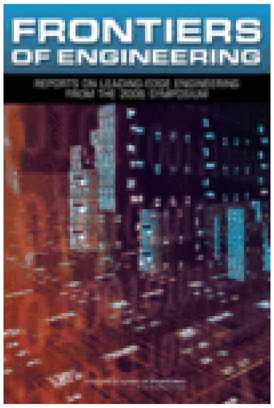


**Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2006 Symposium**

National Academy of Engineering

ISBN: 0-309-66618-X, 202 pages, 6 x 9, (2007)

**This free PDF was downloaded from:**
**http://www.nap.edu/catalog/11827.html**

Visit the National Academies Press online, the authoritative source for all books from the National Academy of Sciences, the National Academy of Engineering, the Institute of Medicine, and the National Research Council:

- Download hundreds of free books in PDF
- Read thousands of books online, free
- Sign up to be notified when new books are published
- Purchase printed books
- Purchase PDFs
- Explore with our innovative research tools

Thank you for downloading this free PDF. If you have comments, questions or just want more information about the books published by the National Academies Press, you may contact our customer service department toll-free at 888-624-8373, visit us online, or send an email to comments@nap.edu.

This free book plus thousands more books are available at http://www.nap.edu.

Copyright © National Academy of Sciences. Permission is granted for this material to be shared for noncommercial, educational purposes, provided that this notice appears on the reproduced materials, the Web address of the online, full authoritative version is retained, and copies are not altered. To disseminate otherwise or to republish requires written permission from the National Academies Press.

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

# FRONTIERS
# OF ENGINEERING

## REPORTS ON LEADING-EDGE ENGINEERING FROM THE 2006 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, N.W. • Washington, D.C. 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 the Air Force Office of Scientific Research, Defense Advanced Research Projects Agency, Department of Defense–DDR&E-Research, National Science Foundation, Ford Motor Company, Microsoft Corporation, Cummins, Inc., and John A. Armstrong.

International Standard Book Number 13: 978-0-309-10339-8
International Standard Book Number 10: 0-309-10339-8

Additional copies of this report are available from The National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20001; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, 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. Wm. A. Wulf 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. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

**www.national-academies.org**

# ORGANIZING COMMITTEE

JULIA M. PHILLIPS (Chair), Director, Physical, Chemical, and Nano Sciences Center, Sandia National Laboratories
APOORV AGARWAL, EVA Engine PMT Leader, Scientific Research Laboratories, Ford Motor Company
M. BRIAN BLAKE, Associate Professor, Department of Computer Science, Georgetown University
TEJAL A. DESAI, Professor of Physiology/Bioengineering, University of California, San Francisco
DAVID B. FOGEL, Chief Executive Officer, Natural Selection, Inc.
HIROSHI MATSUI, Associate Professor, Department of Chemistry, CUNY Graduate Center and Hunter College
JENNIFER K. RYAN, Senior Lecturer, Management Department, College of Business, University College Dublin
WILLIAM F. SCHNEIDER, Associate Professor, Department of Chemical and Biomolecular Engineering, Concurrent in Chemistry, University of Notre Dame
JULIE L. SWANN, Assistant Professor, School of Industrial and Systems Engineering, Georgia Institute of Technology

## Staff

JANET R. HUNZIKER, Senior Program Officer
VIRGINIA R. BACON, Senior Program Assistant

# Preface

In 1995, the National Academy of Engineering (NAE) initiated the Frontiers of Engineering Program, which brings together about 100 young engineering leaders for annual symposia to learn about cutting-edge research and technical work in a variety of engineering fields. The twelfth U.S. Frontiers of Engineering Symposium, at Ford Research and Innovation Center in Dearborn, Michigan, was held on September 21–23, 2006. Speakers were asked to prepare extended summaries of their presentations, which are reprinted in this volume. The intent of this volume, which also includes the text of the dinner speech, a list of contributors, a symposium agenda, and a list of participants, is to convey the excitement of this unique meeting and to highlight cutting-edge developments in engineering research.

## 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 the many developments and challenges in areas that may not be familiar to them.

Every year at the 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 young engineers working in academia, industry, and government in many different engineering disciplines an opportunity to make contacts with and learn from individuals they might 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 will lead to insights that may be applicable in specific disciplines.

The number of participants at each meeting is limited to 100 to maximize opportunities for interactions and exchanges among the participants, 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 the excitement of their work to a technically sophisticated audience with backgrounds in many disciplines. 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 then summarizes the long-term significance of his/her work.

## THE 2006 SYMPOSIUM

The four general topics for the 2006 meeting were: the rise of intelligent software systems and machines, the nano/bio interface, engineering personal mobility for the 21st century, and supply chain management applications with economic and public impact. The rise of intelligent software systems and machines is based on attempts to model the complexity and efficiency of the human brain, or the evolutionary process that created it, with the goal of creating *intelligent systems,* that is, systems that can adapt their behavior to meet the demands of a variety of environments. Speakers in this first session addressed the creation, use, and integration of intelligent systems in various aspects of everyday life in a modern society and suggested future capabilities of machine intelligence. The four talks covered the commercialization of auditory neuroscience, or the development of a machine that can hear; the creation of intelligent agents in games; the co-evolution of the computer and social sciences; and computational cognitive models developed to improve human-robot interactions.

The evolution in engineering that resulted from the harnessing of biomolecular processes, such as self-assembly, catalytic activity, and molecular recognition, was the topic of the session on the bio/nano interface. Two speakers described their work on using biotechnology to solve nanotechnology problems. Their presentations covered biological and biomimetic polypeptide materials and the application of biomimetics to devices. The third and fourth speakers took the opposite approach. They described solving biotechnology problems using nanotechnology. The topics were optical imaging for the in vivo assessment of tissue pathology and the commercialization and future developments in bionanotechnology.

The papers in the session on engineering personal mobility for the 21st century were based on the premise that providing people in the developing world with the same level of personal mobility that people in the developed world enjoy is one of the great challenges for the 21st century. Mobility on that scale must be cost effective, efficient, and environmentally sustainable. Presentations addressed the history and evolution of the availability and expectations of personal mobility, the energy and environmental challenges for current forms of personal mobility, and prospective technologies that could transform personal mobility for this and future generations.

The last session was on supply chain management (SCM) applications with economic and public impact. Although effective SCM is now a significant source of competitive advantage for private companies (e.g., Dell Computer and Wal-Mart), researchers and practitioners have also begun to focus on the public impact of SCM, for example, the relationship between SCM and health care, housing policy, the environment, and national security. The presentations in this session were indicative of the widespread applicability of SCM, such as manufacturing processes, military procurement systems, and public housing policy. The last presentation focused on strategies for dealing with supply chain disruptions.

In addition to the plenary sessions, the participants had many opportunities to engage in informal interactions. For example, they attended a “get-acquainted session,” during which individuals presented short descriptions of their work and answered questions from their colleagues. The entire group was also taken on an informative tour of the Ford Rouge Plant.

Every year, a dinner speech is given by a distinguished engineer on the first evening of the symposium. The speaker this year was W. Dale Compton, Lillian M. Gilbreth Distinguished Professor of Industrial Engineering, Emeritus, Purdue University. His talk, entitled, “The Changing Face of Industrial Research,” included a description of the current situation in industrial research and a discussion of the technical and nontechnical problems facing our country, particularly the need to encourage innovative approaches, which are critical to U.S. competitiveness and national prosperity. The text of Dr. Compton’s remarks is included in this volume.

NAE is deeply grateful to the following organizations for their support of the 2006 symposium: Ford Motor Company, Air Force Office of Scientific Research, Defense Advanced Research Projects Agency, U.S. Department of Defense–DDR&E Research, National Science Foundation, Microsoft Corporation, Cummins, Inc., and Dr. John A. Armstrong. NAE would also like to thank the members of the Symposium Organizing Committee (p. iv), chaired by Dr. Julia M. Phillips, for planning and organizing the event.

# Contents

## THE RISE OF INTELLIGENT SOFTWARE SYSTEMS AND MACHINES

## THE NANO/BIO INTERFACE

**ENGINEERING PERSONAL MOBILITY FOR THE 21ST CENTURY**

**SUPPLY CHAIN MANAGEMENT AND APPLICATIONS WITH ECONOMIC AND PUBLIC IMPACT**

## DINNER SPEECH

## APPENDIXES

# THE RISE OF INTELLIGENT SOFTWARE SYSTEMS AND MACHINES

# Introduction

M. Brian Blake
*Georgetown University*
*Washington, D.C.*

David B. Fogel
*Natural Selection, Inc.*
*La Jolla, California*

The human brain is the most powerful computer *not* developed by man. The complexity, performance, and power dissipation of the human brain are all unmatched in traditional computer and software systems. For decades, scientists have attempted to model the complexity and efficiency of the human brain, or the evolutionary process that created it, in other words, to create *intelligent systems* that can adapt their behavior to meet goals in a variety of environments. In this session, we examine the creation, use, and integration of intelligent systems in our lives and offer insights into the future capabilities of machine intelligence.

# Commercializing Auditory Neuroscience

Lloyd Watts
*Audience Inc.*
*Mountain View, California*

In a previous paper (Watts, 2003), I argued that we now have sufficient knowledge of auditory brain function and sufficient computer power to begin building a realistic, real-time model of the human auditory pathway, a machine that can hear like a human being. Based on extrapolations of computational capacity and advancements in neuroscience and psychoacoustics, a realistic model might be completed in the 2015–2020 time frame. This ambitious endeavor will require a decade of work by a large team of specialists and a network of highly skilled collaborators supported by substantial financial resources. To date, from 2002 to 2006, we have developed the core technology, determined a viable market direction, secured financing, assembled a team, and developed and executed a viable, sustainable business model that provides incentives (expected return on investment) for all participants (investors, customers, and employees), in the short term and the long term. So far, progress has been made on all of these synergistic and interdependent fronts.

## SCIENTIFIC FOUNDATION

The scientific foundation for Audience Inc. is a detailed study of the mammalian auditory pathway (Figure 1), completed with the assistance of eight of the world's leading auditory neuroscientists. Our approach was to build working, high-resolution, real-time models of various system components and validate those models with the neuroscientists who had performed the primary research.

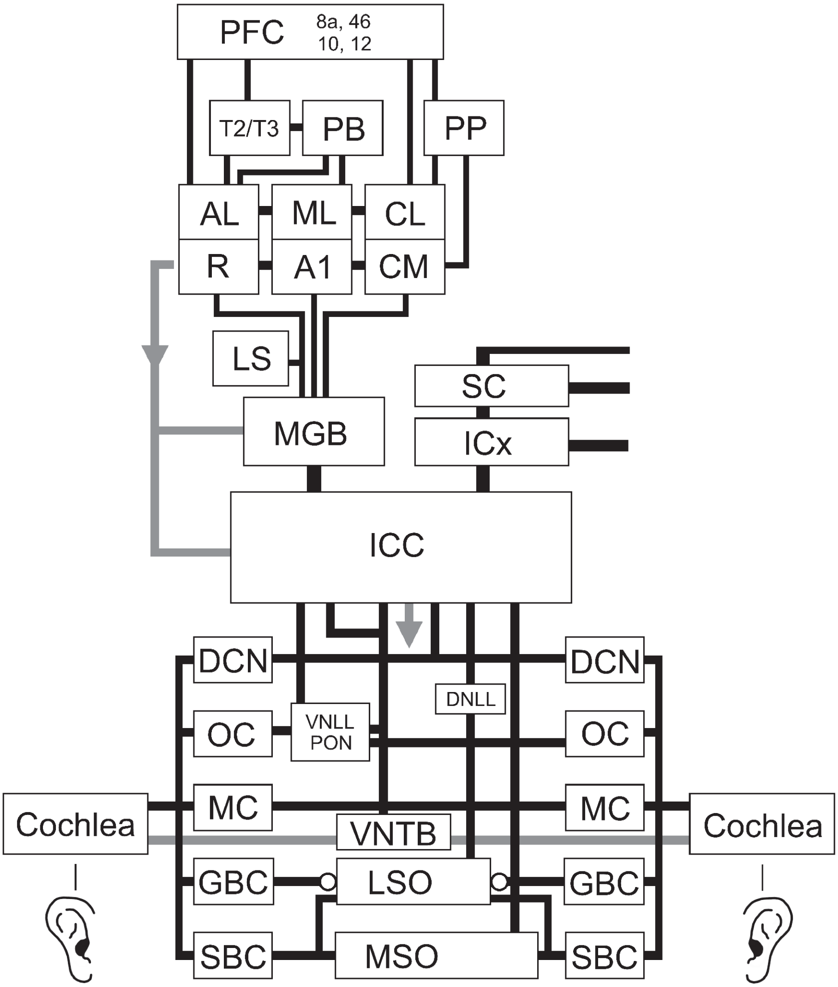


FIGURE 1 A highly simplified diagram of the mammalian auditory pathway. Adapted from Casseday et al., 2002; LeDoux, 1997; Oertel, 2002; Rauschecker and Tian, 2000; and Young, 1998.

