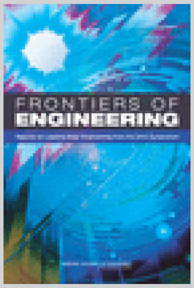


Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2012 Symposium

National Academy of Engineering

ISBN
978-0-309-31281-3

182 pages
6 x 9
PAPERBACK (2013)

FRONTIERS OF ENGINEERING

Reports on Leading-Edge Engineering from the 2012 Symposium

NATIONAL ACADEMY OF ENGINEERING
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS • 500 Fifth Street, NW • Washington, DC 20001

NOTICE: This publication has been reviewed according to procedures approved by a National Academy of Engineering report review process. Publication of signed work signifies that it is judged a competent and useful contribution worthy of public consideration, but it does not imply endorsement of conclusions or recommendations by the NAE. The interpretations and conclusions in such publications are those of the authors and do not purport to represent the views of the council, officers, or staff of the National Academy of Engineering.

Funding for the activity that led to this publication was provided by General Motors, The Grainger Foundation, Microsoft Research, Defense Advanced Research Projects Agency, Air Force Office of Scientific Research, Department of Defense ASD(R&E)–Research Directorate-STEM Development Office, and Cummins Inc. This material is also based upon work supported by the National Science Foundation under Grant No.1202935. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. In addition, the content of this publication does not necessarily reflect the position or the policy of the Government and no official endorsement should be inferred.

International Standard Book Number-13: 978-0-309-31281-3
International Standard Book Number-10: 0-309-31281-7

Additional copies of this report are available for sale from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

ORGANIZING COMMITTEE

KRISTI ANSETH (*Chair*), Howard Hughes Medical Institute Investigator and Distinguished Professor of Chemical and Biological Engineering, University of Colorado, Boulder
KAREN BURG, Hunter Endowed Chair and Professor, Department of Bioengineering, Clemson University
LI-TE CHENG, Software Engineer, Google
MICHAEL DEGNER, Senior Technical Leader, Electric Machine Drives, Electrification Research and Advanced Engineering, Ford Motor Company
ALI KHADEMHOSSEINI, Associate Professor, Harvard Medical School
SANJEEV NAIK, Engineering Group Manager, Advanced Engine Controls, General Motors
BEN SAWYER, Co-founder, Digitalmill
DAVID SHOLL, Michael Tennenbaum Family Chair and GRA Eminent Scholar, School of Chemical and Biomolecular Engineering, Georgia Institute of Technology
ARMIN SOROOSHIAN, Assistant Professor, Department of Chemical and Environmental Engineering, University of Arizona

Staff

JANET R. HUNZIKER, Senior Program Officer
VANESSA LESTER, Program Associate

Preface

This volume highlights the papers presented at the National Academy of Engineering's 2012 U.S. Frontiers of Engineering Symposium. Every year, the symposium brings together 100 outstanding young leaders in engineering to share their cutting-edge research and technical work. The 2012 symposium was held September 13-15, and hosted by General Motors at the GM Technical Center in Warren, Michigan. Speakers were asked to prepare extended summaries of their presentations, which are reprinted here. The intent of this book is to convey the excitement of this unique meeting and to highlight cutting-edge developments in engineering research and technical work.

GOALS OF THE FRONTIERS OF ENGINEERING PROGRAM

The practice of engineering is continually changing. Engineers today must be able not only to thrive in an environment of rapid technological change and globalization, but also to work on interdisciplinary teams. Cutting-edge research is being done at the intersections of engineering disciplines, and successful researchers and practitioners must be aware of developments and challenges in areas that may not be familiar to them.

At the 2½-day U.S. Frontiers of Engineering Symposium, 100 of this country's best and brightest engineers, ages 30 to 45, have an opportunity to learn from their peers about pioneering work being done in many areas of engineering. The symposium gives early career engineers from a variety of institutions in academia, industry, and government, and from many different engineering disciplines, an opportunity to make contacts with and learn from individuals they would not meet in the usual round of professional meetings. This networking may lead to collaborative work and facilitate the transfer of new techniques and approaches. It is hoped that the exchange of information on current developments in many fields of engineering will lead to insights that may be applicable in specific disciplines and thereby build U.S. innovative capacity.

The number of participants at each meeting is limited to 100 to maximize opportunities for interactions and exchanges among the attendees, who are chosen through a competitive nomination and selection process. The topics and speakers for each meeting are selected by an organizing committee of engineers in the same 30- to 45-year-old cohort as the participants. Different topics are covered each year, and, with a few exceptions, different individuals participate.

Speakers describe the challenges they face and communicate the excitement of their work to a technically sophisticated but non-specialized audience. Each speaker provides a brief overview of his/her field of inquiry; defines the frontiers of that field; describes experiments, prototypes, and design studies that have been completed or are in progress, as well as new tools and methodologies, and limitations and controversies; and summarizes the long-term significance of his/her work.

THE 2012 SYMPOSIUM

The four general topics covered at the 2012 meeting were climate engineering, vehicle electrification, serious games, and engineering materials for the biological interface. The climate engineering session described how artificially modifying Earth's systems could combat changes in the planet's radiative balance caused by human activities. The first speaker provided an overview of climate engineering and the considerations before such an intervention is made. This was followed by presentations on removing carbon dioxide from the atmosphere through mechanical or natural means; the role that atmospheric aerosols play in climate engineering and recent field projects on the basic science and physics of cloud brightening; and methods of climate engineering, in particular, potential effects of simulated volcanic eruptions.

As described in the second session, global warming, sustainability, and national security concerns are driving investment in vehicle electrification. Research in this area is focused on technology enablers such as energy storage systems, electric machine drives, and electrical system integration and control. Speakers covered recent improvements in automobile electrical energy storage systems where reducing the cost, size, and weight are key challenges; research in improved magnetic materials used in electric machine drives, including reduction of critical materials like rare earth elements; the impact of vehicle electrification on electrical transmission and distribution systems; and technical approaches to enhancing vehicle safety.

The term "serious games" describes the application of video game technologies into non-entertainment domains, and it is a medium of many design, engineering, and technical fields. Initially, serious games focused on training, with a second wave of applications focused on therapeutic and health behavior change efforts. The current focus of serious games is innovative crowd sourcing activities that tackle scientific, organizational, and social challenges through video game play. This session began with an overview of the serious games space from a national policy and educational standpoint. This was followed by talks on moving innovative game technology from the lab to the living room, how serious science is being achieved with serious games such as Foldit, and the use of serious games as tools for process optimization and complex problem-solving.

The symposium concluded with a session that focused on the cell-cell or cell-tissue components of the biological interface of, for example, tendon to bone or cartilage to bone. In order to simulate these complex interactions, researchers focus on design of materials, control of cells, and design of bioreactors in which to grow and assess these systems. The three talks in this session described engineering tissue-to-tissue interfaces for the formation of complex tissues; identification and modulation of biophysical signals that control stem cell function and fate; and cultivating 3D tissue systems that better model human biology for drug discovery, personalized medicine, and tissue engineering.

In addition to the plenary sessions, the participants had many opportunities to engage in informal interactions. On the first afternoon of the meeting, participants broke into small groups for “get-acquainted” sessions during which individuals presented short descriptions of their work and answered questions from their colleagues. This helped attendees get to know more about each other relatively early in the program. On the second afternoon, General Motors hosted a “Ride-n-Drive” event where attendees could drive or ride in advanced vehicles such as the plug-in hybrid Volt and concept cars such as the two-seat EN-V that can operate autonomously.

Every year, a distinguished engineer addresses the participants at dinner on the first evening of the symposium. The speaker this year was Dr. Alan I. Taub, professor of materials science and engineering at the University of Michigan and former vice president of global R&D at General Motors, who gave a talk on the reinvention of the automobile for 21st century sustainability.

NAE is deeply grateful to the following organizations for their support of the 2012 U.S. Frontiers of Engineering Symposium:

- General Motors
- The Grainger Foundation
- Defense Advanced Research Projects Agency
- Air Force Office of Scientific Research
- Department of Defense ASD(R&E)-Research
- National Science Foundation. This material is based upon work supported by the National Science Foundation under grant number EFRI-1202935
- Microsoft Research
- Cummins Inc.

NAE would also like to thank the members of the Symposium Organizing Committee (p. iv), chaired by Dr. Kristi Anseth, for planning and organizing the event.

Contents

CLIMATE ENGINEERING

VEHICLE ELECTRIFICATION

SERIOUS GAMES

ENGINEERING MATERIALS FOR THE BIOLOGICAL INTERFACE

APPENDIXES

CLIMATE ENGINEERING

Climate Engineering

David S. Sholl
Georgia Institute of Technology

Armin Sorooshian
University of Arizona

Earth's energy balance is sensitive to the actions of humans in ways that are unprecedented in human history. One clear indicator of anthropogenic activity is the increase of carbon dioxide (CO_2) concentrations from 280 parts per million (ppm) before the Industrial Revolution to approximately 385 ppm today. Climate engineering is the concept of proactively and artificially modifying the earth system in ways that will combat human-induced changes in the planet's radiative balance. Measures that have been considered include modification of the atmosphere's reflectivity through the creation of aerosols, cloud brightening, and large-scale carbon sequestration through targeted changes in ocean chemistry and biology. These methods may have side effects and are characterized by a host of moral issues. This session explores the issues surrounding the impact of humans on the climate system, methods that might combat such perturbations, and the moral and legal issues associated with climate engineering.

Eli Kintisch (MIT and *Science* Magazine) opens with an overview of climate engineering and of considerations necessary before such an intervention is contemplated. Christopher Jones (Georgia Institute of Technology) then discusses aspects of removing CO_2 from the atmosphere through mechanical or natural means. Lynn Russell (Scripps Institute of Oceanography) describes the role of atmospheric aerosols in climate engineering and gives examples of how recent field projects have enhanced understanding of the basic science and physics of cloud brightening. Finally, Ben Kravitz (Pacific Northwest National Laboratory) discusses methods of climate engineering with an emphasis on the potential effects of simulated volcanic eruptions.

Together, these speakers present the state of the art in climate engineering knowledge based on modeling, experimental work, and social science considerations.

Overview of Climate Engineering

Eli Kintisch
Science Magazine
Massachusetts Institute of Technology

Top science institutions around the world, including the US National Academies and the UK Royal Society, have called for studies into deliberate tinkering with the planet's climate or atmosphere to partially offset global warming, a practice known as climate engineering or geoengineering. Various characteristics distinguish the two major types of geoengineering: solar radiation management (SRM; e.g., orbiting sunshades, aerosols sprayed into the stratosphere) and carbon dioxide removal (CDR; e.g., carbon-sucking machines, catalysis of oceanic algal growth). The number of scientists studying both is steadily increasing, and several companies are conducting CDR engineering research, but the United States has yet to follow the lead of a number of European countries that have dedicated programs for geoengineering research.

Efforts at global carbon dioxide pollution abatement remain stalled even as the effects of a warming planet become increasingly apparent. Research findings suggest that the planet may be closer to global tipping points, such as the release of methane from permafrost, than previously thought. As the global climate crisis intensifies, taboos once held by scientists and policymakers are falling by the wayside. Adaptation, the organized response to a warming planet and its myriad local impacts, was once viewed by top officials as a distraction from the main priority of mitigating global greenhouse gas emissions. Now local and national governments around the world are creating plans to respond and adapt to warmer temperatures, higher seas, more pervasive drought, and other environmental challenges.

Geoengineering is a radical form of adaptation. The publication in 2006 of a controversial paper by Nobel Prize winner Paul Crutzen titled "Albedo Enhancement by Stratospheric Sulfur Injections: A Contribution to Resolve a Policy

Dilemma?" both jumpstarted the discussion of geoengineering and lent credibility to an idea that had until then existed largely in the shadows of academia.

Every serious researcher or policy expert who studies climate engineering, including Crutzen, believes that cutting greenhouse gas emissions is at least as important as developing geoengineering technologies, if not more urgent.

It is useful to consider abatement, carbon dioxide removal, and solar radiation management in proper context with one another. In Figure 1, each large circular element represents a process that drives the next step in the chain. The central items are interventions that mitigate the impact between two linked terms; for example, efficiency lowers the consumption of energy that results from consumption of goods and services. Three abatement steps—using less energy ("conservation"), using energy more efficiently ("efficiency"), and producing energy less carbon-

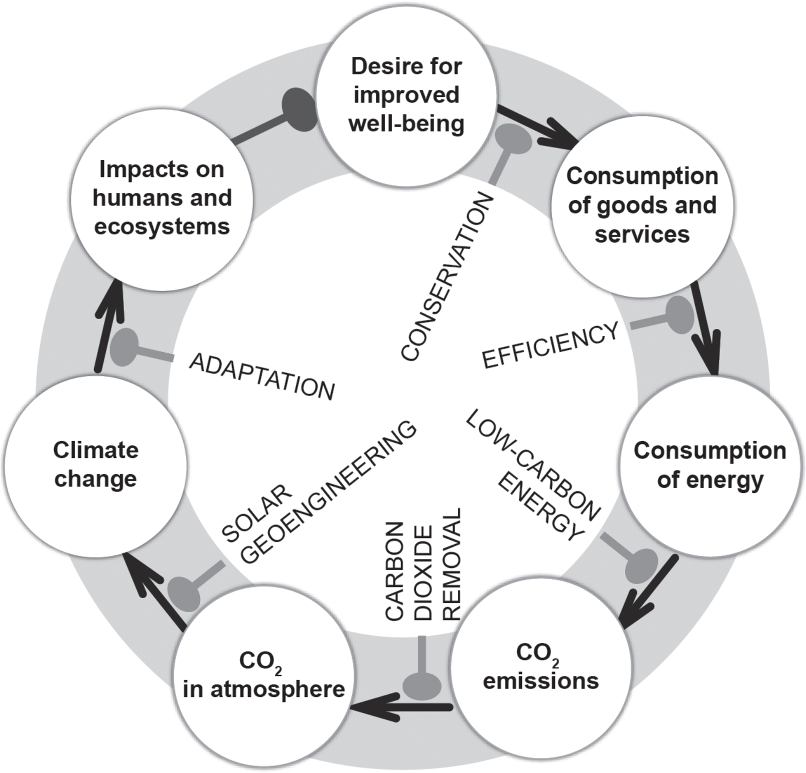


FIGURE 1 Connections among various factors in terms of the climate challenge. See text for discussion. The round-tipped "arrow" between the "impacts on human systems" and "desire for improved well-being" indicates that the former drives the latter. Reprinted with permission from Caldeira et al., 2013.

intensively ("low-carbon energy")—can together lower global greenhouse gas emissions. Next are geoengineering options. The round-tipped "arrows" indicate a more tenuous relationship than the other links.

As shown in Figure 1, CDR goes a step further than abatement. By pulling gases out of the atmosphere it gets at the heart of the problem: if lowering emissions is akin to reducing one's exposure to a virus that causes a fever, CDR is like using an antiviral medication. SRM is one step further still. It does not change the level of CO_2 in the atmosphere but instead serves to reduce its climatic effects. Temperature is the most prominent of those effects, and SRM lowers the planet's thermostat by directly reducing the amount of solar energy absorbed by the planet. Perhaps the metaphorical equivalent is using a cold compress to alleviate fever.

Both CDR and SRM techniques attempt to mimic natural processes that scientists mostly understand. But that is where their similarities end. In their technical aspects, the political dynamics that might govern their deployment, and their feasibility, the differences between them are stark. That's one reason that many scientists try to avoid using the terms "geoengineering" or "climate engineering" to generalize between the two.

PLANETARY SUNBLOCK: SOLAR RADIATION MANAGEMENT

"Fast, cheap, imperfect and uncertain" is how Harvard physicist David Keith (2011), one of the leading thinkers on both methods, describes SRM. The most commonly explored technique for blocking sunlight from the planet is to mimic the natural cooling effect of volcanoes by spreading sulfurous particles in the stratosphere. The following paragraphs explain Keith's characterization.

Fast: The 1991 eruption of the Mount Pinatubo volcano sprayed 5 million tons of sulfur aerosol into the stratosphere as sulfur dioxide, which scattered light away from Earth and cooled the planet by 0.5°C (Kravitz 2013). Modeling studies (e.g., Caldeira and Matthews 2007) suggest that if a similar quantity of sulfur aerosol were artificially injected into the stratosphere, the cooling could be essentially instantaneous.

Cheap: A recent study by an aerospace research firm suggests that the costs of deploying a global SRM scheme to offset anthropogenic warming "are comparable to the yearly operations of a small airline" (McClellan et al. 2010).

Imperfect: A number of modeling studies have suggested various side effects of this technique, including depriving the planet of solar energy that influences rainfall, leading to less precipitation (Ricke et al. 2010). This effect could disrupt the southeast Asian monsoon season or weather in South America, potentially exacerbating droughts.

Uncertain: Many aspects of the climate system are not fully understood, so tinkering with a fundamental variable that drives the system—the amount of solar energy entering it—may have serious unexpected or unintended consequences.

Since Crutzen's landmark paper, research into SRM has evolved from proof-of-concept modeling into more sophisticated efforts. The Geoengineering Model Intercomparison Project (GeoMIP) involves 19 different global climate models. Each has run separate simulations with four standardized scenarios in which solar radiation management is deployed in different ways (Kravitz et al. 2011). Because different climate models employ different assumptions, characteristics, and physics, use of the same initial conditions, the thinking goes, may yield more robust results about the environmental effects of various SRM strategies. One example of the increasingly sophisticated modeling research on stratospheric aerosols is a recent study that found that sulfate aerosols deployed to offset warming caused by a doubling of CO_2 concentrations would make the sky 3 to 5 times brighter—and less blue—than it is currently, which could affect photosynthesis in plants and people's psychological moods (Kravitz et al. 2012).

In the United States, David Keith and Harvard colleague James Anderson, an atmospheric chemist, are planning "to develop in situ experiments to test the risk and efficacy of aerosols in the stratosphere" (Keith 2012).

The most visible effort to explore stratospheric approaches through actual experimentation is the Stratospheric Particle Injection for Climate Engineering (SPICE) project, led by Bristol University and supported by the British government at £1.6 million for 3½ years. Along with ongoing work to design particles and computer modeling, the project originally included a planned field experiment to spray 150 liters of water 1,000 meters in the air to test how a balloon would behave in the wind during spraying, a feasibility test. The field experiment was cancelled because of public concern about lack of regulations on SRM as well as worries over a patent application that one of the research participants had filed before receiving UK funds for the project (Watson 2012).

THINNING THE GREENHOUSE LAYER: CARBON DIOXIDE REMOVAL

Scientists have proposed a variety of techniques for removing CO_2 from the atmosphere. These range from engineering forests to be more carbonaceous, to growing massive algal blooms at sea, to sucking carbon dioxide out of the atmosphere.

Few credible scientists believe CDR techniques to be a panacea. The approach has attracted somewhat less attention and different kinds of controversy than SRM, which Keith (2011) calls "slow, expensive and effective," as explained below.

Slow: Global yearly emissions of CO_2 are 34 million cubic metric tons, resulting in an accumulation of 500 billion tons of anthropogenic CO_2 in the atmosphere. Relying heavily on CDR as part of a climate response strategy means creating a massive industry—perhaps the biggest engineering project in human history—to steadily remove this mass of gas from the atmosphere one molecule at a time.

Expensive: A 2011 study by the American Physical Society concluded that collecting CO_2 directly from the atmosphere "is not currently" economically viable despite "optimistic" technical assumptions (APS 2011). It estimated that the basic cost of a system that could be built today would be about $600/ton, an order of magnitude more than the estimate for low-carbon energy sources.

Effective: CDR methods build off commercial techniques that work in submarines and space shuttles to clean air of CO_2 gas and promise fewer side effects than SRM methods.

A number of startups are focusing on different techniques for CDR. In 2007 Sir Richard Branson launched a $25 million contest called the Virgin Earth Challenge to encourage the development of technologies that "will result in the net removal of anthropogenic, atmospheric greenhouse gases each year for at least ten years without countervailing harmful effects."[1] The 11 contest finalists represent a decent survey of leading commercial entities in this area, including firms that propose to sequester carbon in biochar added to soil, to directly capture atmospheric CO_2 through chemical methods, or to burn biofuels and sequester the resultant CO_2 in the ground.

GEOENGINEERING RESEARCH POLICY AND PUBLIC OPINION

Several European governments have supported organized programs to support climate engineering research. The United States has none. Studies on the governance of climate engineering approaches are being conducted by a coalition co-led by the UK Royal Society (SRM Governance Initiative), an Oxford University group on a two-year grant (Climate Geoengineering Governance project), and the European Transdisciplinary Assessment of Climate Engineering project, led by the Institute for Advanced Sustainability Science in Potsdam, Germany.

Meanwhile, work on the ethics of climate engineering has yielded, among other things, the so-called "Oxford Principles," proposed to restrict research into SRM and CDR (Rayner et al. 2009). They include the following guidelines:

- That SRM be regulated as a public good
- That the public be involved in research related to SRM decisions, including field experiments
- That research plans and results be transparent and shared publicly
- That bodies independent of researchers studying climate engineering assess the environmental and socioeconomic impacts of research
- That decisions about deploying technology on a global scale be made only when "robust governance structures" to oversee such efforts are in place.

---

[1] Virgin Earth Challenge announcement; posted online at www.virgin.com/subsites/virginearth/ (accessed July 28, 2012).

A number of expert panels (e.g., Long et al. 2011) have urged the United States to create a dedicated research program in this area. But although the National Science Foundation has supported a handful of studies on SRM, and funds from various agencies have supported work applicable to CDR approaches, there is no integrated, organized effort in the federal government.

Several studies exploring public opinion on climate engineering technologies have been published. In August 2011 Cardiff University released results of a quantitative public engagement research project involving about 35 people that met for a day and a half. "Very few people were unconditionally positive about either the idea of geoengineering or the proposed [SPICE] field test. However, most were willing to entertain the notion that the test as a research opportunity should be pursued" (Parkhill and Pidgeon 2011).

An Internet poll of 3,105 American, Canadian, and British individuals published in 2011 found that 8% and 45% of respondents, respectively, correctly defined the interchangeable terms "geoengineering" and "climate engineering" (Mercer et al. 2011). In the same survey, respondents were asked to rate statements from 1, for "strongly disagree," to 4, for "strongly agree." For the statement "If scientists find that Solar Radiation Management can reduce the impacts of global warming with minimal side effects, then I would support its use," the average response was 3.01. The statement "Solar Radiation Management will help the planet more than it will hurt it" received an average response of 2.49. The results suggest that geoengineering could be viewed favorably by the public.

CONCLUSION

As the world's population contends with the challenge of climate change, respected scientists will continue studying climate engineering as part of a suite of responses—the most important of which is the immediate curtailing of greenhouse gas emissions. For policymakers and researchers in this area, the following considerations will have to be taken into account: the need to address risks inherent to the two types of climate engineering through research despite a lack of dedicated funding for such work in the United States; the conduct of such studies, including possible field studies, in an ethical way; and ongoing, open debate on the study and use of climate engineering while mindful of public opinion, still nascent, on the prospect of deploying the technology.

REFERENCES

APS [American Physical Society]. 2011. Direct Air Capture of CO_2 with Chemicals. A Technology Assessment for the APS Panel on Public Affairs. Available online at www.aps.org/policy/reports/assessments/upload/dac2011.pdf (accessed November 9, 2012).

Caldeira K, Matthews HD. 2007. Transient climate-carbon simulations of plantetary geoengineering. Proceedings of the National Academy of Sciences of the United States of America 104:9949–9954.

Caldeira K, Bala G, Cao L. 2013. The Science of Geoengineering. Annual Review of Earth and Planetary Sciences 41.

Crutzen PJ. 2006. Albedo enhancement by stratospheric sulfur injections: A contribution to resolve a policy dilemma? Climatic Change 77:211–221.

Keith D. 2011. Towards a Federal Research Program on Solar Radiation Management. Presentation at the American Meteorological Society. Washington DC, April 12. Available online at www.ametsoc.org/atmospolicy/climatebriefing/Keith.pdf (accessed August 5, 2012).

Keith D. 2012. "Can Geoengineering Save the World?" ScienceLive online chat, May 30. http://news.sciencemag.org/sciencenow/2012/05/live-chat-can-geoengineering-sav.html (accessed November 13, 2012).

Kravitz B. 2013. Climate engineering with stratospheric aerosols and associated engineering parameters. In Frontiers of Engineering 2012: Reports on Leading-Edge Engineering from the 2012 Symposium. Washington, DC: National Academies Press.

Kravitz B, Robock A, Boucher O, Schmidt H, Taylor KE, Stenchikov G, Schulz M. 2011. The Geoengineering Model Intercomparison Project (GeoMIP). Atmospheric Science Letters 12:162–167.

Kravitz B, MacMartin DG, Caldeira K. 2012. Geoengineering: Whiter skies? Geophysical Research Letters 39:L11801.

Long JCS, Rademaker S, Anderson JG, Benedick R, Caldeira K, Chaisson J, Goldson D, Hamburg SP, Keith DW, Lehman R, Lowy F, Morgan MG, Sarewitz D, Schelling TC, Shepherd JG, Victor DG, Welan D, Winickoff DE. 2011. Geoengineering: A National Strategic Plan for Research on the Potential Effectiveness, Feasibility, and Consequences of Climate Remediation Technologies. Washington: Bipartisan Policy Center.

Mercer AM, Keith DW, Sharp JD. 2011. Public understanding of solar radiation management. Environmental Research Letters 6:044006.

McClellan J, Sisco J, Suarez B, Keogh G. 2010. Geoengineering Cost Analysis: Final Report. Cambridge MA: Aurora Flight Sciences Corporation.

Parkhill KA, Pidgeon NF. 2011. Public engagement on geoengineering research: Preliminary report on the SPICE deliberative workshops. Technical Report. Understanding Risk Group Working Paper, 11-01. Cardiff University School of Psychology, Wales.

Rayner S, Redgwell C, Savulescu J, Pidgeon N, Kruger T. 2009. Memorandum on draft principles for the conduct of geoengineering research. UK House of Commons Science and Technology Committee enquiry into the Regulation of Geoengineering.

Ricke K, Morgan G, Allen M. 2010. Regional climate response to solar-radiation management. Nature Geoscience 3:8.

Watson M. 2012. Testbed news, May 16. Available online at http://thereluctantgeoengineer.blogspot.com/2012/05/testbed-news.html (accessed August 12, 2012).

Removing Carbon Dioxide from the Atmosphere: Possibilities and Challenges of Air Capture

Christopher W. Jones
Georgia Institute of Technology

More than 150 years of fossil fuel combustion has increased the global atmospheric carbon dioxide (CO_2) concentration from approximately 280 ppm in preindustrial times to almost 400 ppm today. Given the link between the rising atmospheric CO_2 concentration and global climate change, the world is now "carbon-constrained," and scientists, engineers, segments of the public, and scientifically literate policymakers are pushing for rapid development of alternative energy sources. However, the coupling of population growth and an ever-higher global standard of living means that energy demand will continue to increase. Although a worldwide effort is focused on development and deployment of renewable energy technologies, development is outpaced by energy demand. For this reason, fossil energy will continue to supply the preponderance of global energy for generations—and global CO_2 emissions will keep rising, hastening climate change.

CO_2 emissions from fossil fuel combustion are associated with three broad categories: (1) electricity production from coal- or gas-fired power plants (33–50% of total), (2) land, air, or sea transportation (~33%), and (3) other industrial uses. Global climate and energy strategies addressing anthropogenic emissions have focused on capturing the CO_2 emitted from the world's largest point sources, coal-fired power plants. This can be done in a variety of ways, for example by modification of existing plants to capture the CO_2 produced (i.e., postcombustion capture, or PCC) or by designing new plants that enable more efficient CO_2 capture. However, these approaches, even if widely adopted, would address only the 33–50% of CO_2 emissions associated with large point sources.

The most difficult CO_2 emissions to address are those associated with transportation. Onboard CO_2 capture from mobile sources such as automobiles and airplanes is currently impractical. While the electrification of passenger vehicles is (very slowly) shifting some energy use for transportation to large electricity-generating point sources, some mobile CO_2 sources, such as planes, will likely never be electrified. Thus, alternative technologies for addressing CO_2 emissions from mobile sources are needed.

DIRECT CAPTURE OF CO_2 FROM AIR ("AIR CAPTURE")

In 1999, Klaus Lackner first proposed the widespread development and deployment of devices that extract CO_2 directly from the atmosphere as a way to address global CO_2 emissions and climate change (Lackner et al. 1999). Although initially considered an alternative to capture from large point sources, the direct capture of CO_2 from the air, or "air capture," is generally considered a complementary technology to point source capture. Implementation of the two technologies together could allow long-term use of fossil energy while slowing or mitigating the impacts of anthropogenic CO_2 emissions on climate change. Furthermore, unlike other climate mitigation options—often described as geoengineering, whereby humans tinker with the planet to influence climate—CO_2 capture from air may be a safer option, a form of traditional pollution control.

Why, then, have PCC and air capture not been widely implemented? Because, in the absence of a price on emitted carbon, there is no incentive for the private sector to adopt such technologies. A recent study published by the US Department of Energy suggests that 90% of the coal-fired power plants in the United States could implement PCC at a cost of approximately \$60 per ton of CO_2 captured (Nichols 2011). However, as an emerging technology, there are far fewer detailed technoeconomic descriptions of air capture processes, and the limited reports offer a wide array of estimated costs. One study of air capture processes based on CO_2 absorption using basic alkaline hydroxide solutions suggested costs of \$500–1,000/ton CO_2 (House et al. 2011), whereas an evaluation of a second-generation technology based on use of supported amine adsorbents estimated costs closer to \$100/ton CO_2 (Kulkarni and Sholl 2012).

TECHNICAL CHALLENGES OF AIR CAPTURE

Most gas separation processes considered for air capture are based on CO_2-absorbing liquids or CO_2-adsorbing solids, with the overall process passing through cyclical stages of adsorption and desorption, as shown in Figure 1. This approach is common and used in a variety of scalable gas separation technologies. However, compared to most large-scale gas separation processes, air capture has a unique challenge associated with the ultradilute nature of CO_2 (~400 ppm) in the atmosphere. Yet it also has some key advantages over PCC; for example, it can be located anywhere in the world, because ambient air is largely uniform in composition; in addition, impurities in fossil fuel exhaust (e.g., nitrogen and sulfur

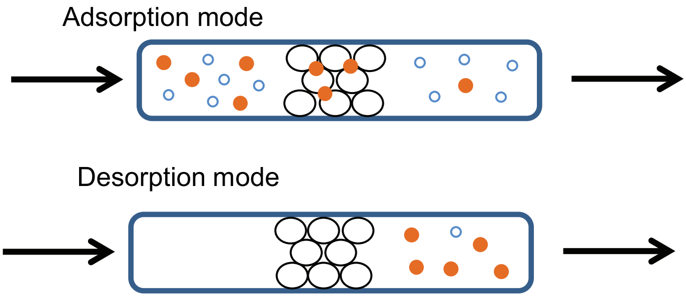


FIGURE 1 Gas separation processes based on adsorption onto solid are cyclical in nature. An adsorption cycle (top) selectively removes some gas species (small solid sphere) by adsorption on a solid (large shaded sphere), yielding a purified exhaust, followed by a desorption cycle (bottom) that liberates a concentrated product, most often induced by a pressure and/or temperature change.

oxides) are present in ambient air at a much lower level, precluding the need for gas pretreatment to remove these species. This allows for siting processes at locations appropriate for CO_2 use or sequestration, negating the need for transport of concentrated CO_2 in pipelines over long distances.

There are five important criteria for an economically scalable air capture process.

1. Because of the low concentration of CO_2 in air, very large volumes of air must be moved through the process—about 125 times and 375 times more than for CO_2 capture from a natural gas- or coal-fired power plant, respectively, assuming an equivalent capture fraction. Thus, to prevent excessive energy requirements for gas movement, the process must have very low pressure drops associated with the air flow.
2. Also associated with the low ambient CO_2 concentration, the process must use materials and/or fluids with high CO_2 capture capacities, such as those with a very high density of adsorption sites and/or very strong CO_2-adsorbent interactions.
3. Favorable adsorption kinetics are important to enable short cycle times (long cycle times lead to impractical plant sizes associated with large inventories of adsorption media).
4. Because absorption and adsorption are exothermic processes, the removal of CO_2 from the capture media for concentration is endothermic and can require significant energy input. This regeneration energy must be provided at low cost, ideally in the form of low-grade waste heat.
5. Finally, the process equipment and adsorption media must have a suitably long lifetime because the above factors will make air capture a capital-

intensive process with large plant sizes compared to many traditional gas separation processes.

AIR CAPTURE VIA ADSORPTION ON AMINES

A wide variety of CO_2-adsorbing materials have been considered for use in PCC processes. In contrast, the scope of materials for air capture applications is dramatically decreased because processes must operate near ambient temperature and pressure and offer good adsorption capacities under ultradilute conditions. Supported amine materials, a class of solids functionalized with organic amine sites, are the only materials available that offer large CO_2 capacities under air capture conditions and operate near ambient temperature. Several research groups have recently reported the suitability of such materials for CO_2 capture from ultradilute gases (e.g., Goeppert et al. 2012).

One process that shows the potential to meet the five key criteria involves supported amine adsorbents coated on a high-surface-area structured contactor (an object that contacts flowing fluid), such as a ceramic monolith. Such contactors are already produced on a large scale for use in catalytic exhaust gas clean-up and offer a low-cost route to high surface areas with low pressure drops. Flow of air through the adsorbent-lined channels at high velocity allows for rapid CO_2 adsorption kinetics. Desorption is achieved by flowing low-grade saturated steam (70–105°C) through the monoliths and over the adsorbent layer, providing both a thermal and concentration driving force for desorption. Steam in this temperature range can be obtained as low-grade waste heat from a variety of industrial processes or produced via solar-thermal heating at low cost. The concentrated CO_2 product is produced from the steam/CO_2 mixture via condensation or compression. Finally, the robustness of the monolith contactor offers promise for long-term stability of adsorbent materials.

The above description is only one process possibility; undoubtedly other promising approaches are being actively researched as well.

OUTLOOK FOR AIR CAPTURE

CO_2 capture from ambient air, or “air capture,” is an emerging technology that, if deployed on a large scale alongside traditional postcombustion capture, could play a critical role in stabilizing or even reducing global atmospheric CO_2 levels. But its development is almost entirely in the hands of private enterprise: although almost $7 billion federal dollars have been spent on research and development on methods to capture carbon from large point sources, the total federal investment for CO_2 capture from air may be as little as $300,000. Thus essentially all investment in air capture technologies in the United States has been private, and, in the absence of a carbon tax, initial deployments of such technologies will be focused on profitable industrial use, such as in greenhouses or enclosed algal bioreactors, or for enhanced oil recovery. Without policy changes or a reallocation of federal research dollars to this area, initial deployments will focus on profit-generating applications, and technological developments that might support implementation of air capture as a climate change mitigation strategy will be deferred to the future.

DISCLOSURE

The author collaborates with and has a financial interest in Global Thermostat, LLC, a company actively engaged in commercializing technology for CO_2 capture from ultradilute gases.

REFERENCES

Goeppert A, Czaun M, Prakash GKS, Olah GA. 2012. Air as the renewable carbon source of the future: An overview of CO_2 capture from the atmosphere. Energy and Environmental Science 5:7833–7853.

House KZ, Baclig AC, Ranjan M, van Nierop EA, Wilcox J, Herzog HJ. 2011. Economic and energetic analysis of capturing CO_2 from ambient air. Proceedings of the National Academy of Sciences of the United States of America 51:20428–20433.

Kulkarni AR, Sholl DS. 2012. Analysis of equilibrium-based TSA processes for direct capture of CO_2 from air. Industrial and Engineering Chemistry Research 51:8631–8645.

Lackner KS, Ziock H-J, Grimes P. 1999. Carbon dioxide extraction from air: Is it an option? Proceedings of the 24th International Conference on Coal Utilization and Fuel Systems, Clearwater, FL.

Nichols C. 2011. Coal-Fired Power Plants in the United States: Examination of the Costs of Retrofitting with CO_2 Capture Technology, Revision 3. DOE/NETL Report 402/102309. Washington, DC: Department of Energy, National Energy Technology Laboratory.

Offsetting Climate Change by Engineering Air Pollution to Brighten Clouds

Lynn M. Russell
Scripps Institution of Oceanography
University of California, San Diego

Natural, industrial, and residential combustion produces both aerosols that cool Earth and CO_2 that warms it, and the amount of combustion worldwide has increased substantially since the invention of the steam engine as well as with the increase in populations relying on wood and char burning. Natural and early man-made combustion processes emitted aerosols and CO_2 roughly proportionally, although the ratios of emission types were dependent on burning conditions. In the wake of smog-induced respiratory-health-related deaths in London in 1952, and ensuing legislation in favor of limiting emissions in the United States and Europe, "air quality engineering" was developed to reduce combustion-related aerosol emissions. But the reductions—without corresponding reductions in CO_2 emissions—led to more warming (with some offset for reductions in absorbing aerosol emissions). One approach to "climate engineering" is to undo these reductions in aerosol emissions in a way that avoids the health and visibility impacts of pollution but still allows for particles to cool Earth both by reflecting sunlight directly and by brightening clouds (which magnify the scattering of light with water). The engineering challenge with this approach is that clouds are the least understood component of the climate system, and current models cannot reliably predict their formation and properties. Recent research in the Eastern Pacific Emitted Aerosol Cloud Experiment (E-PEACE) 2011 illustrates that judicious selection of the meteorological regime and the size and composition of particle emissions can achieve substantial cooling effects. However, socioeconomic questions about climate engineering remain—such as the possibility that, if implemented, sudden cessation of enhanced particle emissions could exacerbate the climate effects on ecosystems and might interfere with oceanic and terrestrial ecosystem processes—thus requiring cautious and comprehensive research.

BACKGROUND

The fundamental physics that control the global mean surface temperature are well understood: about one-third of the incoming solar radiation is reflected back to space by Earth's albedo and the remaining two-thirds is absorbed at the surface, then emitted as longwave energy. In this way the incoming and outgoing energy at the top of the atmosphere largely balances the energy leaving, after partially trapping some of the energy by the greenhouse effect of atmospheric water vapor and clouds as well as greenhouse gases (IPCC 2007). These interactions constitute the Earth's radiative energy balance, as illustrated in Figure 1. Higher surface mean temperatures (T_{surf}) are due to the greenhouse effect, caused by the man-made release of CO_2 and other greenhouse gases. Increasing albedo (α) can offset CO_2-enhanced greenhouse warming by increasing the shortwave reflection of clouds. And clouds can be brightened to increase their reflectance by adding aerosol particles, which increase the number and decrease the mean size of cloud droplets. An example of such brightening is provided by the "ship tracks" created by the emissions of cargo ships crossing the Pacific Ocean, as shown in Figure 2.

Keeping in mind that maintaining global mean surface temperature does not imply that regional temperatures or precipitation patterns are kept constant, engineering the global mean surface temperature to reduce changes from present-day conditions could be sufficient to alleviate some of the most severe effects of global warming. Adding aerosol is straightforward, since particle production is a side effect of most combustion processes as well as a result of vaporization of liquids in condensable conditions. The real challenge in engineering aerosol particles to offset climate change by brightening clouds is predicting how the earth system, and in particular its clouds, will affect the albedo response to increased particles.

