through even a very large document and highlight in minutes to seconds any terms that you are interested in. For our purposes, hepatitis B is yellow, tocilizumab purple, and   prophylaxis is red. I also use a usefulness scale. Did I think the abstract had any value when I went through and reviewed it?

That is what the output looks like in Acrobat. There were 10 hits on tocilizumab, 22 on hepatitis B, 4 on prophylaxis, for a total of 36 hits. This is what I2E looks like. I2E does not do hits, but it is really good at connecting dots. Out of the above hits, I found two that I thought were somewhat valuable. A usefulness score of 6 is not bad, but it is not great.

Is there any correlation with these results between the Adobe Process and using I2E? The Acrobat process identified 20 abstracts, of which two were somewhat valuable, I2E identified 9 abstracts of which 2 were somewhat valuable. What I really want to stress is time. Creating the corpus and reviewing the data with the Acrobat method: 2 hours 50 minutes. With Embase: 21 minutes. Creation of I2E index for query searching 5 minutes. Creating the I2E query: 1 minute 32 seconds. Running the I2E query: 6.48 seconds. The results of I2E are pretty close to the results of the Acrobat process.

Project 8:
Learning to act with recurrent neural networks

Team members: Natalia Díaz Rodríguez, Yad Faeq, Bogdana Rakova, and others

PRESENTATION BY YAD FAEQ

https://youtu.be/l-V5Iy2oSyg

PROPOSAL

In the general reinforcement learning setting, an agent needs to learn an optimal policy for achieving goals in an environment, while general purpose agents need to exhibit this ability across a diverse range of environments. Our aim is to design a recurrent neural network architecture that uses a hierarchy of LSTM units, an external memory, and learned compositions of modules to achieve transfer learning and avoid catastrophic forgetting. This functionality could be applied to the design and control of nanoscale systems.

Transcript

As the title says, we are learning to act with recurrent neural networks. In the current space for reinforcement learning, an agent learns to adapt to a cyclic envi-

FORESIGHT INSTITUTE

ronment. So, we have an agent that learns to adapt to a specific task that needs to be achieved. However, a general purpose agent, in most cases cases encountered, needs to generalize long enough to learn over different environments. Our aim is to design an LSTM network, an architecture that avoids catastrophic forgetting, and it uses transfer learning to adapt to different environments. For example, an agent that will learn to adapt to one environment will learn to control, for example, a macroscopic scanning control agent. We want to design an architecture that learns to adapt to the different environments, it will rebalance inputs, and achieve designated tasks using transfer learning.

When you are in a lab, some of the experiments can be very expensive. If you miss something during scanning, this agent will assist the person doing the scanning to avoid such errors. This will save time and the cost of replicating the experiment. Other things that would be nice to have, since are mission learning practitioners, would be others with additional expertise, to look over the real world applications of what we are building, and along the lines of building such products, one of the most important things is the actual GPU, the actual hardware that we use, so we are looking forward in terms of funding to have such hardware provided.

.
References:

"Progressive Neural Networks" AA Rusu, NC Rabinowitz, G Desjardins, H Soyer, J Kirkpatrick, K Kavukcuoglu, R Pascanu, R Hadsell. Google DeepMind, London, UK. (Sep 2016; this version, v3). https://arxiv.org/abs/1606.04671

"Reuse of Neural Modules for General Video Game Playing" A Braylan, M Hollenbeck, E Meyerson, R Miikkulainen. Department of Computer Science, The University of Texas at Austin. (Submitted on 4 Dec 2015). https://arxiv.org/abs/1512.01537

"On Learning to Think: Algorithmic Information Theory for Novel Combinations of Reinforcement Learning Controllers and Recurrent Neural World Models" Juergen Schmidhuber. The Swiss AI Lab. (Submitted on 30 Nov 2015)
https://arxiv.org/abs/1511.09249

"Learning to Communicate to Solve Riddles with Deep Distributed Recurrent Q-Networks" JN Foerster, YM Assael, N de Freitas, S Whiteson. University of Oxford, United Kingdom; Canadian Institute for Advanced Research, CIFAR NCAP Program, Google DeepMind. (Submitted on 8 Feb 2016). https://arxiv.org/abs/1602.02672

Data & Resources

OpenAI gym for data - https://gym.openai.com/docs

"OpenAI Gym is a toolkit for developing and comparing reinforcement learning algorithms. It makes no assumptions about the structure of your agent, and is compatible with any numerical computation library, such as TensorFlow or Theano. You can use it from Python code, and soon from other languages."

Resources: GPU instances from EC2 for research, Titan X GPU

https://www.google.com/#q=GPU+instances+from+EC2+for+research%2C+Titan+X+GPU

07. Conclusion

"AI for Scientific Progress: Bringing Digital Control to Physical Matter," was a highly interactive workshop convened to develop proposals for AI-assisted methods to bridge research bottlenecks in atomic precision, reflecting Foresight's vision of a future of synergistic technologies. This report covered the resulting ambitious research proposals, many of which are now open to collaboration and funding. If you are interested in helping the advancement of high-impact, beneficial research like this, there are several things you can do: get this report in front of the right people to make things happen, spread the word that our most powerful lever to improve the state of the world is novel technology, support Foresight Institute to extend the frequency and scale of our efforts. Get in touch to find out more: foresight@foresight.org.

Background Reading

The Artificial Intelligence Landscape

Beginnings

The appearance of usable computers during the 1950s afforded the opportunity to build and test experimental and theoretical disciplines around theories of intelligence that philosophers had been playing with for more than two millennia. Early systems incorporated symbolic logic, heuristics to link knowledge and actions, and theories of computation, probability, and knowledge representation. The connections among propositional logic, Turing's theory of computation, and the basic physiology of neurons in the brain led to both the logicist and the connectionist approaches to AI. This early work led to a two-month workshop at Dartmouth during the summer of 1956 for those interested in automata theory, neural nets (Fig. 1), and the study of intelligence. At this workshop, Allen Newell and Herbert Simon introduced Logic Theorist, a reasoning program that was able to prove most of the theorems in Chapter 2 of Russell and Whitehead's Principia Mathematica. For the next 20 years, the ten attendees from this workshop, plus their students and colleagues at Carnegie Mellon University, MIT, Stanford and IBM, dominated the new field, which became known as artificial intelligence, from the suggestion of the workshop organizer John McCarthy (2).

"The two main traditions in the beginning of AI are the logicist and the connectionist approaches to AI."
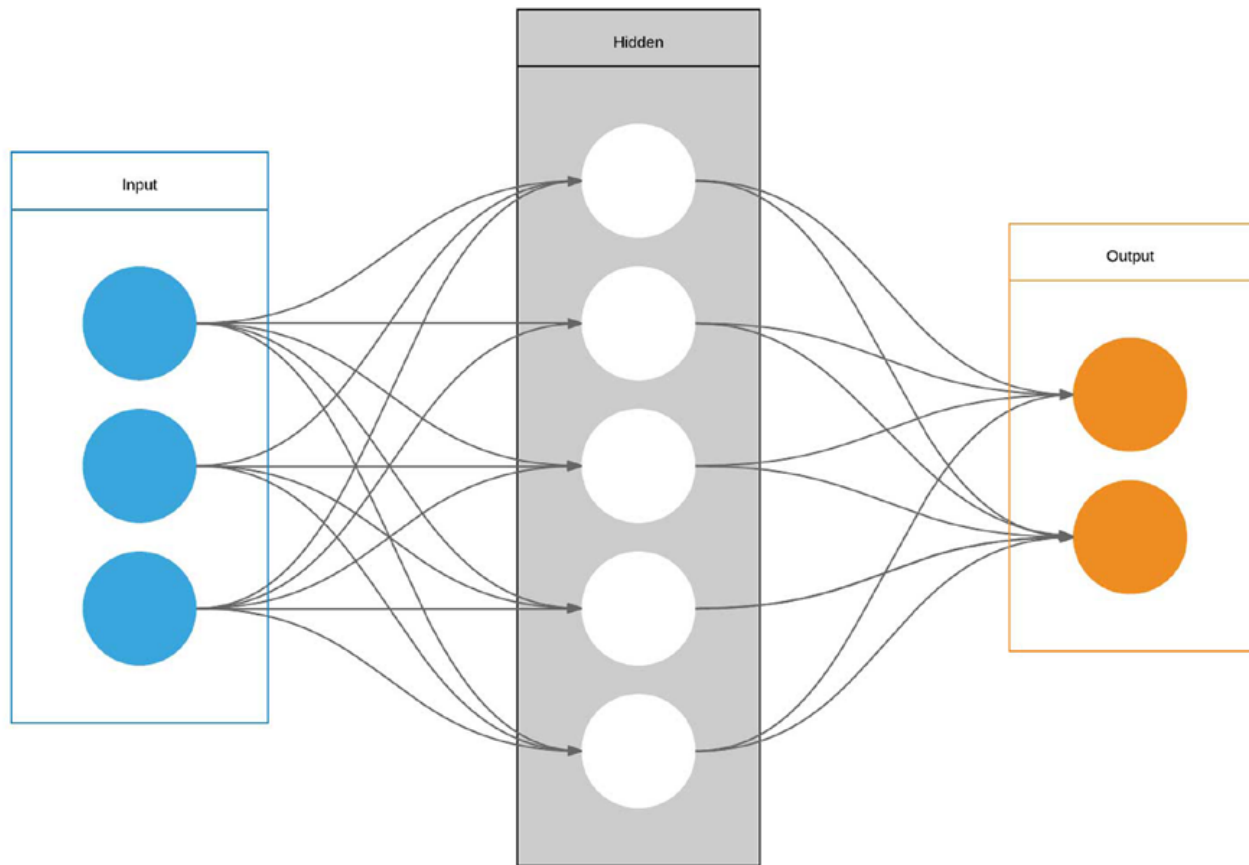
FORESIGHT INSTITUTE

Figure 1

"An artificial neural network is an interconnected group of nodes, akin to the vast network of neurons in a brain. Each circular node represents an artificial neuron and an arrow represents a connection from the output of one neuron to the input of another."
https://commons.wikimedia.org/wiki/File:Artificial_Neural_Network.jpg
Image credit: Author - LearnDataSci, learndatasci.com. Permission is granted to share and remix copy under the terms of the Creative Commons Attribution-Share Alike 4.0 International license.