The basic model-building began in 1998 and continued through 2002, just when personal computers crossed the 1 GHz mark, which meant that, for the first time in history, it was possible to build working, real-time models of real brain-system components, *in software,* on consumer computer platforms. Early demonstrations in 2001 and 2002 included high-resolution, real-time displays of the cochlea; binaural spatial representations, such as interaural time differences (ITDs) and interaural level differences (ILDs); high-resolution, event-based correlograms; and a demonstration of real-time, polyphonic pitch detection, all based on well-established neuroscience and psychoacoustic findings.

## MARKET FOCUS AND PRODUCT DIRECTION

In the early years of the company, we explored many avenues for commercialization. After a two-year sojourn (from 2002 to 2004) into noise-robust speech recognition, we re-assessed the market and determined that the company's greatest commercial value was in the extraction and reconstruction of the human voice, a technology that could be used to improve the quality of telephone calls made from noisy environments. This insight was driven by the enormous sales in the cell-phone market and the need for cell-phone users to be heard clearly when they placed calls from noisy locations. At that point, work on speech recognition was de-emphasized, and the company began to focus in earnest on commercializing a two-microphone, nonstationary noise suppressor for the mobile telephone market.

## TECHNOLOGY

Figure 2 is a block diagram of Audience's cognitive audio system, which is designed to extract a single voice from a complex auditory scene. The major elements in the system are: Fast Cochlea Transform™ (FCT), a characterization process, a grouping process, a selection process, and Inverse FCT.

- FCT provides a high-quality spectral representation of the sound mixture, with sufficient resolution and without introducing frame artifacts, to allow the characterization of components of multiple sound sources.
- The characterization process involves computing the attributes of sound components used by human beings for grouping and stream separation. These attributes include: pitches of constituent, nonstationary sounds; spatial location cues (when multiple microphones are available), such as onset timing and other transient characteristics; estimation and characterization of quasistationary background noise levels; and so on. These attributes are then associated with the raw FCT data as acoustic tags in the subsequent grouping process.
- The grouping process is a clustering operation in low-dimensionality spaces to "group" sound components with common or similar attributes into a single auditory stream. Sound components with sufficiently dissimilar attributes are associated with different auditory streams. Ultimately, the streams are tracked through time and associated with persistent or recurring sound sources in the auditory environment. The output of the grouping process is the raw FCT data associated with each stream and the corresponding acoustic tags.
- The selection process involves prioritizing and selecting separate auditory sound sources, as appropriate for a given application.

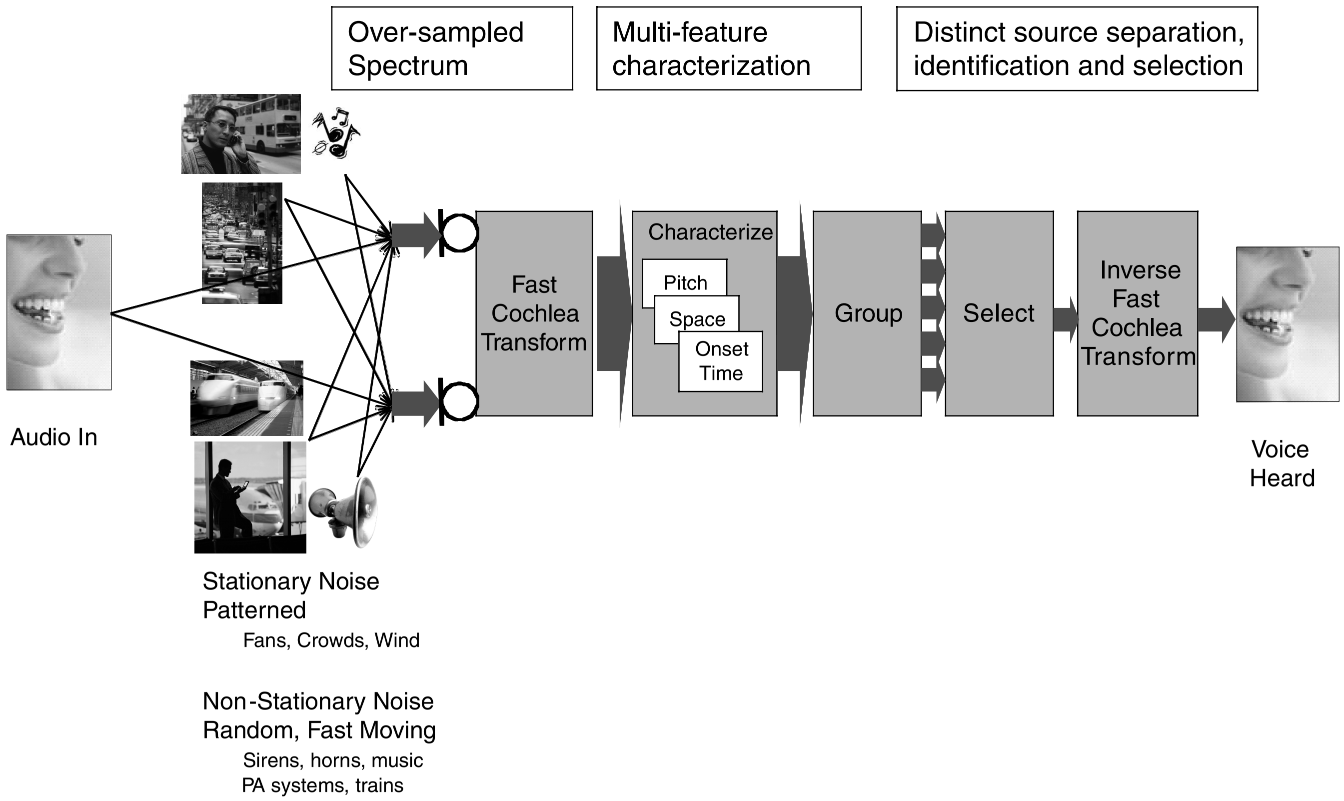


FIGURE 2 Architecture of the cognitive audio system. Source: Audience Inc.

- Inverse FCT telephony applications involve reconstructing and cleaning up the primary output of the system to produce a high-quality voice. Inverse FCT converts FCT data back into digital audio for subsequent processing, including encoding for transmission across a cell-phone channel.

## TECHNICAL DETAILS

### Fast Cochlea Transform™

FCT, the first stage of processing, must have adequate resolution to support high-quality stream separation. Figure 3 shows a comparison of the conventional fast Fourier transform (FFT) and FCT. In many applications, FFT is updated every 10 ms, giving it coarse temporal resolution, as shown in the right half of the FFT panel. FCT is updated with every audio sample, which allows for resolution of glottal pulses, as necessary, to compute periodicity measures on a performant basis as a cue for grouping voice components.

Because of the way FFT is often configured, it provides poor spectral resolution at low frequencies; very often, the following processor (such as a back-

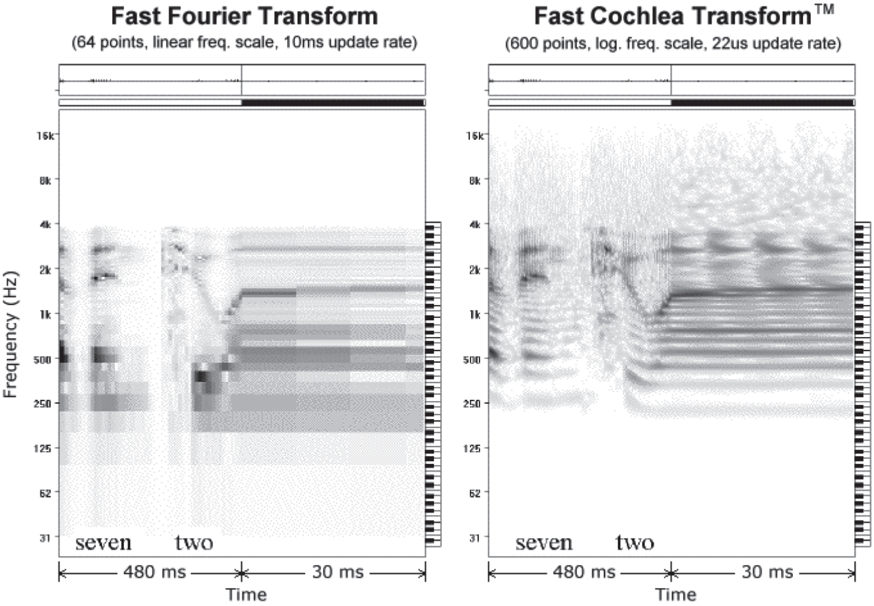
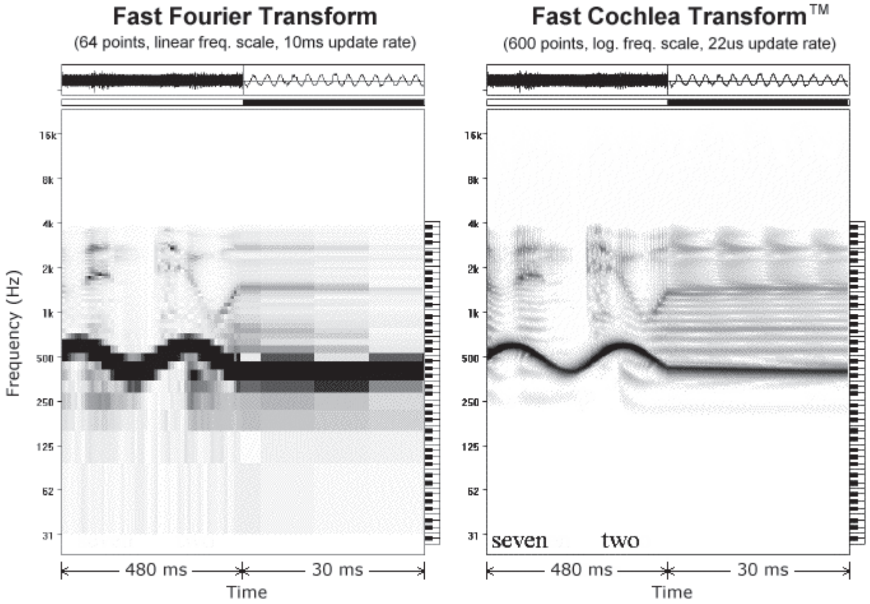


FIGURE 3 Comparison of fast Fourier transform and Fast Cochlea Transform™. Source: Audience Inc.