RECENT MODEL SIMULATIONS OF CLOUD BRIGHTENING

Model simulations have established the climate impacts of distributing enough particles to modify enough clouds to offset sufficient global warming to delay or lessen some of the effects expected in Earth's changing climate (Latham 1990, 2002; Latham et al. 2008). Some schemes focus on a perceived need for engineering and development of new technology, such as Flettner rotors and high-efficiency seawater atomization (Salter et al. 2008). Other studies use detailed global modeling investigations to show what fraction of clouds are brightened, with more aggressive increases in brightening resulting in exacerbation of climate in some regions even as others are improved (Rasch et al. 2009). Global simulations have also shown that where clouds are targeted is important because some choices result in exacerbation of drought conditions in some regions (Korhonen et al. 2010; Rasch et al. 2009). In addition, recent studies have investigated the com-

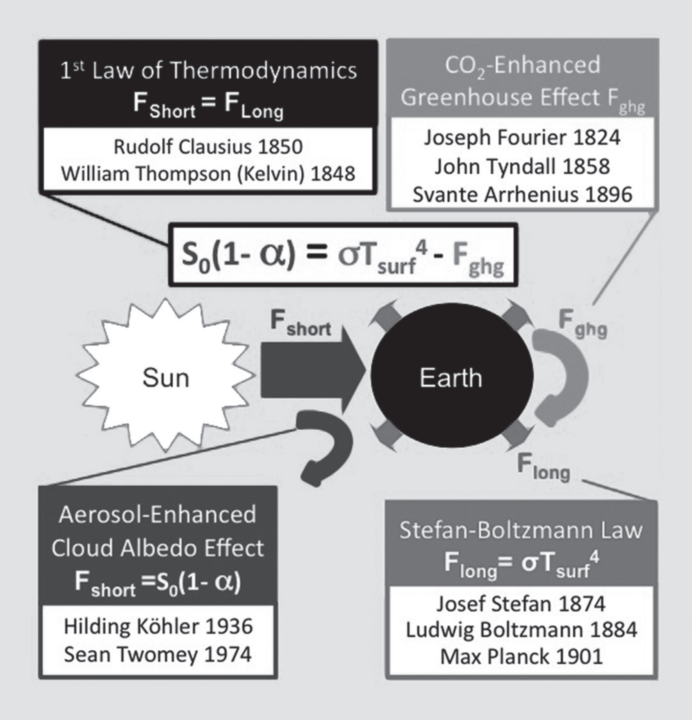


FIGURE 1 Diagram illustrating a simplified version of Earth's energy balance between incoming shortwave radiation (F_{short}) from the sun and outgoing longwave radiation (F_{long}) from the surface, showing the two underlying equations (the first law of thermodynamics and the Stefan-Boltzmann equation) and two enhanced effects (the CO_2-enhanced greenhouse effect and the aerosol-enhanced cloud albedo effect), along with the key scientists who made seminal contributions in each area. The Stefan-Boltzmann is constant σ, Earth's albedo is α, the global mean surface temperature (approximating Earth as a black body) is T_{surf}, and F_{ghg} is the radiation reemitted to Earth by greenhouse gases (and other longwave-absorbing components in the atmosphere).

plexities of aerosol cloud interactions, including the damping of cloud brightening by reductions in cloud supersaturation (Korhonen et al. 2010) and by overlapping plumes[1] of particles (Wang et al. 2011).

However, aerosol-cloud-radiation interactions are widely held to be the largest single source of uncertainty in projections of climate change due to increasing

---

[1] Cloud brightening is nonlinear, so two plumes of particles that overlap each other do not typically produce twice as much brightening.

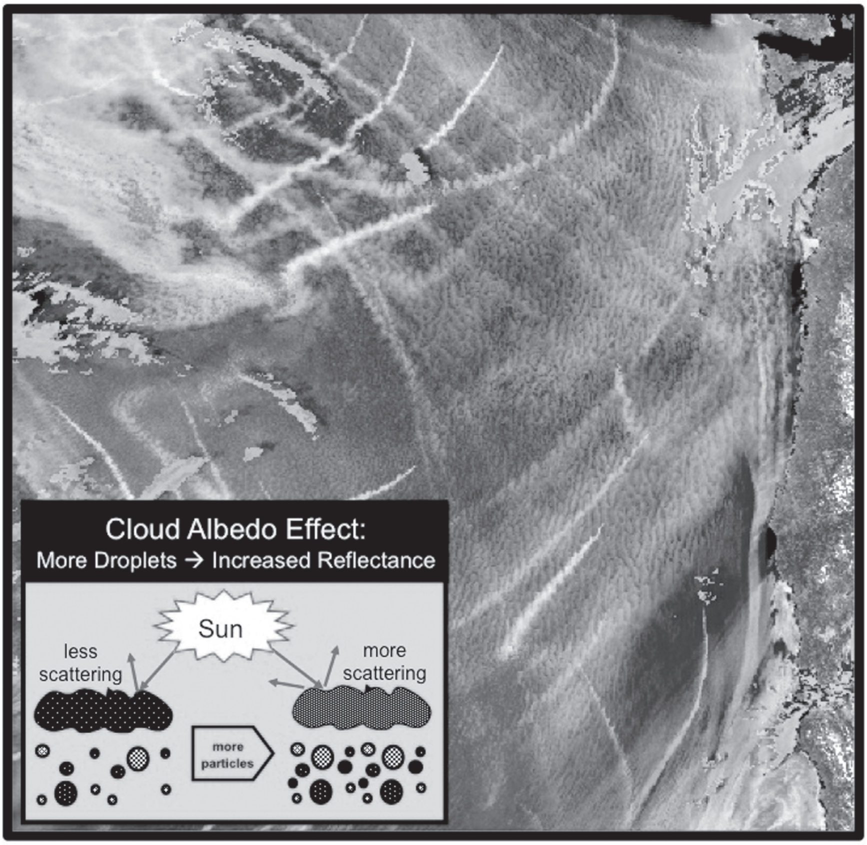


FIGURE 2 Enhanced multispectral image of the west coast of North America on June 21, 1994, from the NOAA Advanced Very High Resolution Reflectance (AVHRR) satellite, showing reflectance of low clouds (gray) with tracks from cargo ships (white). High clouds (light gray with shadow), land (dark gray on right side), and clear sky over ocean (along coast at right). Compiled by Kurt Nielsen of the Naval Postgraduate School (reproduced with permission).

anthropogenic emissions. The underlying causes of this uncertainty in modeled predictions are the gaps in fundamental understanding of cloud processes (IPCC 2007). Although there has been significant progress with both observations and models, and the qualitative aspects of the indirect effects of aerosols on clouds are well known, the quantitative representation of these processes is nontrivial and limits the ability to represent them in global climate models. Current global models lack (1) accurate aerosol particle activation, with associated implications for the profiles of supersaturation, vertical velocity, liquid water content, and drop distribution; (2) realistic microphysical growth and precipitation processes that control the formation and impacts of drizzle on cloud structure, lifetime, and particle concentration; and (3) eddy-based transport processes that control the effects of entrainment on cloud thickness and lifetime as well as the dispersion of aerosol plumes. These basic scientific issues have not been addressed by climate models or by climate engineering proposals that involve perturbing marine stratocumulus; the following section describes work by our multi-institution collaboration to address them.

NEW EXPERIMENTAL EVIDENCE OF CLOUD BRIGHTENING

To learn more about the cloud physical processes that affect aerosol-cloud-radiation interactions, we designed E-PEACE 2011 as a targeted aircraft campaign with embedded modeling studies, using the Center for Interdisciplinary Remotely Piloted Aircraft Studies (CIRPAS) Twin Otter aircraft and the R/V *Point Sur* in July 2011 off the coast of Monterey, California, with a full payload of instruments to measure particle and cloud number, mass, composition, and water uptake distributions (Russell et al. 2013; Shingler et al. 2012). Three central aspects of the collaborative E-PEACE design are described below, followed by highlights of the findings.

1. Controlled particle sources were used to separate particle-induced feedback effects from natural variability. We have investigated three types of sources of different particle sizes and compositions to characterize specific aspects of aerosol-cloud interactions: (1) ship-emitted particles at rates of 10^{16} to 10^{18} s^{-1} with dry diameters between 50 and 100 nm (Coggon et al. 2012), (2) shipboard smoke generator particles at rates of 10^{11} to 10^{13} s^{-1} with dry diameters between 50 nm and 1 µm, and (3) aircraft-based milled, coated salt particles at rates of 10^{9} s^{-1} with dry diameters between 3 and 5 µm. The shipboard smoke generators are shown in Figure 3.
2. Satellite observations showed that not all ship tracks cause cloud brightening (Chen et al. 2012), indicating a variety of cloud feedback responses to increased particle concentrations. These observations were compared to the features predicted by large eddy simulations and aerosol-cloud parcel modeling of the impacts of turbulence, precipitation, and other cloud processes on the number concentration and size distribution of cloud drops (Lu and Seinfeld 2005, 2006; Russell et al. 1999).
3. The track from the controlled emission of smoke-generated particles demonstrated efficient cooling of clouds at very low warming cost, using existing technology and minimal resources. We noted that cooling outweighed warming by a factor of 50 on the day that a track was observed

FIGURE 3 Operation of smoke generators aboard the research vessel (R/V) *Point Sur* for E-PEACE in July 2011.

> (Russell et al. 2013). This cooling effect exceeds that of commercial shipping, for which track-making ships induce twice as much cooling as warming (on days when tracks formed).

One of the most interesting results of E-PEACE was the activation into cloud droplets of smoke particles composed almost entirely of organic constituents (Shingler et al. 2012). This result was surprising because many organic components are hydrophobic and do not serve as effective cloud nuclei at supersaturations below 0.2%. The large diameter of smoke particles makes it possible for them to activate with fewer soluble constituents.

A second finding is the formation of tracks from the smoke particles in cloud-covered marine boundary layers. The organic smoke particles not only activated to cloud droplets but also did so in sufficient numbers to form a track with a detectable increase in brightness. However, there was a range of brightening observed for the many different tracks formed by particle emissions from fairly similar cargo ships (Chen et al. 2012), indicating that cloud feedback processes play an important role in determining cloud brightening.

The third important finding of E-PEACE was the frequency of low clouds with multiple layers, which reduce the impact of particles on clouds. Tracks did not form every day because of either the absence or structure of clouds near the ship. To produce a track with a significant albedo effect, the cloud layer needed to be uniform and single-layered. In addition, for rapid mixing of particles into the cloud layer, the layer needed to be below approximately 500 m. During the 12 days of the E-PEACE cruise, multiple cloud layers of 100 to 1,000 m were present on more than half. Since particles emitted by ships on the ocean surface are usually transported only to the lowest cloud layer, their modification of droplet distributions does not appreciably change the albedo seen from above the top cloud layer. In such cases, particles have little effect on the radiation balance. The presence of low cloud layers overlying the layers affected by the smoke particles resulted in a low frequency of track formation. This finding is significant because it shows the need for representing small-scale cloud structure in global climate models in order to improve predictions of aerosol-induced cloud albedo changes.

IMPLICATIONS FOR CLIMATE ENGINEERING

The E-PEACE results provide a proof of concept that cloud brightening to reduce global mean warming is possible, with existing, decades-old technology, for some cloud conditions (but it will not reduce drought or ocean acidification). Track formation requires sufficient particle production to increase droplet number by 100 to 300 cm^{-3} over well-mixed boundary layers 100 m to 600 m high and spanning track widths of several kilometers. Cargo ships and portable smoke generators can both easily emit 10^{16} to 10^{18} s^{-1}, which is sufficient at wind speeds of up to 10 m s^{-1} to make tracks in unpolluted marine air. The advantage of smoke generators for climate engineering is that the lower fuel consumption by the much smaller ship has a substantially lower CO_2 cost, making cooling more efficient.

However, the radiative effects are not the only ones to be considered before deploying on a large scale (Russell et al. 2012). In particular, careful research is needed to assess the impacts of particle deposition on ocean and downwind terrestrial ecosystems; sustained changes in particle deposition could have deleterious impacts on ocean and land biota. Furthermore, shifts in precipitation patterns and direct radiation at the surface, if substantial, could affect crop production. And implementation of cloud brightening in regions near susceptible human populations could affect health.

CONCLUSIONS

Although the technology for particle emission and distribution exists, the engineering required for cloud brightening is hardly trivial. The most critical challenges to engineering the design of large-scale cloud brightening are (1) cloud feedback processes that affect the cloud response to aerosol enhancements and reduce the expected brightening, (2) multilayered clouds that mask changes in underlying clouds, and (3) ecosystem impacts of particle deposition (Russell et al. 2012). These issues require region-specific observations and small-scale, short-duration testing to determine realistic constraints for modeling. In addition, although particle production is feasible with existing technology, there are ample opportunities for optimizing the efficiency of particle emission processes and for minimizing their ecosystem impacts.

Knowledge of aerosol-cloud interactions remains sufficiently uncertain that consideration of their use for climate engineering is premature. Substantial advances are needed in understanding of aerosol and cloud physics to quantify their role in climate change, and such advances require experimental as well as modeling studies. If such studies demonstrate the effectiveness of particles for cloud brightening, it may be possible to use this method to offset some of the warming to the global mean surface temperature caused by greenhouse gases.

GOING FORWARD

The seriousness of the consequences of global warming merits research into the possibility of using cloud brightening for climate engineering. However, while cloud brightening will target atmospheric emissions outside of national boundaries (since offshore marine stratocumulus have some of the largest impact on albedo) in areas that largely lack environmental regulations, any large-scale implementation should involve multinational agreement and cooperation, as well as compensation for unexpected and harmful consequences. Furthermore, as with any solar reflection method that does not also reduce greenhouse gases, once initiated the cessation of cooling would likely cause accelerated warming as the system returns to the nonmasked warming (Russell et al. 2012).

In summary, while cloud brightening could be appropriate to prevent tipping points (such as massive sea ice loss, which some predict may occur as early as 2015[2]), implementation of cloud brightening to offset climate warming should be considered as an option only after sufficient research is devoted to better constraining aerosol-cloud-radiation interactions.

ACKNOWLEDGMENTS

E-PEACE was supported by the National Science Foundation under grant AGS-1013423. Thanks to Kurt Nielsen of the Monterey Naval Postgraduate School for compiling the satellite image in Figure 2.

---

[2] See for example an article in the September 14, 2012, issue of *The Guardian*, “Arctic expert predicts final collapse of sea ice within four years”; available online at www.guardian.co.uk/environment/2012/sep/17/arctic-collapse-sea-ice?newsfeed=true (accessed November 9, 2012).

REFERENCES

Chen Y-C, Christensen MW, Xue L, Sorooshian A, Stephens GL, Rasmussen RM, Seinfeld JH. 2012. Occurrence of lower cloud albedo in ship tracks. Atmospheric Chemistry and Physics 12:8223–8235.

Coggon MM, Sorooshian A, Wang Z, Metcalf AR, Frossard AA, Lin JJ, Craven JS, Nenes A, Jonsson HH, Russell LM, Flagan RC, Seinfeld JH. 2012. Ship impacts on the marine atmosphere: Insights into the contribution of shipping emissions to the properties of marine aerosol and clouds. Atmospheric Chemistry and Physics 12:8439–8458.

IPCC [Intergovernmental Panel on Climate Change]. 2007. Climate change 2007: The physical science basis. Contribution of Working Group I to the fourth assessment report. In IPCC (Intergovernmental Panel on Climate Change), ed. Solomon S, Qin D, Manning M, Chen Z, Marquis M, Avery KB, Tignor M, Miller HL. Cambridge UK: Cambridge University Press.

Korhonen H, Carslaw KS, Romakkaniemi S. 2010. Enhancement of marine cloud albedo via controlled sea spray injections: A global model study of the influence of emission rates, microphysics and transport. Atmospheric Chemistry and Physics 10:4133–4143.

Latham J. 1990. Control of global warming. Nature 347:39–340.

Latham J. 2002. Amelioration of global warming by controlled enhancement of the albedo and longevity of low-level maritime clouds. Atmospheric Science Letters 3:52–58.

Latham J, Rasch P, Chen CC, Kettles L, Gadian A, Gettelman A, Morrison H, Bower K, Choularton T. 2008. Global temperature stabilization via controlled albedo enhancement of low-level maritime clouds. Philosophical Transactions of the Royal Society A–Mathematical Physical and Engineering Sciences 366:3969–3987.

Lu ML, Seinfeld JH. 2005. Study of the aerosol indirect effect by large-eddy simulation of marine stratocumulus. Journal of the Atmospheric Sciences 62(11):3909.

Lu ML, Seinfeld JH. 2006. Effect of aerosol number concentration on cloud droplet dispersion: A large-eddy simulation study and implications for aerosol indirect forcing. Journal of Geophysical Research Atmospheres 111:D02207.

Rasch PJ, Latham J, Chen CC. 2009. Geoengineering by cloud seeding: Influence on sea ice and climate system. Environmental Research Letters 4:045112–045119.

Russell LM, Seinfeld JH, Flagan RC, Ferek RJ, Hegg DA, Hobbs PV, Wobrock W, Flossmann AI, O'Dowd CD, Nielsen KE, Durkee PA. 1999. Aerosol dynamics in ship tracks. Journal of Geophysical Research Atmospheres 104(D24):31077.

Russell LM, Rasch PJ, Mace GM, Jackson RB, Shepherd J, Liss P, Leinen M, Schimel D, Vaughan NE, Janetos AC, Boyd PW, Norby RJ, Caldeira K, Merikanto J, Artaxo P, Melillo J, Morgan MG. 2012. Ecosystem impacts of geoengineering: A review for developing a science plan. Ambio 41:350–369.

Russell LM, Sorooshian A, Seinfeld JH, Albrecht BA, Nenes A, Ahlm L, Chen Y-C, Coggon M, Craven JS, Flagan RC, Frossard AA, Jonsson H, Jung E, Lin JJ, Metcalf AR, Modini R, Mülmenstädt J, Roberts GC, Shingler T, Song S, Wang Z, Wonaschütz A. 2013. Eastern Pacific Emitted Aerosol Cloud Experiment (E-PEACE). Bulletin of the American Meteorological Society (in press).

Salter S, Sortino G, Latham J. 2008. Sea-going hardware for the cloud albedo method of reversing global warming. Philosophical Transactions of the Royal Society A–Mathematical Physical and Engineering Sciences 366:3989–4006.

Shingler T, Dey S, Sorooshian A, Brechtel FJ, Wang Z, Metcalf AR, Coggon M, Mülmenstädt J, Russell LM, Jonsson HH, Seinfeld JH. 2012. Characterisation and airborne deployment of a new counterflow virtual impactor inlet. Atmospheric Measurement Techniques 5:1259–1269.

Wang H, Rasch PJ, Feingold G. 2011. Manipulating marine stratocumulus cloud amount and albedo: A process-modelling study of aerosol-cloud-precipitation interactions in response to injection of cloud condensation nuclei. Atmospheric Chemistry and Physics 11:4237–4249.

Climate Engineering with Stratospheric Aerosols and Associated Engineering Parameters

Ben Kravitz
Pacific Northwest National Laboratory

Climate engineering with stratospheric aerosols, an idea inspired by large volcanic eruptions, could cool Earth's surface and thus alleviate some of the predicted dangerous impacts of anthropogenic climate change. However, the effectiveness of climate engineering to achieve a particular climate goal, and any associated side effects, depend on certain aerosol parameters and how the aerosols are deployed in the stratosphere. Through the examples of sulfate and black carbon aerosols, this paper examines "engineering" parameters—aerosol composition, aerosol size, and spatial and temporal variations in deployment—for stratospheric climate engineering. The effects of climate engineering are sensitive to these parameters, suggesting that a particle could be found or designed to achieve specific desired climate outcomes. This prospect opens the possibility for discussion of societal goals for climate engineering.

BACKGROUND

Large volcanic eruptions cause cooling of Earth's surface by creating a layer of stratospheric sulfate aerosols that scatter incoming solar radiation. The 1991 eruption of Mount Pinatubo, which injected approximately 20 teragrams (Tg) of sulfur dioxide (SO_2) into the stratosphere, caused global cooling by 0.5°C for the next 12 months (Soden et al. 2002). This and other eruptions have inspired study of a method of climate engineering: the deliberate creation of a layer of stratospheric aerosols to cool the planet (Budyko 1974).

In addition to surface cooling, large tropical eruptions such as that of Mount Pinatubo have other important effects. They induce patterns of winter warming over continents in the Northern Hemisphere, a dynamical response of the atmospheric circulation to stratospheric heating by the aerosols (Shindell et al. 2001; Stenchikov et al. 1998). At the same time, the summer monsoons in India and East Asia are weakened by a smaller temperature gradient between the Indian Ocean and the Asian continent and reduced evaporative flux from the Indian Ocean (Boos and Kuang 2010; Manabe and Terpstra 1974; Oman et al. 2006). Furthermore, an increase in available photochemical surfaces provided by the aerosols catalyzes ozone loss (Kinnison et al. 1994).

High-latitude eruptions, such as that of Katmai in 1912, have somewhat different climate effects. Winter warming does not occur, and weakening of the Indian summer monsoon is more prominent (Oman et al. 2006). The aerosols have a shorter atmospheric lifetime (~8 months) than in tropical eruptions (~12 months) since both the aerosols' travel from the tropics to the poles and midlatitude storm tracks (where they are removed) account for much of the lifetime of stratospheric aerosols injected into the tropics.

The time of year of the eruption also plays a critical role in determining climate impacts. Aerosols injected in the winter at high latitudes will have reduced radiative effects because of less sunlight and will also be removed from the stratosphere more quickly, in part due to large-scale deposition (Kravitz and Robock 2011).

SIMULATIONS

Climate engineering with stratospheric sulfate aerosols has been studied repeatedly with climate models. Simulations in which globally averaged temperature is returned to a reference state show that the tropics are slightly overcooled and that high latitudes, particularly the Arctic, are warmer than in the reference case (Govindasamy and Caldeira 2000; Kravitz et al. 2013). As distinct from the impacts of large tropical volcanic eruptions, Northern Hemisphere continents do not show winter warming patterns for climate engineering with stratospheric sulfate aerosols (Robock et al. 2008), a method that cools the surface more than the rest of the troposphere, stabilizing the lower atmosphere and weakening the hydrologic cycle (Bala et al. 2008). Studies have not yet revealed whether summer monsoon weakening is a robust feature of climate model response to this method of climate engineering.

Simulated climate effects depend on the method of climate engineering, namely stratospheric sulfate aerosols that are similar to the aerosols from the Mount Pinatubo eruption. Such aerosols have particular compositions (approximately 75% sulfuric acid and 25% water) and sizes (effective radius of ~0.5 μm) (Rasch et al. 2008). They are also assumed to be injected above the equator and distributed through an altitude of 16–25 km. If any of these parameters changes, the radiative and climate effects will likely change as well.

This paper presents some options for "engineering" aerosol parameters, specifically composition and size, as well as the latitude and time of year of aerosol injection. The discussion focuses largely on examples involving sulfate and black carbon aerosols, with less attention to particles designed to optimize particular radiative and climatic outcomes. The concluding section addresses societal implications of various potential choices.

ENGINEERING PARAMETERS

Composition: Sulfate vs. Black Carbon Aerosols

Sulfate aerosols scatter nearly 100% of visible and ultraviolet light, whereas black carbon aerosols are excellent absorbers. Although both types of aerosols will prevent some amount of solar radiation from reaching the surface if placed in the stratosphere, black carbon will cause significant stratospheric heating; for example, 1 Tg of black carbon aerosols (0.08 µm radius) in the lower stratosphere has been simulated to cause more than 20°C of stratospheric heating (Kravitz et al. 2012). Conversely, the eruption of Mount Pinatubo created 29 times the aerosol loading and produced 2–3°C of stratospheric heating (Stenchikov et al. 2002), which increased Arctic zonal winds, forcing a positive mode of the Arctic oscillation.

The reactions governing catalytic ozone loss are temperature dependent, so stratospheric heating would also cause stratospheric ozone loss (Groves et al. 1978), particularly in the Arctic, but in the Antarctic evaporation of polar stratospheric clouds would slow ozone loss in the region. The addition of photochemical surfaces to the stratosphere would promote ozone loss from both sulfate and black carbon aerosols; stratospheric climate engineering with 2 Tg S yr^{-1} would delay recovery of the Antarctic ozone hole by 30–70 years (Tilmes et al. 2008).

Black carbon aerosols (typical radius 0.08 µm) cause more cooling per unit mass than volcanic sulfate aerosols. Stratospheric injection of 1 Tg yr^{-1} black carbon aerosols has been simulated to cause 0.4°C of surface cooling (Kravitz et al. 2012). In contrast, an injection rate of as much as 5 Tg SO_2 per year would increase the cooling to 0.6°C (Robock et al. 2008).

Stratospheric aerosols will fall into the troposphere within a few years. The amount of additional rain acidity resulting from climate engineering with 5 Tg SO_2 per year would likely be insufficient to cause damage to most ecosystems (Kravitz et al. 2009), but black carbon is toxic and causes respiratory impairment (Baan et al. 2006). Moreover, if black carbon lands on snow or bright surfaces, it lowers the albedo of those surfaces and the planet retains more solar radiation, exacerbating global warming (Vogelmann et al. 1988).

Size

Depending on aerosol composition, particle size may be somewhat predetermined. Sulfate aerosols tend to coagulate, and SO_2 can condense onto existing particles. Both factors tend to increase particle size. Sulfate aerosols are most efficient at scattering when the particles are small (~0.1 μm radius). As they grow larger, so does their infrared effect; above ~2 μm in radius, infrared effects overwhelm the scattering effects and become net absorbing particles. In addition, larger particles have a greater fall speed and thus a lower atmospheric lifetime. Simulations have shown that increasing the size of black carbon particles by 50% reduced surface cooling by more than a factor of 2 (Kravitz et al. 2012).

The inclusion of aerosol microphysics in simulations increases the amount of SO_2 needed from 5 Tg per year (Robock et al. 2008) to more than 50 Tg (English et al. 2011; Heckendorn et al. 2009; Pierce et al. 2010). One proposal to overcome microphysical limitations is direct condensation of sulfuric acid vapor to produce a monodisperse distribution of small sulfate aerosols, but this idea is untested (Pierce et al. 2010). Black carbon aerosols tend not to coagulate in ways that alter their radiative properties and are generally smaller. Moreover, in the stratosphere, they could be heated by the sun and self-loft, increasing the fall distance and thus atmospheric lifetime of the particles (Pueschel et al. 2000; Rohatscheck 1996).

Spatial/Temporal Distribution

Longitude of stratospheric aerosol injection is largely irrelevant, as the general circulation of the atmosphere will evenly distribute the aerosols across all longitudes within a matter of weeks. Conversely, the radiative and climatic impacts of climate engineering are quite sensitive to the latitude and altitude of the particles. Surface cooling from stratospheric aerosol climate engineering tends to increase with the stratospheric altitude of the aerosols, in part due to longer atmospheric lifetime (Ban-Weiss et al. 2012; Kravitz et al. 2012).

Variation in solar radiation reductions by latitude and season results in modest improvements (<6%) in global temperature and precipitation residuals (climate engineering minus reference case) compared to a uniform solar reduction, but targeting regions with the highest residuals results in improvements in these regions by up to 30% (MacMartin et al. 2012). Aerosols injected extratropically tend to remain in the hemisphere of injection, and stratospheric sulfate aerosol climate engineering in only one hemisphere can shift the Intertropical Convergence Zone, a band of equatorial precipitation, potentially causing Sahelian greening or drying (J.M. Haywood and A. Jones, UK Met Office, personal communication, March 31, 2012).

DESIGNED PARTICLES

Changing the "engineering" parameters for sulfate and black carbon can "fine-tune" the climate effects of stratospheric aerosol climate engineering to some extent, but some side effects are unavoidable. For example, although they are excellent scatterers, sulfate aerosols mostly scatter light forward, whereas cooling is achieved by scattering sunlight back to space. Black carbon absorbs solar radiation, keeping the energy in the atmosphere. And the optimal sulfate aerosol size may not be achievable because of coagulation.

These concerns suggest that it may be necessary to design particles in order to achieve desired aerosol parameters such as, for example, a perfectly scattering particle that photophoretically levitates at 50 km in altitude (Keith 2010). Although the climate effects of such a particle are unknown, as are its side effects on stratospheric chemistry and atmospheric circulation, it may be possible to create particles that take advantage of certain properties and alleviate some side effects or shortcomings.

CONCLUSIONS

The goal of mapping "engineering" parameters for stratospheric aerosol particles and their climate effects is to, eventually, be able to address the question of what society might want climate engineering to do. For example, if societal goals are primarily to preserve Arctic sea ice, climate engineering could be done by injecting sulfate aerosols into the Arctic troposphere during spring. If the primary goal is to cool the planet while avoiding any increase in rain acidity, perhaps black carbon would be preferable to sulfate. If the primary and side effects of climate engineering can be chosen, the foundation is laid for discussions to determine climate goals.

Such a discussion will not have clear answers, though, as the goals of climate engineering do not depend solely on climatology. There are multiple stakeholders with myriad values that encompass scientific, social, political, legal, ethical, and personal dimensions, and there is no clear method of synthesizing and addressing these issues on a global scale. Moreover, assuming that a method of climate engineering could be designed and chosen, society will need to decide how much climate engineering will be done.

The choice of "engineering" parameters has not been fully explored, and there are many uncertainties in the predicted impacts of climate engineering. Dedicated research is needed to develop relevant engineering tools and enhance understanding in these areas. The purpose here is to illustrate areas of a potential research agenda that could be useful in choosing methods and goals of climate engineering.

REFERENCES

Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, Cogliano V. 2006. Carcinogenicity of carbon black, titanium dioxide, and talc. The Lancet Oncology 7(4):295–296.

Bala G, Duffy PB, Taylor KE. 2008, Impact of geoengineering schemes on the global hydrological cycle. Proceedings of the National Academy of Sciences 105(22):7664–7669.

Ban-Weiss GA, Cao L, Bala G, Caldeira K. 2012. Dependence of climate forcing and response on the altitude of black carbon aerosols. Climate Dynamics 38(5-6):897–911.

Boos WR, Kuang Z. 2010. Dominant control of the South Asian monsoon by orographic insulation versus plateau heating. Nature 463:218–222.

Budyko MI. 1974. Climate and Life. New York: Academic Press.

English JM, Toon OB, Mills MJ, Yu F. 2011. Microphysical simulations of new particle formation in the upper troposphere and lower stratosphere. Atmospheric Chemistry and Physics 11:9303–9322.

Govindasamy B, Caldeira K. 2000. Geoengineering Earth's radiation balance to mitigate CO_2-induced climate change. Geophysical Research Letters 27(14):2141–2144.

Groves KS, Mattingly SR, Tuck AF. 1978. Increased atmospheric carbon dioxide and stratospheric ozone. Nature 273:711–715.

Heckendorn P, Weisenstein D, Fueglistaler S, Luo BP, Rozanov E, Schraner M, Thomason LW, Peter T. 2009. The impact of geoengineering aerosols on stratospheric temperature and ozone. Environmental Research Letters 4:045108.

Keith DW. 2010. Photophoretic levitation of engineered aerosols for geoengineering. Proceedings of the National Academy of Sciences 107(38):16428–16431.

Kinnison DE, Grant KE, Connell PS, Rotman DA, Wuebbles DJ. 1994. The chemical and radiative effects of the Mount Pinatubo eruption. Journal of Geophysical Research 99:25705–25731.

Kravitz B, Robock A. 2011. The climate effects of high latitude eruptions: The role of the time of year. Journal of Geophysical Research 116:D01105.

Kravitz B, Robock A, Oman L, Stenchikov G, Marquardt AB. 2009. Sulfuric acid deposition from stratospheric geoengineering with sulfate aerosols. Journal of Geophysical Research 114:D14109.

Kravitz B, Robock A, Shindell DT, Miller MA. 2012. Sensitivity of stratospheric geoengineering with black carbon to aerosol size and altitude of injection. Journal of Geophysical Research 117:D09203.

Kravitz B, Caldeira K, Boucher O, Robock A, Rasch PJ, Alterskjær K, Bou Karam D, Cole JNS, Curry CL, Haywood JM, Irvine PJ, Ji D, Jones A, Lunt DJ, Kristjánsson JE, Moore J, Niemeier U, Schmidt H, Schulz M, Singh B, Tilmes S, Watanabe S, Yoon J-H. 2013. Climate model response from the Geoengineering Model Intercomparison Project (GeoMIP). Journal of Geophysical Research, submitted.

MacMartin DG, Keith DW, Kravitz B, Caldeira K. 2012. Managing tradeoffs in geoengineering through optimal choice of non-uniform radiative forcing. Nature Climate Change, available online at www.nature.com/nclimate/journal/vaop/ncurrent/full/nclimate1722.html.

Manabe S, Terpstra TB. 1974. The effects of mountains on the general circulation of the atmosphere as identified by numerical experiments. Journal of the Atmospheric Sciences 31:3–42.

Oman L, Robock A, Stenchikov GL, Thordarson T. 2006, High-latitude eruptions cast shadow over the African monsoon and the flow of the Nile, Geophysical Research Letters, 33, L18711.

Pierce JR, Weisenstein DK, Heckendorn P, Peter T, Keith DW. 2010. Efficient formation of stratospheric aerosol for climate engineering by emission of condensible vapor from aircraft. Journal of Geophysical Research 37:L18805.

Pueschel RF, Verma S, Rohatschek H, Ferry GV, Boiadjieva N, Howard SD, Strawa AW. 2000. Vertical transport of anthropogenic soot aerosol into the middle atmosphere. Journal of Geophysical Research 105(D3):3727–3736.

Rasch PJ, Tilmes S, Turco RP, Robock A, Oman L, Chen C-C, Stenchikov GL, Garcia RR. 2008. An overview of geoengineering of climate using stratospheric sulfate aerosols. Philosophical Transactions of the Royal Society A, 366:4007–4037.

Robock A, Oman L, Stenchikov GL. 2008. Regional climate responses to geoengineering with tropical and Arctic SO_2 injections. Journal of Geophysical Research 113:D16101.

Rohatschek H. 1996. Levitation of stratospheric and mesospheric aerosols by gravitophoresis. Journal of Aerosol Science 27(3):467–475.

Shindell DT, Schmidt GA, Miller RL, Rind D. 2001. Northern Hemisphere winter climate response to greenhouse gas, ozone, solar, and volcanic forcing. Journal of Geophysical Research 106(D7):7193–7210.

Soden BJ, Wetherald RT, Stenchikov GL, Robock A. 2002. Global cooling after the eruption of Mount Pinatubo: A test of climate feedback by water vapor. Science 296(5568):727–730.

Stenchikov GL, Kirchner I, Robock A, Graf H-F, Antuña JC, Grainger RG, Lambert A, Thomason L. 1998. Radiative forcing from the 1991 Mount Pinatubo volcanic eruption. Journal of Geophysical Research 103(D12):13837–13857.

Stenchikov GL, Robock A, Ramaswamy V, Schwarzkopf M Daniel, Hamilton K, Ramachandran S. 2002. Arctic Oscillation response to the 1991 Mount Pinatubo eruption: Effects of volcanic aerosols and ozone depletion, Journal of Geophysical Research, 107(D24), 4803.

Tilmes S, Müller R, Salawitch R. 2008. The sensitivity of polar ozone depletion to proposed geoengineering schemes. Science 320(5880):1201–1204.

Vogelmann AM, Robock A, Ellingson RG. 1988. Effects of dirty snow in nuclear winter simulations. Journal of Geophysical Research 93(D5):5319–5332.

VEHICLE ELECTRIFICATION

Vehicle Electrification

Michael W. Degner
Ford Motor Company

Sanjeev Naik
General Motors

The automobile industry is undergoing one of the most dramatic and rapid transformations in its history. This transformation, toward vehicle electrification, is being driven by concerns about global warming, sustainability, and national security. In response to these concerns, many countries have implemented regulations that mandate dramatic improvements in automobile fuel economy over the next 10 to 15 years—more than double that of just a few years ago. And in efforts to meet these regulations, automobile manufacturers are developing advanced powertrain technologies, of which hybrid electric, plug-in hybrid electric, and battery electric vehicles (HEVs, PHEVs, and BEVs) are some of the most promising.

The concept of vehicle electrification is not new. In fact, battery electric vehicles were the most common type of vehicle in the late 1800s and early 1900s. But they were quickly displaced by improved internal combustion engines, which have dominated the automobile powertrain landscape for more than 100 years. That dominance is now being challenged by vehicle electrification, thanks to dramatic improvements in energy storage systems, electric machine drives, and electrical system integration and control.

The speakers in this session discuss some of these advances, research to make further improvements, and some of the challenges that need to be addressed to enable widespread vehicle electrification.

The first speaker, Jeff Sakamoto (Michigan State University), describes efforts to improve automobile electrical energy storage systems. Reducing the cost, size, and weight of such systems is a key challenge preventing the widespread adoption of PHEVs and BEVs, which show the most promise to dramatically reduce the world's dependence on petroleum. His paper reports recent improvements in battery technologies (particularly with the use of lithium-ion technology), industry targets required to enable widespread adoption of plug-in vehicles, and some of the ongoing research to meet these targets.

In the second talk, Matthew Willard (Case Western Reserve University), outlines the challenges and research under way to develop improved magnetic materials, which are used in electric machine drives. Magnetic materials of both the hard (permanent magnets) and soft (electrical steels and magnetic cores) type are critical in the design of high-performance electric machines and power electronic converters. This presentation covers the desired material characteristics, some of the research challenges to develop better materials, and efforts to reduce the use of critical, strategic materials—rare earths—in these magnetic materials.

Widespread adoption of plug-in electric vehicles could represent a significant increase in the load on the electrical transmission and distribution system. Arindam Maitra (Electric Power Research Institute) presents the results of an EPRI study on the potential impacts of this increased load and the changes necessary in the electrical transmission and distribution systems to address such impacts.

Today's drivers have high expectations for the safety, reliability, comfort, and connectivity in their vehicles. The final paper by Rahul Mangharam (University of Pennsylvania) discusses technical approaches to enhance vehicle safety through remote diagnostics, networking, recalls, and software upgrades.

Keeping Up with Increasing Demands for Electrochemical Energy Storage

Jeff Sakamoto
Michigan State University.

The interest in vehicle electrification is unprecedented. Several automotive manufacturers are producing or planning to produce hybrid electric vehicles (HEVs), plug-in hybrid electric vehicles (PHEVs), and fully or battery electric vehicles (BEVs). Lithium-ion (Li-ion) battery technology is the current leading candidate to meet the near- and medium-term needs for electric vehicles. Leveraging considerable growth and development from the manufacturing of batteries for microelectronics, Li-ion technology has advanced significantly in the last decade. However, the leap from small-scale microelectronic batteries (tens of watt hours) to electric vehicle battery packs (tens of kilowatt hours) is not trivial. Performance metrics such as cost/kilowatt hour, specific energy (Wh/kg), specific power (kW/kg), safety, and cycle life are considerably more demanding for electric vehicles than for laptops and cell phones. Electric vehicles (EVs) show promise in minimizing reliance on fossil fuels, but their widespread use will likely require a revolutionary advance in energy storage technology. Research in sophisticated and efficient power electronics, battery/cell telemetry, safety, thermal management, and schemes to recycle/reuse EV batteries can help to establish a solid foundation for the development and use of EVs. This article provides an overview of energy storage technology for vehicle electrification, highlights challenges, and discusses opportunities at the frontiers of battery research.

THE NEED FOR ADVANCED ENERGY STORAGE

In terms of sustainability, minimizing dependence on fossil fuels and reducing CO_2 emissions are compelling arguments to electrify vehicles. And from a practical perspective, EVs can take advantage of existing infrastructure for electrical power production and transport—infrastructure that will soon be bolstered by efforts to augment renewable energy production whose primary byproduct is electrical power (e.g., through photovoltaic cells and wind turbines).

If electrical energy becomes the preferred form of energy, electrochemical energy storage is a natural fit. In contrast, hydrogen fuel cell technology requires an entirely new infrastructure to efficiently produce hydrogen and then transport, store, and reconvert it to electrical energy.