Early work on theorem provers, first order logic, problem solvers, knowledge representation, the invention of the AI programming language Lisp, time-sharing computer systems, and neural networks eventually ran into problems as methods that worked for simple demonstrations failed on more complex real-world problems. For example, lack of contextual knowledge of specific domains led to failures, such as the famous example of the language translation program that translated the English sentence "The spirit is willing but the flesh is weak" into Russian, and then retranslated it back to English, giving the result "The vodka is good but the meat is rotten." Another problem was "combinatorial explosion", when a program had to deal with more than few dozen facts (2).

A famous 1969 book, Perceptrons by Marvin Minsky and Seymour Papert, showed that certain simple neural networks (perceptrons, Fig. 2) were not capable of proving certain logical predicates. Since the book pointed out that the simplest type of neural networks that were most popular with researchers at the time had some limitations, it discouraged work on neural networks for a time. However, what was clear was that more complex networks (back-propagation learning algorithms for multi-layer networks) could compute any possible Boolean function. This book pushed research first in one direction and then in another, a complex topic that was explored by a peer-reviewed sociological study of scientific development published in 1996 (ref. 2, 3).
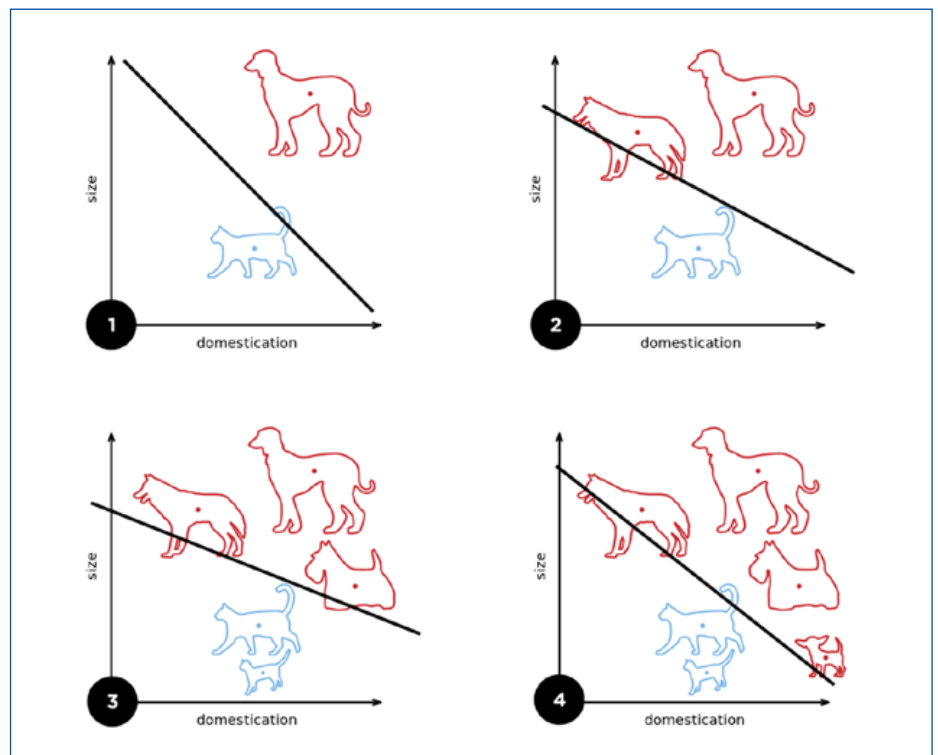

Figure 2

"A diagram showing a perceptron updating its linear boundary as more training examples are added."
https://en.wikipedia.org/wiki/Perceptron.
Image Credit: Elizabeth Goodspeed. This file is licensed under the Creative Commons Attribution-Share Alike 4.0 International license.
https://commons.wikimedia.org/wiki/File:Perceptron_example.svg

> *"Knowledge-based expert systems depend upon special-purpose rules while neural networks are modeled on the brain."*

A fresh breeze for AI

The early approaches described above used only limited knowledge about specific domains. More successful expert systems were developed by using more complete knowledge bases for narrow knowledge domains, such as solving molecular structure using information from mass spectrometer experiments. The expertise of such knowledge-intensive systems derived from large numbers of special-purpose rules. The commercial success of expert systems was followed by renewed work on neural networks. Physicists applied techniques of statistical mechanics, used to analyze large collections of atoms, to optimize large collections of neurons.

Over the past 30 years, AI programs have found increasing success with complex real-world applications. One example is speech recognition programs based on hidden Markov models which use rigorous mathematical theory and training on a large body of real speech data. Previously unthinkable, some systems have come to surpass humans at performing a narrow set of tasks by acting rationally according to laws of probability and decision theory, rather than by imitating human experts (2): A chess-playing program that eventually defeated the world champion human player in 1997 (4), a question-answer program that defeated two former champions of the quiz show Jeopardy! in 2011 (5), and a computer program to play the board game Go that defeated the world champion in 2016 (6). Other less famous, but remarkable successes of deep neural networks bear resemblance to the way humans process information: Microsoft's ResNet architecture, which won all international image recognition competitions in 2015, built a neural net that was ~150 layers deep, instead of the common ~8 layers. The depth was made possible by allowing the net to skip certain layers when not needed, using layers only when needed, similarly selective to how the human brain operates.

Another example in this space is Google Translate's zero shot translation that was revealed in late 2016. While being trained to translate a multitude of languages, the system builds a body of shared "translation knowledge" between a multitude of languages. Surprisingly, this translation knowledge even allows the system to translate between a language pair it has not been trained on yet. Whether this common bridge representation classifies as its own language, i.e. an "interlingua" that is a predecessor of human consciousness is debatable. However, this style of learning at least resembles human learning and explains why human individuals who are bilingual are better at learning new languages because they draw on knowledge of multiple languages to learn new ones.
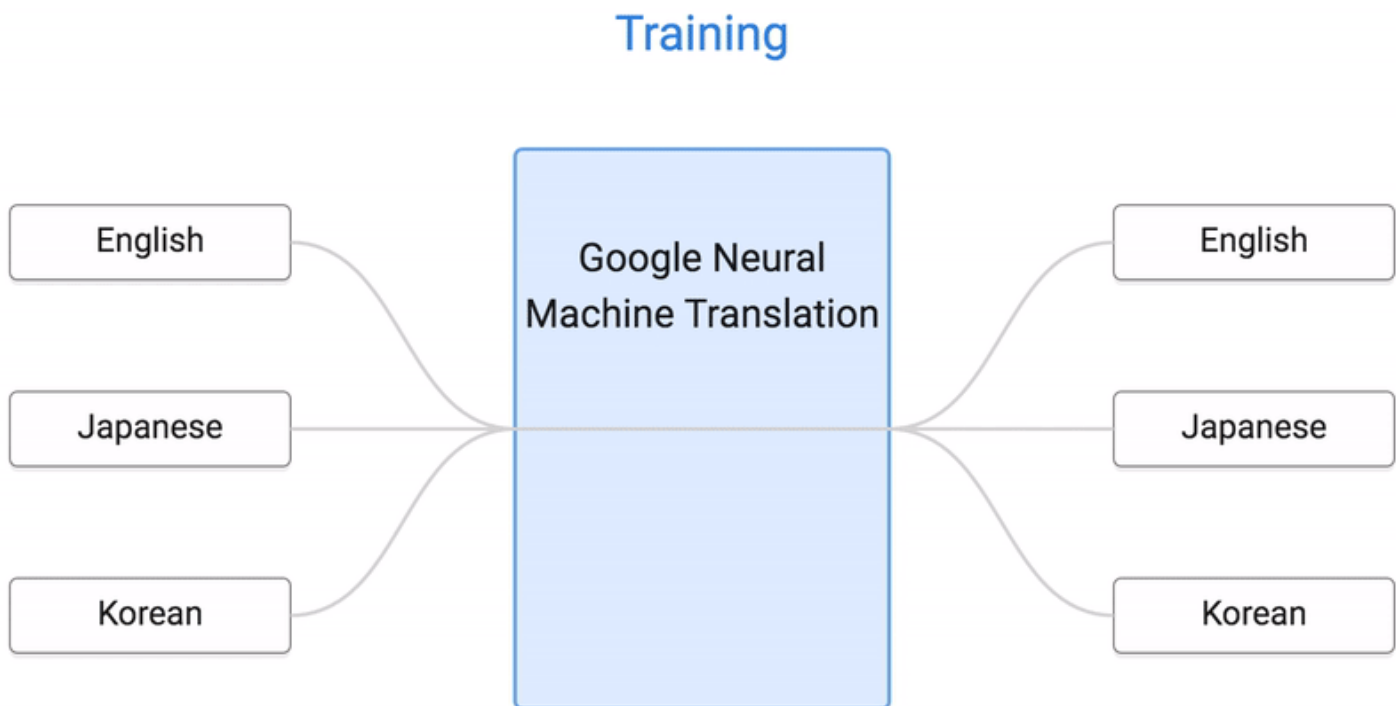
FORESIGHT
INSTITUTE

Training

Figure 3

Illustration of Google's zero-shot learning, taken from Google's research blog.
https://research.googleblog.com/2016/11/zero-shot-translation-with-googles.html

Narrow AI vs. Artificial General Intelligence

An introduction to AI conducted by the AI researchers present at the workshop for the benefit of the nanotechnologists urged for the distinction of the terms Narrow AI and Artificial General Intelligence (AGI), which are often conflated in media coverage and public discourse.