Fast Fourier Transform
(64 points, linear freq. scale, 10ms update rate)

Fast Cochlea Transform™
(600 points, log. freq. scale, 22us update rate)

FIGURE 4 Multistream separation demonstration (speech + siren). Note that the Fast Cochlea Transform creates a redundant, oversampled representation of the time-varying auditory spectrum. We have found this is necessary to meet the joint requirements of perfect signal reconstruction with no aliasing artifacts, at low latency, with a high degree of modifiability in both the spectral and temporal domains. Source: Audience Inc.

end speech recognizer) is only interested in a smooth estimate of the spectral envelope. FCT, however, is designed to give high-resolution information about individual resolved harmonics so they can be tracked and used as grouping cues in the lower formants.

High resolution is even more important in a multisource environment (Figure 4). In this example, speech is corrupted by a loud siren. The low spectro-temporal resolution of the frame-based FFT makes it difficult to resolve and track the siren, and, therefore, difficult to remove it from speech. The high spectro-temporal resolution of FCT makes it much easier to resolve and track the siren as distinct from the harmonics of the speech signal. The boundaries between the two signals are much better defined, which results in high performance in the subsequent grouping and separation steps.

## Characterization Process

The polyphonic pitch algorithm is capable of resolving the pitch of multiple speakers simultaneously and detecting multiple musical instruments simulta-

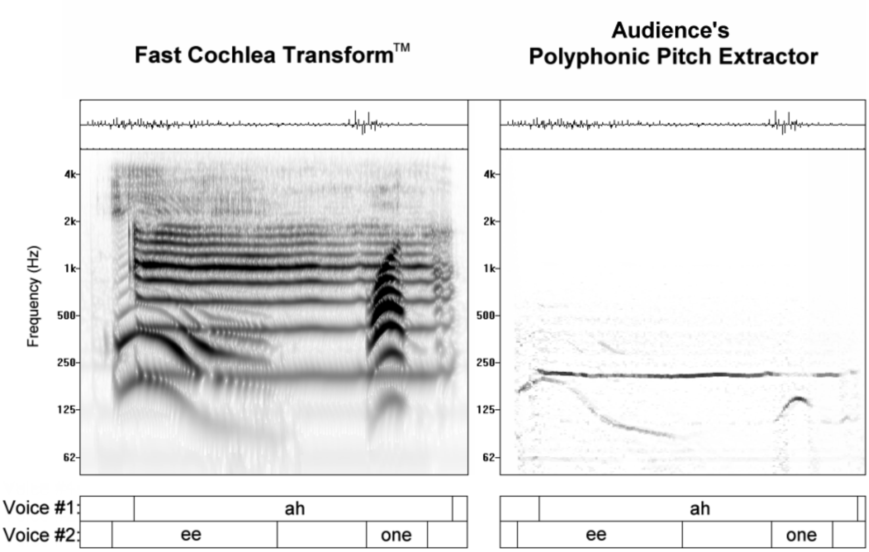


FIGURE 5 Polyphonic pitch for separating multiple simultaneous voices. Source: Audience Inc.

neously. Figure 5 shows how the simultaneous pitches of a male and female speaker are extracted. Spatial localization is valuable for stream separation and locating sound sources, when stereo microphones are available. Figure 6 shows the response of binaural representations to a sound source positioned to the right of the stereo microphone pair.

Figure 7 shows an example of stream separation in a complex audio mixture (voice recorded on a street corner with nearby conversation, noise from a passing car, and ringing of a cell phone) in the cochlear representation. After sound separation, only the voice is preserved.

## Inverse Fast Cochlea Transform

After sound separation in the cochlear (spectral) domain, the audio waveform can be reconstructed for transmission, playback, or storage, using the Inverse FCT. The Inverse FCT combines the spectral components of the FCT back into a time-domain waveform.

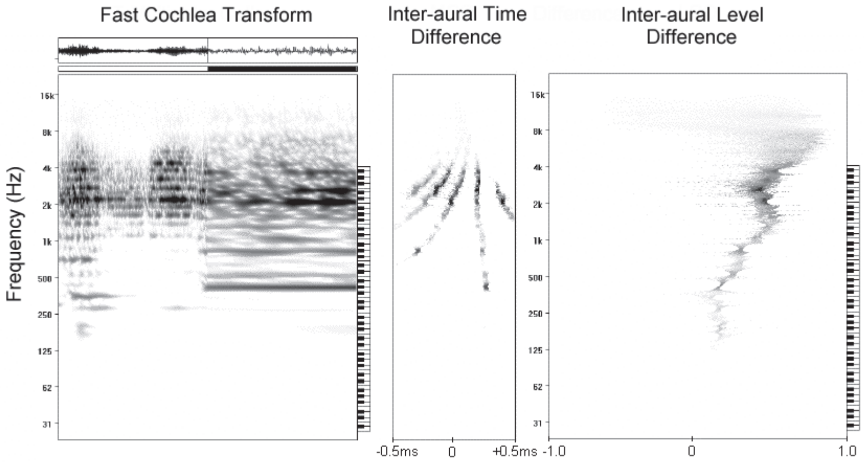


FIGURE 6 Response of the cochlear model and computations of ITD and ILD for spatial localization. Source: Audience Inc.

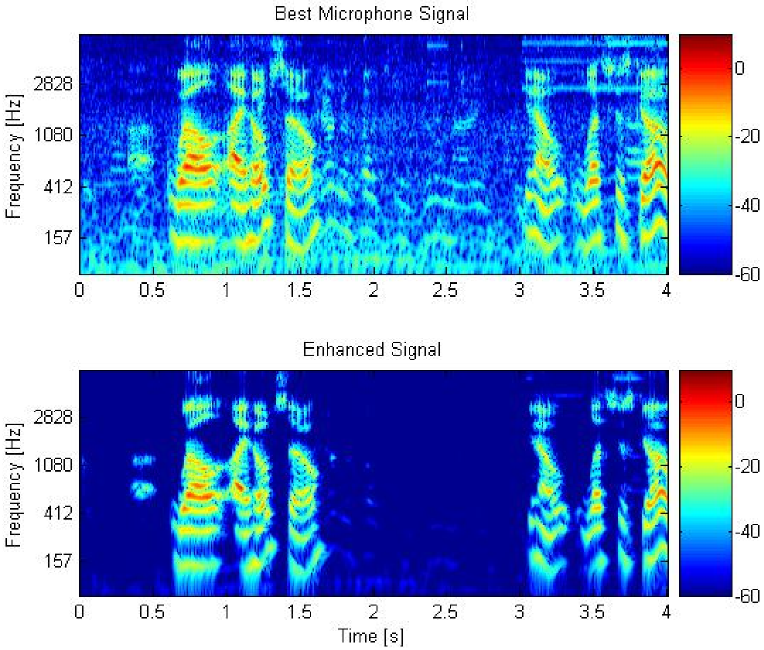


FIGURE 7 Separation of a voice from a street-corner mixture, using a handset with real-time, embedded software. Top panel: mixture of voice with car noise, another voice, and cell-phone ring-tone. Bottom panel: isolated voice. Source: Audience Inc.

## PRODUCT DIRECTION

So far, the company's product direction has remained true to the original goal of achieving commercial success by building machines that can hear like human beings. Along the way, we have found points of divergence between what the brain does (e.g., computes with spikes, uses slow wetware, does not reconstruct audio) and what our system must do to be commercially viable (e.g., compute with conventional digital representations, use fast silicon hardware, provide inverse spectral transformation). In general, however, insights from our studies of the neuroscience and psychoacoustics of hearing have led to insights that have translated into improved signal-processing capacity and robustness.

## PRODUCT IMPLEMENTATION

In the early days of the company, I assumed it would be necessary to build dedicated hardware (e.g., integrated circuits or silicon chips) to support the high computing load of brain-like algorithms. Therefore, I advised investors that Audience would be a fabless semiconductor company with a strong intellectual property position (my catch-phrase was "the nVidia of sound input"). Because the project was likely to take many years and implementation technology changes quickly, Paul Allen, Microsoft cofounder and philanthropist, advised us in 1998 to focus on the algorithms and remain flexible on implementation technology (personal communication, 1998). Eight years later, in 2006, his counsel continues to serve the company well.

As we enter the market with a specific product, we are finding acceptance for both dedicated hardware solutions and embedded software solutions, for reasons that have less to do with computational demands than with the details of integrating our solution into the existing cell-phone platform (e.g., the lack of mixed-signal support for a second microphone). So, the company is a fabless semiconductor company after all, but for very different reasons than I expected when the company was founded in 2000.

## REFERENCES

Casseday, J., T. Fremouw, and E. Covey. 2002. Pp. 238–318 in Integrative Functions in the Mammalian Auditory Pathway, edited by D. Oertel, R. Fay, and A. Popper. New York: Springer-Verlag.

LeDoux, J.E. 1997. Emotion circuits in the brain. Annual Review of Neuroscience 23: 155–184.

Oertel, D. 2002. Pp. 1–5 in Integrative Functions in the Mammalian Auditory Pathway, edited by D. Oertel, R. Fay, and A. Popper. New York: Springer-Verlag.

Rauschecker, J., and B. Tian. 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proceedings of the National Academy of Sciences 97(22): 11800–11806.

Watts, L. 2003. Visualizing Complexity in the Brain. Pp. 45–56 in Computational Intelligence: The Experts Speak, edited by D. Fogel and C. Robinson. New York: IEEE Press/John Wiley & Sons.

Young, E. 1998. Pp. 121–158 in The Synaptic Organization of the Brain, 4th ed., edited by G. Shepherd. New York: Oxford University Press.

## FURTHER READING

Bregman, A. 1990. Auditory Scene Analysis. Cambridge, Mass.: MIT Press.

Shepherd, G. 1998. The Synaptic Organization of the Brain, 4th ed. New York: Oxford University Press.

# Creating Intelligent Agents in Games

Risto Miikkulainen
*The University of Texas at Austin*

Games have long been a popular area for research in artificial intelligence (AI), and for good reason. Because games are challenging yet easy to formalize, they can be used as platforms for the development of new AI methods and for measuring how well they work. In addition, games can demonstrate that machines are capable of behavior generally thought to require intelligence without putting human lives or property at risk.

Most AI research so far has focused on games that can be described in a compact form using symbolic representations, such as board games and card games. The so-called good old-fashioned artificial intelligence (GOFAI; Haugeland, 1985) techniques work well with symbolic games, and to a large extent, GOFAI techniques were developed for them. GOFAI techniques have led to remarkable successes, such as Chinook, a checkers program that became the world champion in 1994 (Schaeffer, 1997), and Deep Blue, the chess program that defeated the world champion in 1997 and drew significant attention to AI in general (Campbell et al., 2002).

Since the 1990s, the field of gaming has changed tremendously. Inexpensive yet powerful computer hardware has made it possible to simulate complex physical environments, resulting in tremendous growth in the video game industry. From modest sales in the 1960s (Baer, 2005), sales of entertainment software reached $25.4 billion worldwide in 2004 (Crandall and Sidak, 2006). Video games are now a regular part of many people's lives, and the market continues to expand.

Curiously, very little AI research has been involved in this expansion. Many video games do not use AI techniques, and those that do are usually based on relatively standard, labor-intensive scripting and authoring methods. In this and other respects, video games differ markedly from symbolic games. Video games often involve many agents embedded in a simulated physical environment where they interact through sensors and effectors that take on numerical rather than symbolic values. To be effective, agents must integrate noisy input from many sensors, react quickly, and change their behavior during the game. The AI techniques developed for and with symbolic games are not well suited to video games.

In contrast, machine-learning techniques, such as neural networks, evolutionary computing, and reinforcement learning, are very well suited to video games. Machine-learning techniques excel in exactly the kinds of fast, noisy, numerical, statistical, and changing domains that today's video games provide. Therefore, just as symbolic games provided an opportunity for the development and testing of GOFAI techniques in the 1980s and 1990s, video games provide an opportunity for the development and testing of machine-learning techniques and their transfer to industry.

## ARTIFICIAL INTELLIGENCE IN VIDEO GAMES

One of the main challenges for AI is creating intelligent agents that can become more proficient in their tasks over time and adapt to new situations as they occur. These abilities are crucial for robots deployed in human environments, as well as for various software agents that live in the Internet or serve as human assistants or collaborators.