To put into perspective the amount of energy consumed by the transportation sector, of the total 2.85×10^{16} watt-hours (1 Quad = 2.93×10^{14} watt-hours) of energy used by the United States in 2011 27.7% (7.91×10^{15} watt-hours) went to transportation (Figure 1).[1] However, due to the relatively low chemical-to-mechanical energy conversion efficiency of internal combustion engine (ICE) technology, the ratio of serviceable to rejected energy is disproportionately low compared to other energy use sectors.

If EVs can improve energy efficiency in the short term and the technology for non-fossil-fuel-based/renewable electrical power generation can be realized in the long term, the benefits to our country's current and future sustainability are clear. Assuming the latter, the following discussions focus on electrical energy storage, specifically batteries.

CHALLENGES FOR ELECTROCHEMICAL ENERGY STORAGE AND USE IN EVS

Defining the ideal battery for EVs is complicated because of the numerous powertrain configurations involved in HEVs, PHEVs, and BEVs; for example, the capacity (kWh), power (kW), and cycle life can be considerably different for an HEV compared to a BEV (Khaligh and Li 2010). To simplify discussion, this article focuses on BEVs with battery characteristics that can power a four-seat vehicle for approximately 100 miles on a single charge, criteria favorable for widespread adoption.[2] Figure 2 shows the necessary performance attributes of an effective EV battery.

Vehicle range is determined by the amount of energy stored in the battery and the rate at which the energy is expended to propel the vehicle. A 23 kWh battery used to power a ~70 kW electric motor is believed to be sufficient to achieve a range of about 160 km. The mass and volume of the battery should be minimized to reduce the vehicle mass while maximizing vehicle cabin volume, respectively.

---

[1] These data and the accompanying figure are from the Lawrence Livermore National Laboratory website, https://flowcharts.llnl.gov/content/energy/energy_archive/energy_flow_2011/LLNLUSEnergy2011.png, accessed November 9, 2012.

[2] Whether this BEV performance standard is specifically required to significantly impact energy consumption is not yet known, but agencies and auto companies generally agree with this definition (Bruce et al. 2012; CCC 2012; Thackeray et al. 2012).

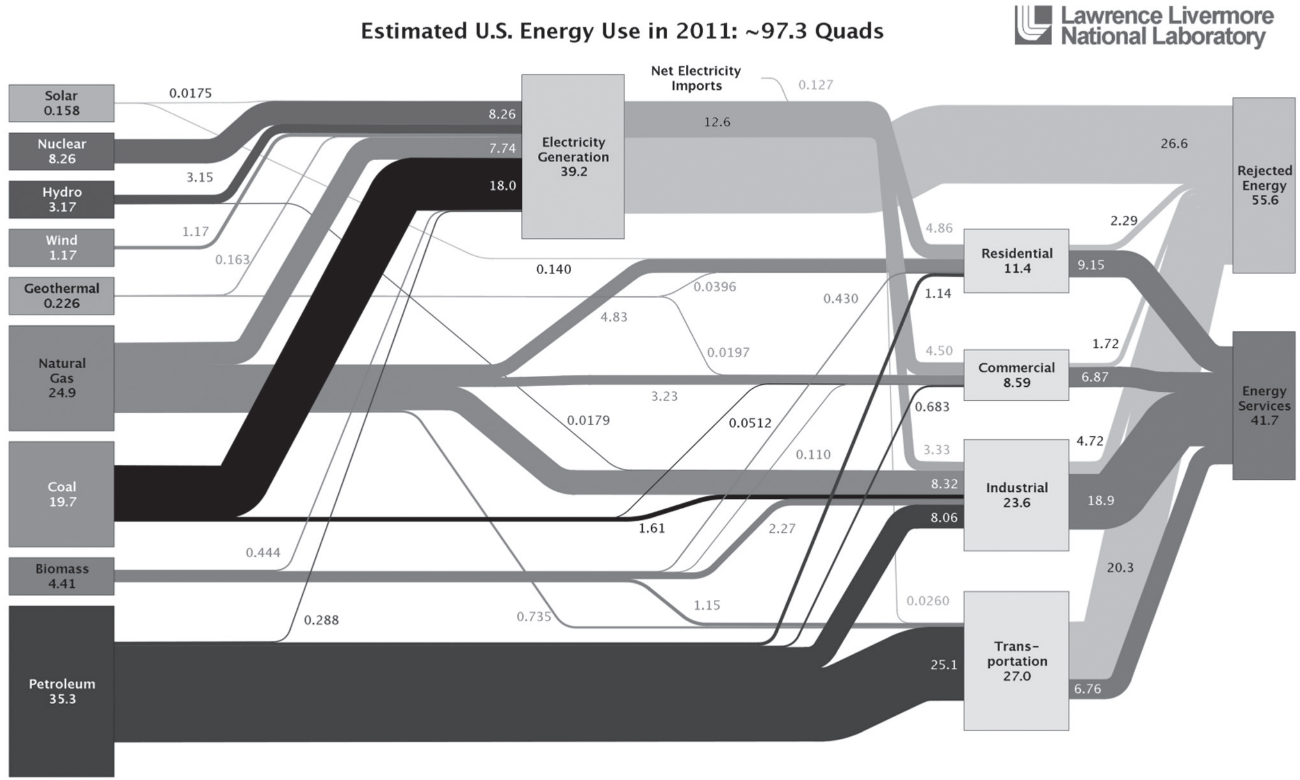


FIGURE 1 Energy use in the United States in 2011.

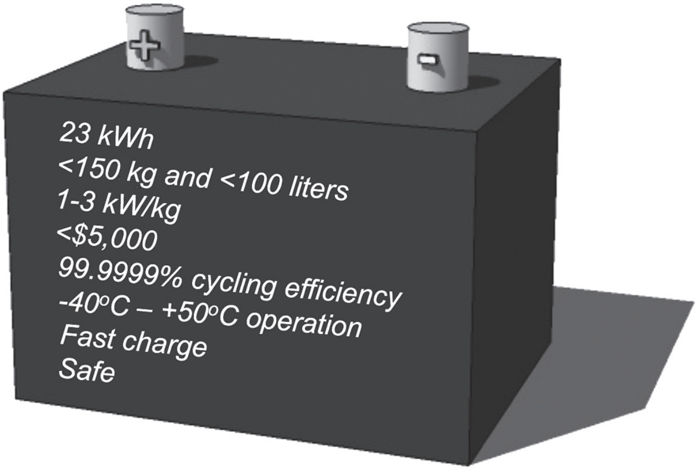


FIGURE 2 Battery performance criteria to power the next generation of battery electric vehicles.

Vehicle acceleration is determined by specific power (kW/kg) or how quickly the stored energy can be extracted per unit mass of battery. A common metric is in the single to multi-kW/kg range (e.g., 1–3 kW/kg).

Replacement of the ICE powertrain with an electric powertrain should not considerably add to the vehicle cost, and the cost of the battery pack should be less than $5,000.

Ensuring consistent, long-term vehicle range requires a charging efficiency of 99.9999% such that approximately 80% of the original battery capacity is available at the end of the vehicle's life.

Widespread use of BEVs will entail operation in dramatically different climates, so the battery must be capable of operating at relatively low and high ambient temperatures.

Although it is difficult to quantify how fast is fast enough, the issue of range anxiety may be addressed if a battery pack can be charged at a charging station as quickly as a gasoline tank can be filled at a gas station.

Last, and perhaps most importantly, the battery technology must be safe and reliable.

LI-ION BATTERIES

Of the battery chemistries available today (Figure 3), Li-ion has the highest specific energy (Tarascon and Armand 2001) and is the only technology capable of meeting the criteria shown in Figure 2. While other energy storage technologies such as supercapacitors, flywheels, and compressed air are in development, only Li-ion batteries are mature enough or meet the necessary criteria or both (Dunn et al. 2011). Li-ion batteries also have the distinct advantage of both intrinsically high cell voltage (>3 V) and the capacity to store lithium ions in the solid state, resulting in high specific energy and low cell volume (energy density), respectively.

In a typical Li-ion cell (Figure 4), lithium ions are shuttled, with relatively high efficiency, between the anode and cathode via a liquid Li-ion electrolyte (EVSAE 2012). Graphite (in powder form) is by far the most common anode that reversibly uptakes and releases lithium ions between graphene sheets. The cathode consists of a ceramic of nominal formula $LiMO_2$ (in powder form), where M stands for a transition metal such as cobalt (Co), manganese (Mn), or nickel (Ni) that can change valence states upon insertion/extraction of Li-ions. During

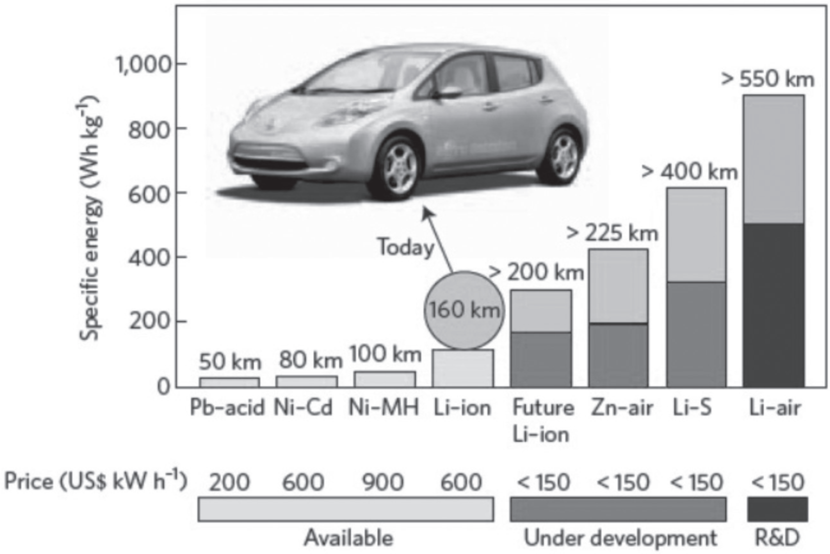


FIGURE 3 Comparison of battery technologies currently available and under development (Bruce et al. 2012). Darker shading in the bars indicates the specific energy values in laboratory-scale prototypes; lighter shading indicates the range of anticipated specific energy values for Li-S and Li-air technologies, respectively. Cd, cadmium; Li, lithium; MH, metal hydride; Ni, nickel; Pb, lead; Zn, zinc.

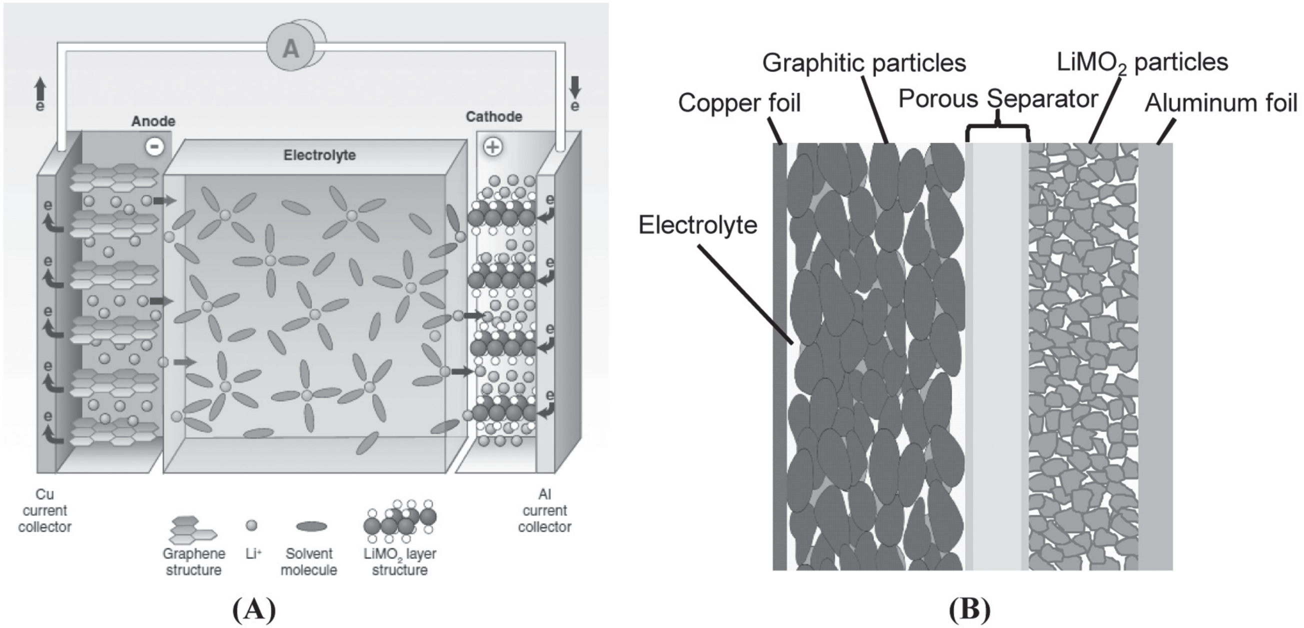


FIGURE 4 Schematic of a lithium (Li)-ion cell: (A) at the atomic scale (A, amperes; reprinted with permission from Tarascon and Armand 2001) and (B) at the microscopic scale (adapted from EVSAE 2012).

discharge, it is more energetically favorable for the graphite anode to release its Li-ions and for the cathode uptake Li-ions to reduce the M valence charge (e.g., M^{4+} to M^{3+}). This shuttling of lithium ions from the anode to the cathode is accompanied by the simultaneous passing of electrons through an external circuit to power the electric vehicle.

Since their invention in 1991 by Sony and professor John B. Goodenough, Li-ion batteries have been integrated into cell phones, laptop computers, and other microelectronics (Figure 5). And some of the first Li-ion-powered EVs were not terrestrial but instead were vehicles sent to survey the surface of Mars in 2003 through NASA's Mars Exploration Program (Huang et al. 2000). The Mars Exploration Rover Li-ion batteries started development in 1996 and were flight qualified and implemented in 2003, a testament to how quickly Li-ion battery technology can progress.

In 2008, a combination of factors led to a significant push to improve vehicle fuel efficiency, resulting in a rapid transformation of the auto industry with an emphasis on vehicle electrification. In 2011 GM rolled out the PHEV Volt and Nissan the BEV Leaf, and in 2012 Ford started selling the BEV Ford Focus.

These past and recent successes are impressive, but Li-ion battery packs still require considerable reductions in cost as well as increases in specific energy to extend vehicle range. The following section presents a materials perspective on opportunities in electrochemical energy storage and milestones whose achievement will address these issues.

OPPORTUNITIES IN ELECTROCHEMICAL ENERGY STORAGE

Unlike lead (Pb)-acid, nickel-cadmium (Ni-Cd), and nickel–metal hydride (Ni-MH) battery technologies, Li-ion technology performance has room for improvement, as shown in Figure 3. Advanced electrode and cell designs and electrode material breakthroughs (Thackeray et al. 2012) may enable a doubling in energy density and a fourfold reduction in cost compared to available Li-ion technology. Eventually, however, Li-ion technology improvements will crest, requiring a breakthrough in battery technology to approach the cost target (~$150/kWh) and the range of an ICE powertrain vehicle (>400 km).

Several research and government agency reports (e.g., Bruce et al. 2012; CCC 2012) present complementary near-term roadmaps to guide battery research and development over the next two decades. Three milestones extrapolated from these roadmaps illustrate the frontiers of battery development, with substantial steps in 2015 and 2020 followed by a revolutionary leap in 2030.

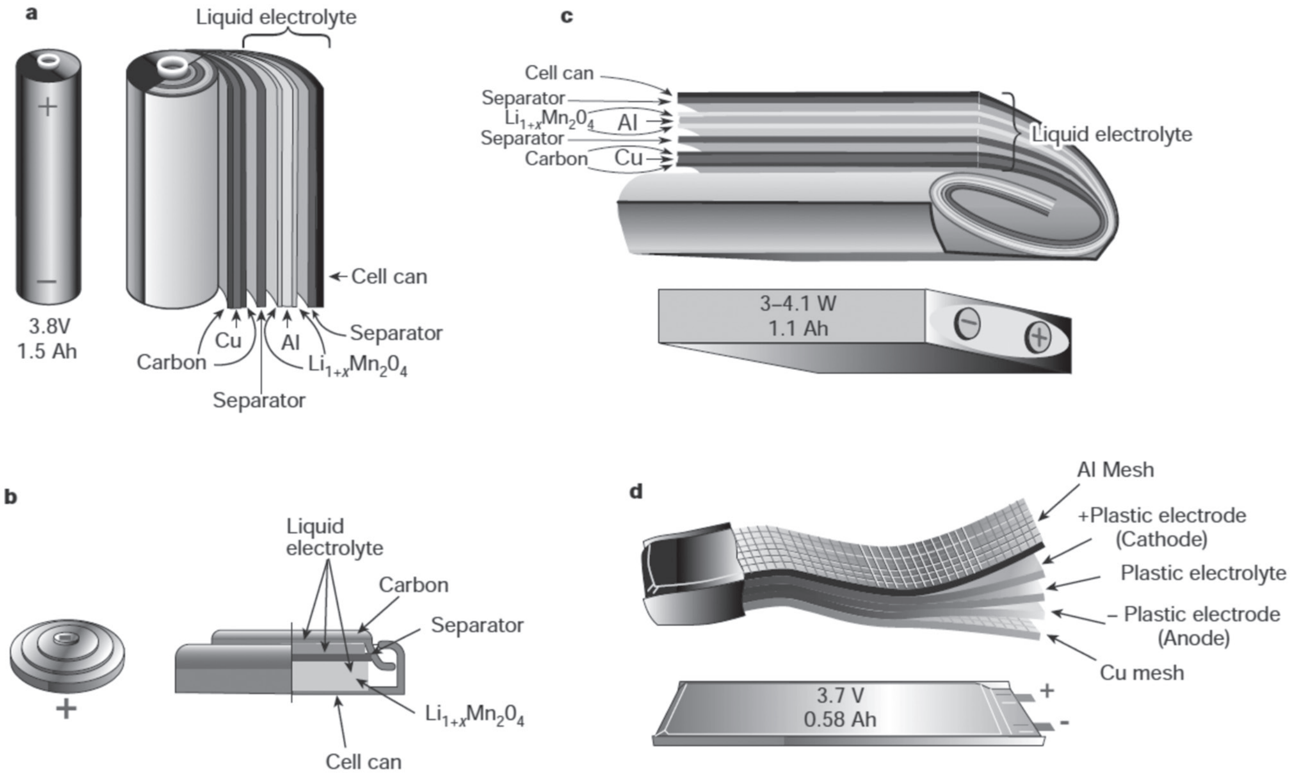


FIGURE 5 Li-ion batteries come in a variety of designs: (a) spiral wound cell, (b) button or watch cell, (c) prismatic cell, and (d) solid polymer (electrolyte) battery (Tarascon and Armand 2001). Ah, amp-hour; Al, aluminum; Cu, copper; V, volt.

2015 MILESTONE: OPTIMIZE CURRENT MATERIALS AND CELL COMPONENT DESIGN

In the short term the focus is on optimization of materials and conventional liquid electrolytes. At present, approximately 50% of a battery pack mass is dead weight (Johnson and White 1998). For example, in the cell cross-section shown in Figure 4b, only the graphite anode and $LiMO_2$ particles store lithium and therefore energy; the electrical current–collecting foils, electrolyte, separator, and hermetic container do not store energy.

Increasing the mass/volume fraction of active material is one strategy to improve specific energy. Making thicker, less porous electrodes is a popular approach to achieve this, but thicker and less porous active electrode layers impede the transport of ions in the electrolyte and thus reduce power (Buqa et al. 2005).

Furthermore, the nonuniform current in thicker electrodes can cause metallic lithium to deposit on the anode and oxygen gas to be released from certain $LiMO_2$ cathodes, which can be a safety hazard in the presence of heat and flammable electrolyte solvents. These challenges can be addressed through research on advanced electrode designs, powder processing, and coating technologies (DOE 2010).

Cycle life is another concern that requires attention. A passivation layer forms on the surface of a graphitic anode particle during the solid electrolyte interphase (SEI). As lithium intercalates and deintercalates from graphite particles, the corresponding swelling and contraction create fissures in the SEI, resulting in the continuous and irreversible consumption of lithium and diminishing capacity retention. Again, improved electrode designs to homogenize charge flow could address this concern, as could the development of new electrolytes and/or electrolyte additives to make the SEI more robust.

Economies of scale will probably not play a significant role in minimizing cost per kilowatt hour ($/kWh) (Bruce et al. 2012; CCC 2012) by 2015. Rather, new materials with appreciably better performance and lower cost are needed to bring costs down to the target of approximately $150/kWh.

2020 MILESTONE: ELECTRODE AND ELECTROLYTE MATERIALS BREAKTHROUGHS

The 2020 milestone focuses on reducing cost rather than increasing specific energy, although it is hoped that the latter will increase by more than a factor of two. Once the electrode and cell design have been optimized, increases in the specific energy will require new electrode and complementary electrolyte materials that can store more lithium or charge-per-unit mass/volume and that have higher voltage (energy = amps × volts × time). If the new materials can be made at comparable or lower cost, a byproduct of increased specific energy will be a commensurate decrease in cost/kWh (Figure 3).

Alloying anodes such as silicon (Si)- or tin (Sn)-based electrodes will likely constitute the next generation of Li-ion battery anodes (Thackeray et al. 2012). The term "alloying" is used to describe the reversible, electrochemical reaction between lithium and a pure element such as silicon or tin.

Specific capacity (milliamp hours per gram, or mAh/g), which refers to the amount of lithium that an electrode can uptake and release, is commonly used where one mole (6.94 grams) of lithium can provide 26.8 Ah of electrical charge. Graphitic anodes have a theoretical specific capacity of 372 mAh/g, and silicon and tin have specific capacities of 4,009 and 960 mAh/g, respectively, making the interest in these anodes apparent.

However, a >300% change in volume accompanies the uptake and release of lithium from silicon and tin, creating significant mechanical stresses that cause decrepitation and poor cycle life (Deshpande et al. 2010). One solution is to reduce the powder particle size from the typical micron scale to the nano scale and thus decrease the magnitude of strain. Creating nano Si wires with <100 nm dimensions, originally demonstrated by the Cui group (Wu et al. 2012), reduces the overall strain to minimize decrepitation and improve cycle life. Another approach is to increase cycle life by embedding Si or Sn particles in an elastic or compliant carbon matrix to create an encapsulation effect (Zhao et al. 2011). Envia Systems recently announced a 400 Wh/kg Li-ion cell pack using Si-based anodes, but it has yet to be commercialized (Thackeray et al. 2012). Advanced materials processing and materials engineering could play a major role in optimizing alloying electrode performance and reducing cost.

On the cathode side, there are two promising approaches. First, the cathode system, a composite layered structure, enables the full extraction of one molecular unit of lithium, or $x = 1$ per formula unit of $x\mathrm{Li_2MnO_3}(1-x)\mathrm{LiMO_2}$ (M = Mn, Ni, Co) (Thackeray et al. 2012). This type of material, developed by Thackeray and colleagues at Argonne National Laboratory, can deliver nearly twice the specific capacity compared to conventional $\mathrm{LiMO_2}$ cathodes.

There are a few practical concerns associated with this material strategy, however. For example, the lithium must come from the anode (which is not the case with conventional $\mathrm{LiMO_2}$ cathodes) and the charging voltage (4.6 V) is outside the stability window of most conventional electrolytes, resulting in diminished cycle life.

The second approach involves increasing the cathode reaction voltage from about 4.0 V to approximately 5.0 V to result in a 20% increase in specific energy, provided the specific capacity is comparable to that of conventional cathodes. Examples of high-voltage cathodes include $\mathrm{LiMn_{1.5}Ni_{0.5}O_2}$ and $\mathrm{LiMPO_4}$ (M = Co, Ni) (Allen et al. 2011), both of which are relatively mature compared to the composite layered cathodes described above, but the lack of stable electrolytes limits their widespread implementation.

Higher cell voltage (cathode side) and a stable SEI (anode side) with advanced anodes both require significant improvements in electrolytes. One approach is to integrate additives to conventional electrolytes to improve the high-voltage (cathode) stability. The Kang group achieved this by increasing the electrolyte stability to enable the use of $LiCoPO_4$ (4.8 V) cathodes (von Cresce and Kang 2011). A completely different approach involves a solid electrolyte material breakthrough using a ceramic rather than liquid electrolyte. The advantages could include higher stability (0 to >6 V) and perhaps safety as a flammable liquid electrolyte is replaced by a highly thermal and chemically stable ceramic.

A class of ceramics referred to as "fast-ion conductors" conducts lithium ions about as fast as a conventional liquid electrolyte. Additionally, these ceramics have negligible electronic conductivity and the Li-ion conductivity improves with increasing temperature. Recent examples of promising solid electrolytes include sulfur (S)-based (Kamaya et al. 2011) and oxide-based electrolytes (Murugan et al. 2007; Rangasamy et al. 2011) that exhibit Li-ion conductivities comparable to conventional liquid electrolytes.

Next-generation Li-ion batteries will require new materials for anodes, cathodes, and electrolytes. Advanced materials and ceramic processing technology based on lessons learned from the 2015 milestone will play a key role in achieving the 2020 milestone. The development of new electrolyte materials, in particular, will advance progress toward the 2030 milestone of enabling new battery chemistry beyond Li-ion technology.

2030 MILESTONE: BEYOND LI-ION BATTERIES

If electric powertrains are to replace ICE technology, without raising concerns about cost or range, a new battery technology is required (Bruce et al. 2012). Three of the most popular battery chemistries that represent the frontier of energy storage are Li-S, Zn-air, and Li-air (the metal air batteries are actually semifuel cells, but for brevity and consistency they are referred to as batteries). Because the challenges related to Zn-air technology are relatively well known (Lee et al. 2011), the focus here is on Li-S and Li-air batteries, which are not as well understood.

Li-S is attractive because of its high theoretical energy density (2,199 Wh/l), high theoretical specific energy (2,567 Wh/kg), and the low cost and abundance of sulfur (Bruce et al. 2012). Factoring in the mass of the electrolyte, electrical current–collecting foils, packaging, and other features, the practical specific energy is reduced to approximately 600 Wh/kg, which is still considerably higher than that of advanced Li-ion batteries. In a Li-S cell, elemental lithium and sulfur are the reactants, a nonaqueous electrolyte shuttles lithium ions between electrodes, and, because sulfur does not have sufficient electrical conductivity, a specific porous carbon (Ji et al. 2009) is added to increase the effective electrical conductivity of the S-cathode.

Two challenges remain: (1) prevention of deleterious mechanisms that result from the formation of soluble Li-S compounds during cycling and (2) achievement of a stable/cyclable Li-electrolyte interface, a challenge since the 1980s when it led to the demise of rechargeable lithium metal anode batteries.

Li-air batteries are of two types: nonaqueous and aqueous (Bruce et al. 2012); the "air" in question is the source of oxygen and, for the aqueous batteries, water vapor. Nonaqueous batteries involve the reaction of lithium with oxygen gas (O_2) to form Li_2O_2. (The reference to "air" may be a bit misleading since both water vapor and carbon dioxide must be excluded from the reaction/cell in the nonaqueous configuration.) During discharge, lithium is transported through a nonaqueous electrolyte and reacts with O_2 in the presence of a porous carbon network and a catalyst to form solid precipitates of Li_2O_2. The theoretical energy density of this system is (3,436 Wh/l) and the theoretical specific energy is (3,505 Wh/kg). Some of the key challenges for nonaqueous Li-air batteries are (1) development of an oxygen-permeable membrane that excludes carbon dioxide and water vapor, (2) development of effective cathode electrodes that prevent pore occlusion resulting from the formation of solid byproducts during discharge, and (3) effective integration of catalysts to improve reaction kinetics.

In the second type of Li-air battery an aqueous electrolyte is used to transport lithium ions into a carbon cathode electrode to form lithium hydroxide (LiOH) during discharge. At lower concentrations LiOH is soluble in the electrolyte, whereas at higher concentrations (i.e., greater degrees of discharge) it precipitates out as a solid. The theoretical energy density of the aqueous variant is (2,234 Wh/l) and the theoretical specific energy is (3,582 Wh/kg). Some of the challenges that remain for aqueous Li-air technology are technologies to (1) protect the lithium metal anode from the aqueous electrolyte using a ceramic electrolyte membrane, (2) prevent reactions with carbon dioxide from ambient air, and (3) prevent pore and electrolyte interface occlusion when/if LiOH precipitates at higher depths of discharge. Although there are few examples of advanced prototypes, the projected specific energy for both Li-air variants is expected to be about 1,000 Wh/kg.

The majority of the challenges involve the discovery of new materials and development of an electrolyte to enable the use of metallic lithium anodes. The need for ceramic electrolytes that protect the lithium metal anode is one aspect common to Li-air and Li-S technology. In addition to poor cycle stability, excess lithium is required to counter the effects of poor cycling efficiency. For example, two- to fourfold excess lithium may be necessary, thus reducing the energy density. One recent material breakthrough (Murugan et al. 2007) identified a new class of ceramic oxide electrolyte that is believed to exhibit the unprecedented combination of stability against lithium with high, room-temperature ionic conductivity (Figure 6).

In addition to new electrolytes, advanced catalyst and catalyst support electrodes, similar to those found in fuel cells, are required to improve rechargeability and power.

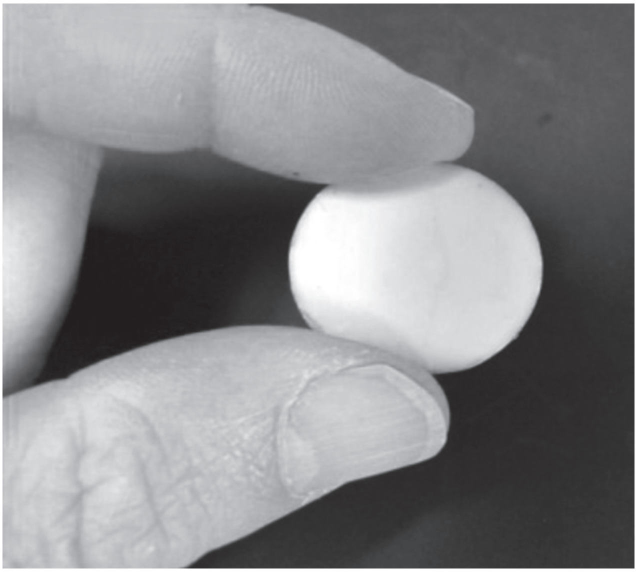


FIGURE 6 Ceramic electrolytes may enable the use of metallic lithium (Li) anodes. A new ceramic electrolyte referred to as LLZO ($Li_7La_3Zr_2O_{12}$) exhibits the unprecedented combination of high ionic conductivity and stability against metallic lithium and air. Shown above, a prototypical LLZO membrane fabricated in the Sakamoto lab using unique powder synthesis and sintering technology.

CONCLUSIONS

There is a compelling need for advanced electrochemical energy storage to power the next generation of electric vehicles. Furthermore, interest in Li-ion technology is on the rise, if growing attendance at the five-year-old annual symposium “Beyond Li-ion” is any evidence. But although Li-ion batteries offer substantial performance advantages over previous battery technologies, range capacity and cost are major challenges to overcome by 2015. Better electrode, cell, and pack design, together with advanced manufacturing and power electronics, will establish a solid foundation for future EV technology.

By 2020, material and electrolyte breakthroughs are expected to provide moderate improvements in BEV range—and dramatic reductions in cost. Anodes that are cheap (based on Si or Sn) are expected to uptake and release more lithium per unit mass. On the cathode side, the focus will be on increasing the voltage and lithium uptake and release per unit mass. Developing higher-stability liquid and solid electrolytes will complement higher-voltage cathodes and efforts to revolutionize energy storage in the long term (2030).

Provided the necessary electrical infrastructure is in place by 2030, a breakthrough in electrochemical energy storage is required if ICE technology is to be replaced by BEVs. Li-air or Li-S batteries may be the high specific energy, low-cost technology of the future, but significant materials and engineering challenges must first be overcome. Solving the lithium metal anode–electrolyte interface stability issue; developing novel catalyst/catalyst support cathodes; and creating stable, semipermeable solid electrolytes require further research and development if Li-air and Li-S technologies are to mature.

The frontiers of electrochemical energy storage are exciting from multiple perspectives, and are likely to generate significant engineering research and development opportunities in the coming decades.

REFERENCES

Allen J, Jow TR, Wolfenstine J. 2011. Improved cycle life of Fe-substituted $LiCoPO_4$. Journal of Power Sources 196(20):8656–8661.

Bruce PG, Freunberger SA, Hardwick LJ, Tarascon JM. 2012. Li-O_2 and Li-S batteries with high energy storage. Nature Materials 11:19–29.

Buqa H, Goers D, Holzapfel M, Spahr ME, Novak P. 2005. High rate capability of graphite negative electrodes in lithium-ion batteries. Journal of the Electrochemical Society 152(2):A474–A481.

CCC [Climate Change Committee]. 2012. Cost and performance of EV batteries. Cambridge UK: Element Energy Limited.

Deshpande R, Cheng YT, Verbrugge MW. 2010. Modeling diffusion-induced stress in nanowire electrode structures. Journal of Power Sources 195:5081–5088.

DOE [Department of Energy]. 2010. Energy storage R & D annual progress report. Washington.

Dunn B, Kamath H, Tarascon J-M. 2011. Electrical energy storage for the grid: A battery of choices. Science 334:928–935.

EVSAE [Society of Automotive Engineers International Vehicle Electrification]. 2012. Lithium in the limelight: Li-ion battery research heats up around the globe. June. Available online at www.nxtbook.com/nxtbooks/sae/12EVSD0626/index.php.

Huang CK, Sakamoto J, Wolfenstine J, Surampudi S. 2000. The limits of low-temperature performance of Li-ion cells. Journal of the Electrochemical Society 147:2893–2896.

Ji X, Lee KT, Nazar LF. 2009. A highly ordered nanostructured carbon-sulphur cathode for lithium-sulphur batteries. Nature Materials 8:500–506.

Johnson BA, White RE. 1998. Characterization of commercially available lithium-ion batteries. Journal of Power Sources 70:48–54.

Kamaya N, Homma K, Yamakawa Y, Hirayama M, Kanno R, Yonemura M, Kamiyama T, Kato Y, Hama S, Kawamoto K, Mitsui A. 2011. A lithium superionic conductor. Nature Materials (letters) 10:682–686.

Khaligh A, Li Z. 2010. Battery, ultracapacitor, fuel cell, and hybrid energy storage systems for electric, hybrid electric, fuel cell, and plug-in hybrid electric vehicles: State of the art. IEEE Transactions on Vehicular Technology 59(6):2806–2814.

Lee JS, Kim ST, Cao R, Choi N-S, Liu M, Lee KT, Cho J. 2011. Metal-air batteries with high energy density: Li-air versus Zn-air. Advanced Energy Materials 1:34–50.

Murugan R, Thangadurai V, Weppner W. 2007. Fast lithium ion conduction in garnet-type $Li_7La_3Zr_2O_{12}$. Angewandte Chemie International Edition 46:7778–7781.

Rangasamy E, Wolfenstine J, Sakamoto J. 2011. The role of Al and Li concentration on the formation of cubic garnet solid electrolyte of nominal composition $Li_7La_3Zr_2O_{12}$. Solid State Ionics 206:28–32.

Tarascon J-M, Armand M. 2001. Issues and challenges facing lithium batteries. Nature 414:359–367.

Thackeray MM, Wolverton C, Isaacs ED. 2012. Electrical energy storage for transportation: Approaching the limits of, and going beyond, lithium-ion batteries. Energy and Environmental Science 5:7854–7863.

von Cresce A, Kang X. 2011. Electrolyte additive in support of 5 V Li-Ion chemistry. Journal of the Electrochemical Society 158(3):A337–A342.

Wu H, Chan G, Choi JW, Ryu I, Yao Y, McDowell MT, Lee SW, Jackson A, Yang Y, Hu L, Cui Y. 2012. Stable cycling of double-walled silicon nanotube battery anodes through solid-electrolyte interphase control. Nature Nanotechnology (letters) 7:310–315.

Zhao X, Hayner CM, Kung MC, Kung HH. 2011. In-plane vacancy-enabled high-power Si-graphene composite electrode for lithium-ion batteries. Advanced Energy Materials 1(6):1079–1084.

Stronger, Lighter, and More Energy Efficient: Challenges of Magnetic Material Development for Vehicle Electrification

MATTHEW A. WILLARD
Case Western Reserve University

A fundamental transformation of the transportation sector in the United States is under way (Electrification Coalition 2009). In parallel with advances in renewable energy resources for power generation, the rising use of electric and hybrid electric vehicles (EVs and HEVs) is starting to change the future of civilian transportation. Similar efforts are moving forward for "more electric" ships, aircraft, and other military technologies.

Most people don't think of magnetic materials in association with EV/HEVs, but they play an important role in improving the efficiency and performance of devices in electric power generation, conditioning, and conversion (Chau and Chan 2007).[1] In fact, many functions in modern vehicles would not be possible without advanced magnetic materials: they are used in safety features, engines, controls, braking, and in motors and actuators used for fans, pumps, wipers, and locks.

In transportation technologies, reliability, power density, and overall energy capacity are essential. Yet, despite the vast array of vehicle technologies that use magnets, the unique power systems used to supplement (or replace) the internal combustion engine remain a challenge for magnet designers. Specifically, improvements are needed in "permanent" magnetic materials for motors and generators and in "soft" magnetic materials for inverters and power electronics. In addition, the availability of critical materials for permanent magnets is a growing concern.

---

[1] Conversion refers to changes from DC (from batteries) to AC (from generators) and vice versa. Conditioning refers to filtering of electric power. These are accomplished by inverters or power electronics in EVs.

The following sections address state-of-the-art magnetic materials and briefly describe research to improve their performance.

DEFINITION OF TERMS

Two parameters determine magnetic performance: saturation magnetization ("magnetization") and coercive field ("coercivity"). Magnetization is the density of magnetic moments in a ferromagnetic material. A material with higher magnetization can produce larger external magnetic fields than a same-sized material with lower magnetization, and by the same token requires less material to achieve the same magnetic field.

In addition to the level of saturation magnetization, some magnetic materials, called "soft" magnets, require the application of an external magnetic field to align their magnetic moments and others, called "hard" or permanent magnets, produce significant magnetic field without an applied field. The distinguishing characteristic between these classes of materials is their coercivity, the intensity of the magnetic field required to reduce the magnetization to zero.

Hard and soft magnetic materials were refined during the 20th century to provide optimal performance for applications in which magnetization is either very resistant to switching when a magnetic field is applied (i.e., hard or permanent magnets) or easily switched when a magnetic field is applied (i.e., soft magnets) (Gutfleisch et al. 2011). Hard magnetic materials are characterized by large coercivities (more than ~10 kiloamperes per meter, or kA/m) and greater energy storage, making them useful for motor and generator applications. Soft magnetic materials, which have a low value of coercivity (less than ~400 amperes per meter, or A/m), are used in applications that require easy switching, such as induction motors, inverters, and power electronics (Emadi et al. 2008).

Coercivities available today span 8 orders of magnitude between the softest and hardest magnetic materials (Figure 1). Progress in the development of magnetic materials has been accomplished with jumps in performance when new materials are introduced, followed by incremental steps as compositions and processing steps are refined to provide the best microstructures and phase combinations.[2]

The following sections focus on each class of magnetic material and some of the current technological issues being addressed by researchers.

PERMANENT MAGNETS

Rare Earths and Their Challenges

High-performance permanent magnets typically used in EVs and HEVs are made of rare earth elements (largely neodymium [Nd] and dysprosium [Dy]), a

---

[2] A phase is a chemically distinct region in a material that possesses uniform physical properties.