While narrow AI systems are focused on doing one specific task very well, often better than a human, the AGI Society (http://www.agi-society.org/) defines Artificial General Intelligence as an "emerging field aiming at the building of thinking machines; that is, general-purpose systems that can perform any intellectual task as

FORESIGHT
INSTITUTE

> **"Major AI effort today is deep neural nets trained on massive amounts of data labeled by humans."**

well as the human mind (and perhaps ultimately well beyond human general intelligence)." While success in AGI-related research lags behind success in narrow AI, it will likely render narrow AI research obsolete once achieved.

Initially, the term "Artificial Intelligence," as coined by John McCarthy, was used to mean Artificial General Intelligence, but as more companies appropriated the term AI to describe their narrow AI methods, the term AGI was introduced to distinguish the two. An interesting parallel exists to the field of nanotechnology;The term nanotechnology initially described atomically precise technology before it got appropriated by media and industry to mean "very small" -- and the terms "atomic precision" and "molecular manufacturing" were introduced to distinguish the original meaning of nanotechnology from the common use of the word. The difference between current AI systems and an ultimate AGI is best illustrated using the above mentioned, well-known narrow AI examples of Watson and AlphaGo.

IBM's Watson uses rule-based systems and symbolic logic based systems, and more recently, neural nets. Since neural nets need to be trained on massive amounts of data labeled by humans, each new application requires doing something specific to prepare the system: prepare rules, collect and label the data set, train and retrain, etc. The main disadvantage is that Watson's AI is a black box that provides little insight into how it works. It is not possible to reasonably extract from the system why it produced the result that it did. If a specific bug is distorting the output, it is hard to improve performance by finding and fixing the bug because most of its work is sub-symbolic. One needs to train the system again using a different labeled data set and hope for better results. While there are computer-assisted approaches to tuning neural nets, e.g., "hyper parameter optimization" for finding the most promising arrangement of networks, currently, each problem is still its own story.

The black-box characteristic is problematic for a more general approach to intelligence in AI, but also poses problems for human-AI interaction. In situations in which providing reliable reasoning is essential, e.g., medical applications, the lack of evidence provided by machine learning results causes doctors to distrust the AI's recommendation. Nevertheless, a few weeks after this workshop was held, a conference paper presented "a new way to train neural networks so that they provide not only predictions and classifications but rationales for their decisions." The system learns "to extract pieces of input text as justifications – rationales – that are tailored to be short and coherent, yet sufficient for making the same prediction." (8). Whether this post-rationalization inhibits intelligence is debatable, but illustrates a move to-

FORESIGHT
INSTITUTE

ward a more general approach to the problem.

Another example of a very successful narrow AI program that is limited to the area on which it is trained is Google's AlphaGo. It may exceed a human's expertise in the domain of AlphaGo games, but cannot easily transfer this knowledge if the rules of the game are changed or if it was to play a different game altogether. Interesting work toward a more general approach to AI was recently published by researchers at OpenAI. Their software platform, Universe, was created to test different AI-frameworks on a diverse set of games and website. The goal is to measure how generalizable the framework's intelligence already is, but also to train the framework to allow knowledge transfer between different games and websites. The ultimate goal of AGI is an agent that understands the general context of any situation, not only of games and websites, and adapts to them as needed, as would a human.

Paths to AGI

AGI was not considered extensively at this workshop, since its successful development would likely revolutionize every area of human endeavour in ways that deys current predictions. Discussion indicated, however, that most advances in the research field of AGI would highly speed up progress in important areas of science and technology, e.g., increasing generality in intelligence might allow an agent to read and understand very technical scientific literature via Natural Language Processing and draw inferences between different research areas. This capability could allow automated generation of non-trivial scientific hypotheses.

A diversity of paths to AGI are currently being explored. Some approaches want to build up deep neural nets to become general, some are purely symbolic, and some endeavor to put symbolic approaches and sub-symbolic nets together.

A good overview of the current state of AGI research and the potential prospects and challenges arising from AGI can be gleaned from the following publications:

- "Mapping the Landscape of Human-Level Artificial General Intelligence" Sam S. Adams, Itamar Arel, Joscha Bach, Robert Coop, et al. AI MAGAZINE, 2012.
  http://www.aaai.org/ojs/index.php/aimagazine/article/view/2322/2269

- "Artificial General Intelligence" Ben Goertzel, Cassio Pennachin. Springer, 2007. PDF:
  http://sanghv.com/download/soft/machine%20learning,%20artificial%20intelligence,%20mathematics%20ebooks/AI/artificial%20general%20intelligence%20(2007).pdf

"Neural networks are black boxes that provide little insights as to why they work."

"AGI (artificial general intelligence) systems equals or exceeds general human intelligence while narrow AI systems perform well on a subset of tasks."

"Increasingly general AI, especially for effective natural language processing could significantly accelerate progress in important areas of science and technology"

FORESIGHT
INSTITUTE

■ "Artificial General Intelligence: Concept, State of the Art, and Future Prospects" Ben Goertzel. Journal of Artificial General Intelligence, 2014. PDF: https://www.degruyter.com/downloadpdf/j/jagi.2014.5.issue-1/jagi-2014-0001/jagi-2014-0001.xml

■ "Autonomous Technology and the Greater Human Good" Steve Omohundro. Taylor and Francis, 2014. PDF: http://www.tandfonline.com/doi/full/10.1080/0952813X.2014.895111%20

■ "Deep-Learning Reading Papers Roadmap" GitHub: https://github.com/songrotek/Deep-Learning-Papers-Reading-Road-map

■ "Intelligence Explosion" Luke Muehlhauser. http://intelligenceexplosion.com/

■ "Machine Ethics" Michael Anderson, Susann Leigh Anderson. Cambridge, 2007.

■ "Research Priorities For Beneficial AI" Future of Life Institute. PDF https://futureoflife.org/data/documents/research_priorities.pdf

■ "Smarter Than Us" Stuart Armstrong. MIRI: https://intelligence.org/smarter-than-us/

■ "Superintelligence: Paths, Dangers, Strategies." Nick Bostrom. Oxford University Press, 2014. Purchase at: https://www.amazon.com/Superintelligence-Dangers-Strategies-Nick-Bostrom/dp/0198739834

■ " The AI Revolution: The Road to Superintelligence" Tim Urban. waitbutwhy, 2015 : http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html

■ "The Value Learning Problem" Nate Soares, MIRI 2016. https://intelligence.org/files/ValueLearningProblem.pdf

FORESIGHT
INSTITUTE

The Atomic Precision Landscape

One example of why atomically precise manufacturing is such an attractive proposal is illustrated by a recent theoretical study, which presents a new design paradigm that both vastly simplifies the design of mechanical nanocomputers, and relies solely on mechanisms with very low friction (9). Because only rotary joints are used for movement, a mechanical computer based on this design is estimated to provide 1012 giga floating point operations per second (GFLOPS) per watt. The increase in energy efficiency compared to conventional computers would be a factor of millions, or even billions. This would not only decrease energy use for computation, but would greatly decrease heat from energy dissipation in applications requiring extensive real time computation. For example, extensive application of swarms of medical nanorobots might otherwise be limited by the heat produced by onboard computation. Fig. 4 provides a model of a component of such a mechanical molecular computer.
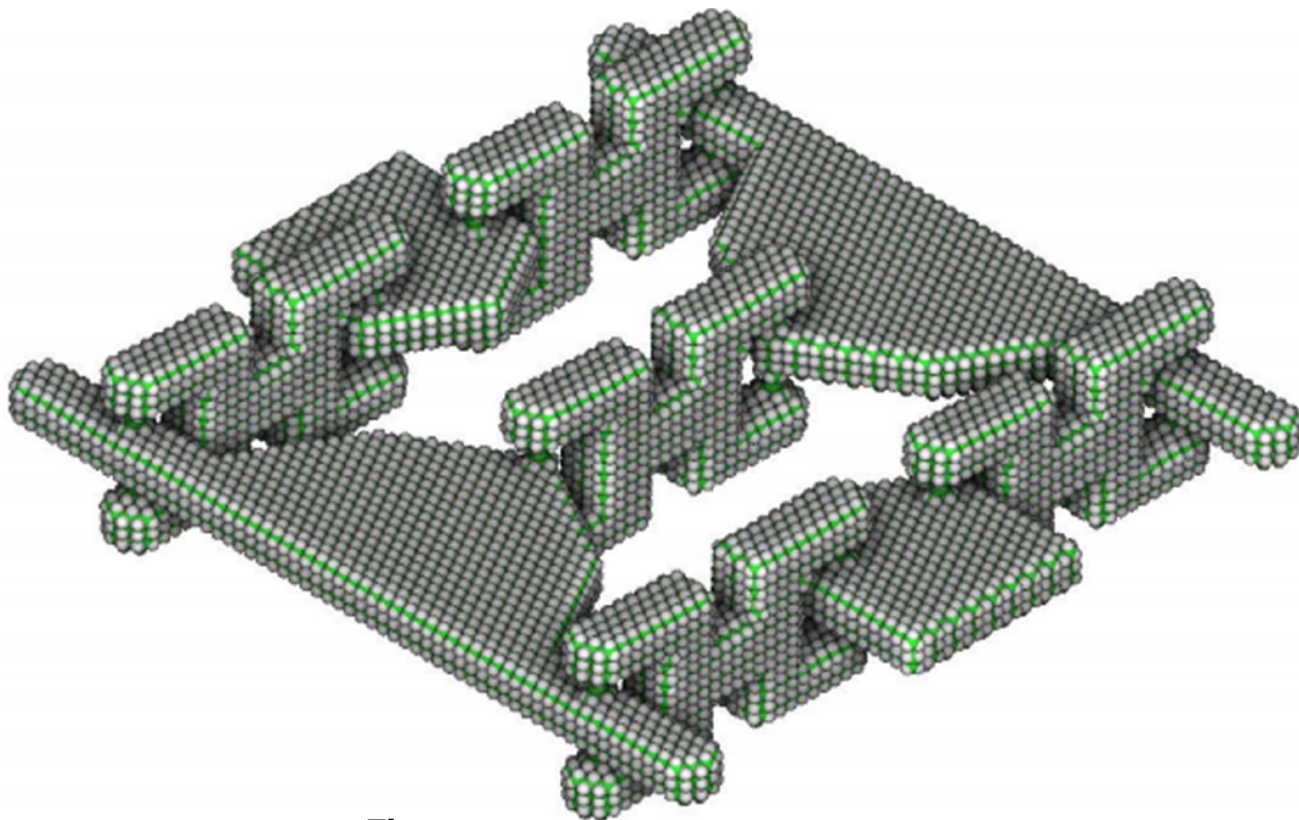