Although current technology is still not sufficiently robust to deploy such systems in the real world, they are already feasible in video games. Modern video games provide complex artificial environments that can be controlled and carry less risk to human life than any real-world application (Laird and van Lent, 2000). At the same time, video gaming is an important human activity that occupies millions of people for countless hours. Machine learning can make video games more interesting and reduce their production costs (Fogel et al., 2004) and, in the long run, might also make it possible to train humans realistically in simulated, adaptive environments. Video gaming is, therefore, an important application of AI and an excellent platform for research in intelligent, adaptive agents.

Current video games include a variety of high-realism simulations of human-level control tasks, such as navigation, combat, and team and individual tactics and strategy. Some of these simulations involve traditional AI techniques, such as scripts, rules, and planning (Agre and Chapman, 1987; Maudlin et al., 1984), and a large part of AI development is devoted to path-finding algorithms, such as A*-search and simple behaviors built using finite-state machines. AI is used to control the behavior of the nonplayer characters (NPCs, i.e., autonomous computer-controlled agents) in the game. The behaviors of NPCs, although sometimes impressive, are often repetitive and inflexible. Indeed, a large part of the gameplay in many games is figuring out what the AI is programmed to do and learning to defeat it.

Machine learning in games began with Samuel's (1959) checkers program, which was based on a method similar to temporal-difference learning (Sutton, 1988). This was followed by various learning methods applied to tic-tac-toe, backgammon, go, Othello, and checkers (see Fürnkranz, 2001, for a survey). Recently, machine-learning techniques have begun to appear in video games as well. For example, Fogel et al. (2004) trained teams of tanks and robots to fight each other using a competitive co-evolution system, and Spronck (2005) trained agents in a computer role-playing game using dynamic scripting. Others have trained agents to fight in first- and third-person shooter games (Cole et al., 2004; Hong and Cho, 2004). Machine-learning techniques have also been applied to other video game genres, from Pac-Man (Lucas, 2005) to strategy games (Bryant and Miikkulainen, 2003; Yannakakis et al., 2004).

Nevertheless, very little machine learning is used in current commercial video games. One reason may be that video games have been so successful that a new technology such as machine learning, which would fundamentally change the gaming experience, may be perceived as a risky investment by the industry. In addition, commercial video games are significantly more challenging than the games used in research so far. They not only have large state and action spaces, but they also require diverse behaviors, consistent individual behaviors, fast learning, and memory of past situations (Gomez et al., 2006; Stanley et al., 2005).

## NEUROEVOLUTION

The rest of this article is focused on a particular machine-learning technique, neuroevolution, or the evolution of neural networks. This technique not only promises to rise to the challenge of creating games that are educational, but also promises to provide a platform for the safe, effective study of how intelligent agents adapt.

Evolutionary computation is a computational machine-learning technique modeled after natural evolution (Figure 1a). A population of candidate solutions are encoded as strings of numbers. Each solution is evaluated in the task and assigned a fitness based on how well it performs. Individuals with high fitness are then reproduced (by crossing over their encodings) and mutated (by randomly changing components of their encodings with a low probability). The offspring of the high-fitness individuals replace the low-fitness individuals in the population, and over time, solutions that can solve the task are discovered.

In neuroevolution, evolutionary computation is used to evolve neural net-

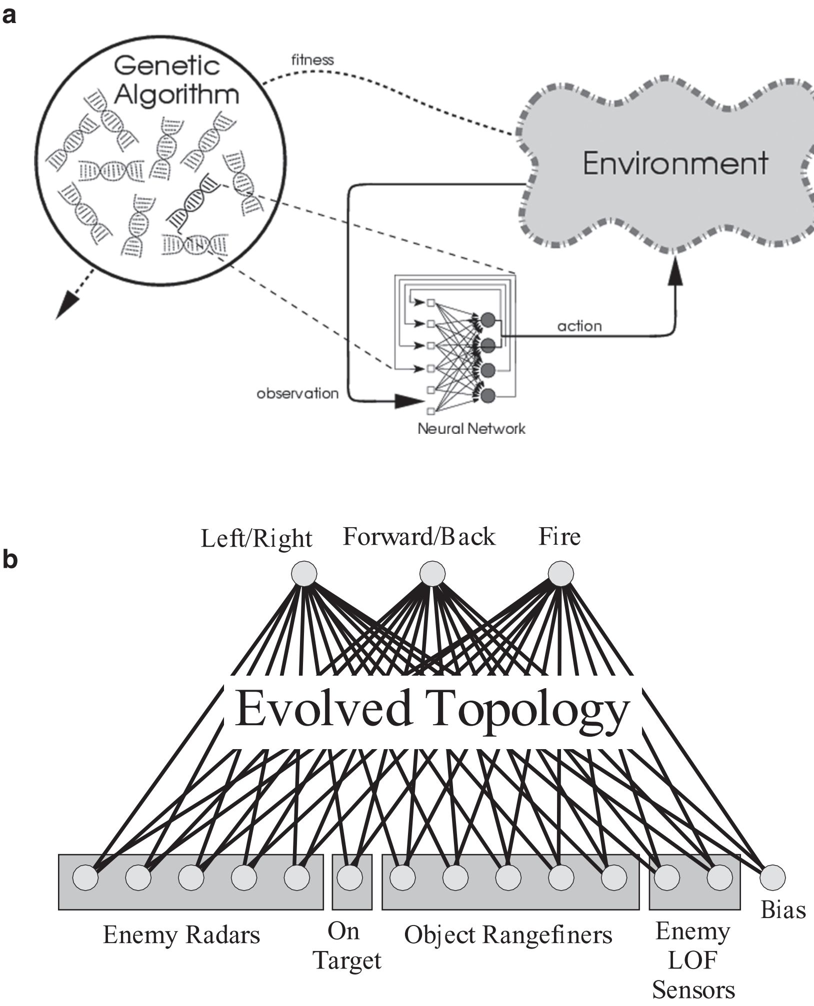


FIGURE 1 a. Evolving neural networks. Solutions (such as neural networks) are encoded as chromosomes, usually consisting of strings of real numbers, in a population. Each individual is evaluated and assigned a fitness based on how well it performs a given task. Individuals with high fitness reproduce; individuals with low fitness are thrown away. Eventually, nearly all individuals can perform the task. b. Each agent in neuroevolution receives sensor readings as input and generates actions as output. In the NERO video game, the network can see enemies, determine whether an enemy is currently in its line of fire, detect objects and walls, and see the direction the enemy is firing. Its outputs specify the direction of movement and whether or not to fire. In this way, the agent is embedded in its environment and must develop sophisticated behaviors to do well. For general neuroevolution software and demos, see *http://nn.cs.utexas.edu*.

work weights and structures. Neural networks perform statistical pattern transformation and generalization, and evolutionary adaptation allows for learning without explicit targets, even with little reinforcement. Neuroevolution is particularly well suited to video games because (1) it works well in high-dimensional spaces, (2) diverse populations can be maintained, (3) individual networks behave consistently, (4) adaptation takes place in real time, and (5) memory can be implemented through recurrency (Gomez et al., 2006; Stanley et al., 2005).

Several methods have been developed for evolving neural networks (Yao, 1999). One particularly appropriate for video games is called neuroevolution of augmenting topologies (NEAT; Stanley and Miikkulainen, 2002), which was originally developed for learning behavioral strategies. The neural networks control agents that select actions in their output based on sensory inputs (Figure 1b). NEAT is unique in that it begins evolution with a population of small, simple networks and *complexifies* those networks over generations, leading to increasingly sophisticated behaviors.

NEAT is based on three key ideas. First, for neural network structures to increase in complexity over generations, a method must be found for keeping track of which gene is which. Otherwise, it will not be clear in later generations which individuals are compatible or how their genes should be combined to produce offspring. NEAT solves this problem by assigning a unique historical marking to every new piece of network structure that appears through a structural mutation. The historical marking is a number assigned to each gene corresponding to its order of appearance over the course of evolution. The numbers are inherited unchanged during crossover, which allows NEAT to perform crossover without expensive topological analysis. Thus, genomes of different organizations and sizes remain compatible throughout evolution.

Second, NEAT speciates the population, so that individuals compete primarily within their own niches instead of with the population at large. In this way, topological innovations are protected and have time to optimize their structures. NEAT uses the historical markings on genes to determine the species to which different individuals belong.

Third, unlike other systems that evolve network topologies and weights, NEAT begins with a uniform population of simple networks with no hidden nodes. New structure is introduced incrementally as structural mutations occur, and only those structures survive that are found to be useful through fitness evaluations. This way, NEAT searches through a minimal number of weight dimensions and finds the appropriate complexity level for the problem. This process of complexification has important implications for the search for solutions. Although it may not be practical to find a solution in a high-dimensional space by searching that space directly, it may be possible to find it by first searching in lower dimensional spaces and complexifying the best solutions into the high-dimensional space.

As is usual in evolutionary algorithms, the entire population is replaced with each generation in NEAT. However, in a real-time game or simulation, this would seem incongruous because every agent's behavior would change at the same time. In addition, behaviors would remain static during the large gaps between generations. Therefore, in order to apply NEAT to video games, a real-time version of it, called rtNEAT, was created.

In rtNEAT, a single individual is replaced every few game ticks. One of the poorest performing individuals is removed and replaced with a child of parents chosen from among the best-performing individuals. This cycle of removal and replacement happens continually throughout the game and is largely invisible to the player. As a result, the algorithm can evolve increasingly complex neural networks fast enough for a user to interact with evolution as it happens in real time. This real-time learning makes it possible to build machine-learning games.

## MACHINE-LEARNING GAMES

The most immediate opportunity for neuroevolution in video games is to build a "mod," a new feature or extension, to an existing game. For example, a character that is scripted in the original game can be turned into an adapting agent that gradually learns and improves as the game goes on. Or, an entirely new dimension can be added to the game, such as an intelligent assistant or tool that changes as the player progresses through the game. Such mods can make the game more interesting and fun to play. At the same time, they are easy and safe to implement from a business point of view because they do not change the original structure of the game. From the research point of view, ideas about embedded agents, adaptation, and interaction can be tested with mods in a rich, realistic game environment.

With neuroevolution, however, learning can be taken well beyond game mods. Entirely new game genres can be developed, such as machine-learning games, in which the player explicitly trains game agents to perform various tasks. The fun and challenge of machine-learning games is to figure out how to take agents through successive challenges so that in the end they perform well in their chosen tasks. Games such as Tamagotchi "Virtual Pet" and Black & White "God Game" suggest that interaction with artificial agents can make for viable and entertaining games. In NERO, the third such game, the artificial agents adapt their behavior through sophisticated machine learning.

## THE NERO GAME

The main idea of NERO is to put the player in the role of a trainer or drill instructor who teaches a team of agents by designing a curriculum. The agents are simulated robots that learn through rtNEAT, and the goal is to train them for military combat.

The agents begin the game with no skills but with the ability to learn. To

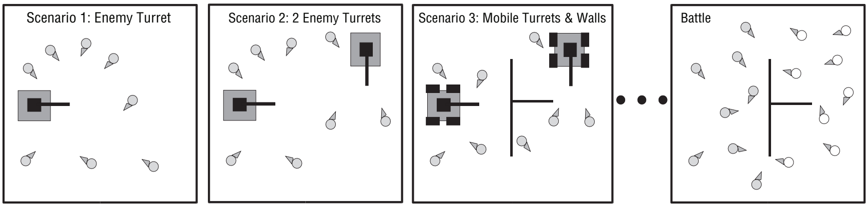


FIGURE 2 A sample training sequence in NERO. The figure depicts a sequence of increasingly difficult training exercises in which agents attempt to attack turrets without getting hit. In the first exercise, there is only a single turret; additional turrets are added by the player as the team improves. Eventually walls are added, and the turrets are given wheels so they can move. Finally, after the team has mastered the hardest exercises, it is deployed in a battle against another team. For animations of various training and battle scenarios, see *http://nerogame.org*.

prepare them for combat, the player must design a sequence of training exercises and goals. Ideally, the exercises will be increasingly difficult so that the team begins by learning basic skills and then gradually builds on them (Figure 2). When the player is satisfied that the team is well prepared, the team is deployed in a battle against another team trained by another player, allowing the players to see if their training strategies pay off.