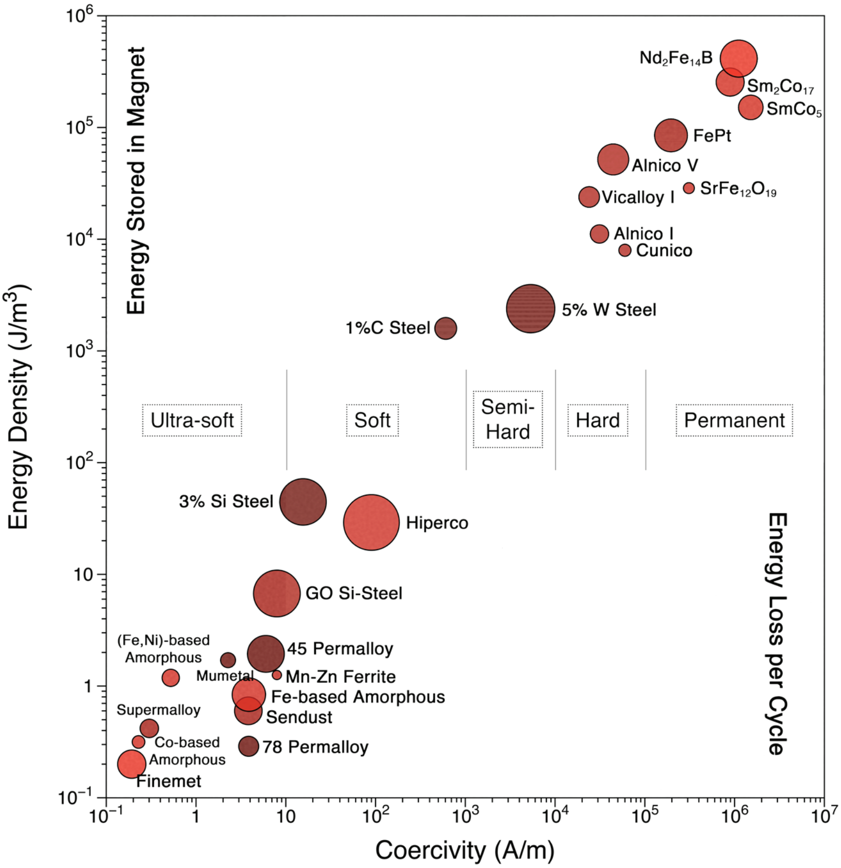


FIGURE 1 Energy density plotted against coercivity for state-of-the-art magnetic materials. The materials in the top part of the figure are used in applications where the magnetization is fixed in the material, resulting in energy storage in the magnet (i.e., permanent magnet). The materials in the bottom part of the figure are used in applications where the magnetization is switched frequently, resulting in an energy loss per cycle (i.e., soft magnets). Desirable characteristics are maximum energy storage (in permanent magnets) or minimum energy loss (in soft magnets) per cycle. The circle size is proportional to the size of the material's magnetization. A/m = amperes per meter.

magnetic transition metal element (iron [Fe]), and a metalloid element (boron [B]). The rare earth elements provide a considerable magnetocrystalline anisotropy and are responsible for the energy storage capacity of these alloys, and iron provides a relatively large magnetization. These alloys have been refined over the past 30 years into premier permanent magnet materials with the largest energy storage capability (Figure 1, top). However, the growing market and in some cases the scarcity of some rare earth elements have driven recent research efforts to consider alternative materials or ways of reducing the amount of rare earth elements in permanent magnets (Alonso et al. 2012; Sugimoto 2011).

Light rare earth elements (e.g., Nd, praseodymium [Pr], lanthanum [La]) are not actually rare—their natural abundances are similar to those of copper and nickel. Why then is there concern about rare earths? First, they are difficult to separate from each other as they bond through 4-d electrons, which are the same for all rare earth elements. Second, they are very reactive with oxygen, adding to the difficulty of refining them as metals. Third, the heavy rare earths (e.g., Dy, terbium [Tb]), which are essential to extend use to the 200°C required for EV/HEV operation, are scarcer than the light rare earths (Gutfleisch et al. 2011). Finally, Chinese companies dominate mining for all rare earths and exports are expected to decline in the coming years because these resources are used entirely for Chinese domestic products. For these reasons, alternative, rare earth–free materials have become a topic of intense research.

Alternatives to Rare Earths?

Theoretically, nanocomposite magnetic materials consisting of a mixture of finely divided regions of soft and hard magnetic phases (less than 10 nm in diameter) can improve energy storage while reducing the rare earth content of the alloy. However, simply mixing the typically available powders (5–10 µm in diameter) of hard and soft magnetic materials and pressing them together to full density will not produce the desired improvement. Rather, the powders would have to be nanoparticles no larger than 15 nm in order to achieve the required microstructure and concomitant improvement in energy storage. If a nanocomposite microstructure were produced, permanent magnets could provide energy storage of ~1.1 MJ/m^3—nearly a factor of 3 more than available today—and at 10% of the required amount of rare earths! But since the hypothesis in 1991 of this type of alloy (Kneller and Hawig 1991), it has not yet been demonstrated in the bulk form necessary for motors and generators.

Other, more radical ideas include the complete elimination of rare earths in favor of a variety of unusual substitute materials such as manganese aluminide (MnAlC), manganese bismuthide (MnBi), iron nitride ($Fe_{16}N_2$), and cobalt carbides (Co_3C/Co_2C). Figure 2 illustrates the differences in magnetization and coercivity of these and other materials.

SOFT MAGNETS

Soft magnetic alloys do not suffer from the same critical materials problem that plagues rare earth permanent magnets. However, the trend to miniaturization of soft magnetic components while maintaining energy efficiency is important as

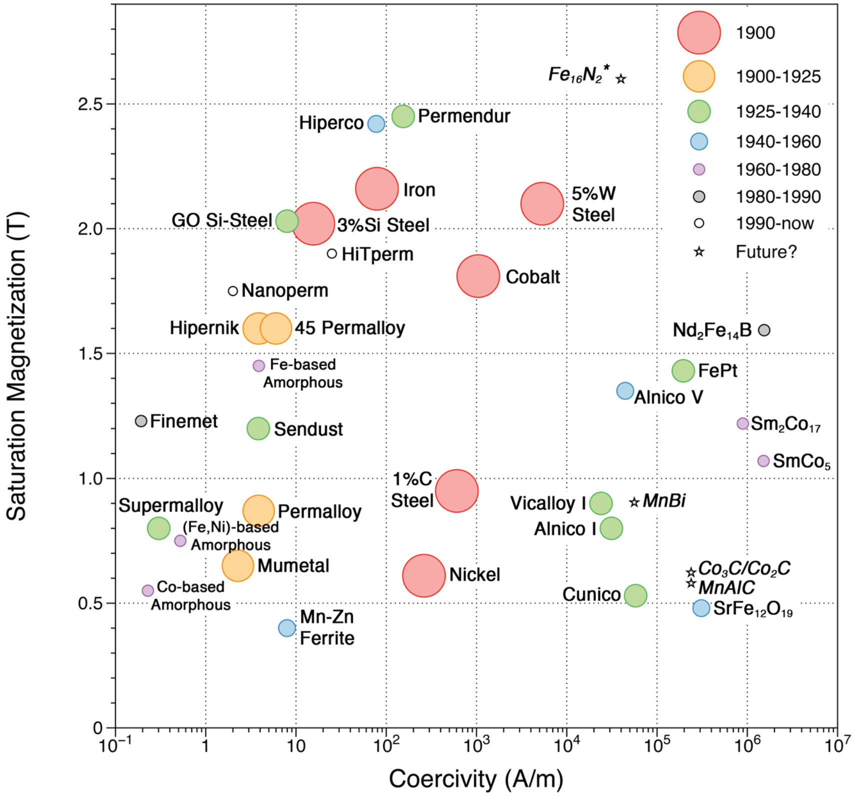


FIGURE 2 Saturation magnetization plotted against coercivity for state-of-the-art magnetic materials. Symbol size and color identify discovery date. Magnetic softness improves to the left, and magnetic hardness improves to the right. The second star next to $Fe_{16}N_2$ designates an estimate from thin film values. This figure appears in color in the online posting of this article at http://www.nap.edu/catalog.php?record_id=18185.

electrical generation and conversion technologies require more power (Willard and Daniil 2013).

Miniaturization can be achieved by increasing magnetization and/or switching frequency. The latter affords the most significant reduction, but the materials lose some energy as heat (i.e., core losses) during each switching cycle, resulting in energy inefficiency (see Figure 1, bottom). These core losses appreciably increase with switching frequency, so conventional materials do not perform well under these conditions. Suitable candidates for high-frequency use are amorphous and nanocrystalline alloys, which possess low coercivity and high electrical resistivity. Higher switching frequencies are of increasing importance for vehicle electrification when power electronics are considered for conditioning and conversion because higher frequencies enable greater component miniaturization.

Recent advances in nanocrystalline soft magnetic alloys provide materials that are energy efficient to hundreds of kHz, with larger magnetization than comparable amorphous alloys and good thermal stability (to 500°C in some cases) (Willard et al. 2012). However, mechanical brittleness and difficulties with processing in ambient air have limited the widespread use of these materials.

OUTLOOK

Improvements in soft magnetic properties through continued development in processing-microstructure-property relationships will provide the premier materials of the future. For nanocrystalline soft magnetic alloys, refinement of compositions to enhance energy efficiency (i.e., reduce core losses), mechanical performance (i.e., reduce brittleness), and air-processability are expected to advance this technology in the near future. For permanent magnets, advances in nanostructured composite materials, produced in bulk with crystallographic texture (i.e., crystal alignment), will show the most near-term technology improvement.

With new, rare earth–free options being explored extensively, the future of magnetic materials for vehicle electrification can be viewed optimistically. This is especially true with current research interests in low-cost $Fe_{16}N_2$ and high-anisotropy MnAl compounds. With a considerable investment, as was made for rare earth permanent magnets over the past 30 years, a rare earth–free material capable of operating at temperatures to 200°C is certainly possible in the next 10 to 15 years.

REFERENCES

Alonso E, Sherman AM, Wallington TJ, Everson MP, Field FR, Roth R, Kirchain RE. 2012. Evaluating rare earth element availability: A case with revolutionary demand from clean technologies. Environmental Science and Technology 46:3406–3414.

Chau KT, Chan CC. 2007. Emerging energy-efficient technologies for hybrid electric vehicles. Proceedings of the IEEE 95(4):821–835.

Electrification Coalition. 2009. Electrification Roadmap: Revolutionizing Transportation and Achieving Energy Security. Washington. Available online at www.electrificationcoalition.org.

Emadi A, Lee YJ, Rajashekara K. 2008. Power electronics and motor drives in electric, hybrid electric, and plug-in hybrid electric vehicles. IEEE Transactions on Industrial Electronics 55(6):2237–2245.

Gutfleisch O, Willard MA, Brück E, Chen CH, Sankar SG, Liu JP. 2011. Magnetic materials and devices for the 21st century: Stronger, lighter, and more energy efficient. Advanced Materials 23:821–824.

Kneller EF, Hawig R. 1991. The exchange-spring magnet: A new material principal for permanent magnets. IEEE Transactions on Magnetics 27(4):3588–3600.

Sugimoto S. 2011. Current status and recent topics of rare-earth permanent magnets. Journal of Physics D: Applied Physics 44:064001.

Willard MA, Daniil M. 2013. Nanocrystalline soft magnetic alloys: Two decades of progress. In: Buschow KHJ, ed. Handbook of Magnetic Materials, chapter 4. Amsterdam: Elsevier. In press.

Willard MA, Daniil M, Knipling KE. 2012. Nanocrystalline soft magnetic materials at high temperatures: A perspective. Scripta Materialia 67:554–559.

Analysis of Projected Impact of Plug-in Electric Vehicles on the Distribution Grid

Arindam Maitra
Electric Power Research Institute

INTRODUCTION

A new era of plug-in electric vehicles (PEVs) has begun. Nissan and General Motors launched production PEVs in December 2010, and Ford, Mitsubishi, Toyota, Tesla, and others have announced plans to introduce such vehicles to the US market. With the rapidly approaching commercialization of plug-in hybrid (PHEVs) and battery electric vehicles (BEVs) as well, utilities need to ensure that they can support customers' use of such vehicles by preparing for the installation of residential, commercial, and private infrastructure in their service territories and by managing the impact of these new loads on the electric distribution system.

In light of these developments and needs, the Electric Power Research Institute (EPRI) initiated a multiyear project (EPRI 2012; Maitra et al. 2009; Taylor et al. 2009) with 19 utilities to understand PEV system impacts in the United States and Canada. This paper provides an overview of the study, presents the results relevant to the US analysis, and summarizes the conclusions.

STUDY METHODOLOGY AND GENERAL ANALYSIS FRAMEWORK

The study methodology was designed to capture potential near-term distribution system impacts in response to increased customer load. Assuming a near-term planning horizon (1 to 5 years), only the characteristics of first-generation PEVs are considered. Specifically, PEVs are modeled as simple loads whose characteristics are mainly dictated by customer behavior; controlled dispatching or vehicle-to-grid operations are not included (for a discussion of the latter, see Mangharam 2013). Growth in the base load is also not included because no particular planning year is evaluated in any given scenario. Finally, only residential customers are considered, as initial adopters are expected to charge at their residence.

The project included an assessment of PEV charging effects on specific circuits in a utility's distribution system—typically one or two representative feeders per utility—based on detailed simulations of distribution systems, customer load characteristics, and potential electric vehicle (EV) penetration and charging profiles. The results of the simulations were combined to develop summaries of general concerns and to identify assets most likely to be at risk, conditions that could require additional monitoring to avoid problems, and the impacts of different charging profiles.

As part of a PEV distribution impact collaborative project, EPRI developed a novel methodology to evaluate the impact of PEVs on distribution systems. The study methodology was designed to capture potential distribution system impacts in response to customer adoption of the new load type and was applied to 36 radial distribution feeders. The analytical framework was developed to evaluate the impacts of PEVs on distribution system thermal loading, voltage regulation, transformer loss of life, unbalance, distribution system losses, and harmonic distortion levels. These impacts are primarily determined by the assumed location of PEVs throughout the distribution network, the time of day that PEVs are expected to charge from the system, and the magnitude and duration of the charge cycle. In order to determine both system-level and individual component–level impacts, the framework provides for both deterministic and stochastic consideration of these key spatial and temporal variables (Figure 1). Specifically, the analysis identifies assets at risk of being affected and the likelihood and severity of impact.

- **Asset deterministic analysis** examines each asset's capacity to serve additional demand compared to the worst-case projected PEV demand under the defined scenario. Each asset's capacity is determined via the circuit models and the projected PEV load is derived from probabilistic evaluations of PEV characteristics and number of customers served from each asset.
- **Stochastic analysis** projects likely impacts considering the full projected diversity of the PEV charging through randomly generated system scenarios that model PEV charging and system response over a full calendar year. PEV load location and temporal demands are randomly determined using the PEV characteristic probability distributions discussed above. Results from the simulation and analysis of hundreds of these randomly generated cases provide indications of likely impacts and their severity.

MARKET PENETRATION AND CLUSTERING

The study is based on projected market penetration 1 to 5 years after PEV commercialization. Although the total penetration is assumed to be small, possible

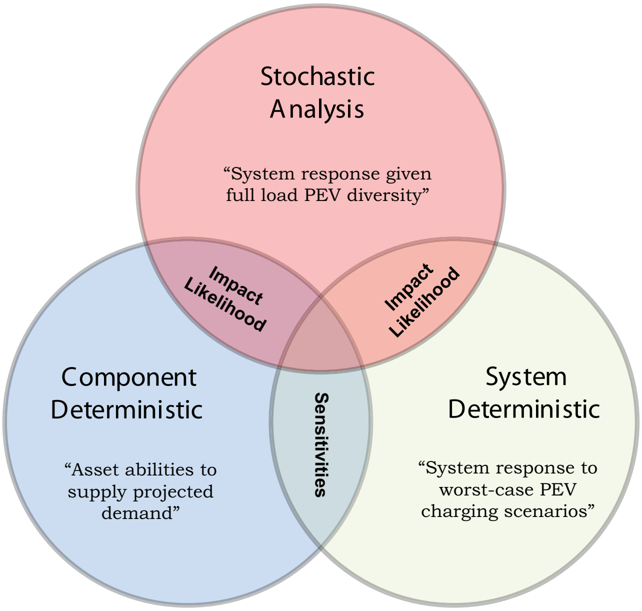


FIGURE 1 System impact analysis framework.

high localized concentrations are possible. Using known distribution system circuit information, PEV charge characteristics, and likely customer behaviors to construct models of system conditions, the analysis framework considers the following principal factors that define PEV loading on distribution systems: PEV market penetration levels per utility customer class (residential, commercial); different PEV charge spectrums (battery type, charger efficiency) and profiles; time profiles and likely customer charging habits; and battery state of charge based on miles driven.

To evaluate circuits from 19 utility operating territories, PEV adoption levels in the range of 2–25% were used. It's important to note that, even for low overall customer PEV adoption rates, based on system configuration and assumed customer adoption probabilities PEV clusters will occur randomly throughout the system, as shown in Figure 2. Each PEV is represented by a circle, and as PEVs are introduced at the same location they are spaced like petals on a flower. Detailed analysis from 36 circuits in 19 utility operating territories revealed a penetration pattern that resembles sparse clusters that are nonuniform, centered on early adopter neighborhoods. Several of these distribution system segments have older homes and are capacity constrained. Higher penetration rates, of course, increase the potential for larger and more numerous clusters. Although these clusters may indicate an increased risk of higher than average loading levels, clustering alone does not signify the likelihood of negative impact because other PEV load characteristics must be taken into account.

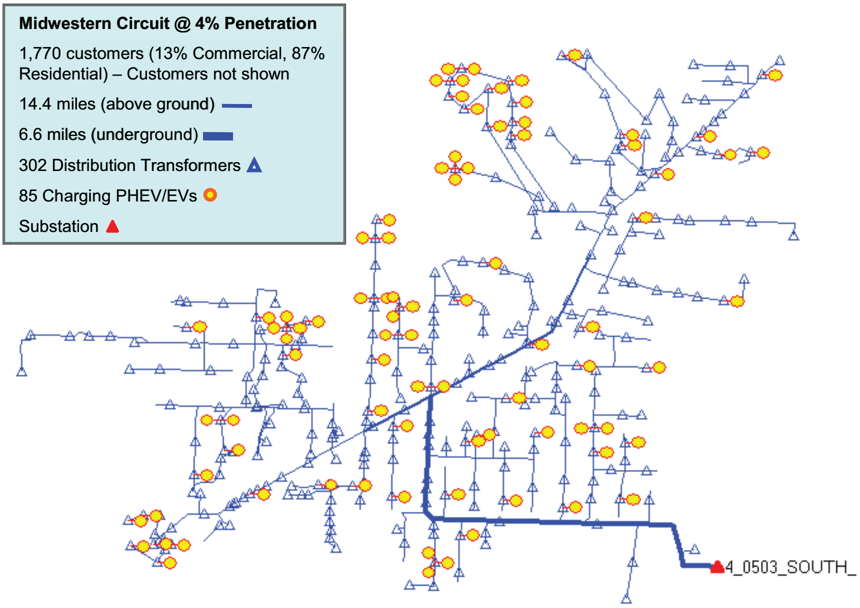


FIGURE 2 Sample daisy plots illustrating clustering at 4% penetration level. EV = electric vehicle; PHEV = plug-in hybrid electric vehicle.

CHARGING INFRASTRUCTURE

There are several ways to recharge PEVs at power levels ranging from less than 1 kilowatt (kW) to as much as 250 kW and at charging times of less than 30 minutes to more than 24 hours. Most residential and public charging will occur at power levels ranging from less than 1 kW to as much as 19.2 kW, with full charge times of 3–8 hours.

Charging is grouped into two classifications based on whether the electricity delivered is alternating current (AC) or direct current (DC). AC charging is governed by SAE Recommended Practice J1772 and has two classification levels in North America. Level 1 charging delivers 120 volts AC (VAC), and the electric vehicle supply equipment (EVSE[1]) generally consists of a self-contained cordset that terminates in a standard NEMA 5-15R plug compatible with any standard 120 volt household outlet. Level 2 charging delivers 208–240 VAC and requires a permanently connected EVSE. Level 1 AC charging is generally limited to

---

[1] EVSE can be defined as "The conductors, including the ungrounded, grounded, and equipment grounding conductors, the electric vehicle connectors, attachment plugs, and all other fittings, devices, power outlets or apparatuses installed specifically for the purpose of delivering energy from the premises wiring to the electric vehicle" (NEC 1996).

1.44 kW, Level 2 can reach 19.2 kW, and most vehicles and installations use 3.3–6.6 kW. Both Level 1 and Level 2 charging use the same connector design at the vehicle, and most vehicles can charge at either voltage through the same charge port.

Instead of an onboard charge port, DC charging, often referred to as “fast charging,” converts AC electricity to DC and directly charges the vehicle battery through an offboard charging station. BEVs have been designed and tested for DC charging at rates of 50–60 kW; the maximum charging power depends on the battery chemistry and system design.

Most electric vehicles are expected to charge at power levels below 7 kW (although the residential charging standard can reach levels of 19.2 kW, or 80 amps at 240 volts). PHEVs can easily recharge overnight at Level 1 (120 V, 1.2 kW) or Level 2 (240 V, 3.3 kW). The specific impacts on a feeder will depend on the design and loading practices for various components of the feeder and characteristics of PEVs in the area.

Charging Patterns

The timing of PEV charging can have either positive or negative impacts on electric generation and transmission systems. A significant amount of PEV charging coincident with the system peak would create a need for additional generation, whereas charging performed consistently during off-peak hours could reduce system costs.

Figure 3 compares the maximum charge powers for Level 1 (120 V) and Level 2 (240 V) EVs to average peak summer demand for households in eight US cities with different climates. Likely implementations of residential Level 2 charging range from a 15 amp circuit (12 amp continuous, 2.88 kW) to a 100 amp circuit (80 amp continuous, 19.2 kW). Higher-capacity EVSE installations are more likely to affect the local distribution system.

It is often assumed that EV charging could create a large load coincident with the peak. However, according to data from the National Personal Transportation Survey,[2] vehicles do not all connect at the same time. Figure 4 shows the distribution of home arrival times (on a 24-hour clock) for average American drivers. Even during the peak hour of 5–6 PM (17–18 on the x-axis), only about 12% of drivers arrive home.

Aggregate Feeder Loading Analysis

Characterization of PEV load diversity’s influence on the system requires examination of the total additional loading expected at the substation (head of

---

[2] US Department of Transportation Federal Highway Administration, www.fhwa.dot.gov/ctpp/jtw/contents.htm.

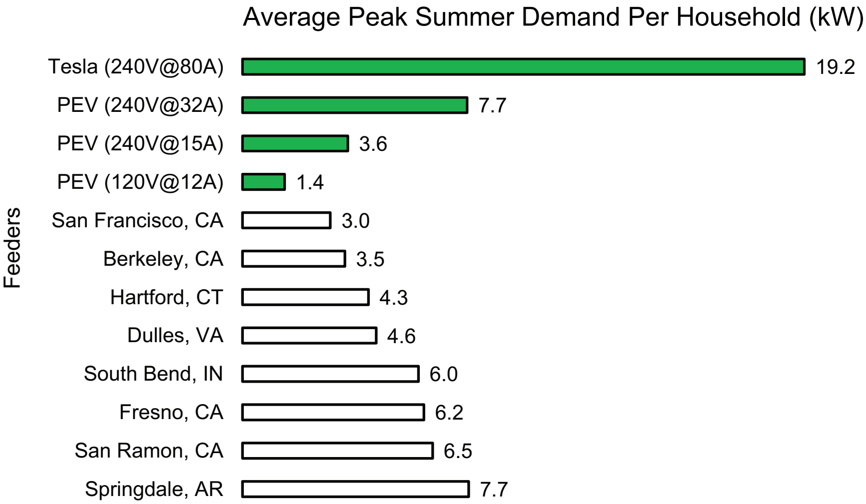


FIGURE 3 Comparison of power consumption for AC levels 1 and 2 charging and for average peak summer household demand in eight US cities. A, amp; kW, kilowatt; PEV, plug-in electric vehicle; V, volt.

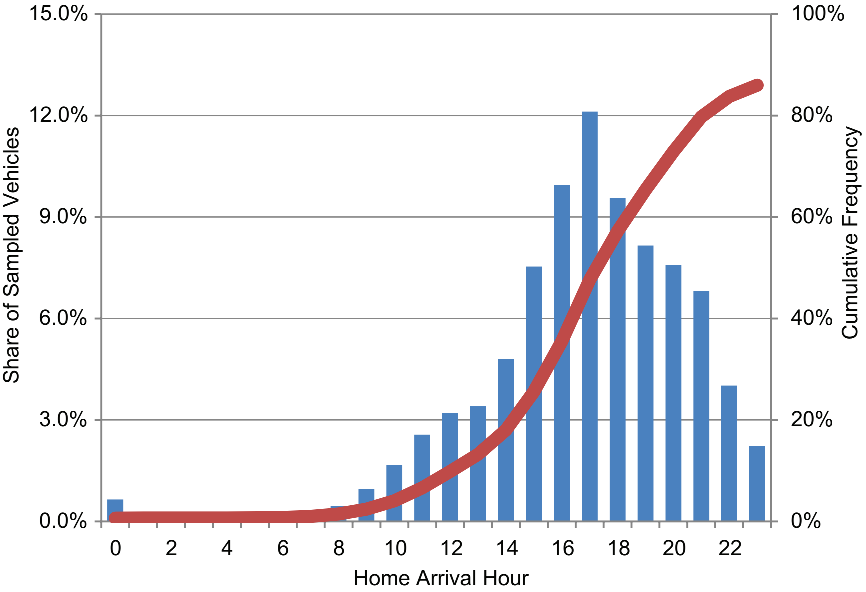


FIGURE 4 Home arrival time distribution.

the feeder) for each circuit. There are uncertainties in the expected makeup of PEVs, charging patterns served from each feeder, and customer habits, but they can be reasonably bounded at the aggregate level for the substation transformer. The study results showed that, based on typical daily driving statistics, the average energy delivered to a midsize sedan during a charge is 5–8 kWh and that for different vehicle mixes the aggregate on-peak load will vary between 700 W and 1100 W per PEV.

Charging patterns at the aggregate level correlate with statistical driving patterns, according to data from the National Household Transportation Survey (NHTS; Vyas et al. 2009). Potential hours of PEV connection to the distribution grid were derived from the likely residential customer home arrival times shown in Figure 4. It is possible to estimate aggregate hourly demand on the substation transformer by coupling NHTS statistics with daily customer driving distance patterns, PEV types (e.g., Chevy Volt, Nissan Leaf, Ford Focus, Mitsubishi iMieV), electrical charger characteristics, and different charging profiles that can be used to control charging.

Figure 5 shows a plausible case for vehicle charging based on a fleet of extended-range electric vehicles (E-REVs, as represented by the Chevy Volt; 30%), blended PEVs (represented by the Ford Escape; 50%), and BEVs (represented by the Nissan Leaf; 20%), all with 7.68 kW chargers that begin charging at full power upon arriving home. Although the charging occurs at peak load,

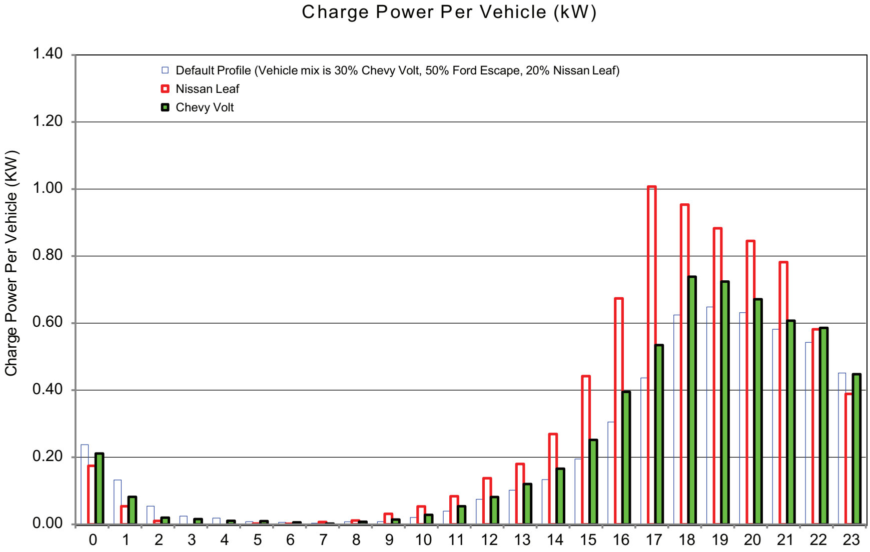


FIGURE 5 Aggregate power demand for uncontrolled vehicle charging.

it uses about 0.7–1.0 kW per vehicle. Other vehicle mixes, with more PEVs or lower-power chargers, will decrease the vehicle charging peak. Similarly, higher-power chargers will increase the vehicle charging peak but the charging will finish sooner.

SYSTEM IMPACTS

Correlating expected demand against asset capacity will provide a strong indicator of the number and type of assets most at risk of exceeding their thermal ratings due to PEV adoption. Peak capacity is determined from the peak load power flow solution and each component's specified thermal ratings. The EPRI analysis shows that higher charging levels/rates (6.6 kW versus 3.3 kW versus 1.4 kW) have a greater impact on transformer capacity, as illustrated in Figure 6.

The calculated peak hour remaining capacities for an example circuit are plotted in Figures 7 and 8 as a function of the number of customers served. Each asset is evaluated against projected PEV demands and its remaining capacity plotted as an individual point and sorted based on customers served and asset class. Projected demands are superimposed as lines for the three market penetration levels (2%, 4%, or 20%) examined. The estimated maximum PEV demand

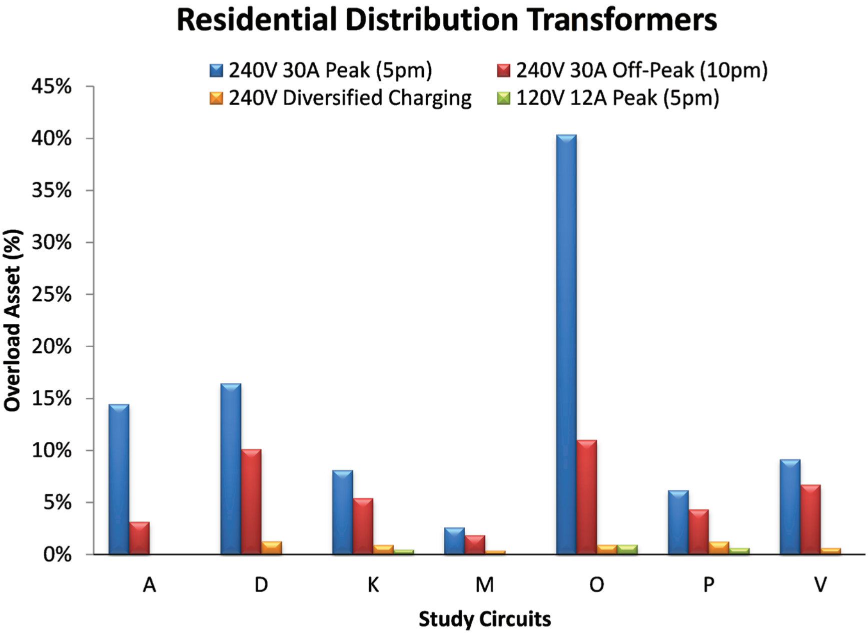


FIGURE 6 PEV charge levels have a stronger impact compared to charge time.

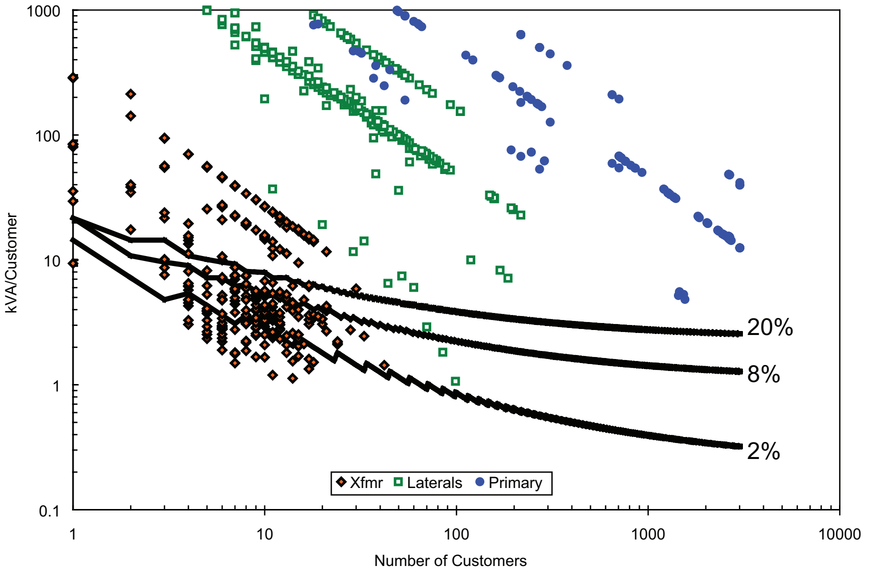


FIGURE 7 Feeder asset thermal overload risk evaluation for 240V 30A PEV charging.

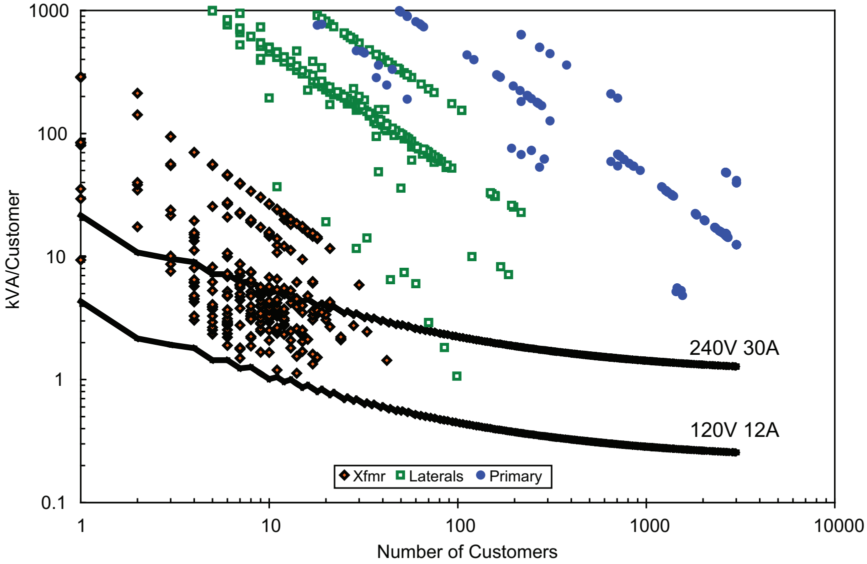


FIGURE 8 Service transformer overload risk evaluation 120V 12A and 240V 30A PEV charging.

is also plotted, permitting the ready identification of assets that are at risk of impact. Assets whose remaining capacity falls above the projected demand are unlikely to be affected by 2%, 4%, or 20% PEV market penetration. Given the 99.99% confidence level used for the P-test and the conservative construction of the maximum projected demand lines, the probability of exceeding the thermal ratings of these assets is less than 0.01%. Thus, assets above the maximum project lines are unlikely to be impacted (where the asset's remaining capacity exceeds the projected PEV demand lines) and can be quickly identified for different PEV penetration levels.

As PEV market penetration increases so does the potential for system impacts (although such impacts cannot be discounted even for penetrations as low as 2%). As expected, the number of assets that fall below the projected maximum PEV demand line increases with the penetration level. Furthermore, the nature of the asset capacities in relation to the maximum PEV demand lines clearly indicates that the impact of PEV adoption will most likely first appear on service transformers. Not surprisingly, transformers with the lowest capacity per customer are the most susceptible.

It is important to note that these circuit models, based on allocation of customer load per transformer kVA, may limit the accuracy of the projections because they do not capture innate variations in transformer loadings. Thus transformers that may be heavily loaded in the field cannot be completely discounted from being overloaded due to PEV charging. In the analysis described here, impact likelihood is determined through stochastic simulations of the circuit operation over a full calendar year for projected PEV penetration levels. In each case, PEVs of specific types are randomly assigned to customer locations according to defined probability distribution function (PDF) and an hourly demand profile for the year is developed from the charge time and remaining charge PDFs. This process is repeated for each penetration level and the simulated results are aggregated to provide an indication of impact likelihood (the analysis also accounts for other system impacts such as steady-state voltage changes and losses). The stochastic analyses are also designed to enable identification of the particular system and PEV conditions that result in a negative impact to the system or asset.

CONCLUSIONS

The results of the EPRI study show the following:

- The extent of system impacts depends on PEV penetration and charging behaviors of PEV adopters.
- The expected aggregate addition to system peak loads is 700–1,000 watts per PEV in a given utility territory.
- Short-term impacts for most utilities should be minimal and localized. There is a possibility, however, of isolated and more severe impacts on

some distribution transformers and secondary service lines, particularly in neighborhoods with older distribution systems and underground systems.

- PEV charge rate, or level, is the PEV characteristic that most influences the overload risks posed to service transformers from PEV adoption. Increased charge durations, due to larger battery sizes, can also impact thermal aging.
- Each transformer's remaining capacity per customer is one of the strongest indicators of the potential risk that a transformer may exceed its thermal ratings. This metric incorporates a number of key factors including the transformer's existing demand, thermal ratings, and the number of customers served.
- Assets near the load are most susceptible to system overloads from PEV clusters as the potential benefit of spatial diversity decreases.
- PEV clustering will occur randomly throughout a system. While it may indicate an increased risk of higher than average loading levels, PEV clustering alone does not signify the likelihood of negative impact as other PEV load characteristics must also be taken into account.
- Transformers characterized by low capacity per customer are the most likely to be affected by PEV adoption. In particular, transformers lower than 25 kVA are expected to be the most susceptible to overloading as they typically have lower amounts of capacity, which can be quickly consumed by one or more PEVs.

REFERENCES

EPRI [Electric Power Research Institute]. 2012. Understanding the Grid Impacts of Plug-in Electric Vehicles–Phase 1 Study. TR 1024101. December.

Maitra A, Kook K, Taylor J, Giumento A. 2009. Evaluation of PEV loading characteristics on Hydro-Quebec's distribution system operations. Electric Vehicle Symposium (EVS)24, Stavanger, Norway, May 13-16. Available online at www.cars21.com/knowledge/papersView/47.

Mangharam R. 2013. The car and the cloud: Automotive architecture for 2020. In Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2012 Symposium. Washington, DC: National Academies Press.

NEC [National Electrical Code]. 1996. NEC Handbook, NFPA 70. Article 625, Electrical Vehicle Charging System Equipment, Section 625-2. Quincy MA: National Fire Protection Association.

Taylor J, Maitra A, Alexander M, Brooks D, Duvall M. 2009. Evaluation of the impact of PEV loading on distribution system operations. IEEE Power Engineering Society, Calgary, July.

Vyas A, Wang M, Santini D, Elgowainy A. 2009. Analysis of the 2001 National Household Transportation Survey in support of the PHEV project to evaluate impacts on electricity generation and GHG emissions. Unpublished information.

The Car and the Cloud
Automotive Architectures for 2020

Rahul Mangharam
University of Pennsylvania

Over the past two decades, automobile journey durations have doubled. Furthermore, travelers increasingly use their vehicle as a mobile office, meeting room, and even living room. With this evolution, informational and entertainment needs in the vehicle now transcend mechanical, electronic, and software boundaries to include services for the driver, passengers, and the vehicle itself.