Figure 4

A molecular model of a diamond-based lock, as proposed by Merkle et al. (9). Carbon atoms are green; hydrogen atoms are gray.
Copyright 2016, Institute for Molecular Manufacturing.
[Permission to use granted by Ralph Merkle, 20161117]

> *"Design for an atomically precise nanomechanical computer manufactured from diamond could increase energy efficiency one million-fold."*

> *"Proteins can be designed de novo to pre-determined atomically precise geometries."*

How could complex parts like the lock illustrated above be manufactured to atomic precision? When Richard Feynman in 1959 first proposed "maneuvering things atom by atom" to manufacture with atomic precision, he proposed no details for the systems of molecular machines that would do the manufacturing (10). In 1981, K. Eric Drexler, inspired by biological molecular machinery, proposed a molecular engineering path toward atomically precise manufacturing, beginning with designing proteins to fold in a predetermined way so that a linear polymer folds under the influence of a constellation of weak (non-covalent) chemical bonds to give a specific geometry of atoms in 3-dimensional space. Within seven years of Drexler's proposal, reports appeared in the literature of proteins designed de novo that were substantially more stable than comparable natural proteins of similar sizes (11).

By 2003, a novel protein fold had been designed with atomic level accuracy, that is, it was found experimentally to be extremely stable and to have the designed structure (12). Progress in protein design has accelerated substantially over the past few years; now, folds and functions can be produced beyond those used by biology (13). For example, one paper published at the end of 2015 demonstrated that existing proteins of a certain structural class (the helix-loop-helix-loop structural motif) oc-
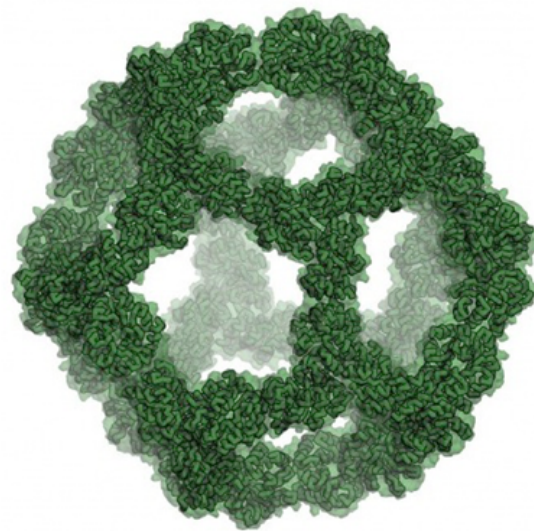

Figure 5

The design model of the icosahedral nanocage shows its large, empty volume (25 nm diameter and an interior volume of about 3000 nm3). "Design of a hyperstable 60-subunit protein icosahedron" Yang Hsia, Jacob B. Bale, Shane Gonen et al. Nature 535, 136–139 (07 July 2016) doi:10.1038/nature18010. [http://www.nature.com/nature/journal/v535/n7610/full/nature18010.html.
See also "Assembling a large, stable, icosahedral protein molecular cage" http://www.foresight.org/nanodot/?p=7154. *Credit:* University of Washington Institute for Protein Design "Self-assembling protein icosahedral shell designed" http://hsnewsbeat.uw.edu/story/%E2%80%8Bself-assembling-protein-icosahedral-shell-designed

FORESIGHT INSTITUTE

"Proteins can be designed de novo with folds not seen in nature, and with more stable folds than natural proteins."

cupy only a small fraction of the sequence and structure space so that it is possible to design novel proteins (not found in nature) with precisely specified geometries (14)

In addition, proteins can now be designed to make large, artificial structures (Fig. 5).

Another approach to using peptide bonds between amino acids to arrange atoms and molecular functions in 3D space, without requiring a long polymer to fold in a specific way, was developed by Dr. Christian Schafmeister, now at Temple University. Intermediate size bis-peptides are produced by solid phase peptide synthesis joining small molecules by two rather than one peptide bonds to give a shape-programmable molecule. Shape-programmable molecules could provide 3D scaffolds upon which to build rationally designed catalysts that could function to build novel molecules using reactions not found in natural biology (Fig. 6). More recently, a conformationally restrained spirocyclic scaffold built using bis-peptides has been used to position two functional groups, a phenol alcohol and a carboxylic acid, to enhance the rate of the Claisen rearrangement by a factor of 58 over the background reaction (15).
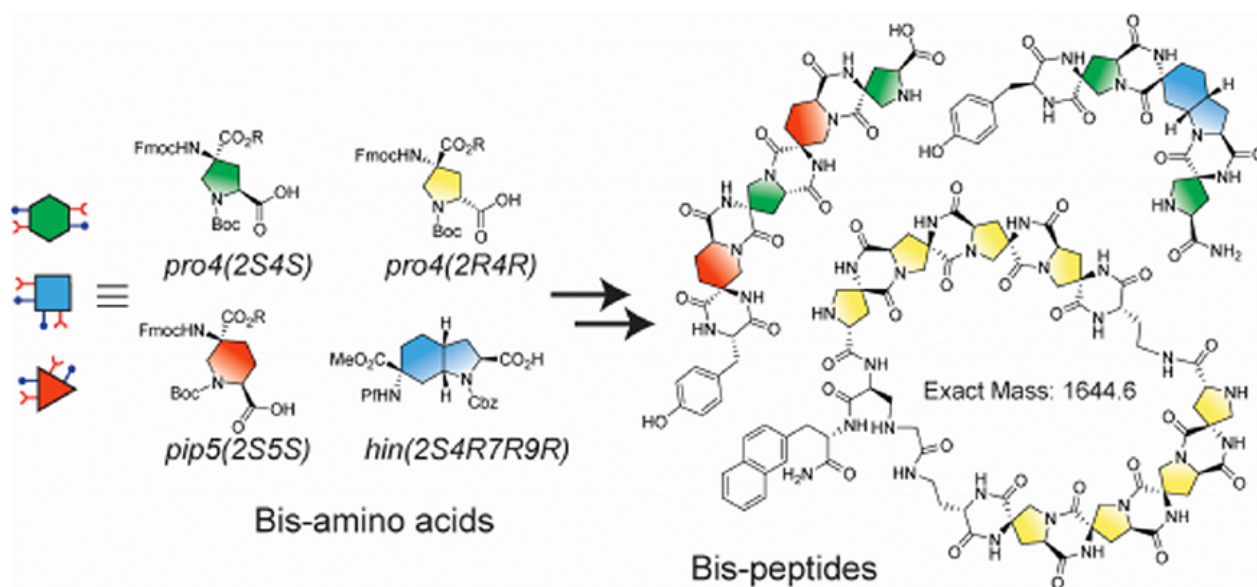

Figure 6

Bis-amino acids are coupled through pairs of bonds to create bis-peptides. The three example bis-peptides on the right have all been synthesized. Fig. 5 of "A Path to a Second Generation Nanotechnology" C. E. Schafmeister.
http://foresight.org/roadmaps/Nanotech_Roadmap_2007_WG_Proc.pdf. Paper 12. Presented at Productive Nanosystems: Launching the Technology Roadmap, October 9-10, 2007
http://foresight.org/Conferences/NanosysBrochure-edited.pdf.

"Shape-programmable peptides provide 3D scaffolds upon which to build rationally designed catalysts."

"Branched DNA junctions make possible complex 2D and 3D structures and devices."