The challenge is to anticipate the kinds of skills that might be necessary for battle and build training exercises to hone those skills. A player sets up training exercises by placing objects on the field and specifying goals through several sliders. The objects include static enemies, enemy turrets, rovers (i.e., turrets that move), flags, and walls. To the player, the sliders serve as an interface for describing ideal behavior. To rtNEAT, they represent coefficients for fitness components. For example, the sliders specify how much to reward or punish agents for approaching enemies, hitting targets, getting hit, following friends, dispersing, etc. Each individual fitness component is normalized to a Z-score (i.e., the number of standard deviations from the mean) so all components can be measured on the same scale. Fitness is computed as the sum of all components multiplied by their slider levels, which can be positive or negative. Thus, the player has a natural interface for setting up a training exercise and specifying desired behavior.

Agents have several types of sensors (Figure 1b). Although NERO programmers frequently experiment with new sensor configurations, the standard sensors include enemy radars, an “on target” sensor, object range finders, and line-of-fire sensors. To ensure consistent evaluations, agents all begin in a designated area of the field called the factory. Each agent is allowed to spend a limited amount of time on the field during which its fitness can be assessed. When time on the field expires, the agent is transported back to the factory, where another evaluation begins.

Training begins by deploying 50 agents on the field. Each agent is controlled by a neural network with random connection weights and no hidden nodes, which is the usual starting configuration for NEAT. As the neural networks are replaced in real time, behavior improves, and agents eventually learn to perform the task the player has set up. When the player decides performance has reached a satisfactory level, he or she can save the team in a file. Saved teams can be reloaded for further training in different scenarios, or they can be loaded into battle mode.

In battle mode, the player discovers how well the training has worked. Each player assembles a battle team of 20 agents from as many different trained teams as desired, possibly combining agents with different skills. The battle begins with two teams arrayed on opposite sides of the field. When one player presses a "go" button, the neural networks take control of their agents and perform according to their training. Unlike training, however, where being shot does not cause damage to an agent's body, agents in battle are destroyed after being shot several times (currently five). The battle ends when one team is completely eliminated. In some cases, the surviving agents may insist on avoiding each other, in which case the winner is the team with the most agents left standing.

Torque, a game engine licensed from GarageGames (*http://www.garagegames.com/*), drives NERO's simulated physics and graphics. An important property of Torque is that its physics is slightly nondeterministic so that the same game is never played twice. In addition, Torque makes it possible for the player to take control of enemy robots using a joystick, an option that can be useful in training.

Behavior can be evolved very quickly in NERO, fast enough so that the player can be watching and interacting with the system in real time. The most basic battle tactic is to seek the enemy aggressively and fire at it. To train for this tactic, a single static enemy is placed on the training field, and agents are rewarded for approaching the enemy. This training requires that agents learn to run toward a target, which is difficult because they start out in the factory facing in random directions. Starting with random neural networks, it takes on average 99.7 seconds for 90 percent of the agents on the field to learn to approach the enemy successfully (10 runs, $sd = 44.5$ s).

Note that NERO differs from most applications of evolutionary algorithms in that the quality of evolution is judged from the player's perspective based on the performance of the entire population, instead of the performance of the population champion. However, even though the entire population must solve the task, it does not converge to the same solution. In seek training, some agents evolve a tendency to run slightly to the left of the target, while others run to the right. The population diverges because the 50 agents interact as they move simultaneously on the field at the same time. If all of the agents chose exactly the same path, they would often crash into each other and slow each other down, so agents naturally take slightly different paths to the goal. In other words, NERO is a massively parallel, coevolving ecology in which the entire population is evaluated together.

Agents can also be trained to avoid the enemy, leading to different battle tactics. In fact, rtNEAT is flexible enough to devolve a population that has converged on seeking behavior into its complete opposite, a population that exhibits avoidance behavior. For avoidance training, players control an enemy robot with a joystick and run it toward the agents on the field. The agents learn to back away to avoid being penalized for being too near the enemy. Interestingly, they prefer to run away from the enemy backward so they can still see and shoot at the enemy (Figure 3a). As an interesting combination of conflicting goals, a

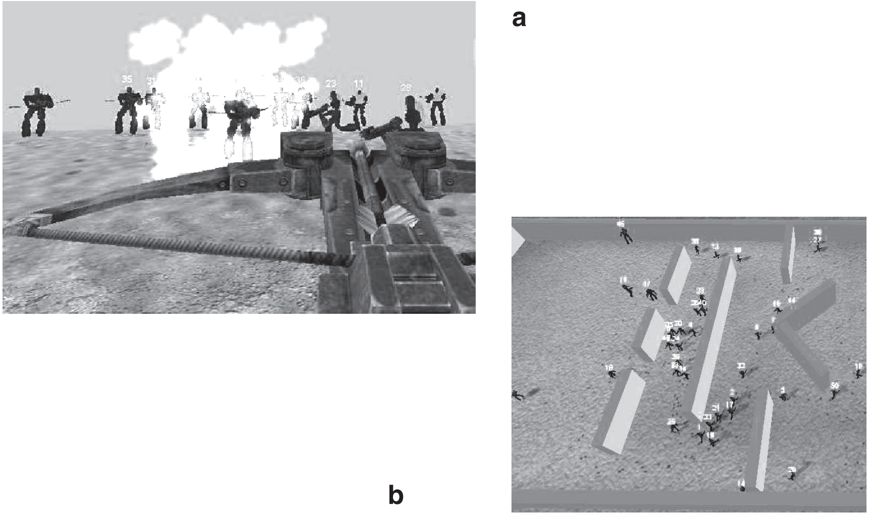


FIGURE 3  Behaviors evolved in NERO. a. This training screenshot shows several agents running away backward while shooting at the enemy, which is being controlled from a first-person perspective by a human trainer with a joystick. This scenario demonstrates how evolution can discover novel and effective behaviors in response to challenges set up by the player. b. Incremental training on increasingly complex wall configurations produced agents that could navigate this complex maze to find the enemy. Remarkably, they had not seen this maze during training, suggesting that they had evolved general path-navigation ability. The agents spawn from the left side of the maze and proceed to an enemy at the right. Notice that some agents evolved to take the path through the top, while others evolved to take the bottom path, suggesting that protecting innovation in rtNEAT supports a range of diverse behaviors with different network topologies. Animations of these and other behaviors can be seen at *http://nerogame.org*.

turret can be placed on the field and agents asked to approach it without getting hit. As a result, they learn to avoid enemy fire, running to the side opposite the bullets and approaching the turret from behind. This tactic is also effective in battle.

Other interesting behaviors have been evolved to test the limits of rtNEAT, rather than specifically preparing troops for battle. For example, agents were trained to run around walls in order to approach the enemy. As performance improved, players incrementally added more walls until the agents could navigate an entire maze (Figure 3b). This behavior was remarkable because it was successful without any path planning.

The agents developed the general strategy of following any wall that stood between them and the enemy until they found an opening. Interestingly, different species evolved to take different paths through the maze, showing that topology and function are correlated in rtNEAT and confirming the success of real-time speciation. The evolved strategies were also general enough for agents to navigate significantly different mazes without further training. In another example, when agents that had been trained to approach a designated location (marked by a flag) through a hallway were attacked by an enemy controlled by the player, they learned, after two minutes, to take an alternative path through an adjacent hallway to avoid the enemy's fire. Such a response is a powerful demonstration of real-time adaptation. The same kind of adaptation could be used in any interactive game to make it more realistic and interesting.

Teams that were trained differently were sometimes surprisingly evenly matched. For example, a seeking team won six out of ten battles, only a slight advantage, against an avoidant team that ran in a pack to a corner of the field next to an enclosing wall. Sometimes, if an avoidant team made it to the corner and assembled fast enough, the seeking team ran into an ambush and was obliterated. However, slightly more often the seeking team got a few shots in before the avoidant team could gather in the corner. In that case, the seeking team trapped the avoidant team and had more surviving numbers. Overall, neither seeking nor avoiding provided a significant advantage.

Strategies can be refined further by observing behaviors during battle and setting up training exercises to improve them. For example, a seeking team could eventually be made more effective against an avoidant team when it was trained with a turret that had its back against the wall. The team learned to hover near the turret and fire when it turned away and to back off quickly when it turned toward them. In this way, rtNEAT can discover sophisticated tactics that dominate over simpler ones. The challenge for the player is to figure out how to set up the training curriculum so sophisticated tactics will emerge.

NERO was created over a period of about two years by a team of more than 30 student volunteers (Gold, 2005). The game was first released in June 2005 at *http://nerogame.org* and has since been downloaded more than 100,000 times. NERO is under continuing development and is currently focused on providing more interactive play. In general, players agree that the game is engrossing and entertaining. Battles are exciting, and players spend many hours perfecting behaviors and assembling teams with just the right combination of tactics. Remarkably, players who have little technical background often develop accurate intuitions about the underlying mechanics of machine learning. This suggests that NERO and other machine-learning games are viable as a genre and may even attract a future generation of researchers to machine learning.

Games like NERO can be used as research platforms for implementing novel machine-learning techniques. For example, one direction for research is to incorporate human knowledge, in terms of rules, into evolution. This knowledge could then be used to seed the population with desired initial behaviors or to give real-time advice to agents during evolution (Cornelius et al., 2006; Yong et al., 2006). Another area for research is to learn behaviors that not only solve a given problem, but solve it in a way that makes sense to a human observer. Although such solutions are difficult to describe formally, a human player may be able to demonstrate them by playing the game himself or herself. An evolutionary learning system can then use these examples to bias learning toward similar behaviors (Bryant, 2006).

## CONCLUSION

Neuroevolution is a promising new technology that is particularly well suited to video game applications. Although neuroevolution methods are still being developed, the technology can already be used to make current games more challenging and interesting and to implement entirely new genres of games. Such games, with adapting intelligent agents, are likely to be in high demand in the future. Neuroevolution may also make it possible to build effective training games, that is, games that adapt as the trainee's performance improves.

At the same time, video games provide interesting, concrete challenges for machine learning. For example, they can provide a platform for the systematic study of methods of control, coordination, decision making, and optimization, within uncertainty, material, and time constraints. These techniques should be widely applicable in other fields, such as robotics, resource optimization, and intelligent assistants. Just as traditional symbolic games catalyzed the development of GOFAI techniques, video gaming may catalyze research in machine learning for decades to come.

## ACKNOWLEDGMENTS

This work was supported in part by the Digital Media Collaboratory of the University of Texas at Austin, Texas Higher Education Coordinating Board, through grant ARP-003658-476-2001, and the National Science Foundation through grants EIA-0303609 and IIS-0083776.

## REFERENCES

Agre, P.E., and D. Chapman. 1987. Pengi: An Implementation of a Theory of Activity. Pp. 268–272 in Proceedings of the 6th National Conference on Artificial Intelligence. Los Altos, Calif.: Morgan Kaufmann.

Baer, R.H. 2005. Videogames: In the Beginning. Springfield, N.J.: Rolenta Press.

Bryant, B.D. 2006. Evolving Visibly Intelligent Behavior for Embedded Game Agents. Ph.D. thesis, University of Texas at Austin. Technical Report AI-06-334.

Bryant, B.D., and R. Miikkulainen. 2003. Neuroevolution for Adaptive Teams. Pp. 2194–2201 in Proceedings of the 2003 Congress on Evolutionary Computation. Piscataway, N.J.: IEEE.