Three trends are emerging in drivers' expectations for their vehicle: (1) continuous connectivity with both the infrastructure (e.g., smart traffic intersections) and other commuters, (2) enhanced levels of productivity and entertainment for the duration of travel, and (3) reduction in cognitive load through semiautonomous operation and automated congestion-aware route planning. To address these demands, vehicles should become more programmable so that almost every aspect of engine control, cabin comfort, connectivity, navigation, and safety will be remotely upgradable and designed to evolve over the lifetime of the vehicle.

Progress toward the vehicle of the future will entail new approaches in the design and sustainability of vehicles so that they are connected to networked traffic systems and are programmable over the course of their lifetime. To that end, our automotive research team at the University of Pennsylvania is developing an in-vehicle programmable system, AutoPlug, an automotive architecture for remote diagnostics, testing, and code updates for dispatch from a datacenter to vehicle electronic controller units. For connected vehicles, we are implementing a networked vehicle platform, GrooveNet, that allows communication between real and simulated vehicles to evaluate the feasibility and application of vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication; the focus in this paper is on its application to safety. Finally, we are working on a tool for large-scale traffic congestion analysis, AutoMatrix, capable of simulating more than 16 million vehicles on any US street map and computing real-time fastest paths for a large subset of vehicles. The tools and platforms described here are free and open source from the author.

PROGRAMMABLE VEHICLES

Vehicles today are built in long design cycles and with electronic architectures that are static in both form and function. Technology adoption is considered only at the beginning of the design cycle, frozen for the lifetime of ownership of the vehicle (~12 years[1]), and often obsolete within 6 years.[2,3] In contrast, the vehicle of the future will be programmable with services for the long-term health and performance of both humans and vehicles.

Electronics and software for engine and cabin controls currently account for over 30% of the cost of an automobile, and this figure is expected to grow as vehicles evolve from mechanical to electronic to software-controlled to service-based mobile cyber-physical system (CPS) platforms. As new automotive electronic architectures are developed to enable remote diagnosis and reprogrammability throughout the life of the vehicle, drivers will be able to choose from a software component marketplace to enhance the safety, performance, and comfort of their vehicle.

Ensuring the safe and correct programming of the new service features is paramount. Automotive plug-and-play devices that communicate to and from the vehicle will allow new classes of services and customization such as online vehicle diagnostics, warranty management, networked infotainment, and integration of applications such as driver behavior and vehicle performance measurements for personalized insurance services.

CONNECTED VEHICLES

Every year, approximately 6.4 million car accidents occur in the United States, typically involving three people (two drivers and one passenger). That translates to roughly 19.2 million Americans injured in car accidents each year, or odds of 1:16 for every individual. Several sources[4] estimate that over 90% of vehicle crashes are due to driver negligence and therefore avoidable (Duri´c and Miladinov-Mikov 2008).

---

[1] Polk.com. 2012. Average age of vehicles reaches record high, January 17. Available online at http://goo.gl/TN5Ow.

[2] US DOE Office of Energy Efficiency and Renewable Energy, Vehicle Technologies Program. 2010. Average Length of Light Vehicle Ownership, May 10. Available online at www1.eere.energy.gov/vehiclesandfuels/facts/2010_fotw622.html.

[3] Polk.com. 2012. Americans are keeping new vehicles an average of nearly six years, February 22. Available online at http://goo.gl/7R3N3.

[4] See, for example, *The Economist*. Look, no hands: Automotive technology: Driverless cars promise to reduce road accidents, ease congestion and revolutionise transport, September 1. Available online at www.economist.com/node/21560989.

A vehicle's "safety bubble" is currently limited to its physical body, with integrated crash and proximity sensors (e.g., ultrasonic, LiDAR, radar). In the vehicle of the future, V2V and V2I wireless communication is expected to enhance safety. Such communication technology, when interfaced with the vehicle's powertrain and using audio and haptic feedback, will be able to issue safety alerts to all approaching vehicles during events such as sudden braking, loss of traction, or airbag deployment. Early warning messages communicated down the highway in a timely "multi-hop" manner (i.e., from one vehicle to another in a few hundred milliseconds) will allow for longer reaction and stopping time and thus prevent a pile-up.

Connected vehicle architectures for such safety-critical automotive systems require much work to ensure security and privacy together with the timely delivery of traffic alerts, warnings, and information updates.

NETWORKED TRAFFIC SYSTEMS

Delays due to traffic congestion cost Americans $78 billion in the form of 4.2 billion lost hours and 2.9 billion gallons of wasted fuel, and 35–55% of these delays are caused by point-based traffic incidents rather than recurring congestion. As the density of vehicles increases, there is a need for large-scale traffic congestion management such that real-time "eco-routing" can be provided to prevent, avoid, and alleviate traffic back-ups. Models and tools for nationwide traffic congestion management, with networked streaming vehicle data, are required to compute the fastest and most eco-friendly routes without new infrastructure costs.

In the Real-Time Systems Lab at Penn, we are investigating the design of such a platform to enable the scaling of traffic network operations to handle data processing for millions of vehicles, estimate and predict congestion, and facilitate route assignment as well as to model traffic operations and disaster response during congestion.

IN-VEHICLE SYSTEMS: REMOTE DIAGNOSTICS, TESTING, AND REPROGRAMMING

More than 20.3 million vehicles were recalled in 2010, many because of software issues related to electronic systems such as cruise control, antilock braking, traction control, and stability control. New and scalable methods are necessary to evaluate such controls in a realistic and open setting.

The increasing complexity of software in automotive systems has resulted in the rise of firmware-related vehicle recalls due to undetected bugs and software faults.[5] In 2009, Volvo recalled 17,614 vehicles because of a software error in the engine-cooling fan control module that could result in engine failure and possibly

---

[5] IEEE Spectrum. 2011. Honda recalls 936,000 more vehicles for electrical and software fixes, September 7. Available online at http://spectrum.ieee.org.

lead to a crash (NHTSA 2009). In August 2011, Jaguar recalled 17,678 vehicles because of concerns that the cruise control might not respond to normal inputs and once engaged could not be switched off.[6] In November 2011, Honda recalled 2.5 million vehicles to update the software that controls its automatic transmissions.[7]

Current automotive systems lack a systematic approach and infrastructure to support postmarket runtime diagnostics for control software (although at least one online source indicates that there is a significant effort to incorporate automotive software testing and verification at the design stage[8]). Once a vehicle leaves the dealership lot, its performance and operation safety are a "black box" to the manufacturers and the original equipment providers.

Furthermore, for the more than 100 million lines of code and 60-plus electronic controller units (ECUs) in a vehicle (Schäuffele and Zurawka 2005), there are only about eight standard diagnostic trouble codes (DTCs) for software and they are extremely general (e.g., "memory corruption"). Of the DTCs for software, none targets the ECU software even though systems such as stability, cruise, and traction control are critical for safety.

In-Vehicle Diagnostics and Recall Management

The current approach to vehicle recalls is reactive: the manufacturer recalls all vehicles of a particular year/make/model only after a problem occurs in a significant number of them. For a software-related recall, the vehicle is taken to a service center and a technician either manually replaces the ECU that has the faulty code or reprograms the ECU code with the new version provided by the manufacturer.

The wait-and-see approach to recalls has a significant cost in both time and money and may have a negative impact on the vehicle manufacturer's reputation. Furthermore, the current recall method relies on word of mouth or the transmission of manually logged information from the service centers to the manufacturer, which takes time—during which a safety-critical system may malfunction.

Consequently, there is an urgent need for systematic postmarket in-vehicle diagnostics for control system software so that issues can be detected early. An in-vehicle system would log sensor values and perform runtime evaluation of the states of the system controls. A remote diagnostic center (RDC) would receive the data (over a network link) to prepare a fault detection and isolation response (Figure 1), in the form of a proposed dynamic diagnostic trouble code (DyDTC) that "observes" the ECUs and system control tasks in question. Once sufficient data are captured, the RDC, using a gray-box model of the vehicle (i.e., with

---

[6] IEEE Spectrum. 2011. Jaguar software issue may cause cruise control to stay on, October 25. Available online at http://spectrum.ieee.org.

[7] Reuters. 2011. Honda recalls 2.5 million vehicles on software issue, August 25. Available online at www.reuters.com.

[8] AUTOSAR (Automotive Open System Architecture); www.autosar.org.

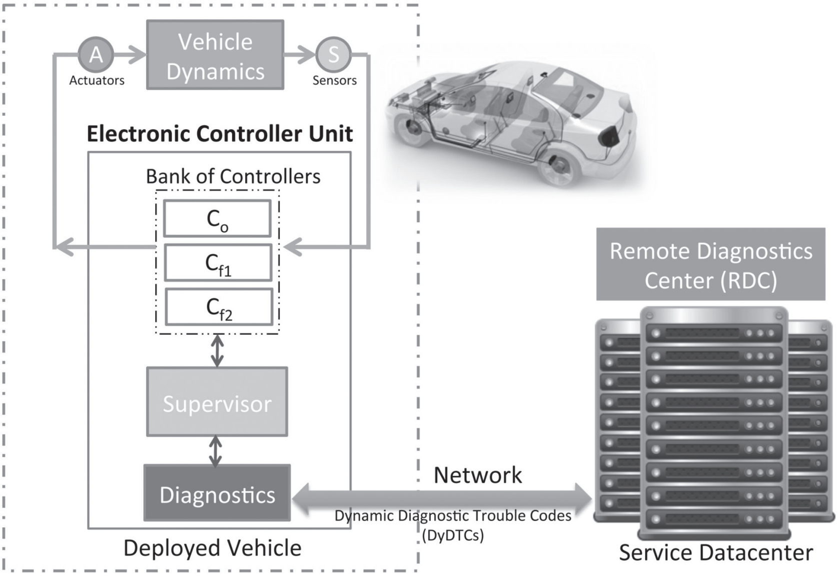


FIGURE 1 Remote diagnostics of automotive control systems showing the vehicle's software architecture (in dashed box) and the remote diagnostics center (RDC) communicating over a network link. The RDC communicates via the onboard "supervisor" with the vehicle control system to observe its state and update its software in the event of an unexpected fault. C_0, C_{f1}, C_{f2} = software-based controllers in the vehicle (e.g., for stability, traction, antilock braking, and cruise control). Using dynamic diagnostic trouble codes (DyDTCs), the RDC observes the state of the vehicle software for postmarket analysis of unanticipated faults.

sensor and control system observation logs), executes system identification to build a model of the vehicle. It then develops a fault-tolerant controller to address the problem and the vehicle is remotely reprogrammed by a code update.

We have developed an early design of such a system, AutoPlug, although we recognize that the approach will be difficult in practice because it would require extensive runtime verification of the updated controller.

Overview of AutoPlug

AutoPlug is an automotive ECU architecture between the vehicle and an RDC to diagnose, test, update, and verify control software. Within the vehicle, we evaluate observer-based runtime diagnostic schemes and introduce a framework for remote management of vehicle recalls. The diagnostic scheme deals with both real- and non-real-time faults, with a decision function to detect and isolate system faults with modeling uncertainties.

We also evaluate the applicability of “opportunistic diagnostics,” where the observer-based diagnostics are scheduled in the ECU’s real-time operating system (RTOS) only when there is slack available in the system (i.e., it can work with existing hardware in vehicles without interfering with current task sets). The performance of this aperiodic diagnostic scheme is similar to that of the standard, periodic scheme under reasonable assumptions. The framework integrates in-vehicle and remote diagnostics and makes vehicle warranty management more cost-effective.

The aim of the AutoPlug architecture, illustrated in Figure 2, is to make the vehicle recall process less reactive with a runtime system for diagnosis of automotive control systems and software. Our focus is on the online analysis of the control system and control software both in the vehicle ECU network and between the vehicle and the RDC. We assume the network link between the two is available.

The runtime system within the vehicle manages:

- **Fault detection and isolation.** Sensor, actuator, and control system states are logged for the specific ECU. The data are analyzed locally, and a summary of the states is transmitted to the RDC.
- **Fault-tolerant controllers.** Once a fault is detected, the high-performance controller is automatically replaced with a backup controller.
- **ECU reprogramming for remote code updates.** Upon receipt of reformulated controller code from the RDC (which will guarantee the stability and safety of the vehicle), the runtime system reprograms the particular controller task(s) with the updated code. This can be done over a cellular or wireless communication link.
- **Patched controller runtime verification.** The updated code is monitored with continuous checks for safety and performance.

While the onboard system provides state updates of the specific controller, the RDC provides complementary support through:

- **Data analysis and fault localization.** By observing sensor and control system operations locally, structured system identification is used to create a model of the vehicle and its control system is evaluated to isolate faulty behavior.
- **Reformulation of control and diagnostic code.** A new controller is formulated for the specific vehicle model and further diagnostic code dispatched.
- **Recall management.** Reformulated controller code is transmitted to the vehicle.

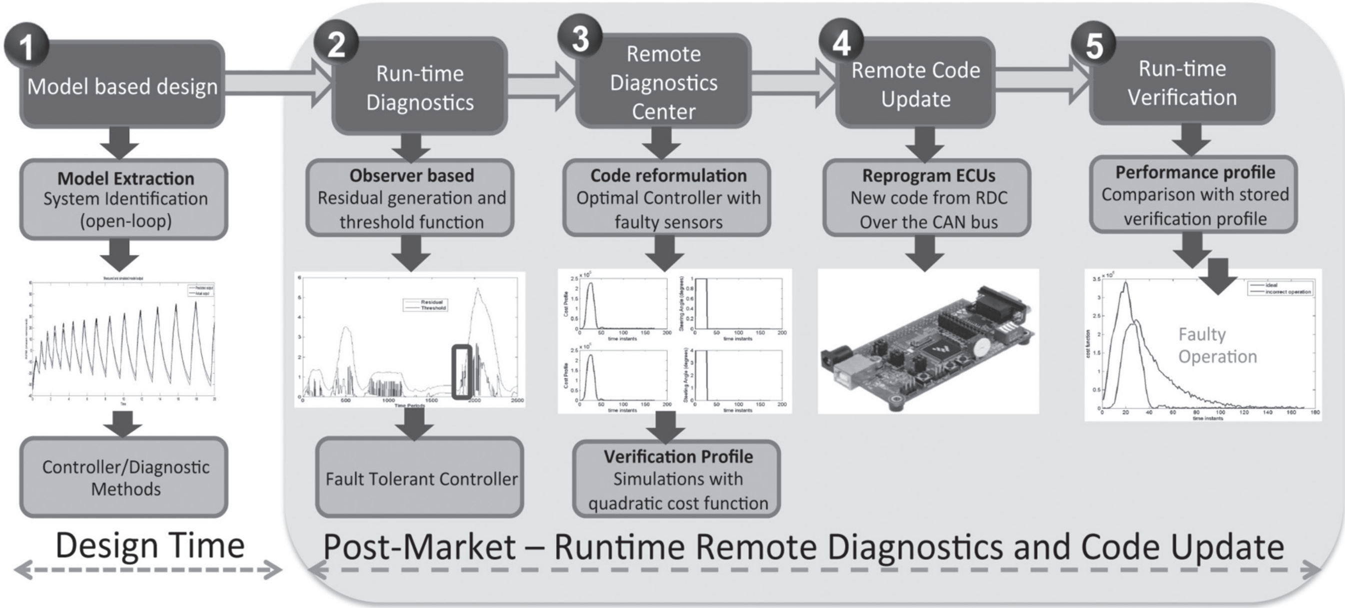


FIGURE 2 End-to-end stages of the AutoPlug automotive architecture. (1) When an unexpected fault is reported, the remote diagnostic center (RDC) sends custom diagnostic code to the vehicle to observe its performance. Using vehicle models developed during the design phase, the RDC safely observes the operation of the software on the vehicle while it is running. Using this information it extracts a new model for the vehicle (perhaps with changes due to wear and tear, faulty sensors, changes in suspension). (2-3) With the updated vehicle model, the control system design is reformulated to correct the faults in the vehicle. (4-5) The RDC remotely updates and verifies the correctness and safety of the reformulated control software. CAN = controller area network; ECU = electronic controller unit.

- **Generation of controller verification profiles.** The updated controller is probed for performance and safety.

The remote diagnostic system is capable of diagnosing and reformulating controllers with real-time faults (e.g., delay, jitter, incorrect sampling rates) and system faults (e.g., stuck-at faults, calibration faults, and noise in sensors/actuators).

AutoPlug Testbed

To design and validate the proposed architecture we developed the AutoPlug testbed, which consists of a hardware-in-loop simulation platform for ECU development and testing (Figure 3). The hardware is in the form of a network of ECUs, interfaced by a controller area network (CAN) bus, on which we implement the control and diagnostic algorithms. Each ECU runs a nano-RK RTOS, a resource kernel (RK) with preemptive priority-based real-time scheduling.

Instead of a real vehicle, we use an open-source racecar simulator, which provides high-fidelity physics-based vehicle models and different road terrains, thus affording both the realism of an actual vehicle and the flexibility to implement our own code. In addition, we can introduce faults not covered by standard DTCs. We have tested basic control algorithms, running as real-time tasks on nano-RK, for antilock braking systems (ABS), traction control, cruise control, and stability control to see that the testbed does indeed perform as a real vehicle would.

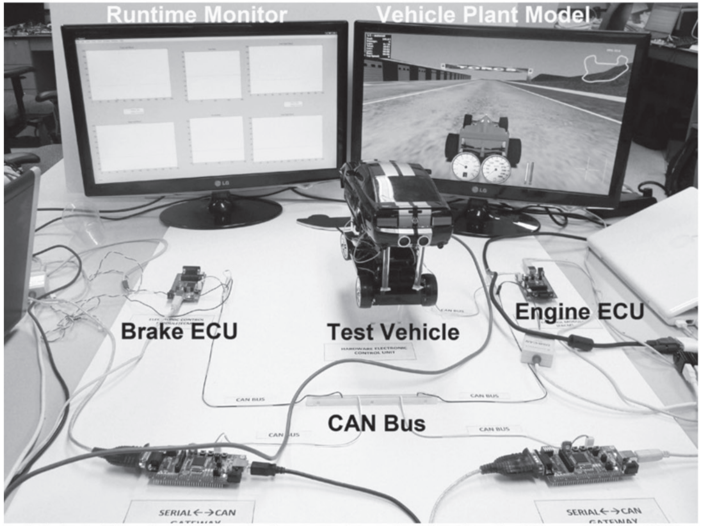
The main contributions of our applied research and development are threefold:

- an architecture that uses both in-vehicle and remote diagnostics for remote recall management of deployed vehicles;
- modification of the traditional observer-based fault detection and isolation scheme for in-vehicle opportunistic diagnosis, as well as an experimental thresholding scheme in the presence of modeling uncertainties; and
- implementation and evaluation of these schemes on real ECUs for hardware-in-loop simulation.

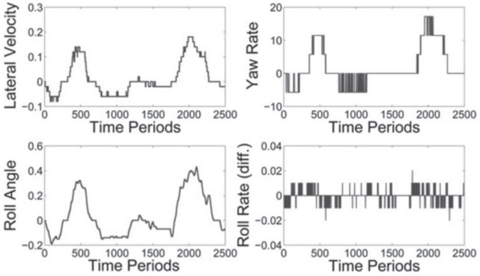
These three features facilitate postmarket diagnostics, testing, and reconfiguration from a remote data center.

VEHICLE-TO-VEHICLE/INFRASTRUCTURE NETWORKING FOR ENHANCED SAFETY

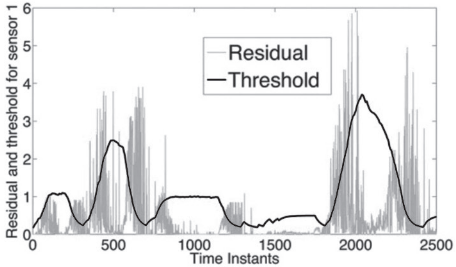
Connected vehicles involve a special class of wireless networks where the maximum relative speeds are in excess of 80 meters per second, the node density can span more than 9,000 vehicles/mi^2, and, most importantly, the dynamics of



(a) AutoPlug hardware-in-loop automotive testbed with multiple ECUs



(b) Real-time monitoring of stability control system with vehicle's lateral velocity, yaw rate and roll angle



(c) Remote diagnostics showing the error detection capabilities

FIGURE 3 (a) AutoPlug hardware-in-loop testbed with real-time monitoring and diagnostics. CAN = controller area network; ECU = electronic controller unit. (b) Real-time monitors for stability controller showing the sensor information of the vehicle dynamics. (c) Analysis of the error signal (i.e., residual) of a particular sensor and its expected values. A smart thresholding scheme is used at the remote diagnostics center to determine the extent of the fault based on the residual signal.

the vehicle, the environment, driver reaction, and interaction with other vehicles are considered in every communication and control decision. Vehicles enabled with programmable short-range wireless networking can communicate with each other and with the infrastructure to enhance the driver's perception of oncoming danger within hundreds of milliseconds and, within seconds or minutes, route the vehicle based on real-time traffic congestion.

With connected vehicles, it is necessary to analyze and validate the effect of incremental deployment of V2V technologies on message delay, coverage, and persistence in the region of interest. Because it is expensive to develop and test experimental protocols on a large fleet of vehicles, there is a need for vehicular network simulators that faithfully model first-order effects of the street topology, vehicle congestion, speed limits, communication channels, and spatiotemporal trends in traffic intensity on the performance and reliability of V2V networking. Once protocols are designed and evaluated through simulation, their performance must be tested with real vehicles and realistic traffic densities. Although it may be possible to deploy a small fleet of vehicles (e.g., a dozen), it is not yet possible to assess the scalability of such protocols in rush-hour bumper-to-bumper vehicle densities.

GrooveNet Connected Vehicle Virtualization Platform

We have developed the GrooveNet vehicular network virtualization platform to simulate thousands of vehicles on any street map and communicate between real and simulated vehicles. GrooveNet supports a variety of models, network and vehicular system interfaces, message types, and operating modes and, by using the same protocols, algorithms, and software implementation in both real and virtual vehicles, facilitates model-based design, model validation, graceful deployment, and rapid prototyping. It works as both a simulator and in-vehicle network platform with connections to the CAN bus and radios using the recently standardized dedicated worldwide spectrum for vehicular communications (IEEE 802.11p/WAVE standard), a GPS unit, and a cellular interface.

Our tests of GrooveNet with a fleet of five vehicles over 400 miles across urban, rural, and suburban terrain show that it has realistic models for car following, communication, mobility, driver types, traffic lights, road-side communication nodes (e.g., wireless stations that transmit updates about traffic lights to enable drivers to adjust their speed accordingly), and other interactive features of real-time driving. Each GrooveNet-enabled vehicle is capable of tight time synchronization via the GPS pulse-per-second signal for time-critical multi-hop communication. Using this platform we will develop a suite of V2V and V2I safety communication protocols to relay traffic incident alerts and warnings of unsafe road conditions in the Philadelphia and Pittsburgh areas.

Simulated and Actual Use of GrooveNet

Figure 4 shows three real vehicles (in the circles), which I refer to here as *R1*, *R2*, and *R3* (from left to right). The first two vehicles are within communication range; *R3*, over a mile away, is not. Thus if a safety alert is triggered by an airbag deployment in *R1*, only *R2* receives the message. To illustrate the progression of the message to approaching vehicles, we simulate virtual vehicles on the same road, each of which will enable a "hop" for the data transmission. *R2* sends the message over a cellular link to the vehicle operations director, which simulates the progression of the message from one to another of the virtual vehicles (V_1, V_2,…) until another real vehicle is in the vicinity of the virtual vehicles. The message thus travels across multiple hops to be received by *R3* over the cellular link as if it were from *R1*. We mask the cellular link's latency by speeding up the simulated communication across the virtual vehicles.

All vehicles follow the same rebroadcast policy, observe the posted speed limit, and obey car following standards. Vehicle density can be increased arbitrarily and its effects observed by a driver in a real vehicle on the road. Varying the number of virtual vehicles enables us to study the performance of the protocols and network algorithms under various densities, driving conditions, and street topologies. As more experimental vehicles become available, we can increase the realism and validation of our models. In the meantime, network virtualization provides the best of both model-based design and real-world validation with rapid prototyping, with only a few real vehicles needed to operate as mobile gateways.

Figure 5 presents a screen shot of GrooveNet implemented in Linux. In the top left panel is the list of simulated and real vehicles with their current position, street speed, and heading (i.e., direction). The top right panel provides a visualization of the current position and heading of vehicles in Pittsburgh, Pennsylvania. Small circles designate vehicles; circles around a dark arrow represent vehicles that rebroadcast an alert message. The bottom panel shows network connectivity between real vehicles via a wireless communication using the 802.11p/WAVE radio interface and between real and virtual vehicles over the cellular network. For this test we drove five real vehicles along Forbes Avenue in Pittsburgh and conducted experiments with more than 4,000 virtual vehicles.

Such hybrid simulation provides application users with an intuitive feel of the impact of communicating vehicle density on packet delivery ratio and event response time, and provides the developer with feedback about accuracy and details needed in the simulation models. This network virtualization will make it possible to answer questions such as: Under what driving conditions and market penetration of networked vehicles will application *A* achieve the desired performance? How does the probability distribution of model *M* compare with the real world? Is the resultant powertrain response safe and under what conditions is it unsafe?

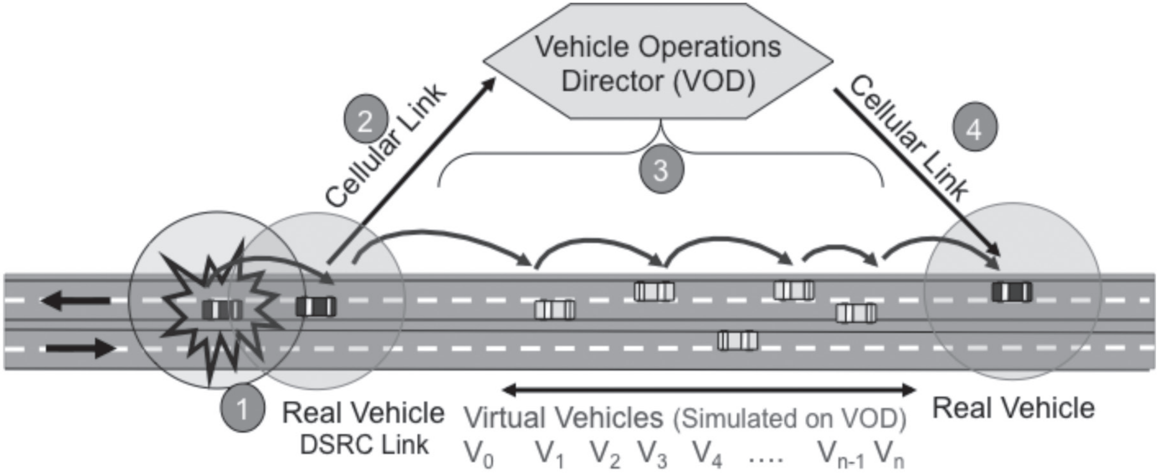
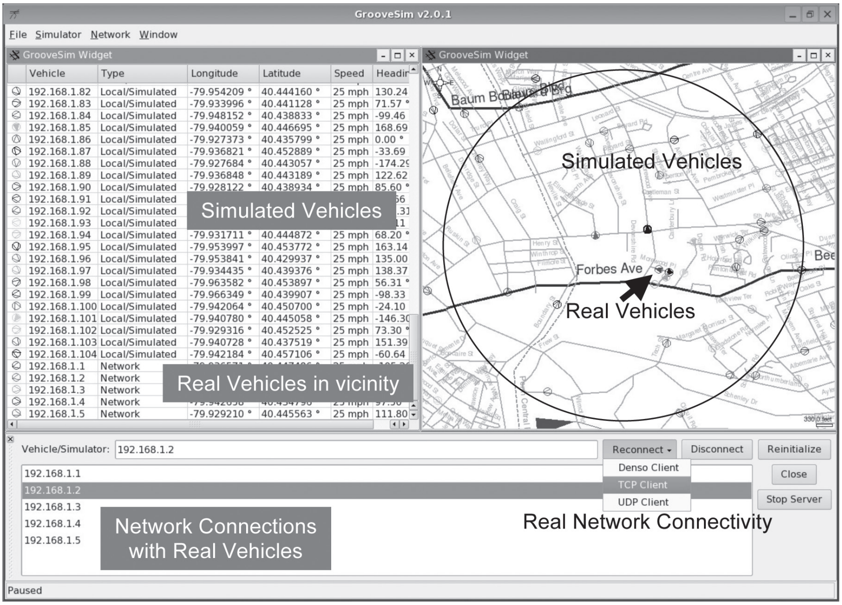


FIGURE 4 Mixed evaluation of real and virtual connected vehicles with the GrooveNet platform. The three vehicles in the circles are real vehicles communicating with short-range wireless communication (using the IEEE 802.11p/WAVE protocol) on the street. The remaining vehicles are simulated to facilitate communication between real and virtual vehicles. This platform allows for scalable and high-fidelity evaluation of vehicle-to-vehicle and vehicle-to-infrastructure network protocols. DSRC = dedicated short-range communication; V = virtual vehicle.



	Vehicle	Type	Longitude	Latitude	Speed	Headir
	192.168.1.82	Local/Simulated	-79.954209 °	40.444160 °	25 mph	130.24
	192.168.1.83	Local/Simulated	-79.933996 °	40.441128 °	25 mph	71.57 °
	192.168.1.84	Local/Simulated	-79.948152 °	40.438833 °	25 mph	-99.46
	192.168.1.85	Local/Simulated	-79.940059 °	40.446695 °	25 mph	168.69
	192.168.1.86	Local/Simulated	-79.927373 °	40.435799 °	25 mph	0.00 °
	192.168.1.87	Local/Simulated	-79.936821 °	40.452889 °	25 mph	-33.69
	192.168.1.88	Local/Simulated	-79.927684 °	40.443057 °	25 mph	-174.2
	192.168.1.89	Local/Simulated	-79.936848 °	40.443189 °	25 mph	122.62
	192.168.1.90	Local/Simulated	-79.928122 °	40.438934 °	25 mph	85.60 °
	192.168.1.91	Local/Simulated	[illegible]	[illegible]	[illegible]	[illegible]
	192.168.1.92	Local/Simulated	[illegible]	[illegible]	[illegible]	[illegible]
	192.168.1.93	Local/Simulated	[illegible]	[illegible]	[illegible]	[illegible]
	192.168.1.94	Local/Simulated	-79.931711 °	40.444872 °	25 mph	68.20 °
	192.168.1.95	Local/Simulated	-79.953997 °	40.453772 °	25 mph	163.14
	192.168.1.96	Local/Simulated	-79.953841 °	40.429937 °	25 mph	135.00
	192.168.1.97	Local/Simulated	-79.934435 °	40.439376 °	25 mph	138.37
	192.168.1.98	Local/Simulated	-79.963582 °	40.453897 °	25 mph	56.31 °
	192.168.1.99	Local/Simulated	-79.966349 °	40.439907 °	25 mph	-98.33
	192.168.1.100	Local/Simulated	-79.942064 °	40.450700 °	25 mph	-24.10
	192.168.1.101	Local/Simulated	-79.940780 °	40.445058 °	25 mph	-146.3
	192.168.1.102	Local/Simulated	-79.929316 °	40.452525 °	25 mph	73.30 °
	192.168.1.103	Local/Simulated	-79.940728 °	40.437519 °	25 mph	151.39
	192.168.1.104	Local/Simulated	-79.942184 °	40.457106 °	25 mph	-60.64
	192.168.1.1	Network	[illegible]	[illegible]	[illegible]	[illegible]
	192.168.1.2	Network	[illegible]	[illegible]	[illegible]	[illegible]
	192.168.1.3	Network	[illegible]	[illegible]	[illegible]	[illegible]
	192.168.1.4	Network	[illegible]	[illegible]	[illegible]	[illegible]
	192.168.1.5	Network	-79.929210 °	40.445563 °	25 mph	111.80

Vehicle/Simulator: 192.168.1.2

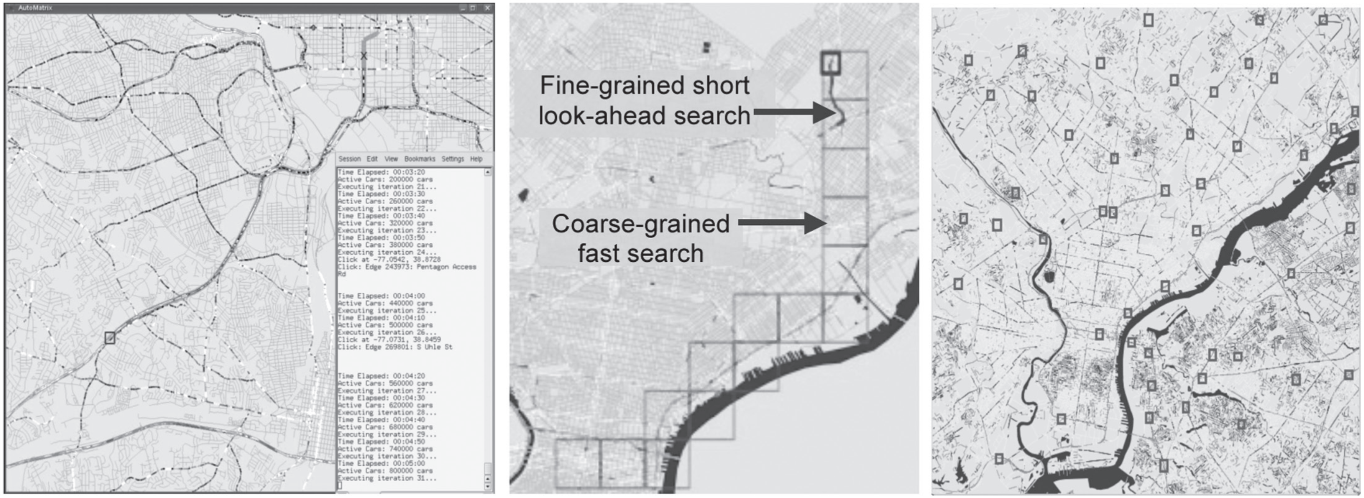
192.168.1.1
192.168.1.2
192.168.1.3
192.168.1.4
192.168.1.5

FIGURE 5 GrooveNet hybrid simulation demonstrating hundreds of virtual vehicles communicating with five real vehicles in the city of Pittsburgh, Pennsylvania. See text for discussion.

TRAFFIC CONGESTION ANALYSIS

To better understand empirical models of traffic congestion in different street topologies across the nation, and to develop sound traffic prediction and congestion-aware fastest-path routing algorithms, it is necessary to analyze large-scale traffic mechanisms. We have developed a traffic analysis tool, AutoMatrix, that simulates and routes over 16 million vehicles on any US street map and provides real-time traffic routing services with hierarchical and synthetic traffic matrices (Figure 6). Using this tool, we are able to investigate the design of adaptive routing strategies, methods to mitigate congestion, and ways to better use traffic network resources. Vehicles are modeled to be car following, have speed variations, communicate periodically, and be capable of multiple distributed and centralized routing algorithms.

AutoMatrix operates on a graphics processing unit (GPU) and so is capable of very large-scale microsimulation and traffic analytics. We have implemented A* routing, which executes each vehicle's search for a fastest path between its origin and destination in a parallel processing manner on the GPU. AutoMatrix is capable of hierarchical routing so routes with different levels of details are possible.



```
Time Elapsed: 00:03:20
Active Cars: 200000 cars
Executing iteration 21...
Time Elapsed: 00:03:30
Active Cars: 260000 cars
Executing iteration 22...
Time Elapsed: 00:03:40
Active Cars: 320000 cars
Executing iteration 23...
Time Elapsed: 00:03:50
Active Cars: 380000 cars
Executing iteration 24...
Click at -77.0542, 38.8728
Click: Edge 243973: Pentagon Access
Rd


Time Elapsed: 00:04:00
Active Cars: 440000 cars
Executing iteration 25...
Time Elapsed: 00:04:10
Active Cars: 500000 cars
Executing iteration 26...
Click at -77.0731, 38.8459
Click: Edge 269801: S Uhle St



Time Elapsed: 00:04:20
Active Cars: 560000 cars
Executing iteration 27...
Time Elapsed: 00:04:30
Active Cars: 620000 cars
Executing iteration 28...
Time Elapsed: 00:04:40
Active Cars: 680000 cars
Executing iteration 29...
Time Elapsed: 00:04:50
Active Cars: 740000 cars
Executing iteration 30...
Time Elapsed: 00:05:00
Active Cars: 800000 cars
Executing iteration 31...
```

FIGURE 6 AutoMatrix real-time traffic congestion modeling and congestion-aware traffic prediction and routing algorithm design. (Left) Traffic congestion simulation showing more than 800,000 vehicles in Washington, DC. (Center) Hierarchical routing showing one vehicle's coarse-grained route (in large boxes), which is determined at the beginning of the trip. The real-time congestion-aware fine-grained fastest route shows the ¾-mile route ahead of the vehicle. (Right) Thousands of vehicles (each small box represents a vehicle), each with unique origins and destinations, routed with real-time congestion-aware fastest-path routes around Philadelphia.

Vehicles can be guided with adaptive routing—the assigned route "responds" to changes in congestion patterns and reroutes the vehicle to the updated fastest path. By modeling point-based congestion, such as blocked lanes due to vehicle breakdowns or accidents, we can model queuing effects as vehicles back up and congestion spreads through the region.

Using these approaches, AutoMatrix has the potential to improve response time to traffic incidents by advising drivers to take the updated fastest path to their destination. We are working to use live traffic congestion data to support the needs of urban transportation operation centers.

CONCLUSION

The future of the automobile lies in the design and development of new vehicles that are programmable, connected vehicles, and networked traffic centers. These efforts are a step toward safer, more efficient, and more enjoyable commuting with automobiles.

REFERENCES

Duri´c P, Miladinov-Mikov M. 2008. Some characteristics of drivers having caused traffic accidents. Medicinski Pregled 61(9-10):464–469. Available online at www.ncbi.nlm.nih.gov/pubmed/19203062.

NHTSA [National Highway Traffic Safety Administration]. 2009. Safety Recalls, ID: 09V218000. Washington, DC: US Department of Transportation. Available online at www.safercar.gov.

Schäuffele J, Zurawka T. 2005. Automotive Software Engineering: Principles, Processes, Methods, and Tools. Warrendale PA: SAE International.

SERIOUS GAMES

Serious Games

LI-TE CHENG
Google

BEN SAWYER
Digitalmill

"Serious games" is the term used to describe the increasing application of video game technologies in nonentertainment domains. From their beginnings in entertainment, video games have grown into a multibillion-dollar industry that has helped advance the state of the art in computer graphics, user interaction, and computational software and hardware. With these advances, new stakeholders began adapting both the technologies and media of video games for training, simulation, education, health, and other uses and areas.

The initial wave of serious games focused on helping students and professionals learn and train. Today, health may be the fastest-growing category of serious games, with applications focused on therapeutic and health behavior change efforts. In addition, a third wave of experimentation is under way to not only educate, exercise, and train people but shape and improve their output and productivity. As in Orson Scott Card's 1985 science fiction story, *Ender's Game* (in which game play manipulates actual military actions), a new generation of serious games focuses on innovative crowd sourcing activities that tackle real-world scientific, organizational, and social challenges through video game play.

Serious games are best understood as a medium of many design, engineering, and technical domains rather than a single specific technology. Although diverse, they share a history as games for entertainment and education. The resulting diverse repertoire includes models, interactive techniques, and aesthetic methods to motivate and support players toward outcomes beyond the emotional experience derived from being entertained.