Another potential path to advanced nanotechnology and artificial molecular machines opened in the late 1980s when Prof. Nadrian Seeman of New York University introduced DNA nanotechnology by demonstrating that nanostructures of various shapes could be built by causing DNA to form branched structures instead of the usual duplex molecule. Several years later Seeman's work on DNA nanotechnology earned him the first Foresight Institute Feynman Prize to be awarded for experimental work (16). Early constructs exhibited topological connectivity, but did not have precisely defined angles (Fig. 7).
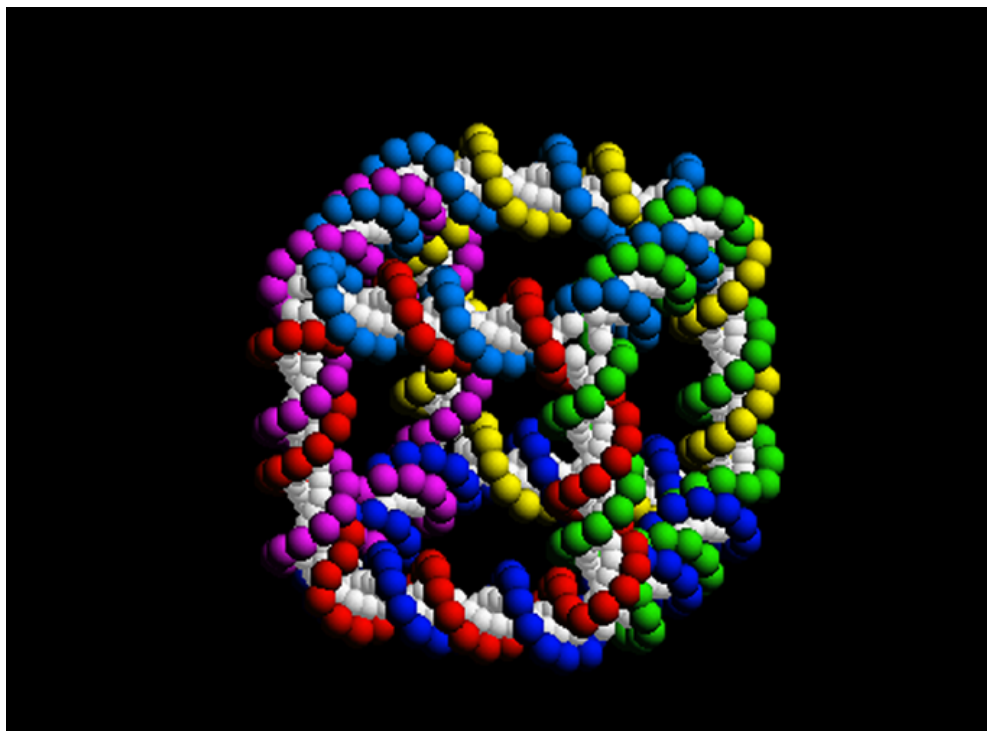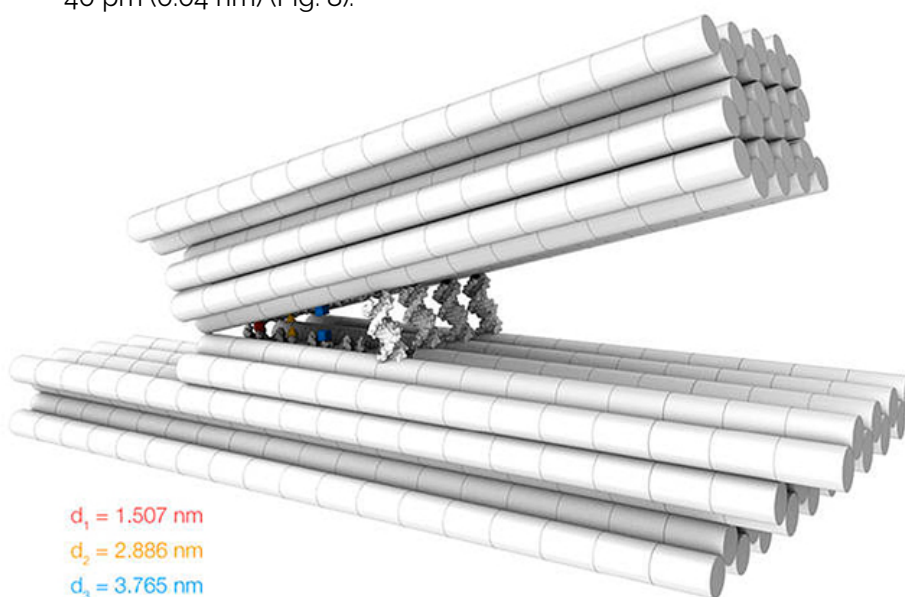

Figure 7

"A DNA Molecule with the Connectivity of a Cube. This representation of a DNA cube shows that it contains six different cyclic strands. Each nucleotide is represented by a single colored dot for the backbone and a single white dot representing the base. Note that the helix axes of the molecule have the connectivity of a cube. However, the strands are linked to each other twice on every edge. Therefore, this molecule is a hexacatenane. To get a feeling for the molecule, follow the front strand around its cycle: It is linked twice to to each of the four strands that flank it, and only indirectly to the strand at the rear. Note that each edge of the cube is a piece of double helical DNA, containing two turns of the double helix."
Source: "New Motifs In DNA Nanotechnology" Nadrian C. Seeman, Hui Wang, Xiaoping Yang, Furong Liu, Chengde Mao et al. Fifth Foresight Conference on Molecular Nanotechnology (November, 1997)
http://foresight.org/Conferences/MNT05/Papers/Seeman/index.html.
Original report of cube: "The synthesis from DNA of a molecule with the connectivity of a cube" J. Chen and N.C. Seeman. Nature 350: 631–633, 1991.

Since then, a steady series of improvements in the technology, including the invention of scaffolded DNA origami in 2006 (17), have brought ever more complex 2D and 3D nanostructures and devices, including molecular walkers, breadboards, and dynamic molecular machines (18). Two recent examples, both from the laboratory of Hendrik Dietz, Technische Universität München, illustrate progress toward atomic precision and dynamic molecular machines based on scaffolded DNA origami.• A manipulator able to position molecules with atomic precision. Despite the fact that DNA double helical segments are 2 nm in diameter, a hinge that is opened or closed by adjusting the length of adjuster helices could be adjusted to a precision of about 40 pm (0.04 nm) (Fig. 8).

$d_1 = 1.507$ nm
$d_2 = 2.886$ nm
$d_3 = 3.765$ nm

Molecular manipulator made from DNA. Dietz Lab/TUM

Figure 8

A DNA origami manipulator comprises two double strand DNA segments connected by a short single strand DNA hinge. A third double strand segment, the adjuster helix, connects the two double helical segments, and is placed at one of three positions from the hinge, and can be varied in length. Four such three-helix assemblies can be placed in parallel to form the molecular manipulator. Depending on the length and position of the adjuster helices, the opening of the manipulator can be controlled in 123 discrete steps, corresponding to a separation between the manipulated molecules of 1.5 to 9.0 nm.
"Placing molecules with Bohr radius resolution using DNA origami" Jonas J. Funke, Hendrik Dietz. *Nature Nanotechnology* **11,** 47–52 (19 October 2015) http://www.nature.com/nnano/journal/v11/n1/full/nnano.2015.240.html. See also "Using DNA nanotechnology to position molecules with atomic precision" http://www.foresight.org/nanodot/?p=6890 *Image source:* credit Dietz Lab, TUM "Nanoscale rotor and gripper push DNA origami to new limits" https://www.tum.de/en/about-tum/news/press-releases/short/article/32983/ [permission to use granted by Prof. Hendrik Dietz <dietz@tum.de> Technische Universität München by email 201120.]

"*A molecular manipulator built using scaffolded DNA origami can position molecules to atomic precision.*"

A former postdoctoral fellow from Dietz's lab, Carlos Castro, now at Ohio State University, demonstrated well-defined programmed motions with DNA nanostructures, thus beginning to fabricate parts for machine designs based upon the way that macroscopic machines work (19). In the most recent demonstration of a dynamic molecular machine, the Dietz lab demonstrated a DNA origami rotor in which the interlocking DNA components rotate freely with respect to one another, propelled by Brownian motion. Although this implementation does not yet include a motor, it demonstrates the feasibility of such a machine, opening the way for active devices (Fig. 9).
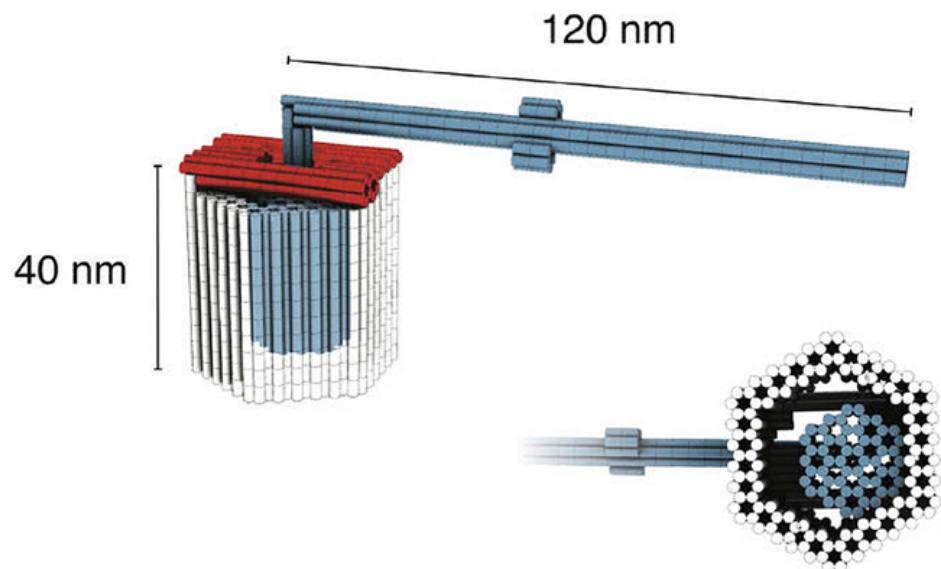
Rotor mechanism assembled from 3-D DNA components. Dietz Lab/TUM

Figure 9

Three multilayer DNA components make up this rotary mechanism. The parts join together with a tight fit and leave just 2 nanometers of play around the axle, allowing the arm to swing but not wobble. "Nanoscale rotary apparatus formed from tight-fitting 3D DNA components" Philip Ketterer, Elena M. Willner, Hendrik Dietz. Science Advances 2(2) e1501209 (19 Feb 2016). DOI: 10.1126/sciadv.1501209 http://advances.sciencemag.org/content/2/2/e1501209. See also "Tightly-fitted DNA parts form dynamic nanomachine" http://www.foresight.org/nanodot/?p=7011.
Image source: credit Dietz Lab, TUM "Nanoscale rotor and gripper push DNA origami to new limits"
https://www.tum.de/en/about-tum/news/press-releases/short/article/32983/

Nanostructures can also be built using DNA's chemical cousin, RNA. RNA can assume a wider range of complex, compact 3D shapes than can DNA, and some RNA molecules can function as catalysts, as can proteins (20). Thus, RNA nanotechnology provides a somewhat different toolkit for constructing functional, atomically precise nanostructures than does DNA nanotechnology.