Campbell, M., A.J. Hoane, Jr., and F.-H. Hsu. 2002. Deep Blue. Artificial Intelligence 134: 57–83.

Cole, N., S.J. Louis, and C. Miles. 2004. Using a Genetic Algorithm to Tune First-Person Shooter Bots. Pp. 139–145 in Proceedings of the 2004 Congress on Evolutionary Computation. Piscataway, N.J.: IEEE.

Cornelius, R., K.O. Stanley, and R. Miikkulainen. 2006. Constructing Adaptive AI Using Knowledge-Based Neuroevolution. Pp. 693–708 in AI Game Programming Wisdom 3, edited by S. Rabin. Revere, Mass.: Charles River Media.

Crandall, R.W., and J.G. Sidak. 2006. Video Games: Serious Business for America's Economy. Entertainment Software Association Report. Available online at: *http://www.theesa.com/files/VideoGame_Final.pdf*.

Fogel, D.B., T.J. Hays, and D.R. Johnson. 2004. A Platform for Evolving Characters in Competitive Games. Pp. 1420–1426 in Proceedings of the 2004 Congress on Evolutionary Computation. Piscataway, N.J.: IEEE.

Fürnkranz, J. 2001. Machine Learning in Games: A Survey. Pp. 11–59 in Machines That Learn to Play Games, edited by J. Fürnkranz and M. Kubat. Huntington, N.Y.: Nova Science Publishers.

Gold, A. 2005. Academic AI and Video Games: A Case Study of Incorporating Innovative Academic Research into a Video Game Prototype. Pp. 141–148 in Proceedings of the IEEE 2005 Symposium on Computational Intelligence and Games. Piscataway, N.J.: IEEE.

Gomez, F., J. Schmidhuber, and R. Miikkulainen. 2006. Efficient Non-linear Control through Neuroevolution. Pp. 654–662 in Proceedings of the European Conference on Machine Learning. Berlin: Springer-Verlag.

Haugeland, J. 1985. Artificial Intelligence: The Very Idea. Cambridge, Mass.: MIT Press.

Hong, J.-H., and S.-B. Cho. 2004. Evolution of Emergent Behaviors for Shooting Game Characters in Robocode. Pp. 634–638 in Proceedings of the 2004 Congress on Evolutionary Computation. Piscataway, N.J.: IEEE.

Laird, J.E., and M. van Lent. 2000. Human-Level AI's Killer Application: Interactive Computer Games. Pp. 1171–1178 in Proceedings of the 17th National Conference on Artificial Intelligence. Menlo Park, Calif.: AAAI Press.

Lucas, S.M. 2005. Evolving a Neural Network Location Evaluator to Play Ms. Pac-Man. Pp. 203–210 in Proceedings of the IEEE Symposium on Computational Intelligence and Games. Piscataway, N.J.: IEEE.

Maudlin, M.L., G. Jacobson, A. Appel, and L. Hamey. 1984. ROG-O-MATIC: A Belligerent Expert System. In Proceedings of the 5th National Conference of the Canadian Society for Computational Studies of Intelligence. Mississagua, Ontario: Canadian Society for Computational Studies of Intelligence.

Samuel, A.L. 1959. Some studies in machine learning using the game of checkers. IBM Journal 3: 210–229.

Schaeffer, J. 1997. One Jump Ahead. Berlin: Springer-Verlag.

Spronck, P. 2005. Adaptive Game AI. Ph.D. thesis, Maastricht University, The Netherlands.

Stanley, K.O., and R. Miikkulainen. 2002. Evolving neural networks through augmenting topologies. Evolutionary Computation 10(2): 99–127.

Stanley, K.O., B.D. Bryant, and R. Miikkulainen. 2005. Real-time neuroevolution in the NERO video game. IEEE Transactions on Evolutionary Computation 9(6): 653–668.

Sutton, R.S. 1988. Learning to predict by the methods of temporal differences. Machine Learning 3: 9–44.

Yannakakis, G.N., J. Levine, and J. Hallam. 2004. An Evolutionary Approach for Interactive Computer Games. Pp. 986–993 in Proceedings of the 2004 Congress on Evolutionary Computation. Piscataway, N.J.: IEEE.

Yao, X. 1999. Evolving artificial neural networks. Proceedings of the IEEE 87(9): 1423–1447.

Yong, C.H., K.O. Stanley, R. Miikkulainen, and I. Karpov. 2006. Incorporating Advice into Evolution of Neural Networks. In Proceedings of the 2nd Artificial Intelligence and Interactive Digital Entertainment Conference. Menlo Park, Calif.: AAAI Press.

# Co-Evolution of Social Sciences and Engineering Systems

ROBERT L. AXTELL
*George Mason University*
*Fairfax, Virginia*

Since the birth of engineering, social and behavioral scientists have played an important role in bringing new technologies to market and designing user interfaces. Many of these technologies have also proven to be invaluable to social and behavioral scientists in their efforts to understand people. In other words, there is a kind of *co-evolution* of engineering systems and social sciences.

Multi-agent systems (MAS), in which a population of quasi-autonomous software objects (agents) interact directly with one another and with their environment for purposes of simulation, control, and distributed computation, is poised exactly at this interface. MAS can be considered social systems, in which each member of a heterogeneous population pursues its own objectives, constrained by its interactions with others. Indeed, ideas from the social sciences, including game theory (e.g., mechanism design), economics (e.g., auction theory), and sociology (e.g., social networks), are increasingly being incorporated into such software systems.

At the same time, social scientists are increasingly using software agents to model social processes, where the dominant approach is to represent each person by a software agent. Such models yield high-fidelity depictions of the origin and operation of social institutions (e.g., financial markets, organizational behavior, and the structure of social norms). They can also be used to understand the differential effects of alternative policies on such institutions.

In short, social systems are systems with multiple agents, and MAS are (increasingly) social systems. The *co-evolution* of social and technological systems means that advances in one field lead to progress in the other, nucleating further improvements in the original field, and so on. This interface between these broadly defined fields of knowledge opens up many opportunities for new research.

## SOCIAL SYSTEMS AS MULTI-AGENT SYSTEMS

Building models in which purposive software objects represent individual people is a way around two classical problems—aggregation and the necessity to assume equilibrium—within conventional mathematical modeling in the social sciences. Because social systems are typically composed of a large number of heterogeneous individuals, mathematical models in the social sciences have been one of two types: (1) aggregate models, in which the heterogeneity of the population is either assumed away (e.g., representative agent models) or averaged away by looking only at mean behavior (e.g., systems dynamics models); and (2) models written at the level of individuals, in which "solution" of the models involves all agents engaging only in equilibrium behavior (e.g., Nash equilibria in game theory, Walras-Arrow-Debreu equilibria in economics) and all dynamic paths by which such equilibria might be achieved are neglected. It is clear how the agent approach fixes aggregate models by fully representing individuals. The agent-based approach also remedies the second problem by letting agents interact directly (in general these are out-of-equilibrium interactions); equilibrium is attained only if a path to it is realized from initial conditions.

MAS grew up in the mid-1990s and combined with so-called artificial life (ALife) models, giving rise to agent-based approaches in the social sciences. As the capacity of computer hardware increased exponentially, more sophisticated agent models could be built, using either more cognitively complex agents or a larger number of simple agents, or both. Thus, large agent populations were soon realized in practice leading naturally to the metaphor of an artificial society (Builder and Bankes, 1991).

In modeling an artificial society, a population of objects is instantiated and permitted to interact. Typically, each object represents one individual and has internal data fields that store the specific characteristics of that individual. Each object also has methods of modifying its internal data, describing interactions, and assessing its self-interest (i.e., it can rank the value to itself of alternative actions). This quality of self-interestedness, or purposefulness, makes the objects into agents.

Conventional mathematical models in the social sciences rely heavily on a suite of heroic assumptions that are false empirically and, arguably, do more harm than good as benchmarks. There are four main ways agent-based computing can be used to relax these assumptions. First, mainstream economics makes much of a "representative agent," conceiving the entire economy as simply a scaled-up version of a single decision maker. This specification is easy to relax computationally. Second, economics models normally consider only rational agent behavior, whereby optimal behavior can be deduced by all agents for all time. Not surprisingly, in a MAS of any complexity, such deductions are computationally intractable and cannot be implemented in practice. Thus, models often resort to bounded rationality. Third, modeling conventions have often dictated that agents not interact directly with other individuals but interact either indirectly through aggregate variables or perhaps through some idealized interaction topology (e.g., random graph, lattice). In agent-based computing, however, any topology, including empirically significant networks, can be easily implemented to mediate agent interactions. Finally, equilibrium is the focal point for all solution concepts in the social sciences. Whether equilibrium obtains or not in an agent-based system, the dynamics matter and are fully modeled.

All of the social sciences—anthropology (Axtell et al., 2002; Diamond, 2002, 2005; Kohler and Gumerman, 2000); geography (Gimblett, 2002); social psychology (Kennedy et al., 2001; Latane et al., 1994; Nowak et al., 2000); sociology (Gilbert and Doran, 1994; Gilbert and Conte, 1995; Flache and Macy, 2002; Macy and Willer, 2002); political science (Axelrod, 1984; Kollman et al., 1992; Cederman, 1997; Lustick et al., 2004); economics (Arifovic and Eaton, 1995; Arifovic, 1996; Kollman et al., 1997; Tesfatsion, 1997; Kirman and Vriend, 2000; Luna and Stefansson, 2000; Allen and Carroll, 2001; Arifovic, 2001; Bullard and Duffy, 2001; Luna and Perrone, 2001; Tesfatsion, 2002, 2003; Arifovic and Masson, 2004; Axtell and Epstein, 1999; Axtell et al., 2001; Young, 1998); finance (Palmer et al., 1994; Arthur et al., 1997; Lux, 1998; LeBaron et al., 1999; Lux and Marchesi, 1999; LeBaron, 2000, 2001a, 2001b, 2002, 2006); organizational science (Carley and Prietula, 1994; Prietula et al., 1998); business (Bonabeau and Meyer, 2001; Bonabeau, 2002); many areas of public policy (Axtell and Epstein, 1999; Moss et al., 2001; Saunders-Newton, 2002; Bourges, 2002; Janssen, 2002; Rauch, 2002); transportation science and policy (Nagel and Rasmussen, 1994; Nagel and Paczuski, 1995; Nagel et al., 1998; Gladwin et al., 2003); public health/epidemiology (Wayner, 1996); demography (Kohler, 2001); and the military (Ilachinski, 2004)—have more or less active research programs using agent computing. Although the nature of these applications is idiosyncratic within particular fields, they are unified methodologically in the search for agent specifications that yield empirically observed (or at least empirically plausible) social behavior.

## MULTI-AGENT SYSTEMS AS SOCIAL SYSTEMS

Not only has agent computing changed the practice of the social sciences, but the social sciences have altered the face of MAS. Certain social science methods have been adopted by computer and information scientists not only at the research frontier, but also in commercial systems. In the same way that social scientists have reworked the MAS paradigm for their own ends, developers have adapted extant social science methods to specific problems in their domains.

The role of agents within computer and information science has been primarily to enhance the function of distributed, decentralized systems. For example, in designing a new computer network, each individual node in the network might be given the ability to manage its own resource allocation, based on information about the overall load on the network. This might be done cooperatively or competitively (Huberman, 1987; Miller and Drexler, 1988). Similarly, a well-designed network should function properly regardless of the topology of how the machines are hooked together. Thus, ideas from graph theory and social network theory—each node can be thought of as socially interactive—have been relevant and put to good use.

Basic research on agent systems has been amplified and extended beyond the academic community by the exigencies of e-commerce. The prospect of automated bargaining, contracting, and exchange among software agents has driven investigators to explore the implications of self-interested agents acting autonomously in computer networks and information technology servers.