The speakers in this session present developments in the serious games field to show that video games and their technologies represent a new strategic tool for engineers to use in future projects.

There were four speakers in the session, and two of their papers are included in this volume. Richard Marks (Sony Computer Entertainment) talks about getting innovative game technology out of the lab into the living room and explains how cutting-edge technology can create new experiences to expand the gaming audience. Phaedra Boinodiris (IBM) illustrates the utility of serious games for businesses in addressing the increasingly complex global environment and offers pointers for the selection and design of a game. At the meeting, Kurt Squire (University of Wisconsin-Madison) discussed the serious games space from a national policy standpoint and as an educator. And to indicate how serious science is being achieved with serious games, Zoran Popovic (University of Washington) described his experiences using crowd sourcing games to tackle scientific challenges.

Moving Innovative Game Technology from the Lab to the Living Room

RICHARD MARKS
Sony Computer Entertainment R&D

Video games have become a giant industry, with global revenues in 2011 estimated at well over $60 billion. They enjoy a mass-market audience while at the same time riding the bleeding edge of technology. Because video games are entertainment, they offer a unique launch pad for new technologies: players are supportive and hopeful, and the focus is enjoyment rather than productivity or a high level of reliability. Game developers can thus make the most of new technologies to explore new experiences.

INNOVATIVE HARDWARE TECHNOLOGIES

Game hardware manufacturers have produced significant innovations in the areas of graphics, computing, display, and input technologies. The interactive nature of video games drives technologies with requirements of both high performance and low latency.

Graphics

An example of game graphics innovation is the Voodoo 3D technology (introduced by 3dfx in 1997), a 3D-only add-on card for PCs that enabled arcade-level visuals. In the same time frame, Nintendo released the Nintendo 64 “Reality” 3D coprocessor, touted as the equivalent of “a high-end Silicon Graphics workstation in your home.” Several years later, Sony created the PlayStation 2 Graphics Synthesizer, which achieved enormous polygon fill rate due to its 2,560-bit width bus to embedded RAM. Games continue to be the driving force in real-time graphics hardware, and are often used as benchmarks for measuring personal computer performance.

Computing

Game-related advances in computing, especially parallel computing, include the Cell processor in the PlayStation 3, a microprocessor that consists of one general-purpose CPU and eight coprocessors to handle streaming computation like that often found in interactive applications. Another example is "cloud gaming." Companies such as OnLive and Gaikai have demonstrated that high-performance gaming is possible using only a thin client by streaming output (effectively a movie) from a powerful server in the cloud. These companies are raising the bar for the types of applications that can be moved to the cloud.

Display

One of the Nintendo 3DS screens uses parallax-barrier technology to present a different image to the left and right eye, effectively creating a 3D image with no need for glasses. The 3DS also includes a slider that lets the player control the strength of the 3D. The PlayStation 3D monitor uses fairly standard LCD shutter glasses to achieve 3D, but it also uses this technology for an innovative "dual view" mode in which two players see different 2D images. In the near future, low-cost head-mounted displays (HMDs) will provide an immersive 3D experience that updates the image seen based on the player's head motion.

Input

Recently, video games have pushed the boundaries of input technology beyond the joystick, gamepad, keyboard, and mouse. A primary reason for this innovation is that earlier advances in graphics and display technology (output) greatly outpaced those in interface technology (input), creating an unbalanced user experience.

Several key technologies for input have enabled the revolution in interfaces. The commercialization of microelectromechanical systems (MEMS) has made possible small, low-cost sensors such as accelerometers, gyros, magnetometers, and pressure and temperature sensors. In conjunction with wireless, high-speed, low-cost, low-power, low-encumbrance digital communication, these microsensors can be used to collect a wide assortment of data for processing. And finally, digital video cameras have become viable as low-cost input devices that (when combined with ever-growing processing capabilities) can provide real-time information about how players are moving their bodies.

Peripherals that use these technologies in various combinations have changed the way video games can be played. The EyeToy digital video camera for

PlayStation 2 was the primary interface for games that explored both enhanced reality (i.e., augmented reality) and video-as-input paradigms (Figure 1). The Nintendo Wii featured a wireless one-handed MEMS-based motion-sensing remote as its primary controller, redefining the way games are played. The PlayStation Eye camera improved on EyeToy, enabling new marker-based augmented reality experiences such as *Eye of Judgment* and *EyePet*. The Microsoft Kinect camera extended video sensing to capture depth information at every pixel, enabling the Xbox to compute the dynamic pose of the player (i.e., skeleton tracking). Kinect also included a microphone array to enable voice control without needing to hold or wear a microphone. PlayStation Move combined MEMS inertial sensing and digital camera sensing into a single system that provides high-precision, high-speed, six-degree-of-freedom tracking of the one-handed controller. *Wonderbook* uses marker-based technology and PlayStation Move to create an interactive book in which each "page" is printed with a different marker, so a unique interactive augmented reality experience is shown on the television the player flips through the pages (Figure 2).

FROM LAB TO LIVING ROOM

Transitioning new technologies from research to product poses challenges in every industry, and it is no different for video game manufacturers. The following sections describe the trajectories of three consumer products that began as research projects in Sony Computer Entertainment R&D.

EyeToy

EyeToy, a mass-market product that sold more than 10 million units globally, began as a research project to investigate the types of experiences that would be possible by plugging a video camera (webcam) into a video game console (Figure 3). The powerful computation capabilities of consoles align well with what is necessary for real-time video processing and computer vision. EyeToy was essentially a standard webcam, but several design choices made it well suited for interactive experiences. For example, low latency was prioritized, so EyeToy compressed each frame individually in order to transfer over USB 1.1, rather than using an interframe video compression method such as MPEG. In addition, 60 frames/sec was the default output video rate to provide smooth, high-speed tracking.

The goal of EyeToy was to introduce video games to a wider audience via an intuitive interface and improve the interactive experience by directly involving the player visually. The biggest concern for EyeToy was the highly variable lighting in people's homes; unlike the office or laboratory environment, which is almost always well lit, the lighting in many family or living rooms is much less consistent. But because players wanted to enjoy their experience, most were happy to add lighting as necessary to brighten the scene while they played.

FIGURE 1 The EyeToy camera turns the television into an augmented reality mirror. Computer graphics are overlaid onto the live video, and real-time motion detection allows players to interact with the graphics as if they were real—for example, to wipe virtual dirt off the screen or bat away virtual ninjas. Source: Courtesy of Sony Computer Entertainment.

FIGURE 2 Wonderbook uses PlayStation Move and a "book" with different markers on each page to create a rich augmented reality experience. The left image shows reality, and the right image shows an example of what the player might see on the television; in this case, reality is "augmented" to include a magic wand, a flying dragon, and an old book that is burning. Source: Courtesy of Sony Computer Entertainment.

FIGURE 3 The EyeToy camera. Source: Courtesy of Sony Computer Entertainment.

PlayStation Eye

Based on feedback from both players and developers, the PlayStation Eye (Figure 4) improved on EyeToy with the addition of a fixed-focus, low-distortion lens with two choices for field of view, standard or wide. To improve video quality, USB 2.0 high speed was used so the resolution could be quadrupled (to 640×480) and video could be transferred uncompressed to avoid artifacts. And the low-light sensitivity was greatly improved. These technical attributes and the product's low cost have made PlayStation Eye the most widely used camera among computer vision researchers and hobbyists.

PlayStation Move

The high specifications of the PlayStation Eye enabled the creation of PlayStation Move (Figure 5), a one-handed motion controller for PlayStation 3 that incorporates a combination of optical and inertial sensing to provide complete six-degree-of-freedom tracking. The design addresses issues discovered during the development of EyeToy, combining the advantages of a camera-based interface with those of motion sensing and buttons. Game developers thus have absolute position and orientation, linear and angular velocities/accelerations, and button state.

FIGURE 4 The PlayStation Eye camera. Source: Courtesy of Sony Computer Entertainment.

FIGURE 5 The PlayStation Move motion controller: initial prototype and final product. Source: Courtesy of Sony Computer Entertainment.

Tracking the Move involves two major steps: image analysis and sensor fusion. Images from the PlayStation Eye are analyzed to locate the illuminated sphere that sits atop the controller (because the sphere is lit, it can be tracked even in complete darkness). Color segmentation is used to find the sphere in the image, and then a projected sphere model is fit to the image to extract the 3D position of the sphere. The results of the image analysis are fused with inertial sensor data from a 3-axis accelerometer and 3-axis gyroscope to provide the full state using a modified unscented Kalman filter (LaViola and Marks 2010).

CONCLUSION

The continual adoption of new technology has been a key factor in the growth of the video game industry. Advances in graphics, processing, display, and input technologies have both improved existing experiences and enabled new ones that appeal to a wider audience. Looking forward, there is every indication that the video game industry will continue to leverage new technology to help push the boundaries of play.

REFERENCE

LaViola JJ Jr, Marks R. 2010. An introduction to 3D spatial interaction with video game motion controllers. Course presented at ACM SIGGRAPH 2010, Los Angeles.

Playing to Win: Serious Games for Business

Phaedra Boinodiris
IBM

GORDIAN KNOTS

Thousands of years ago in ancient Phrygia, there was a massive mound of tangled ropes that was so impressive, it had a name: the Gordian knot. Legend was that whoever could untangle the great knot would become king of all of Asia.

One can imagine a stream of men arriving on horseback, strolling up to the ropes, cursing loudly as they pulled here and pushed there, inadvertently making the mound even bigger and more tangled.

According to the legend, one day Alexander the Great came into town, jumped off his horse, and with his supernaturally sharp blade sliced through the Gordian knot in one stroke, effectively ending the knot's persistent challenge.

If only today's intractable problems could be solved so simply. As the world has become smaller and more interconnected, people and businesses rely on systems that produce huge quantities of data. Executives face enormous challenges in analyzing these quantities as they seek to transform their business to be more responsive to the global economy. According to a recent Massachusetts Institute of Technology report (Hopkins et al. 2010), company executives are seeking ways to not only visualize data but also run simulations and scenario development in order to learn how their organizations might be more agile. Agile businesses achieve 10–15% higher margins, up to 5% faster revenue growth, and up to 38% higher capital efficiency.[1] Agility requires enterprise visibility, operational dexterity, and process integrity. Visualization, motivation, and collaboration are the components of

---

[1] BTM Business Agility, BTM Corporation 2010 (www.btmcorporation.com).

the solution—the blade that can cut through the massive mounds of data to enable businesses to adapt quickly and compete.

Transformation is not optional but it doesn't come easily to companies. Organizations that wish to be agile must transform their processes and decisions, embrace rapid, adaptable integration, and have a flexible and efficient infrastructure. For example, in a recent report, Gartner states that by 2015 more than 50% of organizations that manage innovation processes will "gamify" those processes—make them more accessible for adoption by engaging and teaching executives, managers, and employees through game techniques.[2]

In this article I explain how organizations can incorporate "serious games" to add value and agility in an increasingly complex environment.

SERIOUS GAMES AND SERIOUS SOLUTIONS

Play is a universal language characterized by enjoyment, established rules, and tangible, clear goals. Digital games can create deep, immersive experiences or quick bursts of excitement. Serious games, designed for a primary purpose other than entertainment, focus on clarifying goals, excising irrelevant information, and developing tangible, measurable improvements in a particular activity or task. They create realistic environments for testing strategies, tactics, theories, and ideas, leveraging the best aspects of games to make modeling, prototyping, experimenting, training, and skill acquisition faster, cheaper, more enjoyable, and more visible.

Whether for a training exercise, supply chain, or cyber defense scenario, smart games techniques can help participants visualize and understand complex systems through video and online gaming, engaging them through competition, teamwork, intrigue, curiosity, and problem solving. These features attract participation, encourage creativity, and help establish a path to collaborative work and analysis.

Although technology has changed the appearance and interactions associated with games, the experience associated with the best games has not changed: the challenge of any game or simulation should match the skills—and test the limits—of the players and the surrounding system in a meaningful, enjoyable way. What better test of game and gamer limits than the most serious challenges facing the world today?

Game Playing to Enhance Business Processes

Business simulations have been around for many years. They allow inputs and, given a set of business rules, produce new outputs. What they lack is the collaborative environment that motivates people to optimize. Keeping a business process locked up in a castle turret with fortified walls does no one any good.

---

[2] "Gartner Says By 2015, More Than 50 Percent of Organizations That Manage Innovation Processes Will Gamify Those Processes." Press Release, April 12, 2011; www.gartner.com/it/page.jsp?id=1629214.

Business processes need to be vetted, stressed, and prioritized by the entire value chain to yield a meaningful return on investment.

Today's 24/7 world is filled with large streams of data in previously unimaginable volumes. Approaches such as serious games techniques allow data to be viewed in different ways that nonexperts can understand, contribute to, and act on. In the summer of 2011, thousands of people helped map the structure of an enzyme that could fight HIV and AIDS by playing a downloadable game called Foldit. Researchers were able to crunch data from players' moves to quickly gain valuable insights into protein folding, critical to the development of treatment options. This project shows the power of collective intelligence when big data are harnessed and analyzed through games.

Serious games are increasingly used to test business scenarios and conduct training in both public- and private-sector organizations and corporations around the world. Business gaming techniques are used to motivate and lead large global, virtual teams and to encourage creative problem solving, load balancing, and complex system (e.g., supply chain) optimization. Cross-genre games and games with natural language interpretation[3] are also growing popular as a means to aid critical thinking in the military. These new techniques can save money, time, and resources while making departments and organizations more agile.

One of the key differentiators of a serious games approach to problem solving is a concentration on process optimization. This focus involves examining the most efficient and effective ways to improve procedures via iterative collaborative gameplay, applying Six Sigma principles. Business process improvement can reduce cost and cycle time by as much as 90% while improving quality by more than 60% (Harrington 1991). Results can also include improvements in margin, capacity, and capital reductions.

In a serious games approach, participants sort and understand real data, analyze real issues, and test real potential solutions, applying variables that can be adjusted and readjusted for different approaches. Game play preserves engagement while focusing players on important concerns and helping transform their assumptions, skills, and behaviors.

With cloud computing infrastructure, organizations can use serious games to improve business processes by solving complex problems collaboratively through predictive modeling and real-time visualization of methods to, for example, reduce costs and cycle times. Gaming systems tap employee and citizen insights and promote collaboration with partners for greater organizational agility.

Game Playing to Enhance National Security

Military, security, and emergency services organizations were early adopters of serious games to help test interagency disaster response scenarios or scale skills

---

[3] This term refers to the means for artificial intelligence to interpret natural (i.e., human) language and respond accordingly.

training beyond the platoon level to tackle complex strategy and operational use. The coordinated and cooperative nature of defense work requires team building and prepares for specific and highly synchronized missions. Potentially hazardous work benefits from simulations in which mistakes can be made without causing actual damage or endangerment and then evaluated for future learning.

Serious games techniques can also help optimize military supply chains. By creating real-time strategy games that enable players to examine how unforeseen events might affect real-world components, departments can help make their supply chains work more reliably and efficiently. Business or industry partners can also be included to tap insights from a wider network. The endproduct becomes a new, executable supply chain process that has been prevetted by the broader value chain.

In a cyber defense scenario, players benefit from competing in opposing roles on offense, defense, and network exploitation, playing as different entities such as countries and organizations. Strategic-level serious games should mimic the mundane and repetitive aspects of a scenario as well as information technology (IT) tasks, business processes, and attacks.

Direct representation of the decision process can be an instructive way to introduce new leaders to their roles and to allow key decision makers to focus on anomalous incidents by automating the common. A cyber security game that includes the possibilities of organizational policy, politics, operating costs, and social engineering will better prepare players for real-world complexities.

With current advances in process optimization, cloud, analytics, and artificial intelligence capabilities, the defense industry has the tools it needs to conduct strategy-level and process optimization gaming. These approaches teach the kinds of abilities needed to solve complex problems, including leading and managing, handling logistics and resources, prioritizing tasks, making sense of rapidly changing data, and learning from mistakes.

FIVE STEPS TO SERIOUS GAMING

Game design and development are constantly under-estimated. Many people assume that all they need to develop a serious game is interns with "game skills." The assumption that someone who plays games would be able to design a good game is completely erroneous. From the development side, there are countless game engines on the market that require highly specialized coding skills.

Development of a serious game requires determination of the measures of its effectiveness, an architecture, specifically designed puzzles and/or experiences, a genre, and a platform.[4] Below are five steps for determining how to approach a serious games project.

---

[4] Definitions of terms relevant to serious games are provided in a glossary at the end of this article.

Step 1: Determine the measures that will prove the game was worth the investment.

The very first question to answer concerns the purpose of the game. Is it to sell things? Is it to teach something? Is it to solve a problem? The game must then be designed in such a way that its effectiveness can be quantifiably measured. It is critical to start here. It may be tempting to do this last, but the answer to this can affect the entire architecture of the game so it is essential to start here.

If the game is to be used to improve sales skills, then the design must include measurements to prove that salespeople who played the game measurably understood their trade better than those who did not participate.

If the game is meant to optimize strategy among a group of participants, the results report must be able to substantiate that the model created by the players is better than one that has been Six Sigma–certified by a consultant.

If the game is meant to teach physics to sixth graders, results must show that they learned at least as well as from traditional methods.

How an organization measures success may directly affect the design of the game and the architecture of the system. It may be helpful to do an “after-action” review (i.e., to assess what players actually did during the gameplay) in an automated fashion to facilitate real-time insight.

Step 2: What is to be taught or conveyed?

Learning points should be documented in as much detail as possible, as in the following examples:

- The car salesman’s 7 steps to a sale are . . .
- The best practice business model associated with a disaster response scenario is . . .

Step 3: What kinds of puzzles or experiences are best suited to the information or lesson to be conveyed?

Simpler puzzles can also be used to explain complex systems such as molecular structures, as in the Foldit example cited earlier.

This is the hardest step of the five, and unfortunately few people realize just how hard it is. Most think they could design a great game. But matching the right puzzle/experience(s) to the learning points documented in Step 2 is difficult. Someone who knows games intimately and across genres should help with this step.

Step 4: Based on the puzzles/experiences, what is the right genre for the game?

Now that the basic design of the game has been determined, what genre does it fall into? Is it a city simulation, first-person adventure, strategy game, simulation-style game, pattern-matching game? Careful study of "flow" (a mental state of operation in which the person performing the activity is fully immersed in a feeling of energized focus; Csikszentmihalyi 1996) in games for entertainment in that genre can yield tips about how to proceed.

City sims, turn-based, and real-time strategy games have proven to be enormously powerful as genres to help explain complex systems. City-building games (city sims) are a genre of strategy computer game where players act as the overall planner and leader of a city, with responsibility for its growth and management. In a turn-based strategy (TBS) game (usually a war game, especially a strategic-level war game), players take turns, as distinct from a real-time strategy game, in which all players participate simultaneously.

Step 5: Knowing the genre and audience, what is the right platform for the game?

It is essential to know the intended audience well if the game is to be effective. How long are participants likely to play? What will motivate them to play? Can the game be standalone or does it need to be integrated with other applications? Does the game need to be Web playable? Mobile? Single or multiplayer? How often does the game's content need to be refreshed?

Once these questions are answered it is time to shop around for the right platform and the right vendor to help with development, if in-house expertise is not available. It will be key to know whether the platform is proprietary to the selected vendor. If it is, then future updates will have to come from this vendor unless it offers a "mod kit," which allows noncoders to access and modify surface components of the game (e.g., prices and product descriptions in a sales game).

CHOOSING THE RIGHT GAME STUDIO TO PARTNER WITH TO MAKE A SERIOUS GAME

It is important at the outset to get to know games and know them well. The ability to speak the language of games is essential to work with and gauge the efficacy of the studio that will make the game.

The best way to learn about games is by playing them across genres. The Game Developer's Conference in San Francisco, E3 (Electronic Entertainment Expo), and the East Coast Game Conference in Raleigh, North Carolina, are fantastic venues to learn about innovation in entertainment games. Why start there instead of a serious games summit? Because entertainment features the newest and most innovative ideas and is most likely to showcase examples of game play and techniques that can be adapted for a serious gaming purpose.

Younger employees tend to be enthusiastic volunteers for help in this area. Their aptitude can be assessed by asking them what their favorite games are and why, especially if they play across genres and can critique their favorite games well. Such employees can be very useful resources in a new serious games program.

When trying to find the right game studio partner, I recommend considering the following questions:

1. Do they "get" games?

Get the bios of the staff members who would work on your serious games project. Are they full of e-learning and instructional designers and no one else? What game engines do the staff have expertise with? Take a look at the games they have developed. Do they look engaging? Did the staff correctly match the right kind of game experience to what they are trying to teach, or did they instead create chocolate-covered broccoli—merely creating an attractive cover (gaming) for the necessary "nutrients" (the material to be learned)? Make sure the people on your project have an understanding of good game design. If they come from the entertainment gaming industry, why did they leave? If they think a great game is a multiple choice questionnaire, *run* toward the exit sign.

2. Do they get serious games?

The team members will need to have enough breadth to take a complex idea and make it accessible and engaging to the participants. If all they know is entertainment games, they may not have the skill set needed to work with serious content. The team's bios should reveal whether the members have what it takes to understand, for example, molecular biology well enough to design an effective protein folding game.

3. What about the proximity of the vendor?

The Internet makes working virtually a lot easier, but there will be times when it will be most helpful to look over the designer's shoulder—literally—during the design process. The selected team *must* be able to understand your vision for the game throughout the entire development process.

4. What types of game genre does the vendor specialize in?

Does the game studio specialize in the genre that makes the most sense for your game? If you are making a next-generation city sim game to explain water management, it doesn't make sense to choose a studio that specializes in first-person shooters.

CONCLUSION

Sophisticated information technology, abundant human capabilities, a growing appreciation of engagement, and the desire for discovery have created a foundation for utilizing games to tackle intractable problems and achieve big changes. But gaming requires more than business leaders interested in adding experience points and digital merit badges to individuals who answer the most emails. It requires an investment of time and resources from a panoply of contributors—scientists, researchers, visionaries, futurists, game designers, game developers, game testers, gamers themselves, citizens, media, political leaders, informed business leaders, artists, science fiction writers, popular science writers, universities, academia, lobbyists, and educators. The gaming community has a responsibility to advocate for games and provide educational opportunities; likewise, business leaders have a responsibility to look beyond stereotypes and learn what games have become—a valuable tool for learning, communicating, and collaborating around important goals.

When well designed, games can not only be extremely adept at explaining complex systems but also motivate people to play using a wide variety of game design tricks. These same tricks can also be used to motivate and reward employees and partners who optimize the core components of the underlying business.

It's up to each organization to grasp just how powerful serious games are—and make the most of them. Game on!

GLOSSARY

architecture: how a game is designed
experience: the flow of the game, what the user encounters through gameplay
genre: a category of game (e.g., puzzle, role play, strategy)
platform: web-based, mobile, console, downloadable executable

REFERENCES

Csikszentmihalyi M. 1996. Creativity: Flow and the Psychology of Discovery and Invention. New York: Harper Perennial.

Harrington HJ. 1991. Business Process Improvement: The Breakthrough Strategy for Total Quality, Productivity, and Competitiveness. New York: McGraw-Hill.

Hopkins MS, Kruschwitz N, LaValle S, Lesser E, Shockley R. 2010. Analytics: The New Path to Value. MIT Sloan Management Review Research Report, vol 52. Cambridge MA: Massachusetts Institute of Technology.

ENGINEERING MATERIALS FOR THE BIOLOGICAL INTERFACE

Engineering Materials for the Biological Interface

KAREN J. L. BURG
Clemson University

ALI KHADEMHOSSEINI
Harvard Medical School

Early biomaterial scientists quickly determined the importance in purposeful design of the interface of biomedical devices in eliciting a desired cellular response, including good tissue integration. Indeed, even with respect to biomedical design, the whole is greater than the sum of the parts; that is, the characteristics of a complex tissue are defined by both the individual components and the relationship between them.

The biological interface, such as that of the connection of tendon or cartilage to bone, includes cell-cell and cell-tissue components, and modeling of this interface with cells and biomaterials can enhance understanding of both normal and repair tissue processes. The functionality of a biological interface may be judged by the response of biomaterials to cells or cells to biomaterials. Bulk tissue repair approaches (i.e., repairs of single tissue types) are relatively simple compared with repairs across interfaces, where one must often consider very diverse tissue properties (e.g., tissue mechanics) and the corresponding interfacial interactions. In attempts to simulate these interactions, researchers have focused on the design of materials, control of cells, and design of bioreactors in which to grow and assess these systems.

This session focuses on the whole and the parts and the methods with which to integrate the two. The speakers, representing academia and industry, review the technical concepts of interfacial engineering as well as the practical concepts and limitations in the translation of ideas to commercial application. Helen Lu (Columbia University) describes engineering tissue-to-tissue interfaces for the formation of complex tissues, David Schaffer (University of California, Berkeley) covers identification and modulation of biophysical signals that control stem cell function and fate, and Matthew Gevaert (Kiyatec) talks about cultivating 3D tissue systems to better mimic relevant events.

Engineering Tissue-to-Tissue Interfaces and the Formation of Complex Tissues

Helen H. Lu
Columbia University

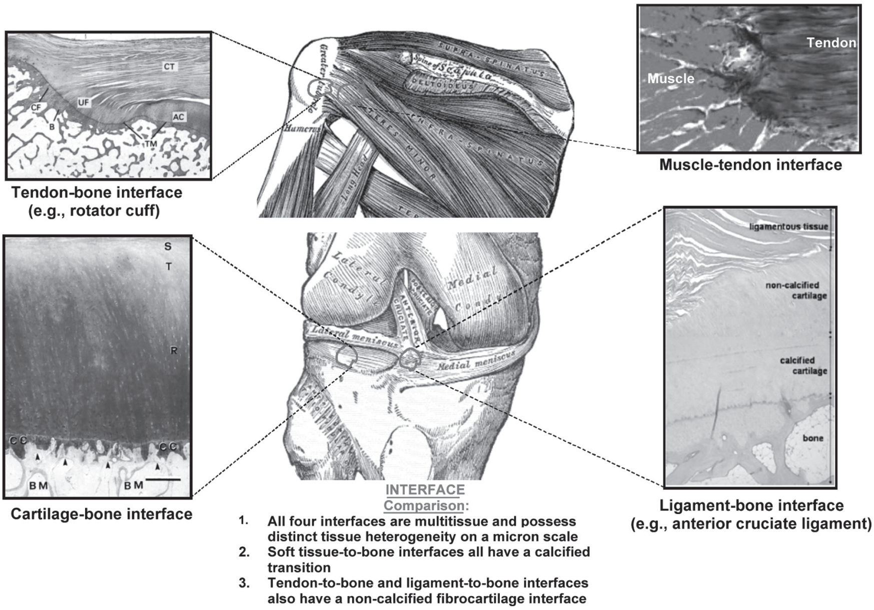
Two significant challenges in the field of tissue engineering are the simultaneous formation of multiple types of tissues and the functional assembly of these tissues into complex organ systems (e.g., the skeletal, muscular, or circulatory systems). These challenges are particularly important for orthopedic regenerative medicine, as musculoskeletal motion requires synchronized interactions among many types of tissue and the seamless integration of bone with soft tissues such as tendons, ligaments, or cartilage. These tissue-to-tissue interfaces are ubiquitous in the body and exhibit a gradient of structural and mechanical properties that serve a number of functions, from mediating load transfer between two distinct types of tissue to sustaining the heterotypic cellular communications required for interface function and homeostasis (Benjamin et al. 1986; Lu and Jiang 2006; Woo et al. 1988). But these critical junctions are prone to injury (from trauma or even exercise and daily activity) and unfortunately do not regenerate after standard surgical repair, thus compromising graft stability and long-term clinical outcome (Friedman et al. 1985; Lu and Jiang 2006; Robertson et al. 1986). Consequently, there is a need for grafting systems that support *biological fixation* or *integrative repair* of soft tissues.

BACKGROUND

Through a combination of cells, growth factors, and/or biomaterials, the principles of tissue engineering (Langer and Vacanti 1993; Skalak 1988) have been readily applied to the formation of a variety of connective tissues such as bone, cartilage, ligament, and tendon both *in vitro* and *in vivo*. More recently, emphasis has shifted from tissue formation to tissue function (Butler et al. 2000), with a focus on imparting biomimetic functionality to orthopedic grafts and enabling their translation to the clinic.

But clinical translation remains elusive as researchers seek to understand how to achieve biological fixation or functional integration of tissue-engineered orthopedic grafts—of bone, ligaments, or cartilage—with each other and/or with the host environment. The challenge is rooted in the complexity of the musculoskeletal system and the structural intricacy of both hard and soft tissues. These tissues, each with a distinct cellular population, must operate in unison to facilitate physiologic function and maintain tissue homeostasis. It is thus not surprising that the transition between various tissue types is characterized by a high level of heterogeneous structural organization that is crucial for joint function.

As shown in Figure 1, ligaments and tendons with direct insertions into bone exhibit a multitissue transition consisting of three distinct but continuous



Tendon-bone interface (e.g., rotator cuff)

Muscle-tendon interface

Cartilage-bone interface

Ligament-bone interface (e.g., anterior cruciate ligament)

INTERFACE Comparison:

1. All four interfaces are multitissue and possess distinct tissue heterogeneity on a micron scale
2. Soft tissue-to-bone interfaces all have a calcified transition
3. Tendon-to-bone and ligament-to-bone interfaces also have a non-calcified fibrocartilage interface

FIGURE 1 Common orthopedic tissue-to-tissue interfaces. Significant structural and compositional homology exists in the orthopedic tissue-to-tissue interfaces of the tendon-bone (Benjamin and Ralphs 1998), muscle-tendon (Larkin et al. 2006), cartilage-bone (Hunziker et al. 2002), and ligament-bone junctions (Iwahashi et al. 2010). Regeneration of these complex junctions is essential for integrative soft tissue repair and treatment of massive, multitissue injuries. Tendon-to-bone interface: AC = articular cartilage, B = bone, CF = calcified fibrocartilage, CT = connective tissue, TM = tidemark, UF = uncalcified fibrocartilage. Cartilage-to-bone interface: BM = bone marrow space, CC = calcified cartilage, R = radial zone, S = superficial zone, T = transitional zone.

regions of ligament, fibrocartilage, and bone (Benjamin et al. 1986; Cooper and Misol 1970; Wang et al. 2006). The fibrocartilage interface is further divided into noncalcified and calcified regions. In light of this complexity, effective tissue engineering must incorporate *strategic biomimicry* or the prioritization of design parameters in order to regenerate the intricate tissue-to-tissue interface and ultimately enable seamless graft integration and functional repair.

MECHANISMS OF INTERFACE REGENERATION

The mechanisms underlying the formation, repair, and maintenance of tissue-to-tissue boundaries are not well understood. In particular, it is not known how distinct boundaries between different types of connective tissues are reestablished after injury. It is likely that mechanical loading (Killian et al. 2012) as well as chemical and biological factors play a role in this complex process.

It has long been observed that when tendon is resutured to its original attachment site, cellular organization resembling that of the native insertion occurs *in vivo* (Fujioka et al. 1998). Investigators have also reported that, although healing after ligament reconstruction does not lead to the reestablishment of the native insertion, a layer of interface-like tissue forms in the bone tunnel (Blickenstaff et al. 1997; Grana et al. 1994; Rodeo et al. 1993). These observations suggest that when trauma or surgical intervention results in nonphysiologic exposure of normally segregated tissue types (e.g., bone or ligament), interactions between the resident cell populations (e.g., osteoblasts in bone, fibroblasts in tendon, stem cells/progenitor cells in both tissues) are critical for initiating and directing the repair response that leads to reestablishment of a fibrocartilage interface between soft tissue and bone.

Specifically, it has been hypothesized that osteoblast-fibroblast interactions mediate interface regeneration through heterotypic cellular interactions that can lead to phenotypic changes or transdifferentiation of osteoblasts and/or fibroblasts (Lu and Jiang 2006). Moreover, these interactions may induce the differentiation of stem cells or resident progenitor cells into fibrochondrocytes and thereby promote the regeneration of the fibrocartilage interface. This hypothesis has been validated using coculture and triculture models of interface-relevant cell populations (Jiang et al. 2005; Wang et al. 2007), models that offer simple and elegant methods to systematically investigate cell-cell interactions (Bhatia et al. 1999; Hammoudi et al. 2010).

When ligament fibroblasts and osteoblasts were cocultured using a model permitting both physical contact and cellular interactions, it was observed that these controlled interactions altered cell growth and upregulated the expression of interface-related matrix markers. These cellular interactions have a downstream effect, either inducing cell transdifferentiation or causing the recruitment and differentiation of progenitor or stem cells for fibrocartilage formation. When this hypothesis was tested in triculture, it was noted that under the influence of osteoblast-fibroblast interactions, stem cells from the bone marrow began to differentiate toward a chondrocyte-like phenotype, producing a matrix similar in composition to that of the interface.

These intriguing findings suggest that heterotypic cellular communications play a regulatory role in the induction of interface-specific markers in progenitor or stem cells, and demonstrate the effects of these interactions in regulating the maintenance of soft tissue-to-bone junctions. The nature of the regulatory cytokines secreted and the mechanisms underlying these interactions are not known, but cell communication is likely to be significant for interface regeneration as well as homeostasis. Therefore the optimal interface scaffold must promote interactions between the relevant cell populations residing in each interface region.

INTERFACE STRUCTURE-FUNCTION RELATIONSHIP AND DESIGN INSPIRATION

From a structure-function perspective, the complex multitissue organization of the soft tissue-to-bone junction is optimized to sustain both tensile and compressive stresses experienced at the ligament-to-bone junction. Numerous characterization studies (Benjamin et al. 1986; Bullough and Jagannath 1983; Matyas et al. 1995; Moffat et al. 2008; Oegema and Thompson 1992; Ralphs et al. 1998; Spalazzi et al. 2004; Thomopoulos et al. 2003; Woo et al. 1988) have revealed remarkable organizational similarities among many tissue-to-tissue interfaces (Figure 1). They often consist of a multitissue, multicell transition and exhibit a controlled distribution of mineral content that, along with other structural parameters such as collagen fiber organization, results in a gradient of mechanical properties progressing from soft tissue to bone.

Direct measurement of interface mechanical properties has been difficult due to the complexity and relatively small scale of the interface, generally ranging from 100 µm to 1 mm in length. Instead, knowledge of insertion material properties has been largely derived from theoretical models.

Moffat and colleagues (2008) recently performed the first experimental determination of the compressive mechanical properties of the anterior cruciate ligament (ACL)-bone interface in a neonatal bovine model. They evaluated the incremental displacement field of the fibrocartilage tissue under the applied uniaxial strain by coupling microcompression with optimized digital image correlation analysis of pre- and postloading images. Deformation decreased gradually from the fibrocartilage interface to bone, and these changes were accompanied by a gradual increase in compressive modulus. The interface also exhibited a region-dependent decrease in strain, and a significantly higher elastic modulus was found for the mineralized fibrocartilage compared to the nonmineralized region. These region-specific mechanical properties enable a gradual transition rather than a sudden increase in tissue strain across the insertion, thereby minimizing the formation of stress concentrations and enabling load transfer from soft to hard tissues.

Given the structure-function dependence inherent in the biological system, these regional changes in mechanical properties are likely correlated to matrix organization and composition across the interface. Partition of the fibrocartilage interface into nonmineralized and mineralized regions likely has a functional significance, as increases in matrix mineral content have been associated with higher mechanical properties in connective tissues.

Evaluation of the insertion site using Fourier transform infrared imaging (Spalazzi et al. 2007) and X-ray analysis revealed an increase in calcium and phosphorous content progressing from ligament to interface and then to bone. A narrow exponential transition in mineral content, instead of a linear gradient of mineral distribution, was detected progressing from the nonmineralized to the mineralized interface regions. Moreover, the increase in elastic modulus progressing from the mineralized to the nonmineralized fibrocartilage interface region was shown to be positively correlated (Moffat et al. 2008) with the presence of calcium phosphate.

These observations have yielded invaluable clues for the design of biomimetic scaffolds for engineering tissue-to-tissue interface. Specifically, a stratified or multiphased scaffold will be essential for recapturing the multitissue organization observed at the soft tissue-to-bone interface. To minimize the formation of stress concentrations, the scaffold should exhibit phase-specific structural and mechanical properties, with a gradual increase in the latter across the scaffold phases. Spatial control of mineral distribution on a stratified scaffold can impart controlled mechanical heterogeneity similar to that of the native interface. Compared to a homogeneous structure, a scaffold with predesigned, tissue-specific matrix inhomogeneity can better sustain and transmit the distribution of complex loads inherent at the multitissue interface.

It is important to bear in mind that the phases of a stratified scaffold must be interconnected and preintegrated with each other, to ensure the formation of *compositionally distinct* yet *structurally contiguous* multitissue regions. Furthermore, interactions between interface-relevant cells serve important functions in the formation, maintenance, and repair of interfacial tissue. Therefore, precise control over the spatial distribution of these cell populations is also critical for multitissue formation and interface regeneration. Consideration of these biomimetic parameters should guide and optimize the design of stratified scaffolds for promoting the formation and maintenance of controlled matrix heterogeneity and interface regeneration.

BIOINSPIRED SCAFFOLD DESIGN FOR INTERFACE TISSUE ENGINEERING

Inspired by the native ACL-to-bone interface, Spalazzi and colleagues (2006, 2008) pioneered the design of a triphasic scaffold (Figure 2C) for the regeneration of this challenging interface. The scaffold's three continuous phases are each engineered for a specific tissue region of the interface: Phase A is a polymer fiber

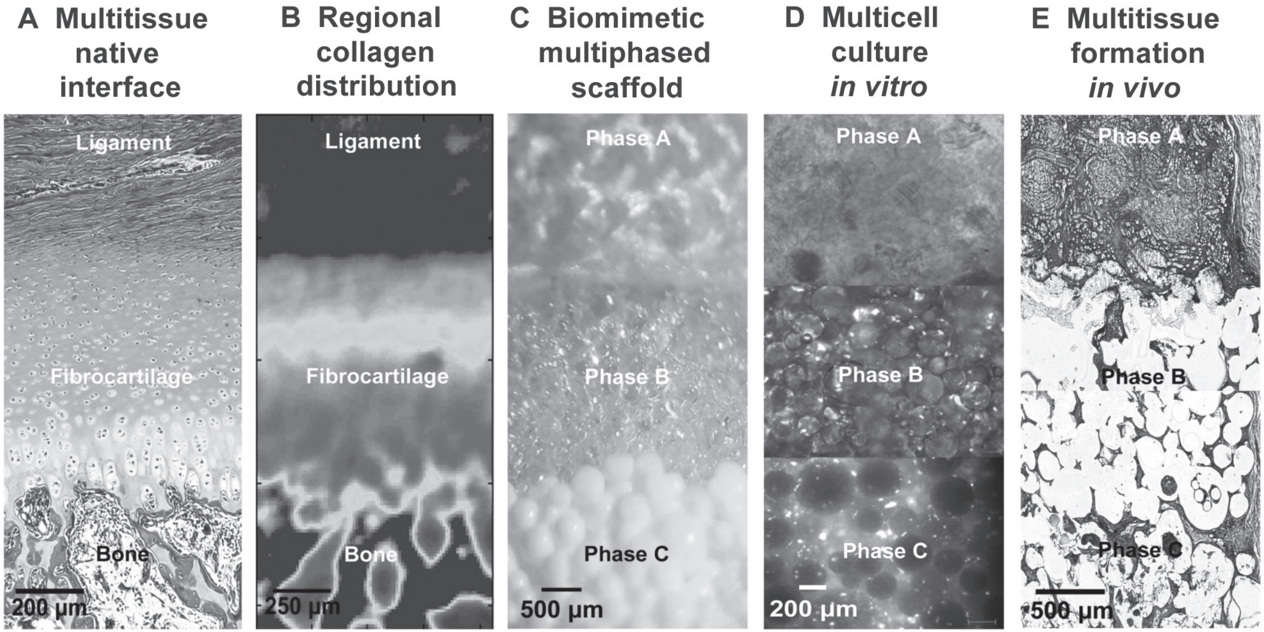


FIGURE 2 Bioinspired stratified scaffold design for interface tissue engineering and integrative soft tissue repair.

mesh for fibroblast culture and soft tissue formation, Phase B consists of polymer microspheres and is designed for fibrochondrocyte culture, and Phase C is composed of sintered polymer-ceramic composite microspheres for bone formation (Lu et al. 2003). The innovative design is in essence a single scaffold system with three compositionally distinct yet structurally continuous phases, all designed to support the formation of multitissue regions across the ligament-bone junction.