Organic synthesis, particularly the invention of mechanically interlocked organic molecules, provides yet another path to sophisticated molecular machine systems. These include the smallest molecular machines possible, using far fewer atoms than molecular machines made from macromolecules, like proteins and nucleic acids. Recent examples include an autonomous chemically fueled molecular motor (21), a molecular arm to transport cargo (Fig. 10), a light-driven molecular machine (22), and two molecular wheels connected by an axle, which rotate in opposite directions, just like the wheels of a car (23). Making them potentially even more useful, addressable molecular machines have been arranged in the cavities of a porous crystal (24). Other examples include a single molecule pump that concentrates small molecules against a gradient (25), and several types of nanocars (26).

"Dynamic molecular machines with interlocking DNA components can be built using scaffolded DNA origami."
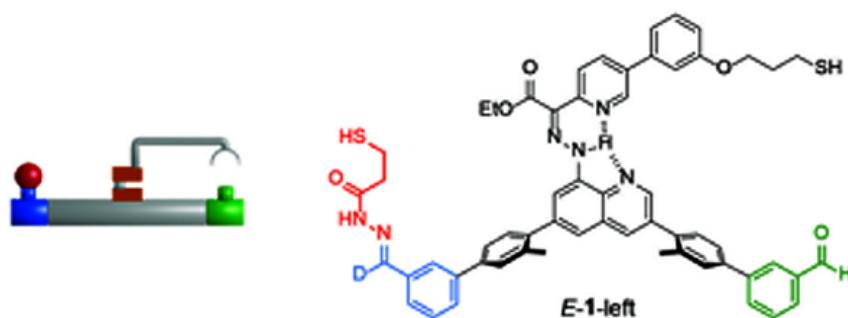

Figure 10

Chemical structure of a molecular robotic arm (shown in black) able to reposition a molecular cargo (shown in red) in either direction from blue-to-green or green-to-blue platform sites. "Pick-up, transport and release of a molecular cargo using a small-molecule robotic arm" Salma Kassem, Alan T. L. Lee, David A. Leigh, Augustinas Markevicius, Jordi Solà. *Nature Chemistry* **8**, 138–143 (21 December 2015) doi:10.1038/nchem.2410.
http://www.nature.com/nchem/journal/v8/n2/full/nchem.2410.html.
See also "Molecular arm grabs, transports, releases molecular cargo"
http://www.foresight.org/nanodot/?p=6937. *Credit:* Leigh Group, University of Manchester
http://www.catenane.net/pages/2015_molecular_transporter.html.
[permission to use granted by Prof. David A Leigh, FRS <david.leigh@manchester.ac.uk>
The University of Manchester by email 20161120.]

> *"A variety of molecular machines can be built from mechanically interlocked organic molecules."*

> *"Dynamic atomic force microscopy enables deterministic manipulation of single atoms on surfaces."*

One of the iconic achievements in the history of nanotechnology occurred in 1989, when a scanning tunneling microscope (STM) tip was used to arrange 35 xenon atoms on a nickel surface to form the IBM corporate logo (27). A number of varieties of scanning probe microscopes proved useful during the following decades for atomically precise construction on surfaces. Non-contact atomic force microscopy (NC-AFM) has demonstrated mechanical manipulation of atoms on semiconductor surfaces (28). Lateral force microscopy provided further improvements in resolution and manipulation (29).

FORESIGHT
INSTITUTE

AI for Scientific Progress: State of the Art

Over the past decade, various flavors of artificial intelligence have been applied to advance scientific understanding and technology development. Many of these have been in biology and medicine, but increasingly, we see application in areas relevant to achieving atomic precision in nanotechnology. It follows a brief description of relevant ongoing research projects by workshop participants that were discussed at the workshop and that inspired the projects put together by participants during the workshop.

AI to plan organic synthesis

One workshop component included the award of the 2016 Feynman Prizes in Theory, Experiment, and the Distinguished Student Award. The 2016 Award for Theory honored workshop participant Bartosz A. Grzybowski who, together with his colleagues, developed a project that signals an imminent improvement in the notoriously challenging, long-term attempt to automate planning synthetic routes for organic chemistry (30). In addition to automating the search for a synthetic route, the program described can be set to follow user-directed priorities, such as the shortest route, the routes using the least expensive reagents, the best-documented reactions, or stipulating that regulated reagents or particular solvents, etc. be avoided.

The paper is acknowledged to have done a promising job converting organic chemistry into a computationally tractable form by dealing with problems in terms of networks and graph theory (31). 'Chematica' not only greatly simplifies cobbling a synthetic route together from reactions reported in the literature, but also deals with how to synthesize a molecule that has never been synthesized before, in structural classes that have not been explored. These are handled by a chemical expert system that explicitly embodies a huge number of rules and the various details that specify the context in which the reaction works or doesn't work. These rules are then used to propose synthesis of new molecules. The various scoring functions that are used to evaluate alternatives at each step of the synthesis are also described.

In discussing the project, a science commentator imagines, "how about taking this new software … and asking it for routes that prioritize the starting MIDA boronates and couplings that the Burke synthesis machine (31) is so good at working with?" [move ref to end of quote]

"Narrow AI can now plan synthetic routes for organic chemistry, including molecules never before synthesized."

FORESIGHT
INSTITUTE

"Computer-Assisted Synthetic Planning" first focuses on the applications of network algorithms to search the universe of known reactions, and introduces scoring functions to evaluate hundreds of millions of synthetic possibilities per second, identifying synthetic routes optimized for various user-specified criteria. Secondly, to deal with completely de novo synthetic planning, the review considers the intellectual connections of synthetic planning with other fields, like chess, and identifies three key differences: (1) the number of rules is much larger (tens of thousands versus order-of-ten for chess), (2) the applicability of any given rule depends on the context, that is, the presence of other chemical groups in the same molecule, and (3) in contrast to chess, there is no well-defined criterion to evaluate "synthetic position" and plan future moves. The third section considers remaining challenges and opportunities for further research.

A set of benchmark problems in planning organic synthesis can be formulated as a planning problem in the Planning Domain Definition Language (PDDL), developed in an attempt to stande AI planning languages. It is demonstrated that only a fraction of the benchmark problems can be encoded using known methods. Organic synthesis is thus proposed as a new challenge for AI planning (32).

Machine Learning for quantum chemistry

The task of explicitly solving the Schrödinger equation is circumvented for a diverse class of small organic molecules by developing a non-linear regression machine learning model for computing molecular atomization energies in chemical compound space. The model is based on a measure of distance in compound space that accounts for stoichiometry and configuration. The results of test calculations show the method to be an order of magnitude more accurate than counting bonds or semi-empirical quantum chemistry (33). The senior author of this paper was a workshop participant.

Predicting new crystal structures in materials research using empirical rules obtained by data mining large amounts of experimental data has not been very effective. However, transferring the concept of heuristic rule extraction to a large library of ab initio calculated results has been demonstrated to be an effective tool for predicting new crystal structures (34). By data-mining density functional quantum mechanical calculations, correlations were found among ab initio calculated energies for 114 different crystal structures in each of 55 binary metal alloys. These correlations were ex-

> *"Machine learning provides an alternative to explicitly solving the Schrödinger equation for a diverse class of small organic molecules."*

tensions of heuristics previously established on the basis of large amounts of experimental structure information. The number of calculations required to obtain a given level of accuracy of crystal structure prediction increased less than linearly with the increase of structures in the library, indicating that data mining of ab initio calculations has promise for practical prediction of crystal structure.

The continued advancement of computational materials science to accurately evaluate electronic and thermodynamic properties of a wide-range of materials requires large open databases. The usefulness of such databases depends upon exact specification of complex quantum mechanical codes, including many details that are seldom reported in publications. The Automatic-Flow (AFLOW) standard for the high-throughput construction of materials science electronic structure databases has been developed to provide standard parameter values for high-throughput computational materials discovery. Parameters include k-point grid density, basis set plane wave kinetic energy cut-off, exchange–correlation functionals, pseudopotentials, DFT+U parameters, and convergence criteria (35).