Because decentralization is an important idea within MAS, ideas from microeconomics and economic general equilibrium theory that focus on decentralization were incorporated into MAS under the rubric "market-oriented programming" (Wellman, 1996). Mechanism design is an approach to the synthesis of economic environments in which the desired performance of a mechanism is specified, and one then figures out what incentives to give the agents in a way that the equilibria (e.g., Nash equilibrium) that are individually rational and incentive compatible achieve the objective. This formalism was developed largely in the 1980s and is today viewed by some as a viable way to design MAS (Kfir-Dahav et al., 2000).

In distributed control, market metaphors have been replaced with actual market models (Clearwater, 1996). Temperature control of a building is an example application (Huberman and Clearwater, 1995) of a MAS that makes explicit use of concepts from economic general equilibrium (e.g., Mas-Collel et al., 1995).

In automated negotiation (e.g., Rosenchein and Zlotkin, 1994), MAS researchers have made extensive use of game theory. In automated contracting, the Contract Net protocol (Weiss, 1999) was an early example of a high-level protocol for efficient cooperation through task sharing in networks of problem solvers. Since then, much more work of this type has been done. More recent work has taken an explicitly social stance, looking for an emergent social order, for example, through the evolution of social orders and customs, institutions, and laws.

## AGENT-BASED TECHNOLOGY AND CO-EVOLUTION

I am aware that portraying agent-based computing as a bridge between engineering and the social sciences may be risky. By touting the apparent effectiveness of a new methodology, there is always a risk that "hype" will overshadow substance and raise unrealistic expectations.

The alternative approach is to paint an evolutionary picture, in which today's new methodologies are seen as logical extensions of adequate but dated conventional methodologies. Thus, progress appears to be natural, with no abrupt changes. This view can be "sold" more easily to existing research communities and is easier to insinuate into conventional discourse.

Evolution or revolution? Continuous change or abrupt change? Smooth transition or phase transition? One is tempted to invoke Kuhn (1962) at this point, but it may be enough to point out that the technical skills required for those who are fomenting change are quite different from those of many current faculty members and those who teach current graduate students. Only a very small subset of social science researchers knows enough about computer science to perform agent-based modeling in their areas of expertise. This is also the major barrier to the systematic adoption of these new techniques—and proof that agent-based modeling constitutes a discontinuous advance.

Assuming that Moore's law will continue to hold true for the next generation (20 to 30 years), the capabilities of agent computing will double every 18 to 24 months, increasing by an order of magnitude each decade. From the social science perspective, this technological revolution will permit the construction of increasingly large models involving greater numbers of progressively more complex agents. When one contemplates the possible desktop hardware of 2020, one can imagine hundreds of gigabytes of ultrafast RAM, fantastic clock and bus speeds, and enormous hard disks. The continuing computer revolution will fundamentally alter the kinds of social science models that can be built. It will also alter the practice of social sciences, as equations give way fully to agents, empirically tested cognitive models arise, and decision models grounded in neuroscience emerge.

It is anyone's guess where co-evolution will lead. Co-evolutionary systems have the capacity to fundamentally alter one another and their environments in novel, creative ways. Thus, speculations for the medium term and long run may look and sound a lot like science fiction.

## REFERENCES

Allen, T.M., and C.D. Carroll. 2001. Individual learning about consumption. Macroeconomic Dynamics 5(2): 255–271.

Arifovic, J. 1996. The behavior of exchange rates in the genetic algorithm and experimental economies. Journal of Political Economy 104(3): 510–541.

Arifovic, J. 2001. Evolutionary dynamics of currency substitution. Journal of Economic Dynamics and Control 25: 395–417.

Arifovic, J., and C. Eaton. 1995. Coordination via genetic learning. Computational Economics 8: 181–203.

Arifovic, J., and P. Masson. 2004. Heterogeneity and evolution of expectations in a model of currency crisis. Nonlinear Dynamics, Psychology, and Life Sciences 8(2): 231–258.

Arthur, W.B., J.H. Holland, B. LeBaron, R. Palmer, and P. Tayler. 1997. Asset Pricing Under Endogenous Expectations in an Artificial Stock Market. Pp. 15–44 in The Economy as an Evolving Complex System II, edited by W.B. Arthur, S.N. Durlauf, and D.A. Lane. Reading, Mass.: Addison-Wesley.

Axelrod, R. 1984. The Evolution of Cooperation. New York: Basic Books.

Axtell, R.L., and J.M. Epstein. 1999. Coordination in Transient Social Networks: An Agent-Based Computational Model of the Timing of Retirement. Pp. 161–183 in Behavioral Dimensions of Retirement Economics, edited by H.J. Aaron. Washington, D.C.: The Brookings Institution Press.

Axtell, R.L., J.M. Epstein, J.S. Dean, G.J. Gumerman, A.C. Swedlund, J. Harburger, S. Chakravarty, R. Hammond, J. Parker, and M.T. Parker. 2002. Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley. Proceedings of the National Academy of Sciences of the U.S.A. 99(supp 3): 7275–7279.

Axtell, R.L., J.M. Epstein, and H.P. Young. 2001. The Emergence of Classes in a Multi-Agent Bargaining Model. Pp. 191–211 in Social Dynamics, edited by S.N. Durlauf and H.P. Young. Cambridge, Mass./Washington, D.C.: MIT Press/Brookings Institution Press.

Bonabeau, E. 2002. Agent-based modeling: methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences of the U.S.A. 99(supp 3): 7280–7287.

Bonabeau, E., and C. Meyer. 2001. Swarm intelligence: a whole new way to think about business. Harvard Business Review 79: 106–114.

Bourges, C. 2002. Computer-Based Artificial Societies May Create Real Policy. Washington Times, May 12, 2002.

Builder, C., and S. Bankes. 1991. Artificial Societies: A Concept for Basic Research on the Societal Impacts of Information Technology. Santa Monica, Calif.: RAND Corporation.

Bullard, J., and J. Duffy. 2001. Learning and excess volatility. Macroeconomic Dynamics 5(2): 272–302.

Carley, K.M., and M.J. Prietula. 1994. Computational Organization Theory. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Cederman, L.-E. 1997. Emergent Actors and World Politics: How States and Nations Develop and Dissolve. Princeton, N.J.: Princeton University Press.

Clearwater, S.H. 1996. Market-Based Control. Hackensack, N.J.: World Scientific.

Diamond, J.M. 2002. Life with the artificial Anasazi. Nature 419: 567–569.

Diamond, J.M. 2005. Collapse: How Societies Choose to Fail or Survive. New York: Allen Lane.

Flache, A., and M.W. Macy. 2002. Stochastic collusion and the power law of learning. Journal of Conflict Resolution 46: 629–653.

Gilbert, N., and R. Conte, eds. 1995. Artificial Societies: The Computer Simulation of Social Life. London: UCL Press.

Gilbert, N., and J. Doran, eds. 1994. Simulating Societies: The Computer Simulation of Social Phenomena. London: UCL Press.

Gimblett, H.R., ed. 2002. Integrating Geographic Information Systems and Agent-Based Modeling Techniques for Simulating Social and Ecological Processes. Santa Fe Institute Studies in the Sciences of Complexity. New York: Oxford University Press.

Gladwin, T., C.P. Simon, and J. Sullivan. 2003. Workshop on Mobility in a Sustainable World: A Complex Systems Approach, Ann Arbor, Michigan, June 20–22, 2003.

Huberman, B.A., ed. 1987. The Ecology of Computation. New York: North-Holland.

Huberman, B.A., and S.H. Clearwater. 1995. A Multi-Agent System for Controlling Building Environments. First International Conference on Multi-Agent Systems, San Francisco, Calif. Cambridge, Mass.: AAAI Press/MIT Press.

Ilachinski, A. 2004. Artificial War: Multiagent-Based Simulation of Combat. Singapore: World Scientific Publishing.

Janssen, M.A., ed. 2002. Complexity and Ecosystem Management: The Theory and Practice of Multi-Agent Systems. Northampton, Mass.: Edward Elgar Publishing Inc.

Kennedy, J., R.C. Eberhart, and Y. Shi. 2001. Swarm Intelligence. San Francisco, Calif.: Morgan Kaufmann.

Kfir-Dahav, N.E., D. Monderer, and M. Tennenholtz. 2000. Mechanism Design for Resource Bounded Agents. Pp. 309–315 in Proceedings of the Fourth International Conference on Multi-Agent Systems. Boston, Mass.: IEEE Computer Society.

Kirman, A.P., and N.J. Vriend. 2000. Learning to Be Loyal: A Study of the Marseille Fish Market. Pp. 33–56 in Interaction and Market Structure: Essays on Heterogeneity in Economics, edited by D. Delli Gatti and A.P. Kirman. Berlin, Germany: Springer-Verlag.

Kohler, H.-P. 2001. Fertility and Social Interactions. New York: Oxford University Press.

Kohler, T.A., and G.J. Gumerman, eds. 2000. Dynamics in Human and Primate Societies: Agent-Based Modeling of Social and Spatial Processes. Santa Fe Institute Studies in the Sciences of Complexity. New York: Oxford University Press.

Kollman, K., J.H. Miller, and S.E. Page. 1992. Adaptive parties in spatial elections. American Political Science Review 86: 929–937.

Kollman, K., J.H. Miller, and S.E. Page. 1997. Political institutions and sorting in a tiebout model. American Economic Review 87(5): 977–992.

Kuhn, T.S. 1962. The Structure of Scientific Revolutions. Chicago, Ill.: University of Chicago Press.

Latane, B., A. Nowak, and J.H. Liu. 1994. Measuring emergent social phenomena: dynamicism, polarization and clustering as order parameters of social systems. Behavioral Science 39: 1–24.

LeBaron, B. 2000. Agent-based computational fiannce: suggested readings and early research. Journal of Economic Dynamics and Control 24: 324–338.

LeBaron, B. 2001a. A builder's guide to agent-based financial martkets. Quantitative Finance 1: 254–261.

LeBaron, B. 2001b. Evolution and time horizons in an agent-based stock market. Macroeconomics Dynamics 5: 225–254.

LeBaron, B. 2002. Short-memory traders and their impact on group learning in financial markets. Proceedings of the National Academy of Sciences of the U.S.A. 99(supp 3): 7201–7206.

LeBaron, B. 2006. Agent-Based Financial Markets: Matching Stylized Facts with Style. Pp. 221–238 in Post Walrasian Macroeconomics: Beyond the Dynamic Stochastic General Equilibrium Model, edited by D.C. Colander. New York: Cambridge University Press.

LeBaron, B., W.B. Arthur, and R. Palmer. 1999. Time series properties of an artificial stock market. Journal of Economics Dynamics and Control 23: 1487–1516.

Luna, F., and A. Perrone, eds. 2001. Agent-Based Methods in Economics and Finance: Simulations in Swarm. Boston, Mass.: Kluwer Academic Publishers.

Luna, F., and B. Stefansson, eds. 2000. Economic Simulations in Swarm: Agent-Based Modeling and Object Oriented Programming. Advances in Computational Economics, volume 14. Boston, Mass.: Kluwer Academic Publishers.

Lustick, I.S., D. Miodownick, and R.J. Eidelson. 2004. Secessionism in multicultural states: does sharing power prevent or encourage it? American Political Science Review 98(2): 209–229.

Lux, T. 1998. The socioeconomic dynamics of speculative markets: interacting agents, chaos and the fat tails of return distributions. Journal of Economic Behavior and Organization 33: 143–165.

Lux, T., and M. Marchesi. 1999. Scaling and criticality in a stochastic multi-agent model of a financial market. Nature 397: 498–500.

Macy, M.W., and R. Willer. 2002. From factors to actors: computational sociology and agent-based modeling. Annual Review of Sociology 28: 143–166.

Mas-Collel, A., M.D. Whinston, and J.R. Green. 1995. Microeconomic Theory. New York: Oxford University Press.