To form the ligament, interface, and bone regions, fibroblasts, chondrocytes, and osteoblasts were seeded onto Phases A, B, and C, respectively. Interactions between these cell types on the stratified scaffold were evaluated both *in vitro* (Spalazzi et al. 2008) and *in vivo* (Spalazzi et al. 2006). Extensive tissue infiltration and abundant matrix deposition were observed, with tissue continuity maintained across scaffold phases. Interestingly, matrix production compensated for the decrease in mechanical properties that accompanied scaffold degradation, and three continuous regions of ligament, interface, and bone-like matrix were formed *in vivo* (Figure 2E).

In addition to stratified scaffolds, there is tremendous interest in designing scaffolds with a gradient of properties—that is, with a relatively gradual and continuous transition in either composition or structural organization, resulting in a linear gradient in mechanical properties (Harris et al. 2006; Seidi et al. 2011; Singh et al. 2008). These novel scaffolds with either a compositional (Erisken et al. 2008; Li et al. 2009) or chemical factor (Phillips et al. 2008; Singh et al. 2010) gradient offer direct regional control and allow for scaffold heterogeneity that mimics the complex native interface. They may thus address the need to recapitulate the complex transition of mechanical and chemical properties that are characteristic of tissue-to-tissue junctions.

Design challenges in engineering biomimetic gradients revolve around scale—how best to recapitulate the micro- to nanoscale gradients that have been reported at the tissue-to-tissue interface. The stratified scaffold approach may represent a simpler strategy, whereby a *gradation* of key compositional and functional properties is preestablished by focusing on forming specific tissue regions of interest and preintegrating them through stratified design. In any case, it is necessary to adopt strategic biomimicry in functional interface scaffold design and to prioritize design parameters for interface regeneration based on the type of interface to be regenerated, the type and severity of injury, and the patient's age and overall health.

In addition to scaffold design, it is expected that cellular contributions will play a pivotal role in mediating the regeneration and homeostasis of the gradation of compositional and mechanical properties at the interface. For example, Ma and colleagues (2009) used cell self-assembly to form bone-ligament-bone constructs by culturing engineered bone segments to ligament monolayers. Paxton and colleagues (2009) also reported promising results when evaluating the use of a polymer ceramic composite and RGD peptide to engineer functional ligament-to-bone attachments.

SUMMARY AND FUTURE DIRECTIONS

The biomimetic interface tissue engineering approach described in this paper is rooted in an in-depth understanding of the inherent structure-function relationship at the tissue-to-tissue interface. The studies discussed indicate that controlling cellular response via coculture, triculture, or growth factor distribution on multiphased scaffolds is a critical emerging strategy to enable the development of local gradients on a physiologically relevant scale.

Many soft tissues connect to bone through a multitissue interface populated by multiple cell types that minimize the formation of stress concentrations while enabling load transfer between soft and hard tissues. In the event of injury or other disruption, reestablishment of tissue-to-tissue interfaces is critical for the formation of multitissue systems and the promotion of integrative tissue repair.

Investigations into the mechanism of interface regeneration have revealed the role of mechanical loading as well as heterotypic cellular interactions in directing the formation, repair, and maintenance of the tissue-to-tissue interface. Moreover, functional and integrative repair may be achieved by coupling both cell- and scaffold-based approaches. The vast potential of stratified scaffold systems is evident because (1) they are designed to support multitissue regeneration by mediating heterotypic cellular interactions and (2) they can be further refined by incorporating well-controlled compositional and growth factor gradients as well as the use of biochemical and biomechanical stimulation to encourage tissue growth and maturation.

Interface tissue engineering will be instrumental for the *ex vivo* development and *in vivo* regeneration of integrated musculoskeletal tissue systems with biomimetic functionality. Yet there remain a number of challenges in this exciting area. These include the need for a better understanding of the structure-function relationship at the native tissue-to-tissue interface and of the mechanisms that govern interface development and regeneration. Furthermore, the *in vivo* host environment and the precise effects of biological, chemical, and physical stimulation on interface regeneration must be thoroughly evaluated to enable the formation and homeostasis of the new interface. Physiologically relevant *in vivo* models are needed to determine the clinical potential of designed scaffolds.

The successful regeneration of tissue-to-tissue interfaces through a bio-inspired approach may promote integrative and functional tissue repair and enable the clinical translation of tissue engineering technologies from bench to bedside. Moreover, by bridging distinct types of tissue, interface tissue engineering will be instrumental for the development of integrated musculoskeletal organ systems with biomimetic complexity and functionality.

REFERENCES

Benjamin M, Ralphs JR. 1998. Fibrocartilage in tendons and ligaments: An adaptation to compressive load. Journal of Anatomy 193(Pt 4):481–494.

Benjamin M, Evans EJ, Copp L. 1986. The histology of tendon attachments to bone in man. Journal of Anatomy 149:89–100.

Bhatia SN, Balis UJ, Yarmush ML, Toner M. 1999. Effect of cell-cell interactions in preservation of cellular phenotype: Cocultivation of hepatocytes and nonparenchymal cells. Journal of the Federation of American Societies for Experimental Biology 13:1883–1900.

Blickenstaff KR, Grana WA, Egle D. 1997. Analysis of a semitendinosus autograft in a rabbit model. American Journal of Sports Medicine 25:554–559.

Bullough PG, Jagannath A. 1983. The morphology of the calcification front in articular cartilage: Its significance in joint function. Journal of Bone and Joint Surgery 65:72–78.

Butler DL, Goldstein SA, Guilak F. 2000. Functional tissue engineering: The role of biomechanics. Journal of Biomechanical Engineering 122:570–575.

Cooper RR, Misol S. 1970. Tendon and ligament insertion: A light and electron microscopic study. Journal of Bone and Joint Surgery (American volume) 52:1–20.

Erisken C, Kalyon DM, Wang H. 2008. Functionally graded electrospun polycaprolactone and beta-tricalcium phosphate nanocomposites for tissue engineering applications. Biomaterials 29:4065–4073.

Friedman MJ, Sherman OH, Fox JM, Del Pizzo W, Snyder SJ, Ferkel RJ. 1985. Autogeneic anterior cruciate ligament (ACL) anterior reconstruction of the knee: A review. Clinical Orthopaedics and Related Research 196:9–14.

Fujioka H, Thakur R, Wang GJ, Mizuno K, Balian G, Hurwitz SR. 1998. Comparison of surgically attached and non-attached repair of the rat Achilles tendon-bone interface: Cellular organization and type X collagen expression. Connective Tissue Research 37:205–218.

Grana WA, Egle DM, Mahnken R, Goodhart CW. 1994. An analysis of autograft fixation after anterior cruciate ligament reconstruction in a rabbit model. American Journal of Sports Medicine 22:344–351.

Hammoudi TM, Lu H, Temenoff JS. 2010. Long-term spatially defined coculture within three-dimensional photopatterned hydrogels. Tissue Engineering Part C Methods 16(6):1621–1628.

Harris BP, Kutty JK, Fritz EW, Webb CK, Burg KJ, Metters AT. 2006. Photopatterned polymer brushes promoting cell adhesion gradients. Langmuir 22(10):4467–4471.

Hunziker EB, Quinn TM, Häuselmann HJ. 2002. Quantitative structural organization of normal adult human articular cartilage. Osteoarthritis Cartilage 10:564–572.

Iwahashi T, Shino K, Nakata K, Otsubo H, Suzuki T, Amano H, Nakamura N. 2010. Direct anterior cruciate ligament insertion to the femur assessed by histology and 3-dimensional volume-rendered computed tomography. Arthroscopy 26:S13–S20.

Jiang J, Nicoll SB, Lu HH. 2005. Co-culture of osteoblasts and chondrocytes modulates cellular differentiation in vitro. Biochemical and Biophysical Research Communications 338:762–770.

Killian ML, Cavinatto L, Galatz LM, Thomopoulos S. 2012. The role of mechanobiology in tendon healing. Journal of Shoulder and Elbow Surgery 21(2):228–237.

Langer R, Vacanti JP. 1993. Tissue engineering. Science 260:920–926.

Larkin LM, Calve S, Kostrominova TY, Arruda EM. 2006. Structure and functional evaluation of tendon–skeletal muscle constructs engineered in vitro. Tissue Engineering 12:3149–3158.

Li XR, Xie JW, Lipner J, Yuan XY, Thomopoulos S, Xia YN. 2009. Nanofiber scaffolds with gradations in mineral content for mimicking the tendon-to-bone insertion site. Nano Letters 9:2763–2768.

Lu HH, Jiang J. 2006. Interface tissue engineering and the formulation of multiple-tissue systems. Advances in Biochemical Engineering/Biotechnology 102:91–111.

Lu HH, El Amin SF, Scott KD, Laurencin CT. 2003. Three-dimensional, bioactive, biodegradable, polymer-bioactive glass composite scaffolds with improved mechanical properties support collagen synthesis and mineralization of human osteoblast-like cells in vitro. Journal of Biomedical Materials Research 64A:465–474.

Ma J, Goble K, Smietana M, Kostrominova T, Larkin L, Arruda EM. 2009. Morphological and functional characteristics of three-dimensional engineered bone-ligament-bone constructs following implantation. Journal of Biomechanical Engineering 131:101017.

Matyas JR, Anton MG, Shrive NG, Frank CB. 1995. Stress governs tissue phenotype at the femoral insertion of the rabbit MCL. Journal of Biomechanics 28:147–157.

Moffat KL, Sun WH, Pena PE, Chahine NO, Doty SB, Ateshian GA, Hung CT, Lu HH. 2008. Characterization of the structure-function relationship at the ligament-to-bone interface. Proceedings of the National Academy of Sciences of the United States of America 105:7947–7952.

Oegema TR Jr, Thompson RC Jr. 1992. The zone of calcified cartilage: Its role in osteoarthritis. In Kuettner KE, Schleyerbach R, Peyron JG, Hascall VA, eds. Articular Cartilage and Osteoarthritis. New York: Raven Press. Pp. 319–331.

Paxton JZ, Donnelly K, Keatch RP, Baar K. 2009. Engineering the bone-ligament interface using polyethylene glycol diacrylate incorporated with hydroxyapatite. Tissue Engineering Part A 15:1201–1209.

Phillips JE, Burns KL, Le Doux JM, Guldberg RE, Garcia AJ. 2008. Engineering graded tissue interfaces. Proceedings of the National Academy of Sciences of the United States of America 105:12170–12175.

Ralphs JR, Benjamin M, Waggett AD, Russell DC, Messner K, Gao J. 1998. Regional differences in cell shape and gap junction expression in rat Achilles tendon: Relation to fibrocartilage differentiation. Journal of Anatomy 193(Pt 2):215–222.

Robertson DB, Daniel DM, Biden E. 1986. Soft tissue fixation to bone. American Journal of Sports Medicine 14:398–403.

Rodeo SA, Arnoczky SP, Torzilli PA, Hidaka C, Warren RF. 1993. Tendon-healing in a bone tunnel: A biomechanical and histological study in the dog. Journal of Bone and Joint Surgery (American volume) 75:1795–1803.

Seidi A, Ramalingam M, Elloumi-Hannachi I, Ostrovidov S, Khademhosseini A. 2011. Gradient biomaterials for soft-to-hard interface tissue engineering. Acta Biomaterialia 7(4):1441–1451.

Singh M, Berkland C, Detamore MS. 2008. Strategies and applications for incorporating physical and chemical signal gradients in tissue engineering. Tissue Engineering Part B: Reviews 14(4):341–366.

Singh M, Sandhu B, Scurto A, Berkland C, Detamore MS. 2010. Microsphere-based scaffolds for cartilage tissue engineering: Using subcritical CO(2) as a sintering agent. Acta Biomaterialia 6:137–143.

Skalak R. 1988. Tissue engineering: Proceedings of a workshop, held at Granlibakken, Lake Tahoe, California, February 26-29. New York: Liss.

Spalazzi JP, Costa KD, Doty SB, Lu HH. 2004. Characterization of the mechanical properties, structure, and composition of the anterior cruciate ligament-bone insertion site. Transactions of the Orthopedic Research Society 29, Poster #1271.

Spalazzi JP, Doty SB, Moffat KL, Levine WN, Lu HH. 2006. Development of controlled matrix heterogeneity on a triphasic scaffold for orthopedic interface tissue engineering. Tissue Engineering 12:3497–3508.

Spalazzi JP, Boskey AL, Lu HH. 2007. Region-dependent variations in matrix collagen and mineral distribution across the femoral and tibial anterior cruciate ligament-to-bone insertion sites. Transactions of the Orthopaedic Research Society (abstract 0891), San Diego.

Spalazzi JP, Dagher E, Doty SB, Guo XE, Rodeo SA, Lu HH. 2008. In vivo evaluation of a multiphased scaffold designed for orthopaedic interface tissue engineering and soft tissue-to-bone integration. Journal of Biomedical Materials Research A 86(1):1–12.

Thomopoulos S, Williams GR, Gimbel JA, Favata M, Soslowsky LJ. 2003. Variations of biomechanical, structural, and compositional properties along the tendon to bone insertion site. Journal of Orthopaedic Research 21:413–419.

Wang IE, Mitroo S, Chen FH, Lu HH, Doty SB. 2006. Age-dependent changes in matrix composition and organization at the ligament-to-bone insertion. Journal of Orthopaedic Research 24:1745–1755.

Wang IE, Shan J, Choi R, Oh S, Kepler CK, Chen FH, Lu HH. 2007. Role of osteoblast-fibroblast interactions in the formation of the ligament-to-bone interface. Journal of Orthopaedic Research 25:1609–1620.

Woo SL, Maynard J, Butler DL, Lyon RM, Torzilli PA, Akeson WH, Cooper RR, Oakes B. 1988. Ligament, tendon, and joint capsule insertions to bone. In Woo SL, Buckwalter JA, eds. Injury and Repair of the Musculoskeletal Soft Tissues. Savannah GA: American Academy of Orthopedic Surgeons. Pp. 133–166.

Identification and Modulation of Biophysical Signals That Control Stem Cell Function and Fate

David V. Schaffer
University of California, Berkeley

Stem cells are defined by two hallmark properties: the ability to self-renew, or divide while maintaining themselves in an immature state, and the capacity to differentiate into one or more specialized cell types. By virtue of these properties, stem cells play central roles in the development and maintenance of tissues throughout the body, and researchers in the biomedical field are increasingly exploring their potential in cell replacement therapies for treating human disease or injury. In particular, stem cells can theoretically be harvested, expanded, and differentiated in culture, and implanted for tissue regeneration. Alternatively, it may be possible to modulate endogenous pools of stem cells for tissue repair. To achieve both a deeper understanding of their natural biological functions and the ability to tap into their promise as next-generation therapeutics, fundamental knowledge is needed about how stem cell behavior is controlled and, specifically, about the processes of self-renewal and differentiation.

BACKGROUND

Populations of stem cells reside in specialized regions of developing and adult tissues that continuously present them with regulatory cues, and this repertoire of signals is collectively referred to as the stem cell niche (Scadden 2006). This niche includes small molecules, proteins, and other components of the extracellular matrix (ECM; the solid phase material that enmeshes most cells in the body), small growth factor and morphogen proteins that may be soluble or immobilized to the ECM, and signals from the surface of neighboring cells (Figure 1).

Thanks to many successful efforts in genetics, developmental biology, and cell biology, it is well recognized that biochemical cues in the niche play criti-

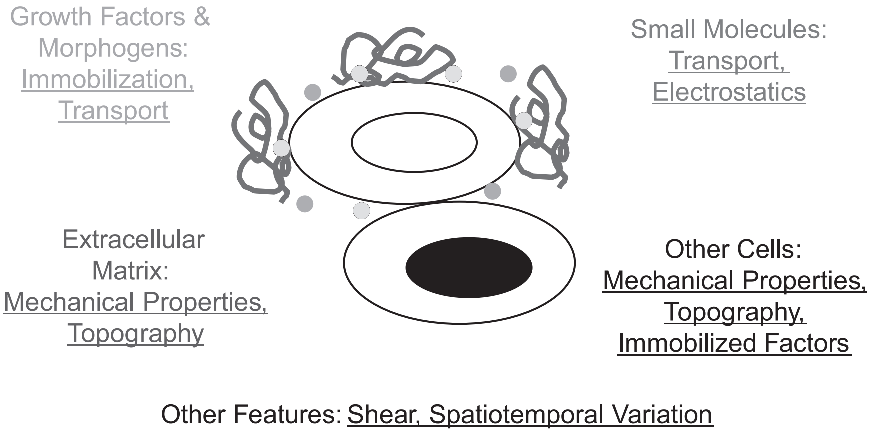


FIGURE 1 Schematic of the stem cell niche. Soluble small molecules, soluble and immobilized growth factor and morphogen proteins, extracellular matrix components, and intercellular components collaborate to regulate stem cell behavior. In addition, numerous physical and engineering principles modulate the manner in which these components present information, including mechanical properties, spatial organization and temporal variation in the presentation of cues, topographical features of the niche on the nano- and microscale, mass transport properties, and electrostatics.

cal roles in regulating stem cell function. However, biology encodes regulatory information to cells not only in the binary absence or presence of a given molecule but also in numerous biophysical aspects of tissues—mechanics, topographical features, electrostatics, biological transport phenomena, and spatiotemporal variation in each of these cues (Figure 1). Thus a major difficulty in studying and manipulating the biophysical properties of the niche is that they are not monogenic but depend on the properties of many molecules and genes.

An emerging theme in stem cell research is to use engineered systems in cell culture—ranging from synthetic materials to microfluidic devices—to systematically vary these biophysical properties and thereby study their effects on stem cells, that is, to provide an “x-axis” in a manner that is not currently possible with genetic approaches. While there are inherent challenges with this paradigm—including establishing the *in vivo* relevance of findings, as well as integrating engineering and biology approaches to explore the underlying mechanisms—these engineering studies have broadened the field’s view of the stem cell niche (Discher et al. 2009; Keung et al. 2010). Furthermore, because of the complexity of their endogenous niches, stem cells are exceedingly difficult to control in culture; therefore, each biophysical property offers a new opportunity to engineer synthetic systems and materials to control stem cell function for regenerative medicine applications.

MECHANOREGULATION OF STEM CELL FUNCTION

There are many mechanical features and processes of tissues that may regulate cell function, including elasticity, viscosity, strain, and others. Landmark work by Engler and colleagues (2006) demonstrated that the lineage choice of differentiating mesenchymal stem cells (MSCs) is strongly influenced by elastic modulus of the surrounding material—i.e., the linear proportionality constant between its strain and stress—such that cells developed into neuron-like cells on soft hydrogels, myoblasts on intermediate stiffnesses, and osteocytes on harder substrates. Subsequent work showed that neural stem cells (NSCs) preferentially differentiate into neurons when cultured on soft materials and astrocytes on hard materials (Saha et al. 2008). Also, a recent study reported that human embryonic stem cell and induced pluripotent stem cell differentiation into neural lineages, but not self-renewal, is mechanosensitive (Figure 2) (Keung et al. 2012). In addition to differentiation, modulus can influence stem cell self-renewal. For example, it was shown that substrate stiffness strongly affects the ability of muscle stem cells to undergo self-renewal in culture and subsequently their capacity to undergo reimplantation into muscle (Gilbert et al. 2010).

In parallel, the regulation of stem cell behavior by extracellular forces requires mechanisms to convert a mechanical cue into a biochemical signal that drives cell fate decisions. ECM protein structures, cell adhesion receptors, the intracellular network of structural proteins known as the cytoskeleton, and key proteins in the nucleus may all serve as mechanosensors. In addition to the stiffness of the cellular microenvironment, shear flow and cyclic strain have both been implicated in regulating the self-renewal and/or differentiation of several classes of stem cells.

Collectively, the studies described above have established the mechanical properties of the stem cell niche as a prominent regulator of fate choice, and offer the promise that mechanical aspects of synthetic materials can be manipulated to better control stem cell fate choice in culture.

TOPOGRAPHICAL AND SHAPE FEATURES OF THE STEM CELL NICHE

In addition to providing resident stem cells with a mechanical milieu, niches provide features that can alter the shape of a cell. On the microscale, ECM and neighboring cells can modulate and even constrain the surface area or volume available to, and therefore the shape of, a cell in a manner important for its function. Likewise, on the nanoscale, ECM proteins often assemble into fibers and other structural features that modulate the topographical features that an adherent cell experiences. Advances in lithography and in materials science have enabled investigators to investigate the effects of these features on stem cell behavior (Kolind et al. 2012).

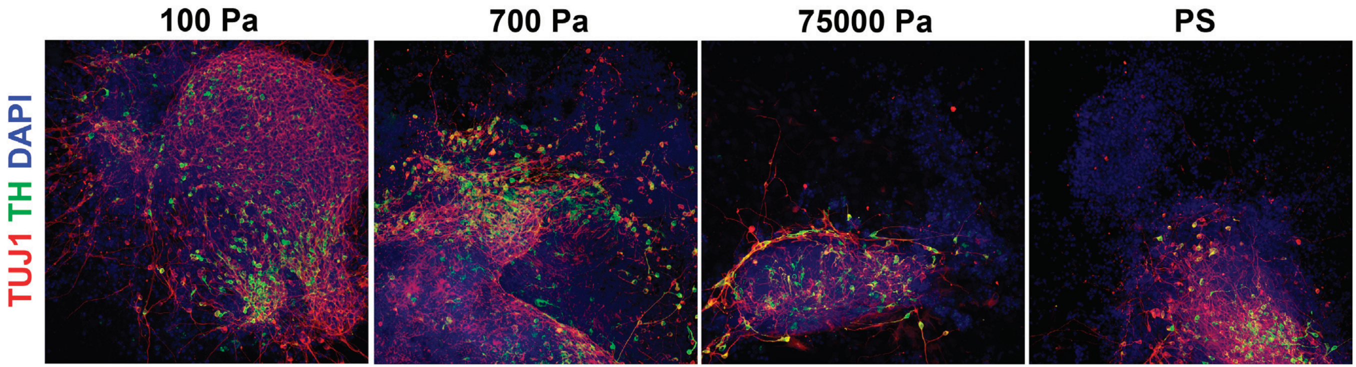


FIGURE 2 Mechanical cues regulate human embryonic stem cell differentiation. Human embryonic stem cells were differentiated into dopaminergic neurons, the cell type that is lost in patients with Parkinson's disease. During the initial phase of differentiation, cells were cultured on polymeric hydrogels of various elastic moduli. The softer materials, coincidentally with stiffnesses characteristic of the brain, led to significantly higher numbers of dopaminergic neurons. Red is the neuronal marker TUJ1, green is the neuronal marker tyrosine hydroxylase (TH), and blue is the nuclear dye DAPI. Pa = Pascal; PS = polystyrene. Adapted from Keung et al. (2012) with permission of The Royal Society of Chemistry. This figure appears in color in the online posting of this article at http://www.nap.edu/catalog.php?record_id=18185.

In seminal work, microcontact printing was used to pattern adhesive islands of different sizes onto a surface (McBeath et al. 2004). When MSCs were seeded onto these substrates, the investigators observed that large islands that enabled cells to spread subsequently promoted osteogenic (bone cell) differentiation, whereas small islands that did not permit substantial cell spreading instead promoted adipogenic (fat cell) differentiation. There has been progress in both extending this principle to other fate choices and elucidating its underlying mechanisms.

In addition to microenvironmental properties that alter cell shape on the micron scale, topographical cues—such as the organization of the ECM into fibers—offer a cell with features that can modulate its shape at the nanometer scale. In early work in this area, culturing NSCs on microgrooves patterned into polystyrene led to significantly higher extents of neuronal differentiation compared to flat surfaces (Recknor et al. 2006). Another study explored the effects of electrospun fibers of polyethersulfone with different dimensions on the NSC behavior and found that small fibers promoted differentiation into one major central nervous system cell type (oligodendrocytes) while larger fibers increased differentiation into neurons (Christopherson et al. 2009). These studies have yielded insights into mechanisms by which structural features in the niche can regulate cell function, and again offer potential opportunities to design biomimetic culture systems that can better control stem cell behavior.

ELECTRIC FIELDS

The role of electrophysiology in the cardiovascular and nervous systems is well appreciated, and a growing body of work has explored the possibility that electric fields may regulate the function of stem cells from these tissues. In initial work, heart muscle precursors became aligned with the direction of an electric field, exhibited a substantial increase in contractile amplitude, and expressed higher levels of various cardiac protein markers compared to cells that were not electrically stimulated (Radisic et al. 2004). Subsequent research has shown that electric fields promote the maturation and differentiation of skeletal muscle precursors (Serena et al. 2008), neural precursors (Ariza et al. 2010), and embryonic stem cells (Kabiri et al. 2012).

FUTURE DIRECTIONS

The application of the physical sciences and engineering to stem cell research has contributed significantly to the development of culture systems to elucidate the basic effects of a biophysical property on cell regulation. Such investigations will greatly benefit from further technological advances, particularly in the development of novel materials whose properties can be varied spatiotemporally to mimic tissue heterogeneity and development. Furthermore, there are considerable additional opportunities for "analysis by synthesis"—engineering systems to emulate and thereby investigate more features of the cellular microenvironment.

Another major need in the field is the development of scalable, safe, and reproducible stem cell culture systems for biomedical translation. Many current culture systems use complex and poorly defined protein mixtures (e.g., serum, matrix) to recreate the complexity of the niche. Basic progress in understanding of key biochemical and biophysical cues can be integrated toward the development of advanced, defined, synthetic culture systems that are in some ways less complicated than current systems containing components derived from animal or human tissue.

Finally, a major challenge in the application of stem cells for tissue engineering and repair is poor cell survival upon implantation into a site of injury or disease, although engineered systems and materials that increasingly integrate biological information to mimic the natural properties of tissue may serve as vehicles that enable cells to better adapt to their new niche after implantation. The integration of biology, physical sciences, and engineering is thus poised to greatly advance stem cell biology and medicine.

REFERENCES

Ariza CA, Fleury AT, Tormos CJ, Petruk V, Chawla S, Oh J, Sakaguchi DS, Mallapragada SK. 2010. The influence of electric fields on hippocampal neural progenitor cells. Stem Cell Reviews 6:585–600.

Christopherson GT, Song H, Mao H-Q. 2009. The influence of fiber diameter of electrospun substrates on neural stem cell differentiation and proliferation. Biomaterials 30:556–564.

Discher DE, Mooney DJ, Zandstra PW. 2009. Growth factors, matrices, and forces combine and control stem cells. Science 324:1673–1677.

Engler AJ, Sen S, Sweeney HL, Discher DE. 2006. Matrix elasticity directs stem cell lineage specification. Cell 126:677–689.

Gilbert PM, Havenstrite KL, Magnusson KEG, Sacco A, Leonardi NA, Kraft P, Nguyen NK, Thrun S, Lutolf MP, Blau HM. 2010. Substrate elasticity regulates skeletal muscle stem cell self-renewal in culture. Science 329:1078–1081.

Kabiri M, Soleimani M, Shabani I, Futrega K, Ghaemi N, Ahvaz HH, Elahi E, Doran MR. 2012. Neural differentiation of mouse embryonic stem cells on conductive nanofiber scaffolds. Biotechnology Letters 34:1357–1365.

Keung AJ, Kumar S, Schaffer DV. 2010. Presentation counts: Microenvironmental regulation of stem cells by biophysical and material cues. Annual Review of Cell and Developmental Biology 26:533–556.

Keung AJ, Asuri P, Kumar S, Schaffer DV. 2012. Soft microenvironments promote the early neurogenic differentiation but not self-renewal of human pluripotent stem cells. Integrative Biology: Quantitative Biosciences from Nano to Macro 4(9):1049–1058.

Kolind K, Leong KW, Besenbacher F, Foss M. 2012. Guidance of stem cell fate on 2D patterned surfaces. Biomaterials 33:6626–6633.

McBeath R, Pirone DM, Nelson CM, Bhadriraju K, Chen CS. 2004. Cell shape, cytoskeletal tension, and RhoA regulate stem cell lineage commitment. Developmental Cell 6:483–495.

Radisic M, Park H, Shing H, Consi T, Schoen FJ, Langer R, Freed LE, Vunjak-Novakovic G. 2004. Functional assembly of engineered myocardium by electrical stimulation of cardiac myocytes cultured on scaffolds. Proceedings of the National Academy of Sciences of the United States of America 101:18129–18134.

Recknor JB, Sakaguchi DS, Mallapragada SK. 2006. Directed growth and selective differentiation of neural progenitor cells on micropatterned polymer substrates. Biomaterials 27:4098–4108.

Saha K, Keung AJ, Irwin EF, Li Y, Little L, Schaffer DV, Healy KE. 2008. Substrate modulus directs neural stem cell behavior. Biophysical Journal 95:4426–4438.

Scadden DT. 2006. The stem-cell niche as an entity of action. Nature 441:1075–1079.

Serena E, Flaibani M, Carnio S, Boldrin L, Vitiello L, De Coppi P, Elvassore N. 2008. Electrophysiologic stimulation improves myogenic potential of muscle precursor cells grown in a 3D collagen scaffold. Neurological Research 30:207–214.

Engineering 3D Tissue Systems to Better Mimic Human Biology

Matthew Gevaert
KIYATEC Inc.

The scientific method—hypothesis-driven design and execution of an experiment—is great . . . except when it could kill you (or me). That's why, for example, there are extensive legal requirements to investigate new pharmaceutical agents using proxies before testing drug toxicity in a human clinical trial. The US Food and Drug Administration requires a combination of nonliving techniques (biochemical assays and *in silico* analysis), *in vitro* models (i.e., cell culture), and animal studies before new compounds may be administered to humans.

Although pharmaceutical and regulatory industries are doing the best they can in the current paradigm, to be blunt it's not going very well. According to recent publications, of all drugs that enter clinical trials, only 12% are eventually approved for use in humans (Paul et al. 2010). In other words, despite best efforts to predict those drug candidates' efficacy and toxicity during preclinical testing, 88% of them fail—usually in terms of their lack of efficacy or unacceptable toxicity—when put to the test in humans.

A new paradigm is needed! And the biggest opportunity lies in cell culture, which typically is still done in a Petri dish (or its derivative, the multiwell plate). This ubiquitous scientific container, first described well over a hundred years ago in the late 19th century (Petri 1887), was already commonplace when cells were first widely cultured in the mid-20th century and remains the standard of cell culture today.

The vast majority of human cell types are adhesion dependent, and after fluid transfer to a Petri dish or well plate they attach to the bottom. Once attached, they normally proliferate and cover the entire bottom surface without stacking, forming a confluent, flat monolayer (shorthanded as "2D" cell culture). As evidenced by usage patterns, normal limitations of this experimental mode (the environment is static, diffusion is passive, constant evaporation alters solute concentrations, frequent media changes are necessary, cell numbers plateau at confluence, the cell experiences stimuli largely unrelated to those it experiences *in vivo*) are viewed as less important than benefits (cells grow well, the approach is cost effective, 2D planes are easily imaged with inexpensive microscopes, existing body of data is 2D, granting agencies still fund it, and the method enables high throughput).

Yet, as I put it in a recent talk to a group of high school STEM whiz kids, "Your Petri Dish Is So 1887."

SIGNIFICANCE

Few people ask (and fewer answer) these basic questions: Do the results of Petri dish–type cell culture experimentation mean anything? Are they at all relevant to the intent of the experiment, which in most cases is to model a process that occurs in the human body? Although the assumption is "yes," in an increasing number of demonstrated cases the answer to these important questions is actually "*no*."

A quantitative way to measure the "behavior" of a cell in culture is its gene expression. In a beautiful demonstration that answers the questions above, a comparison was made of key gene expression profiles of primary human cancers with comparative immortalized epithelial cells in 2D (Ridky et al. 2010). Tellingly, the correlation coefficient between the two datasets was 0.0. But there are much easier and cheaper ways to obtain datasets with exactly zero correlation to the behavior one is trying to characterize than to conduct 2D cell culture experiments!

The tremendous opportunity for improvement lies in the fact that cells are living organisms and can respond dynamically to local stimuli provided by and in their environment. *The solution is to provide a different environment with more of the "right" physical, mechanical, and biochemical stimuli*. Developments that address this challenge will affect much more than *in vitro* modeling of *in vivo* physiology. Aside from the desire to model human beings and the need to minimize the very serious consequences of the scientific method for certain kinds of questions, better *in vitro* systems have enormous implications as both manufacturing methods for implants (e.g., in tissue engineering and regenerative medicine) and as process steps for cell therapy.

ENGINEERING CELL SHAPE THROUGH MATERIAL INTERACTION

As a living entity, each cell has the potential to sense and respond to physical stimulus at each point in all its transecting planes—i.e., its entire surface in three dimensions.

When an adhesion-dependent cell is presented with a flat surface to which it can favorably attach, it tends to maximize its adhesion and adopts a primarily flat morphology. Cells in a 2D paradigm tie up approximately 50% of this interaction capacity with the bottom surface of the well plate, approximately 50% with the liquid environment above the flat cell, and a very small amount in lateral cell-cell interactions.

The fundamental value proposition of "3D" cell culture is to provide a microenvironment in which the potential for physical interaction is distributed in a biologically relevant fashion across the entire surface of the cell. This is normally achieved by culturing cells in a scaffold or matrix material, which can span gels, fibers, or porous solids, among others. Cells in a 3D culture matrix adopt a more complex morphology (e.g., roughly ellipsoid) that is typically much closer to their morphology in their native state—that of a cell in tissue in a living organism.

Does this matter? Again relying on gene expression as a way to measure cell behavior, researchers have documented significant changes in gene expression profiles (recently genomewide) of multiple cell types as a result of 3D relative to 2D cell culture conditions. These changes have been shown to be associated with key biological processes such as tissue development, cell adhesion, immune system activation, and defense response (e.g., Zschenker et al. 2012). Thus, cell morphology is fundamentally deterministic of some important aspects such as cell behavior, signal transduction, protein-protein interaction, and responsiveness to external stimuli. Gene expression profiles in 3D are also shown to have much more relevance to those measured *in vivo* (Birgersdotter et al. 2005; Martin et al. 2008).

In addition to the value of a 3D microenvironment that more effectively models *in vivo* realities, this form-function relationship is also subject to manipulation toward less "natural" ends. Stem cells' differentiation pathway has historically been controlled by soluble factor interactions, either from a second "feeder" layer cell type or as a result of soluble factors added to the cells' media. Surprisingly, forcing a cell into a particular shape (e.g., the stars and flowers shown in Figure 1) by physical confinement can also affect its differentiation pathway even in the absence of soluble factor manipulation (Kilian et al. 2010).

Unfortunately, effectively engineering the 3D microenvironment is not as simple as providing physical interactions in three dimensions. Topography and mechanical stiffness are among biophysical cues in a 3D context that affect cell function. This is proven via either the addition of 3D topography (e.g., grooves, pillars, posts, pyramids, pits) to an otherwise flat surface via microfabrication techniques (wherein the cell is cultured *on* the material) or the incorporation of controlled topography internally and culturing of the cell *in* the material (Nikkhah et al. 2012). Topography can also induce effects that determine stem cell differentiation pathways (Kumar et al. 2012).

Mechanical stiffness affects cell behavior and function, as exemplified by the presence of an "edge effect" in 3D gel scaffolds. Fraley and colleagues (2011) characterized focal adhesions of cells embedded in a 3D collagen gel and reported that tension in the gel decreased with increasing distance from the container surface. Cellular focal adhesions, associated with each cell's cytoskeletal structure, decreased as well. As shown in Figure 1, the authors were able to loosely qualify 2D (cell on surface), 2.5D (cell partially on surface), "3D near" (cell within

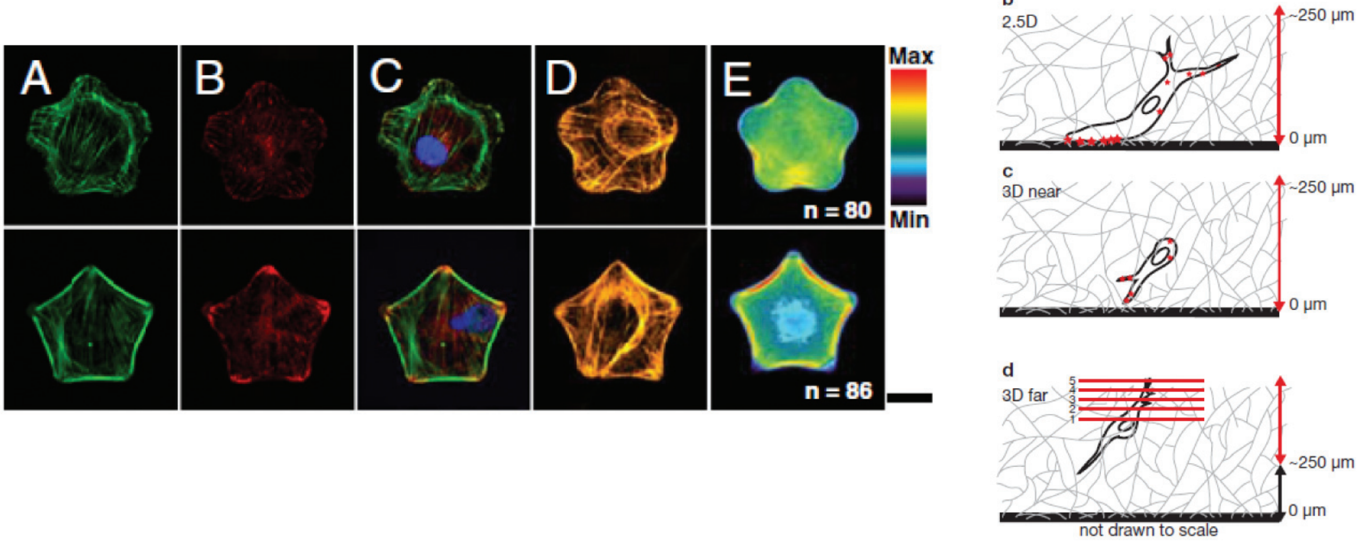


FIGURE 1 (Left) Immunofluorescent images and fluorescent heatmaps of cells in flower (top) and star (bottom) shapes, demonstrating differential cell response to nuanced physical constraints that influence the differentiation pathway. Reprinted with permission from Kilian et al. (2010). (Right) Schematic representation of focal adhesion visualization (stars on cell surface) in live HT-1080 cells cultured at increasing distances (a-d) from the dish bottom, characterizing edge effect in 3D matrix. Reprinted with permission from Fraley et al. (2011). This figure appears in color in the online posting of this article at www.nap.edu/catalog.php?record_id=18185.

250 µm of surface), and "3D far" regions based on the number of focal adhesions per cell.