Machine learning has also been used to facilitate the search for novel thermoelectric materials by bridging the gap between experimental investigations and first principles DFT calculations. Experimental searches have largely been confined to a small set of chemical and structural families. Predicting properties from first principles for a wide range of structures is feasible but very challenging, and has not been used to guide synthesis. An open machine learning recommendations engine based on screening 25,000 known materials was developed and made available on the web. The training set was obtained from a large experimental materials database plus electronic structure data derived from first principles. One limitation is that the volume of high quality training datasets in materials research is much smaller than in the biological sciences, where bioinformatics has been extensively developed. The choice of the descriptor set (low level materials characteristics like crystal structure and chemical composition) for this work proved important. Finally, the recommendation engine was built using the "random forest" algorithm, which uses a collection of weak learners to collectively model complex nonlinear behavior. A material that the recommendation engine scored highly, that was chemically very different from known thermoelectrics, and that was easy to synthesize, was tested and found to exemplify the discovery of a new thermoelectric class deserving of further optimization and further study (36). The authors suggest that this "paradigm could eventually replace trial-and-error and fortuity in the search for new materials across a wide variety of application areas."

Metal-organic frameworks (MOFs) are nanoporous, crystalline materials comprising metal or metal-oxide nodes connected by variously functionalized organic linkers. Reticular chemistry, the rational design of MOFs from molecular building blocks of various different topologies, has been used to design vastly diverse structures, with chemical and geometrical properties optimized for diverse applications. Despite these successes, a noticeable limitation was the lack of a specific topology called 1-D rod topology, in which the metal-oxide secondary building units extend as a rod. MOFs with this topological feature had extremely interesting properties, "including remarkable CO_2 capture in MOF-74", but reticular chemistry studies on 1-D rod MOF's had been largely limited to experimental studies focusing on analogs of MOF-74. The reticular chemistry study reported in (37) combines the creation of MOF-74 analogs, which previously had only been studied experimentally, with a computational method to automate generation of hypothetical analogs of MOFs exhibiting a 1-D rod topology. The approach taken was to vary the chemical composition of the MOF-74 ligand to alter its geometric and chemical properties. MOF-74 requires a novel building algorithm since it has complex connectivity between linkers and 1-D secondary building unit rods. Only 61 of the 60,000,000 chemical species in the database used were identified as feasible substitutes for the MOF-74 linker. One in the set [olsalazine, or 3,3'-azo-bis(6-hydroxybenzoate) salicylic acid] was commercially available and was used to synthesize a MOF-74 analog. This olsalazine analog of MOF-74 was prepared and found to have the predicted crystal structure and CO_2 adsorption behavior, thus validating the new, systematic method for identifying MOFs exhibiting a 1D rod topology. The other 60 analogs identified that could be synthesized exhibit a wide range of surface areas with varied roughness and functionality that could be suited for other separations involving molecules larger than CO_2.

Information acquisition from dynamic force spectroscopy

Most of the tip-sample interactions in scanning probe microscopy are difficult to understand and depend on many parameters. Further, the use of multimodal SPM imaging in recent years, providing multiple channels of information simultaneously, results in huge increases in the amount of information requiring analysis. While the last three decades have seen substantial advances in low noise microscope platforms and improved probes, little attention has been given to the data acquisition process.

One of the most useful modes of SPM for mechanical manipulation of atoms on a surface has turned out to be non-contact atomic force microscopy (nc-AFM), also known as dynamic force microscopy. A specific challenge for information processing

in the case of nc-AFM is the mismatch between pixel acquisition at millisecond time scales and the sub-microsecond time scale of cantilever oscillations, resulting in severe compression of multidimensional dynamic information of a vibrating cantilever to only a few measured parameters. Various non-linearities in the vibrations contain detailed information about sample properties that cannot be ignored or compressed. Multivariate statistical methods based on information theory have been used to acquire and analyze the entire response of the cantilever (38). Only statistically relevant components of the cantilever trajectory, and hence materials functionality, are stored. Principal component analysis establishes dominant behavior types so that correlation functions can identify underlying sources of observed behaviors. This strategy provides information on how the data is structured in space, frequency, and information domains. Thus, data can be explored without information loss or imposed bias.

Progress in dynamic and functional imaging leads to multidimensional data sets containing information on physical and chemical functions along the probe trajectory. The infrastructure for analyzing such data has been developed in the context of medical and satellite imaging. Clustering, unsupervised learning techniques, supervised neural network-based classification, and deep data analysis of physically relevant multivariate statistics data can be used to establish statistically significant correlations in such data sets (39).

Appendix to Project 1

A Hypothetical AI-backed CAD for nanotechnology

Volunteer contribution to project 1 by Jazear Brooks

Summary

In an effort to create an AI-backed computer-aided design (CAD) for nanotechnology, researchers can use either MDT or CaDNAno as the basis for the program --- both are free and open source. MDT has built-in analytical tools, while CaDNAno uses CanDo's finite element analysis. Given that genetic algorithms are limited in their abilities to produce optimal solutions, researchers will likely choose to use a customized, blended AI algorithm. The PDB file format can be used as the basis for generating shareable files containing the following information for each atom: serial number, name, residue sequence number, residue name, chain identifier, and Cartesian coordinates.

CAD Software for Nanotechnology

Much of my research into CAD programs for nanotech designs revealed quite dated software that is no longer generating active participation or recent updates. Perhaps unsurprisingly, Autodesk, makers of AutoCAD, proved to have the strongest offerings. Relevant Autodesk products include the company's Molecular Design Toolkit (MDT), which "enables design at the atomic level with a suite of open source molecular modeling tools and seamless access to cloud computing". As a free and open-source package that includes integrations of other open-source software, MDT offers a mature and well-supported basis for this application.

MDT's analytic functions include both structural and calculation data for atoms, bond topology, primary structures, position, momenta, calculated properties such as potential energy or dipole moments, molecular orbitals, and even geometric analysis methods for analyzing time evolution of angles. While there is extensive documentation on MDT and its various APIs, I was unable to find any granular information on just how the software performs its analyses of macro properties.

For designing three-dimensional nanostructures via DNA origami, CaDNAno is another free and open-source option with integrated analysis functions provided by CanDo and 3D interface functions using Autodesk's Maya software, via the Molecular Maya Toolkit (mMaya).

CaDNAno's integration with CanDo allows the software to perform finite element analysis to provide relative certainty of structures' stability. This analysis offers designers' computational predictions of both 3D shapes and the flexibility of single- and multi-layer structures. According to CanDo, "The thermally induced fluctuations of DNA nanostructures are computed using the equipartition theorem of statistical mechanics and normal mode analysis … Atomic models of DNA nanostructures are generated from 3D solution shapes and thermal fluctuations." If someone would like to view the source code for generating the finite element models, it is available upon request by emailing Danial Dardani at ddardani@mit.edu.

CanDo is a free resource developed by MIT's Laboratory for Computational Biology and Biophysics, and can be used separately from CaDNAa,no if preferred. Cando accepts input from CaDNAdo or Tiamat files.

The specific types of macro analysis that will be most useful to a researcher will, of course, depend on the specific applications of their designs.

AI Algorithms

I reviewed various types of AI algorithms in accordance with the criteria that the algorithm be able to take into account nanosystem requirements and materials and then find the optimal set of actions to take to generate a system. Since we mentioned genetic algorithms specifically at the conference, I researched on their merits and disadvantages in comparison to other types of algorithms. My research revealed that the weaknesses of genetic algorithms are that (1) they won't necessarily find optimal solutions, though they are very good at finding good solutions, and (2) the solutions they provide may vary each time a researcher runs them.

While there are many cheat sheets available to help a researcher choose which algorithm to use, the ultimate answer, as Microsoft says, is "it depends", and "even the most experienced data scientists can't tell which algorithm will perform best before trying them". All of these factors go into the decision-making process: size of the data, quality of the data, nature of the data, what the researcher wants to do with the answer, how the math of the algorithm was translated, and how much time the researcher has. In practice, scientists would need to blend different types of algorithms together in order to achieve a better success rate when finding optimal solutions, such as this example of blending self-organization with a genetic algorithm in order to optimize protein structure predictions.

Assuming that a researcher intends to use genetic algorithms, at least in part, for their project, we can assume that the AI will make errors in which it finds sub-optimal solutions. For the best results, humans will need to customize both the algorithm itself, and provide customized analysis of the algorithm's outputs in order to generate better inputs in an effort to increase the likelihood of the AI finding optimal solutions.

Shareable Data Formats

Practitioners will likely need to use many data formats across the different aspects of the project. For instance, they will likely need to store components and parts in a database format of their choice. If they decide to use CaDNAno, they will manage both .json (design file format) .csv (sequence file format) files, or use a tool to convert these files into the .pdb format. PDB can store the following for each atom: serial number, name, residue sequence number, residue name, chain identifier, and Cartesian coordinates.

Autodesk's Molecular Design Toolkit is a Python library, and many of its functions require a Docker server to run. You can use program database [or is it protein data-bank?](.pdb) assemblies and DNA sequences in Microsoft Deployment Tookit (MDT), as well and read and write structured data file format (.sdf), chemical file format (.xyz) , and crystallographic information file formation (.mcif)files.PDB will likely be a good basis for deriving a researcher's specific format needs. Full documentation on this format is freely available.