Just how "3D" an environment is has very important implications for applications other than modeling. In a recent paper with profound ramifications for cell therapy, investigators demonstrated that a complex response (immunomodulation, e.g., the recruitment of monocytes to an inflamed endothelial monolayer) of cells in 3D was reduced fivefold compared to the same cells on 2D surfaces (Indolfi et al. 2012). The authors observed that the 3D cells had markedly altered cytoskeletal structure with rearranged focal adhesion proteins.

ENGINEERING THE SOLUBLE ENVIRONMENT

In addition to the interaction of a given cell with the materials and other cells surrounding it, the soluble environment has a considerable effect on cell behavior. At the most basic level, it is through the soluble environment that cells receive nutrients and perform basic functions such as respiration and waste elimination. Interference with these basic needs over time compromises the viability of the cell culture. Cells cultured on a 2D surface have nearly 50% of their surface area interacting with the soluble environment and simple, passive diffusion is usually more than sufficient to enable these processes. With frequent media changes to compensate for evaporation, depletion of nutrients, and generation of wastes, compromised viability of 2D cell cultures due to insufficient soluble environment interaction is rarely a concern.

However, inherent in soluble environment interactions and the frequent replacement of cell media is a cyclic change in the media pH and a "feast to famine" dynamic with respect to nutrient access. Media pH in typical 2D cell culture decreases over time (Wu and Kuo 2011) and differing pH levels have been shown to affect cell function (Wu et al. 2007). The removal of "spent" media, containing relatively fewer nutrients and more waste, also removes nonwaste excretions (e.g., proteins), an environmental change that may be directly related to the observable phenomenon that confluent cells in 2D culture do not typically stack but occur as monolayers. In one experiment, researchers began with cells in a typical 2D culture environment and, using a specialized bioreactor that allowed nutrient and waste exchange but preserved insoluble extracellular matrix secretions, created mineralizing, collagenous tissue up to 150 µm thick with as many as 6 cell layers (Dhurjati et al. 2006).

Depending on the density of both the 3D matrix and other cells, a particular cell's soluble environment interactions can be severely compromised and result in muted function or eventually cell necrosis, particularly in the middle of the construct. The window for effective density management is significantly smaller if the *in vitro* model relies only on the passive diffusion that occurs with use of 3D scaffolds in static multiwell plates. The use of perfusion culture systems or bioreactors can minimize or alleviate the deleterious effects and stabilize the soluble environment by avoiding feast-to-famine changes in nutrient availability and maintaining pH. Compared with static conditions, perfusion cell culture has been shown to affect culture morphology and organization (Tomei et al. 2009), increase key enzyme activity (Goldstein et al. 2001), increase mineral deposition and production of protein and cytokines (Gomes et al. 2003; Mercille and Massie 1999), increase cell penetration into and distribution throughout the scaffold (Cimetta et al. 2007; Goldstein et al. 2001; Gomes et al. 2003), increase cell viability especially at the center of cell-scaffold constructs (Cimetta et al. 2007; Mercille et al. 1999), and thus extend the effective duration of the culture experiment.

INCORPORATING BIOLOGICAL SYSTEMS EFFECTS WITH MULTIPLE CELL TYPES

Consideration of a particular cell's interactions with other cells is essential to increasing the correlation of its *in vitro* functions and behavior to an *in vivo* organism. These interactions can take the form of direct cell-to-cell contact or of soluble factor interactions mediated by the environment. Human biology relies on both modes of interaction. Culture-based intercellular interactions among cells of the same type (monocultures) in 3D have been implicitly included in the discussions above, and are at least partially responsible for the morphological changes and functional benefits described.

A second type of intercellular interaction can be modeled via coculture of different types of cells, which can occur in the same culture chamber and create direct cell-to-cell contacts (a mixed coculture) or in multiple, separate chambers with a connected soluble environment through the exchange of soluble factors (a segregated coculture). A coculture and the multicellular biological feedback loop it represents are necessary to reproduce many complex *in vivo* effects, which is not surprising given the many interacting physiological systems that combine to result in complex human biology.

Coculture provides yet another opportunity to engineer greater relevance into an *in vitro* model. As previously described, stem cell differentiation pathways are one of the best known multiple cell type interactions, whereby the differentiation of stem cell "A" is directed (or suppressed) by the presence of soluble factors from cell "B." Rivaling and perhaps surpassing stem cell cocultures for scientific activity are cancer cocultures, particularly cancer-stroma cocultures: it is increasingly being demonstrated that the incorporation of a second cell type materially affects cancer cells in culture (Khodarev et al. 2003) and boosts their relevance to the *in vivo* pathology (Chung et al. 2005; Mahadevan and Von Hoff 2007).

LAYERING COMPLEXITY

Incorporation of any of these themes—3D matrix microenvironment, actively stabilized soluble environment, mixed and segregated cocultures—in an *in vitro* system represents an increase in complexity compared to standard 2D cell culture. Increased complexity is often associated with increased cost and time and decreased efficiency (often measured by throughput). These negative consequences are perhaps the largest reason these promising innovations have not achieved the wide use and rapid commercial uptake initially expected. Yet, in the absence of their purposed integration into drug delivery processes, correlation of *in vitro* models to *in vivo* results is poor, 88% of drug candidates fail in clinical trials, and each successful drug costs approximately $1 billion to develop and launch (Deloitte 2011).

There are nonetheless strong indications of progress toward a new era of *in vitro* models. With biologically derived gel matrices having led the way, there are now many commercially available scaffolds specifically marketed for 3D cell culture. Commonly used in multiwell plates to maintain throughput (but simultaneously limited by their static format), these scaffolds are primed for layering the additional complexities of a stabilized, actively perfused soluble environment and for the clever use of coculture, potentially with multiple matrices matched to cell type.

Basic segregated coculture has become widespread through the use of inserts fitted into the wells of multiwell plates and more recently through mixed but spatially controlled cocultures made possible by 2D microfabrication techniques. Limitations of the first iterations of these innovations include the static nature of multiwell plate culture and (for inserts) a limited range of materials suitable as membranes, but they have established important baselines that will be expanded with the integration of more and better 3D physical and soluble microenvironments.

Finally, early bioreactor systems have demonstrated the clear benefits of perfusion, but their adoption is hampered by high costs per experiment, a requirement for atypical cell culture equipment, and low throughput.

Microfluidic and Mesofluidic Approaches

These first examples of successfully integrating a single innovation theme that acceptably increases complexity (i.e., is worth the tradeoff) have laid the foundation for "layered complexity" approaches that may break new ground in adoption and use. In the United States, recent National Institutes of Health (NIH) and Defense Advanced Research Projects Agency (DARPA) grant solicitations themed around modeling 3 and 10 (respectively) interacting physiological systems were awarded to microfluidics "lab-on-a-chip" submissions (Figure 2, left).

The microfluidics approach embraces perfusion systems at a micrometer scale (the scale of the cells themselves), while layering the complexity of cocultures at various points in the fluidic channels. Benefits include a smaller footprint for the culture chamber device, reduced flow circuit volumes, and the use of micromanufacturing techniques for device manufacturing. This approach has most

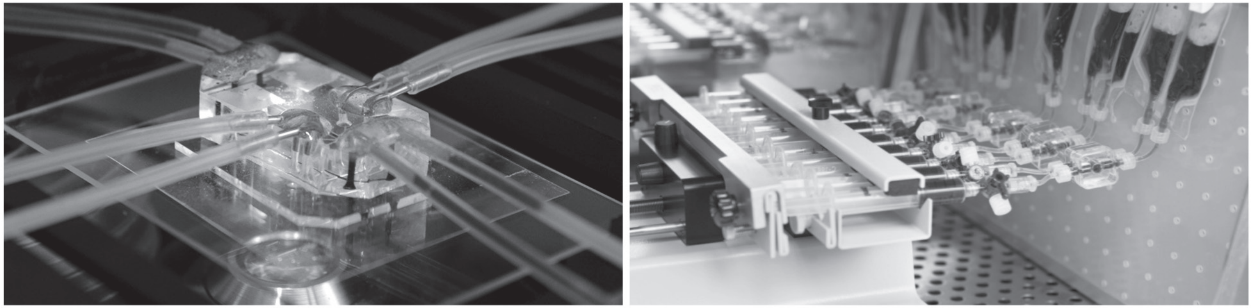


FIGURE 2 Examples of microfluidics (left, printed with permission from the Wyss Institute) and mesofluidics (right, KIYATEC Inc.) “layered complexity” in vitro systems. Both are perfusion based and incorporate 3D microenvironments and coculture interactions, but they differ in scale, manufacturing techniques, potential applications for cell-scaffold constructs, cost, and throughput.

effectively been demonstrated when modeling flat biological barrier models (e.g., gut, lung luminal interfaces) where perfusion takes the form of laminar-type flow over a dense, flat, cell-membrane construct. Technical challenges can include the inability to load and recover scaffolds, "edge effects" of soft matrices, management of cell/matrix density over time in gel matrices, and successful maintenance of constant flow in channels with small dimensions.

Another approach that successfully achieves the desired layered complexity may be thought of as "mesofluidic," with culture chamber dimensions on the scale of millimeters rather than micrometers (Figure 2, right). In contrast to microfluidics, this approach has focused on modeling tissues rather than barriers, and perfusion can take the form of interstitial-type flow through a 3D cell-scaffold construct.

The mesofluidic approach inherits the benefits of more traditional bioreactors, including the highest cell viability over time, best potential to model complexity, and broadest incorporation and recovery of diverse 3D scaffold materials, the latter being an important bridge to biomanufacturing applications such as regenerative medicine, tissue engineering, and some forms of cell therapy. Layered complexity is achieved through inherent accommodation of both mixed and segregated cocultures and more controlled management of the soluble environment through active perfusion. Although cost may be mitigated to the extent that these smaller bioreactor systems can leverage the existing cost structure for 2D cell culture processes,[1] lower throughput in mesofluidic systems remains a tradeoff.

Impacts of Economic and Social Factors

Nontechnical factors are aligning with the emergence of layered-complexity technological approaches. The economic and political environments have changed such that there is an increased focus on the societal value derived from the expenditure of granting agencies' (and ultimately the public's) research monies. It is becoming less acceptable to fund or conduct research that can be demonstrated to have a low, or zero, correlation to the biology being modeled when alternatives with higher correlation exist, even though they are more complex. Funding agencies are increasingly supportive of initiatives that mandate the incorporation of layered complexities.

Contractions in the global pharmaceutical industry have resulted in emphasis on new approaches that both drive down development costs and point toward new understanding of complex biology (and new targets, mechanisms, and pathways). Successful regenerative medicine and cell therapy business models have emerged, heightening the demand for improved *in vitro* manufacturing and quality control processes compliant with Current Good Manufacturing Processes (cGMP). And

---

[1] This cost structure encompasses the costs of commoditized supplies, equipment, instruments, and general infrastructure of traditional cell culture methods.

finally, increasing societal interest (particularly in the European Union) is also driving broader and faster adoption of more complex *in vitro* models and techniques that show promise for refining, reducing, and replacing the use of animals.

CONCLUSIONS

Perception of the value of *in vitro* models is slowly changing to both embrace the need for more relevance and accept the tradeoffs of lower throughput and increased complexity (Table 1). This change is driven by multiple, related dynamics: (1) scientific literature demonstrating the increased relevance of more complex (e.g., 3D, perfused, coculture) cell cultures to *in vivo* biology, especially that of humans, compared to the low relevance of 2D monolayer cultures; (2) the increasing adoption of approaches incorporating single-factor complexity (e.g., 3D environment only), albeit for a limited number of applications; (3) the emergence of "layered complexity"–type approaches whose combined dynamics have begun to enable the modeling of organism-level interactions, with potentially broad application; (4) the unfavorable failure rate and high costs of clinical trials for the pharmaceutical industry, especially given the dearth of new blockbuster drugs; (5) the emergence of viable business models in related industries (regen-

TABLE 1 Evolving Paradigm Through Which the Value of *In Vitro* Models Is Perceived

Traditional Paradigm		New Paradigm	
2D static monolayer cell culture	More complex cell culture: 3D or perfusion or coculture	2D static monolayer cell culture	Layered complexity cell culture (e.g., 3D perfused coculture)
• High throughput	• Lower throughput	• Higher throughput but lower relevance	• Higher relevance but lower throughput
• Acute cost minimization	• Costs more than 2D	• Cost and value of data both matter	• Overall cost reduction potential
• Synch with past data	• Past data disconnect	• Oversimplified	• Managed complexity
• Convenience	• Interesting but impractical	• Use when can get away with	• Use when value justifies cost
NET EFFECT: Very heavy reliance on 2D static monolayer methods with emphasis on throughput and acute cost minimization.		NET EFFECT: Balanced approach that recognizes throughput/relevance tradeoffs and integrates both options. Ultimately reduces overall costs by decreasing late-stage failures (drugs) and/or increasing performance (cell therapy, regenerative medicine).	

erative medicine, cell therapy) that require and can coopt cell culture innovation for cell maturation and processing; and (6) aligned nontechnical trends, including increased emphasis on funding clearly relevant research and on further refining, reducing, and replacing the use of animals.

The combination of these factors results in unprecedented opportunity and provides the required foundation to usher in a new era of better *in vitro* models. As they are implemented, these models will significantly advance understanding of human physiology while simultaneously translating to substantial health and cost benefits.

FINANCIAL DISCLOSURE

Dr. Gevaert is the CEO of, and owns stock in, KIYATEC Inc., a company focused on *in vitro* models with higher correlation to *in vivo* results via convenient and cost-effective perfused 3D cell-based assays.

REFERENCES

Birgersdotter A, Sandberg R, Ernberg I. 2005. Gene expression perturbation in vitro: A growing case for three-dimensional (3D) culture systems. Seminars in Cancer Biology 15:405–412.

Chung LWK, Baseman A, Assikis V, Zhau HE. 2005. Molecular insights into prostate cancer progression: The missing link of tumor microenvironment. Journal of Urology 173:10–20.

Cimetta E, Flaibani M, Mella M, Serena E, Boldrin L, De Coppi P, Elvassore N. 2007. Enhancement of viability of muscle precursor cells on 3D scaffold in a perfusion bioreactor. International Journal of Artificial Organs 30:415–428.

Deloitte. 2011. Measuring the return from innovation: Is R&D earning its investment? Deloitte Annual Review of How the Life Sciences Industry Is Performing in Generating Value from R&D:5.

Dhurjati R, Liu X, Gay CV, Mastro AM, Vogler EA. 2006. Extended-term culture of bone cells in a compartmentalized bioreactor. Tissue Engineering 12:3045–3054.

Fraley SI, Feng Y, Wirtz D, Longmore GD. 2011. Reply: Reducing background fluorescence reveals adhesions in 3D matrices. Nature Cell Biology 13(1):5–7.

Goldstein AS, Juarez TM, Helmke CD, Gustin MC, Mikos AG. 2001. Effect of convection on osteoblastic cell growth and function in biodegradable polymer foam scaffolds. Biomaterials 22:1279–1288.

Gomes ME, Sikavitsas VI, Behravesh E, Reis RL, Mikos AG. 2003. Effect of flow perfusion on the osteogenic differentiation of bone marrow stromal cells cultured on starch-based three-dimensional scaffolds. Journal of Biomedical Materials Research Part A 67A:87–95.

Indolfi L, Baker AB, Edelman ER. 2012. The role of scaffold microarchitecture in engineering endothelial cell immunomodulation. Biomaterials 33:7019–7027.

Khodarev NN, Yu JQ, Labay E, Darga T, Brown CK, Mauceri HJ, Yassari R, Gupta N, Weichselbaum RR. 2003. Tumour-endothelium interactions in co-culture: Coordinated changes of gene expression profiles and phenotypic properties of endothelial cells. Journal of Cell Science 116:1013–1022.

Kilian KA, Bugarija B, Lahn BT, Mrksich M. 2010. Geometric cues for directing the differentiation of mesenchymal stem cells. Proceedings of the National Academy of Sciences of the United States of America 107:4872–4877.

Kumar G, Waters MS, Farooque TM, Young MF, Simon CG Jr. 2012. Freeform fabricated scaffolds with roughened struts that enhance both stem cell proliferation and differentiation by controlling cell shape. Biomaterials 33:4022–4030.

Mahadevan D, Von Hoff DD. 2007. Tumor-stroma interactions in pancreatic ductal adenocarcinoma. Molecular Cancer Therapeutics 6:1186–1197.

Martin KJ, Patrick DR, Bissell MJ, Fournier MV. 2008. Prognostic breast cancer signature identified from 3D culture model accurately predicts clinical outcome across independent datasets. PLoS ONE 3(8):e2994.

Mercille S, Massie B. 1999. Apoptosis-resistant E1B-19K-expressing NS/0 myeloma cells exhibit increased viability and chimeric antibody productivity under perfusion culture conditions. Biotechnology and Bioengineering 63:529–543.

Nikkhah M, Edalat F, Manoucheri S, Khademhosseini A. 2012. Engineering microscale topographies to control the cell-substrate interface. Biomaterials 33:5230–5246.

Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL. 2010. How to improve R&D productivity: The pharmaceutical industry's grand challenge. Nature Reviews Drug Discovery 9:203–214.

Petri RJ. 1887. Eine kleine modification des koch'schen plattenverfahrens. Centralblatt für Bacteriologie und Parasitenkunde 1:279–280.

Ridky TW, Chow JM, Wong DJ, Khavari PA. 2010. Invasive three-dimensional organotypic neoplasia from multiple normal human epithelia. Nature Medicine 16:1450–1455.

Tomei AA, Siegert S, Britschgi MR, Luther SA, Swartz MA. 2009. Fluid flow regulates stromal cell organization and CCL21 expression in a tissue-engineered lymph node microenvironment. Journal of Immunology 183:4273–4283.

Wu M-H, Kuo C-Y. 2011. Application of high throughput perfusion micro 3-D cell culture platform for the precise study of cellular responses to extracellular conditions: Effect of serum concentrations on the physiology of articular chondrocytes. Biomedical Microdevices 13:131–141.

Wu M-H, Urban JPG, Cui ZF, Cui Z, Xu X. 2007. Effect of extracellular pH on matrix synthesis by chondrocytes in 3D agarose gel. Biotechnology Progress 23:430–434.

Zschenker O, Streichert T, Hehlgans S, Cordes N. 2012. Genome-wide gene expression analysis in cancer cells reveals 3D growth to affect ECM and processes associated with cell adhesion but not DNA repair. PLoS ONE 7:e34279.

APPENDIXES

Contributors

Kristi Anseth is a Howard Hughes Medical Institute Investigator and distinguished professor of chemical and biological engineering at the University of Colorado at Boulder. Her research interests lie at the interface between biology and engineering where she designs new biomaterials for applications in drug delivery and regenerative medicine. Dr. Anseth is an elected member of the National Academy of Engineering and the Institute of Medicine.

Phaedra Boinodiris is a serious games program manager at IBM where her work focuses on serious games that push beyond static skills training via inclusion of real data and real processes.

Karen Burg is the Hunter Endowed Chair of Bioengineering and director of the Institute for Biological Interfaces of Engineering at Clemson University. She investigates development of 3D tissue-engineered devices for repairing damaged organs as well as for constructing in vitro tissue test systems (e.g., to test pharmaceuticals or to provide personalized medical treatments). These unique devices comprise cells and a carefully designed, 3D degradable biomaterial. As they develop, in vivo or in vitro, the cells form a 3D tissue as the material degrades.

Li-Te Cheng is a software engineer at Google. His focus is on building new software systems to help people work together. Past systems span games, social media, collaborative software development tools, visualization, virtual reality, augmented reality, and wearable computing.

Michael Degner is the senior technical leader for electric machine drives in the Powertrain Research and Advanced Engineering Laboratory at Ford Motor Company where his research is on electric machine drives for hybrid electric, plug-in hybrid electric, fuel cell electric, and battery electric vehicles. Specific areas of research include design of electric machines, power electronics, control of electric machine drives, and alternative vehicle propulsion systems.

Matthew Gevaert is the CEO and co-founder of KIYATEC (Greenville, South Carolina), a life sciences company enabling better *in vitro* models of complex human biology through perfused 3D cell-based assay services and products. His work is primarily directed toward creating cell cultures with higher correlation to human responses in order to reduce the high failure rate of clinical trials and to help clinicians match cancer patients with the therapies that will be most effective for them. Other research interest areas include cell processing, tissue engineering, regenerative medicine, adult stem cells, technology entrepreneurship, and commercialization.

Ali Khademhosseini is an associate professor at Harvard-MIT Division of Health Sciences and Technology, Brigham and Women's Hospital, and Harvard Medical School. He develops micro- and nanoscale technologies to control cellular behavior with particular emphasis on developing microscale biomaterials and engineering systems for stem cell bioengineering and tissue regeneration.

Christopher Jones is the New-Vision Professor in the School of Chemical and Bioengineering at the Georgia Institute of Technology. His research interests are in the broad areas of materials design and synthesis, catalysis, and adsorption. Specific emphases are on design and understanding of molecular catalysts and catalytic materials for energy applications, fine chemical and pharmaceutical applications, and adsorbents for CO_2 capture.

Eli Kintisch is a contributing correspondent for *Science* magazine and a Knight Science Journalism Fellow at the Massachusetts Institute of Technology. In addition to traditional journalism, he focuses on communicating climate change in artistic and innovative ways. In 2010, he published his first book, *Hack the Planet: Science's Best Hope–or Worst Nightmare–for Averting Climate Catastrophe*.

Ben Kravitz is a postdoctoral research associate in the Atmospheric Sciences and Global Change Division at Pacific Northwest National Laboratory. He studies geoengineering with stratospheric aerosols and details of aerosol scattering using climate models.

Helen Lu is an associate professor in the Department of Biomedical Engineering at Columbia University. Her research seeks to understand how the biological interface between different types of connective tissues are formed and maintained in the body and, more importantly, how to regenerate these distinct tissue-to-tissue boundaries post injury and in complex tissue systems.

Arindam Maitra is the senior program manager of power, delivery, and utilization at the Electric Power Research Institute. He conducts and manages a wide range of research activities and power system studies in the transmission, distribution, power electronics, power quality, smart grid, electric transportation, energy efficiency, and IntelliGrid research areas.

Rahul Mangharam is the Stephen J. Angello Chair and an assistant professor in the Department of Electrical and Systems Engineering at the University of Pennsylvania. His research involves real-time architectures and scheduling algorithms for computing, communication, and coordination of physical computing systems. These include wireless control of industrial automation networks, medical device software and systems, energy-efficient building automation, real-time parallel computation, and networked automotive cyberphysical systems.

Richard Marks is a senior member of the US research and development department of Sony Computer Entertainment, where he focuses on engineering new experiences for video games. His primary focus is interaction technologies such as real-time video processing, sensor fusion, and control theory.

Sanjeev Naik is the engineering group manager in the Advanced Engine Controls Department of General Motors. His work focuses on the application of modern and adaptive control and signal processing techniques to improve the efficiency, performance, and safety of conventional and electrified vehicles.

Zoran Popovic is a professor in the Department of Computer Science and Engineering and director of the Center for Game Science at the University of Washington. His research interests lie primarily in computer graphics, especially in character animation, motion editing, physically based modeling, and modeling and simulation of natural phenomena. He is also interested in nonlinearly constrained optimization, motion planning, and biomechanics.

Lynn Russell is a professor of atmospheric science at the Scripps Institution of Oceanography. She investigates the chemical and physical components of aerosol particles in the earth's atmosphere under pristine marine conditions and in anthropogenically influenced conditions in order to assess their impacts on climate. Her program in aerosol research focuses on the fundamentals of aerosol formation by nucleation and the characterization and evolution of particles in the atmosphere.

Jeff Sakamoto is an associate professor in the Department of Chemical Engineering and Materials Science at Michigan State University where his research focuses on solution-based synthesis of porous materials. The ability to order interconnected porosity at multiple length scales provides a modular experimental platform enabling investigations into the interplay between micro-meso-macro pore morphology and mass/charge transport for energy storage and biomedicine.

Ben Sawyer is the co-founder of Digitalmill, which helps organizations outside the videogame industry such as health, education, defense, workforce development, and science, utilize tools, techniques, and resources from the field of videogames and videogame technologies.

David Schaffer is a professor in the Department of Chemical and Biomolecular Engineering, the Department of Bioengineering, and the Helen Wills Neuroscience Institute at the University of California at Berkeley, where he is also the director of the Berkeley Stem Cell Center. His research program applies engineering principles to solve molecular-level problems that currently impede clinical translation. Specifically, he engineers systems to investigate and control stem cell function for application in neural tissue regeneration. In parallel, he engineers viral gene therapy vehicles to improve their performance in numerous therapeutic applications.

David Sholl is the Michael Tennenbaum Family Chair and GRA Eminent Scholar in the School of Chemical and Biomolecular Engineering at the Georgia Institute of Technology. His research is in computational materials modeling, carbon capture, gas separation membranes, heterogeneous catalysis, and nanostructured surfaces.

Armin Sorooshian is an assistant professor in the Department of Chemical and Environmental Engineering and the Department of Atmospheric Sciences at the University of Arizona. He uses modeling, laboratory and field measurements, and remote sensing data to understand the effects of atmospheric aerosol particles on public health, climate change, the hydrologic cycle, and ecosystems. He also develops instrumentation for aerosol measurements and is interested in the interface of climate science and public policy.

Kurt Squire is an associate professor in the Educational Communications Technology Division of Curriculum and Instruction at the University of Wisconsin-Madison and co-founder and director of the Games, Learning, and Society Initiative. He investigates the potential of videogame-based technologies for systemic change in education, and his work integrates research and theory on digital media (particularly games) with theories of situated cognition in order to understand how to design educational environments in a digital age.

Alan Taub is a professor of materials science and engineering at the University of Michigan where his research is on advanced materials and processing. He is leading an effort to establish a new center within the College of Engineering that will focus on advanced manufacturing of lightweight material structures for automotive and aerospace applications. Previously, Dr. Taub held positions at GE, Ford Motor Company, and most recently as vice president of Global Research and Development at General Motors.

Matthew Willard is an associate professor in the Department on Materials Science and Engineering at Case Western Reserve University. He has been active in the magnetic materials community, performing research in topic areas including nanocrystalline soft magnets, rare earth permanent magnets, ferromagnetic shape memory alloys, and magnetocaloric materials. His research focuses on the intersection of nanoscience and energy technologies, especially where size, weight, and efficiency of magnetic components are paramount.

Program

NATIONAL ACADEMY OF ENGINEERING

2012 U.S. Frontiers of Engineering Symposium
September 13-15, 2012

Chair: Kristi Anseth, University of Colorado, Boulder

CLIMATE ENGINEERING

Organizers: David Sholl and Armin Sorooshian

Overview of Climate Engineering
Eli Kintisch

Removing Carbon Dioxide from the Atmosphere: Possibilities and Challenges of Air Capture
Christopher Jones

Offsetting Climate Change by Engineering Air Pollution to Brighten Clouds
Lynn Russell

Climate Engineering with Stratospheric Aerosols and Associated Engineering Parameters
Ben Kravitz

***

VEHICLE ELECTRIFICATION

Organizers: Michael W. Degner and Sanjeev Naik

Keeping Up with Increasing Demands for Electrochemical Energy Storage
Jeff Sakamoto

Stronger, Lighter, and More Energy Efficient:
Challenges of Magnetic Material Development for Vehicle Electrification
Matthew Willard

Analysis of Projected Impact of Plug-in Electric Vehicles
on the Distribution Grid
Arindam Maitra

The Car and the Cloud: Automotive Architectures for 2020
Rahul Mangharam

***

SERIOUS GAMES

Organizers: Li-Te Cheng and Ben Sawyer

Overview of Serious Games
Kurt Squire

Moving Innovative Game Technology from the Lab to the Living Room
Richard Marks

How Serious Science Is Being Achieved with Serious Games
Zoran Popovic

Playing to Win: Serious Games for Business
Phaedra Boinodiris

***

ENGINEERING MATERIALS FOR THE BIOLOGICAL INTERFACE

Organizers: Karen Burg and Ali Khademhosseini

Engineering Tissues-to-Tissue Interfaces and the Formation of Complex Tissues
Helen Lu

Identification and Modulation of Biophysical Signals That
Control Stem Cell Function and Fate
David Schaffer

Engineering 3D Tissue Systems to Better Mimic Human Biology
Matthew Gevaert

***

DINNER SPEECH

The Reinvention of the Automobile for 21st Century Sustainability
Alan I. Taub

Participants

NATIONAL ACADEMY OF ENGINEERING

2012 U.S. Frontiers of Engineering Symposium
September 13-15, 2012

Pieter Abbeel
Assistant Professor
Department of Electrical Engineering and Computer Science
University of California, Berkeley

Kristi Anseth*
Distinguished Professor of Chemical and Biological Engineering, Professor of Surgery, and HHMI Assistant Investigator
Department of Chemical and Biological Engineering
University of Colorado, Boulder

Andrea Armani
Assistant Professor
Department of Chemical Engineering and Materials Science
University of Southern California

Muhannad Bakir
Associate Professor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Billy Bardin
R&D Director
Feedstocks and Energy Division
The Dow Chemical Company

Halil Berberoglu
Assistant Professor
Department of Mechanical Engineering
University of Texas at Austin

Phaedra Boinodiris**
Serious Games Program Manager
IBM

Alexandra Boltasseva
Assistant Professor
School of Electrical and Computer Engineering
Purdue University

---

*Organizing Committee
**Speaker

David Brumley
Assistant Professor
Department of Electrical and
Computer Engineering
Carnegie Mellon University

Karen Burg*
Hunter Endowed Chair and Professor
Department of Bioengineering
Clemson University

Xi Chen
Associate Professor
Department of Earth and
Environmental Engineering
Columbia University

Li-Te Cheng*
Software Engineer
Google

Gian Colombo
Metallurgist
Research and Development
Carpenter Technology

Xiquan Cui
Senior Engineer
MEMS Research and Innovation Center
Qualcomm Inc.

Michael Degner*
Senior Technical Leader, Electric
Machine Drives
Electrification Research and Advanced
Engineering
Ford Motor Company

Frank DelRio
Mechanical Engineer
Material Measurement Laboratory
National Institute of Standards and
Technology

Zhiqun (Daniel) Deng
Senior Research Scientist
Energy and Environment Directorate
Pacific Northwest National Laboratory

Jennifer Dionne
Assistant Professor
Department of Materials Science and
Engineering
Stanford University

Nathan Domagalski
Senior Research Investigator II
Development Engineering, R&D
Bristol-Myers Squibb

Ayman EL-Refaie
Senior Engineer
Electrical Machines Lab
GE Global Research

Michael Escuti
Associate Professor
Department of Electrical and
Computer Engineering
North Carolina State University

Joelle Frechette
Assistant Professor
Department of Chemical and
Biomolecular Engineering
Johns Hopkins University

Yan Fu
Technical Specialist
Passive Safety Research and Advanced
Engineering
Ford Motor Company

Weiying Gao
Research Associate, Electrical Engineer
Materials Science and Engineering
DuPont

David Garrett
Associate Technical Director
Office of the CTO
Broadcom Corporation

Brian Gerkey
CEO
Open Source Robotics Foundation

Matthew Gevaert**
CEO & Co-founder
KIYATEC Inc

Sayata Ghose
Materials, Process and Physics Engineer
Boeing Research and Technology
Boeing Company

Jordan Green
Assistant Professor
Department of Biomedical Engineering
Johns Hopkins University

Piyush Gupta
Member of Technical Staff
Mathematics of Networks and Communications Research
Bell Labs, Alcatel-Lucent

Robert Hampshire
Assistant Professor of Operations Research and Public Policy
Heinz College
Carnegie Mellon University

Jessica Harrison
Principal Consultant
Electricity Transmission and Distribution
DNV KEMA Energy and Sustainability

Steve Hartmann
Senior Engineering Manager
Medtronic Navigation
Medtronic

Reed Hendershot
Senior Principal Research Engineer
Merchant Gases R&D
Air Products and Chemicals, Inc.

Elizabeth Hillman
Associate Professor
Department of Biomedical Engineering
Columbia University

Elizabeth Hoegeman
CFS WW Manufacturing Leader
Cummins Fuel Systems
Cummins Inc.

Jeremy Hollman
R&D Propulsion Group Lead/SUAS Group Lead
Cambridge Research and Development Center
Aurora Flight Sciences

Mona Jarrahi
Assistant Professor
Department of Electrical Engineering and Computer Science
University of Michigan

Michael Jewett
Assistant Professor
Department of Chemical and Biological Engineering
Northwestern University

Suzette Johnson
Technical Fellow
Cyber Intelligence Division
Northrop Grumman

Christopher Jones**
New-Vision Professor
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Anupama Kaul
Program Director
ECCS Division, Directorate for Engineering
National Science Foundation

Ali Khademhosseini*
Associate Professor
Harvard-MIT Division of Health Sciences and Technology
Harvard University

Eli Kintisch**
Contributing Correspondent
Science Magazine

Scott Klemmer
Associate Professor
Department of Computer Science
Stanford University

LaShanda Teresa Korley
Nord Distinguished Assistant Professor
Department of Macromolecular Science and Engineering
Case Western Reserve University

Ben Kravitz**
Post-doctoral Research Associate
Pacific Northwest National Laboratory

Christopher Kruegel
Associate Professor
Department of Computer Science
University of California, Santa Barbara

T. C. Michael Law
Associate
Geotechnical Engineering
Mueser Rutledge Consulting Engineers

Chunhao Lee
Staff Researcher
Propulsion Systems Research Lab, GM Global R&D
General Motors

Steven Little
Chair, Department of Chemical and Petroleum Engineering, Associate Professor, and Bicentennial Faculty Fellow
Departments of Chemical Engineering, Bioengineering, Immunology, and The McGowan Institute for Regenerative Medicine
University of Pittsburgh

Xiang Liu
Distinguished Member of Technical Staff
Optical Transmission Systems Research
Bell Labs, Alcatel-Lucent

Helen Lu**
Associate Professor
Department of Biomedical Engineering
Columbia University

Jason Lyons
Research Scientist
Functional Additives R&D
Arkema, Inc.

Brian MacCleery
Principal Product Manager for Clean Energy Technology
Product Management
National Instruments

Arindam Maitra**
Senior Project Manager
Power Delivery and Utilization
Electric Power Research Institute

Rahul Mangharam**
Stephen J. Angello Chair and Assistant Professor
Department of Electrical and Systems Engineering
University of Pennsylvania

Richard Marks**
Senior Member
US R&D
Sony Computer Entertainment

Youssef Marzouk
Associate Professor
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology

Kristyn Masters
Associate Professor
Department of Biomedical Engineering
University of Wisconsin-Madison

Meagan Mauter
Assistant Professor
Chemical Engineering/Engineering and Public Policy
Carnegie Mellon University

Jason May
Research Staff Engineer
Microwave Technology, Microelectronics Lab
HRL Laboratories

J. Rhett Mayor
Associate Professor
Department of Mechanical Engineering
Georgia Institute of Technology

Timothy McKnight
Chairman, Oak Ridge National Laboratory Seed Money Program
Office of Institutional Planning and Measurement Science and Systems Engineering Division
Oak Ridge National Laboratory

Michael McMickell
Technical Manager
Engineering
Honeywell Aerospace

W. David Merryman
Assistant Professor
Department of Biomedical Engineering
Vanderbilt University

Rahul Mital
Global Technical Specialist
Aftertreatment-Diesel, GM Global Powertrain Engineering
General Motors

Mohammad Mofrad
Associate Professor
Department of Bioengineering
University of California, Berkeley

Nathan Moody
Project Leader and Technical Staff Member
High Power Electrodynamics
Los Alamos National Laboratory

Sanjeev Naik*
Engineering Group Manager
Advanced Engine Controls
General Motors

Tse Nga Ng
Senior Research Scientist
Electronic Materials and Devices Lab
Palo Alto Research Center

Elisabeth Nguyen
Engineering Specialist
Software Systems Engineering Department
Aerospace Corporation

Nicholas Peters
Senior Scientist
Broadband Optical Networking Department
Applied Communication Sciences

Desiree Plata
Assistant Professor
Department of Civil and Environmental Engineering
Duke University

Zoran Popovic**
Professor and Director, Center for Game Science
Department of Computer Science and Engineering
University of Washington

David Reeder
Principal Engineer
Process Solutions Technology Development Center
Cargill, Inc.

Kate Riggins
Principal Engineer
Personal Care Research and Development
Procter & Gamble Company

Lynn Russell**
Professor
Climate Sciences
Scripps Institution of Oceanography

Jeff Sakamoto**
Assistant Professor
Department of Chemical Engineering and Materials Science
Michigan State University

Ben Sawyer*
Co-founder
Digitalmill

Wallace Sawyer
Professor and Distinguished Teaching Scholar
Department of Mechanical and Aerospace Engineering
University of Florida

David Schaffer**
Professor
Department of Chemical Engineering, Bioengineering, and Neuroscience
University of California, Berkeley

Charles Schroeder
Assistant Professor
Department of Chemical and Biomolecular Engineering
University of Illinois at Urbana Champaign

Stephanie Severance
Technical Project Leader
Off Highway Engineering
Cummins Inc.

Behrouz Shafei
Assistant Professor
Department of Civil and Environmental Engineering
University of Massachusetts, Amherst

David Sholl*
Michael Tennenbaum Family Chair and GRA Eminent Scholar
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Leena Singh
Principal Member of the Technical Staff
Algorithms and Software
Charles Stark Draper Laboratory

Armin Sorooshian*
Assistant Professor
Department of Chemical and Environmental Engineering
University of Arizona

Shukri Souri
Principal and Practice Director
Electrical Engineering and Computer Science
Exponent, Inc.

Kurt Squire**
Associate Professor
Educational Communications and Technology
University of Wisconsin–Madison

Joshuah Stolaroff
Environmental Scientist
Global Security
Lawrence Livermore National Laboratory

Tong Sun
Principal Scientist
Xerox Innovation Group
Xerox Webster Research Center

Stuart Thomas
Global Technology and Licensing Manager
DuPont Cellulosic Ethanol
Industrial Biosciences
DuPont

Kevin Turner
Associate Professor
Department of Mechanical Engineering and Applied Mechanics
University of Pennsylvania

Kimberly Turner
Professor and Chair
Department of Mechanical Engineering
University of California, Santa Barbara

Chris Urmson
Head of Engineering, Self-Driving Car Project
Google [X]
Google

Peter van Beek
Principal Researcher
Systems, Algorithms and Services
SHARP Laboratories of America, Inc.

Sergei Vassilvitskii
Senior Research Scientist
Strategic Technologies
Google

Kuansan Wang
Principal Researcher
Microsoft Research
Microsoft

Kevin Wasson
Senior Project Engineer - Precision
Advanced Engineering
Corning, Inc.

Ulrike Wegst
Associate Professor
Thayer School of Engineering
Dartmouth College

Matthew Willard**
Associate Professor
Department of Materials Science and Engineering
Case Western Reserve University

Jianzhong Wu
Professor
Department of Chemical and Environmental Engineering
University of California, Riverside

Miao Yu
Associate Professor
Department of Mechanical Engineering
University of Maryland

Dinner Speaker

Alan Taub
Professor
Department of Materials Science and Engineering
University of Michigan

Guests

Laura Adolfie
Director, STEM Development Office
Office of the Assistant Secretary of Defense for Research and Engineering

Andrew Chraplyvy
Vice President
Optical Networks Research
Bell Labs, Alcatel-Lucent

William Hayden
Vice President
The Grainger Foundation

Rosemarie Wesson
Acting Director
Emerging Frontiers in Research and Innovation
National Science Foundation

General Motors

Mark Verbrugge
Director, Chemical Sciences and Materials Systems Laboratory
GM Research and Development

Lerinda Frost
Executive Communications and Business Support
GM Research & Development

National Academy of Engineering

Lance A. Davis
Executive Officer

Janet Hunziker
Senior Program Officer

Vanessa Lester
Program Associate