References

1. "Artificial intelligence in nanotechnology" G M Sacha and P Varona. *Nanotechnology* **24** 452002 (11 October 2013)
   http://dx.doi.org/10.1088/0957-4484/24/45/452002
2. *Artificial Intelligence: A Modern Approach*. Stuart J. Russell and Peter Norvig. First edition 1995. ISBN 0-13-103805-2. Prentice-Hall, Inc. Englewood Cliffs, New Jersey.
3. Perceptrons (book)
   https://en.wikipedia.org/wiki/Perceptrons_(book). (accessed Nov. 3, 2016)
4. Deep Blue versus Garry Kasparov
   https://en.wikipedia.org/wiki/Deep_Blue_versus_Garry_Kasparov (accessed Nov. 4, 2016)
5. Watson (computer)
   https://en.wikipedia.org/wiki/Watson_(computer) (accessed Nov. 4, 2016)
6. AlphaGo versus Lee Sedol
   https://en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol (accessed Nov. 4, 2016)
7. Hanson Robotics
   http://www.hansonrobotics.com/ (accessed Nov. 5, 2016)
8. "Making computers explain themselves" MIT press release (Oct. 27, 2016)
   http://www.csail.mit.edu/making_computers_explain_themselves.
   "Rationalizing Neural Predictions" (Conference paper)
   https://people.csail.mit.edu/taolei/papers/emnlp16_rationale.pdf

FORESIGHT INSTITUTE

9. "Molecular Mechanical Computing Systems" Ralph C. Merkle, Robert A. Freitas Jr., Tad Hogg, et al. IMM Report No. 46, April 2016;

   http://www.imm.org/Reports/rep046.pdf

10. "There's Plenty of Room at the Bottom" Richard P. Feynman.

    http://www.zyvex.com/nanotech/feynman.html (accessed Nov. 5, 2016)

11. "Molecular engineering: An approach to the development of general capabilities for molecular manipulation" K. Eric Drexler. *Proc. Natl. Acad. Sci. USA* **78(9):** 5275-5278, September 1981.

    http://www.imm.org/publications/pnas/ (accessed Nov. 5, 2016)

12. ""Design of a Novel Globular Protein Fold with Atomic-Level Accuracy" Brian Kuhlman, Gautam Dantas, Gregory C. Ireton, Gabriele Varani, Barry L. Stoddard, David Baker. *Science* **302**, 13641368- (21 Nov. 2003) DOI: 10.1126/science.1089427

13. For example:

    http://www.foresight.org/nanodot/index.php?s=%22protein+design%22

14. "Exploring the repeat protein universe through computational protein design" TJ Brunette, Fabio Parmeggiani, Po-Ssu Huang, Gira Bhabha *et al. Nature* **528**, 580–584 (24 December 2015) doi:10.1038/nature16162

    http://www.nature.com/nature/journal/v528/n7583/full/nature16162.html.
    See also "De novo protein design space extends far beyond biology"
    http://www.foresight.org/nanodot/?p=6973

15. "Acceleration of an Aromatic Claisen Rearrangement via a Designed Spiroligozyme Catalyst that Mimics the Ketosteroid Isomerase Catalytic Dyad" Matthew F. L. Parker, Sílvia Osuna, Guillaume Bollot, *et al. J. Am. Chem. Soc.* **136**, 3817-3827 (January 23, 2014). doi:10.1021/ja409214c

    http://pubs.acs.org/doi/abs/10.1021/ja409214c

16. "1995 Feynman Prize in Nanotechnology Awarded For Pioneering Synthesis of 3-D DNA Objects" Foresight Update 23, 30 November 1995.

    http://foresight.org/Updates/Update23/Update23.1.html#anchor415574

17. "Folding DNA to create nanoscale shapes and patterns" Paul W. K. Rothemund. *Nature* **440**, 297-302 (16 March 2006) | doi:10.1038/nature04586

    http://www.nature.com/nature/journal/v440/n7082/abs/nature04586.html

18. See, for example these blog posts:

    http://www.foresight.org/nanodot/index.php?s=%22DNA+nanotechnology%22;
    http://www.foresight.org/nanodot/index.php?s=%22Scaffolded+DNA+origami%22;
    http://www.foresight.org/nanodot/index.php?s=breadboard

19. "Programmable motion of DNA origami mechanisms" Alexander E. Marras, Lifeng Zhou, Hai-Jun Su, and Carlos E. Castro. *Proc. Nat. Acad. Sci. USA* **112**: 713–718 (20 January 2015), doi: 10.1073/pnas.1408869112

    http://www.pnas.org/content/112/3/713.abstract.

FORESIGHT INSTITUTE

See also "Structural DNA nanotechnology with programmed motions"
http://www.foresight.org/nanodot/?p=6430.

20. http://www.foresight.org/nanodot/index.php?s=%22RNA+nanotechnology%22

21. "Chemical fuel keeps molecular motor moving"
http://www.foresight.org/nanodot/?p=7148

22. "Single-molecule light-driven nanosubmarine"
http://www.foresight.org/nanodot/?p=6884

23. "One-directional rotation in a new artificial molecular motor"
http://www.foresight.org/nanodot/?p=6856

24. "Addressable molecular machines arranged in a porous crystal"
http://www.foresight.org/nanodot/?p=6765

25. "Single molecule pump concentrates small molecules"
http://www.foresight.org/nanodot/?p=6594

26. "Single molecule nanocar with functional wheels driven by electron tunneling"
http://www.foresight.org/nanodot/?p=4851 "Cool new nanocars"
http://www.foresight.org/nanodot/?p=2952 "Nanocar gets nanomotor via US/Dutch collaboration"
http://www.foresight.org/nanodot/?p=2213

27. "Positioning single atoms with a scanning tunnelling microscope" D. M. Eigler & E. K. Schweizer. Nature 344: 524 - 526 (05 April 1990); doi:10.1038/344524a0
http://www.nature.com/nature/journal/v344/n6266/abs/344524a0.html

28. "Mechanical Vertical Manipulation of Selected Single Atoms by Soft Nanoindentation Using Near Contact Atomic Force Microscopy" Noriaki Oyabu, Oscar Custance, Insook Yi, Yasuhiro Sugawara, and Seizo Morita. Phys. Rev. Lett. 90: 176102 (2003)
http://dx.doi.org/10.1103/PhysRevLett.90.176102

29. "Quantifying Molecular Stiffness and Interaction with Lateral Force Microscopy" Alfred John Weymouth, Thomas Hofmann, Franz J. Giessibl. Science 343:1120-1122 (07 Mar 2014)
http://science.sciencemag.org/content/343/6175/1120 (accessed Nov. 6, 2016)

30. "The Algorithms Are Coming" Derek Lowe. April 12, 2016. In the Pipeline, an editorially independent blog published by *Science Translational Medicine*.
http://blogs.sciencemag.org/pipeline/archives/2016/04/12/the-algorithms-are-coming. The above blog discusses this paper, which introduces the software Grzybowski and his colleagues call "Chematica": "Computer-Assisted Synthetic Planning: The End of the Beginning" Sara Szymkuć, Ewa P. Gajewska, Tomasz Klucznik *et al. Angewandte Chemie* (international ed.) **55**: 5904-5937. 8 April 2016. DOI: 10.1002/anie.201506101
http://onlinelibrary.wiley.com/doi/10.1002/anie.201506101/full.

31. a) http://blogs.sciencemag.org/pipeline/archives/2016/04/12/the-algorithms-are-

coming, and b) "Automated synthesis expands nanotechnology building block reper-
toire" http://www.foresight.org/nanodot/?p=6498

32. "Exploring Organic Synthesis with State-of-the-Art Planning Techniques" Rami Mat-
loob and Mikhail Soutchanski. Proceedings of Scheduling and Planning Applica-
tions woRKshop (SPARK), pages 52-61, held on June 13, 2016 in conjunction with the
26th International Conference on Automated Planning and Scheduling, London, UK.
http://www.cs.ryerson.ca/~mes/publications/MatloobSoutchanskiExploringOrgan-
icSynthesisWithState-of-the-ArtPlanning_SPARK2016.pdf

33. "Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learn-
ing" Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, O. Anatole von Lil-
ienfeld. *Phys. Rev. Lett.* **108**, 058301 (31 January 2012). DOI:
https://doi.org/10.1103/PhysRevLett.108.058301.
Preprint: https://arxiv.org/abs/1109.2618

34. "Predicting Crystal Structures with Data Mining of Quantum Calculations" Stefano
Curtarolo, Dane Morgan, Kristin Persson, John Rodgers, Gerbrand Ceder. *Phys. Rev.
Lett.* **91**, 135503 (September 2003).
https://doi.org/10.1103/PhysRevLett.91.135503

35. "The AFLOW standard for high-throughput materials science calculations" Camilo E.
Calderon, Jose J. Plata, Cormac Toher, et al. Computational Materials Science 108:
233–238 (14 August 2015).
http://dx.doi.org/10.1016/j.commatsci.2015.07.019

36. "Perspective: Web-based machine learning models for real-time screening of ther-
moelectric materials properties" Michael W. Gaultois, Anton O. Oliynyk, Arthur Mar,
Taylor D. Sparks, Gregory J. Mulholland, and Bryce Meredig. *APL Mater.* **4**, 053213 (27
May 2016); doi: 10.1063/1.4952607
http://dx.doi.org/10.1063/1.4952607

37. "*In silico* design and screening of hypothetical MOF-74 analogs and their experimental
synthesis" Matthew Witman, Sanliang Ling, Samantha Anderson, *et al. Chem. Sci.* **7**,
6263-6272 (21 Jun 2016) DOI: 10.1039/C6SC01477A
http://pubs.rsc.org/is/content/articlelanding/2016/sc/c6sc01477a#!divAbstract

38. "Complete information acquisition in dynamic force microscopy" Alexei Belianinov,
Sergei V. Kalinin & Stephen Jesse. *Nature Communications* **6**: 6550 (13 Mar 2015) DOI:
10.1038/ncomms7550.
http://www.nature.com/articles/ncomms7550

39. "Big data and deep data in scanning and electron microscopies: deriving functionality
from multidimensional data sets" Alex Belianinov, Rama Vasudevan, Evgheni Strel-
cov, Anton Ievlev, Chad Steed *et al. Advanced Structural and Chemical Imaging* **1:6** (13
May 2015) DOI 10.1186/s40679-015-0006-6
http://link.springer.com/article/10.1186/s40679-015-0006-6