Miller, M.S., and K.E. Drexler. 1988. Markets and Computation: Open Agoric Systems. Pp. 231–266 in The Ecology of Computation, edited by B.A. Huberman. New York: North-Holland.

Moss, S., C. Pahl-Wostl, and T. Downing. 2001. Agent-based integrated assessment modelling: the example of climate change. Integrated Assessment 2(1): 17–30.

Nagel, K., and M. Paczuski. 1995. Emergent traffic jams. Physical Review E 51: 2909–2918.

Nagel, K., and S. Rasmussen. 1994. Traffic at the Edge of Chaos. Pp. 224–235 in Artificial Life IV, edited by R.A. Brooks and P. Maes. Cambridge, Mass.: MIT Press.

Nagel, K., R. Beckman, and C.L. Barrett. 1998. TRANSIMS for Transportation Planning. Technical Report. Los Alamos, N.M.: Los Alamos National Laboratory.

Nowak, A., R.R. Vallacher, A. Tesser, and W. Borkowski. 2000. Society of self: the emergence of collective properties in self-structure. Psychological Review 107: 39–61.

Palmer, R.G., W.B. Arthur, J.H. Holland, B. LeBaron, and P. Tayler. 1994. Artificial economic life: a simple model of a stock market. Physica D 75: 264–274.

Prietula, M.J., K.M. Carley, and L. Gasser, eds. 1998. Simulating Organizations: Computational Models of Institutions and Groups. Cambridge, Mass.: MIT Press.

Rauch, J. 2002. Seeing around corners. Atlantic Monthly 289: 35–48.

Rosenschein, J.S., and G. Zlotkin. 1994. Rules of Encounter: Designing Conventions for Automated Negotiation among Computers. Cambridge, Mass.: MIT Press.

Saunders-Newton, D. 2002. Introduction: computer-based methods: state of the art. Social Science Computer Review 20(4): 373–376.

Tesfatsion, L. 1997. How Economists Can Get a Life. Pp. 533–564 in The Economy as an Evolving Complex System, Volume II, edited by W.B. Arthur, S. Durlauf, and D.A. Lane. Menlo Park, Calif.: Addison-Wesley.

Tesfatsion, L. 2002. Agent-based eomputational economics: growing economies from the bottom up. Artificial Life 8(1): 55–82.

Tesfatsion, L. 2003. Agent-based computational economics: modeling economies as complex adaptive systems. Information Sciences 149(4): 262–268.

Wayner, P. 1996. Computer Simulations: New-Media Tools for Online Journalism. New York Times, October 9, 1996.

Weiss, G., ed. 1999. Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence. Cambridge, Mass.: MIT Press.

Wellman, M. 1996. Market-Oriented Programming: Some Early Lessons. Pp. 74–95 in Market Based Control: A Paradigm for Distributed Resource Allocation, edited by S.H. Clearwater. Hackensack, N.J.: World Scientific Press.

Young, H.P. 1998. Individual Strategy and Social Structure. Princeton, N.J.: Princeton University Press.

# Using Computational Cognitive Models to Improve Human-Robot Interaction

Alan C. Schultz
*Naval Research Laboratory*
*Washington, D.C.*

We propose an approach for creating more cognitively capable robots that can interact more naturally with humans. Through analysis of human team behavior, we build computational cognitive models of particular high-level human skills that we have determined to be critical for good peer-to-peer collaboration and interaction. We then use these cognitive models as reasoning mechanisms on the robot, enabling the robot to make decisions that are conducive to good interaction with humans.

## COGNITIVELY ENHANCED INTELLIGENT SYSTEMS

We hypothesize that adding computational cognitive reasoning components to intelligent systems such as robots will result in three benefits:

1. Most, if not all, intelligent systems must interact with humans, who are the ultimate users of these systems. Giving the system cognitive models can improve the human-system interface by creating more common ground in the form of cognitively plausible representations and qualitative reasoning. For example, mobile robots generally use representations, such as rotational and translational matrixes, to represent motion and spatial references. However, because this is not a natural mechanism for humans, additional computations must be made to translate between these matrixes and the qualitative spatial reasoning

used by humans. By using cognitive models, reasoning mechanisms, and representations, we believe we can achieve a more effective and efficient interface.

2. Because the resulting system interacts with humans, giving the system behaviors that are more natural and compatible with human behaviors can also result in more natural interactions between human and intelligent systems. For example, mobile robots that must work collaboratively with humans can have less than effective interactions with them if their behaviors are alien or non-intuitive to humans. By incorporating cognitive models, we can develop systems whose behavior can be more easily anticipated by humans and is more natural. Therefore, these systems are more compatible with human team members.

3. One key area of interest is measuring the performance of intelligent systems. We propose that the performance of a cognitively enhanced intelligent system can be compared directly to human-level performance. In addition, if cognitive models of human performance have been developed in creating an intelligent system, we can directly compare the behavior and performance of the task by the intelligent system to the human subject's behavior and performance.

## Hide and Seek

Our foray into this area began when we were developing computational cognitive models of how young children learn the game of hide and seek (Trafton et al., 2005b; Trafton and Schultz, 2006). The purpose was to create robots that could use human-level cognitive skills to make decisions about where to look for people or things hidden by people. The research resulted in a hybrid architecture with a reactive/probabilistic system for robot mobility (Schultz et al., 1999) and a high-level cognitive system based on ACT-R (Anderson and Lebiere, 1998) that made the high-level decisions for where to hide or seek (depending on which role the robot was playing). Not only was this work interesting in its own right, but it also led us to the realization that "perspective taking" is a critical cognitive ability for humans, particularly when they want to collaborate.

## Spatial Perspective Taking

To determine how important perspective and frames of reference are in collaborative tasks in shared space (and also because we were working on a DARPA-funded project to move these capabilities to the NASA Robonaut), we analyzed a series of tapes of a ground controller and two astronauts undergoing training in the NASA Neutral Buoyancy Tank facility for an assembly task for Space Station Mission 9A. When we performed a protocol analysis of these tapes (approximately 4000 utterances) focusing on the use of spatial language and commands from one person to another, we found that the astronauts changed their frame of reference approximately every other utterance. As an example of how prevalent these changes in frame of reference are, consider the following utterance from ground control:

> "… if you come *straight down* from where you are, uh, and uh, kind of peek *down under the rail* on the *nadir side*, by *your right hand*, almost *straight nadir*, you should see the…."

Here we see five changes in frame of reference (highlighted in italics) in a single sentence! This rate of change is consistent with the results found by Franklin et al. (1992). In addition, we found that the astronauts had to adopt others' perspectives, or force others to adopt their perspective, about 25 percent of the time (Trafton et al., 2005a). Obviously, the ability to handle changing frames of reference and to understand spatial perspective will be a critical skill for the NASA Robonaut and, we would argue, for any other robotic system that must communicate with people in spatial contexts (i.e., any construction task, direction giving, etc.).

### Models of Perspective Taking

Imagine the following task, as illustrated in Figure 1. An astronaut and his robotic assistant are working together to assemble a structure in shared space. The human, who because of an occluded view can see only one wrench, says to the robot, "Pass me the wrench." From the robot's point of view, however, two

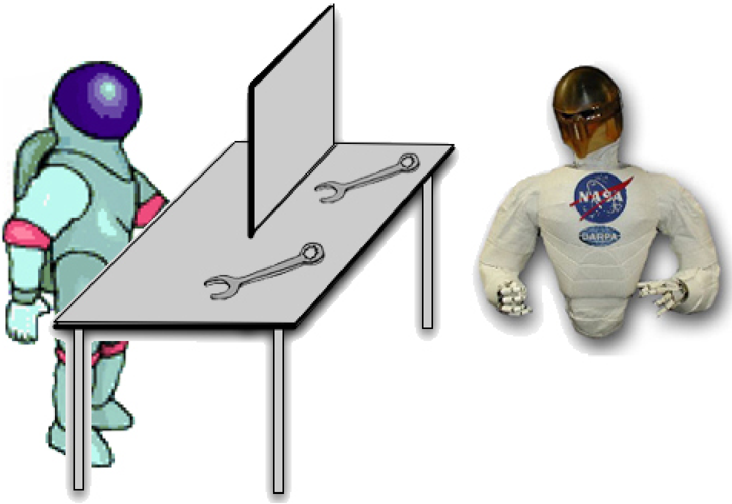


FIGURE 1 A scenario in which an astronaut asks a robot to "Pass me the wrench."

wrenches are visible. What should the robot do? Evidence suggests that humans, in similar situations, will pass the wrench they know the other human can see because this is a jointly salient feature (Clark, 1996).

We developed two models of perspective taking that could handle this scenario in a general sense. In the first approach, we used the ACT-R/S system (Harrison and Schunn, 2002) to model perspective taking using a cognitively plausible spatial representation. In the second approach, we used Polyscheme (Cassimatis et al., 2004) to model the cognitive process of mental simulation; humans tend to simulate situations mentally to solve problems. Using these models, we demonstrated a robot that could solve problems similar to the wrench problem.

## FUTURE WORK

We are now exploring other human cognitive skills that seem important for peer-to-peer collaborative tasks and that are appropriate for building computational cognitive models that can be added to our robots. One new skill we are considering is nonvisual, high-level focus of attention, which helps focus a person's attention on appropriate parts of the environment or situations based on current conditions, task, expectations, models of other agents in the environment, and other factors. Another human cognitive skill we are considering involves the role of anticipation in human interaction and decision making.

## CONCLUSION

Clearly, for humans to work as peers with robots in shared space, the robot must be able to understand the natural human tendency to use different frames of reference and must be able to take the human's perspective. To create robots with these capabilities, we propose using computational cognitive models, rather than more traditional programming paradigms for robots, for the following reasons. First, a natural and intuitive interaction reduces cognitive load. Second, more predictable behavior engenders trust. Finally, more understandable decisions by the robot enable the human to recognize and more quickly repair mistakes. We believe that computational cognitive models will give our robots the cognitive skills they need to interact more naturally with humans, particularly in peer-to-peer relationships.

## REFERENCES

Anderson, J.R., and C. Lebiere. 1998. Atomic Components of Thought. Mahwah, N.J.: Lawrence Erlbaum Associates.

Cassimatis, N., J.G. Trafton, M. Bugajska, and A.C. Schultz. 2004. Integrating cognition, perception, and action through mental simulation in robots. Robotics and Autonomous Systems 49(1–2): 13–23.

Clark, H.H. 1996. Using Language. New York: Cambridge University Press.

Franklin, N., B. Tversky, and V. Coon. 1992. Switching points of view in spatial mental models. Memory & Cognition 20(5): 507–518.

Harrison, A.M., and C.D. Schunn. 2002. ACT-R/S: A Computational and Neurologically Inspired Model of Spatial Reasoning. P. 1008 in Proceedings of the 24th Annual Meeting of the Cognitive Science Society, edited by W.D. Gray and C.D. Schunn. Fairfax, Va.: Lawrence Erlbaum Associates.

Schultz, A., W. Adams, and B. Yamauchi. 1999. Integrating exploration, localization, navigation and planning through a common representation. Autonomous Robots 6(3): 293–308.

Trafton, J.G., and A.C. Schultz. 2006. Children and Robots Learning to Play Hide and Seek. Pp. 242–249 in Proceedings of ACM Conference on Human-Robot Interaction, March 2006. Salt Lake City, Ut.: Association for Computing Machinery.

Trafton, J.G., N.L. Cassimatis, D.P. Brock, M.D. Bugajska, F.E. Mintz, and A.C. Schultz. 2005a. Enabling effective human-robot interaction using perspective-taking in robots. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans 35(4): 460–470.

Trafton, J.G., A.C. Schultz, N.L. Cassimatis, L. Hiatt, D. Perzanowski, D.P. Brock, M. Bugajska, and W. Adams. 2005b. Communicating and Collaborating with Robotic Agents. Pp. 252–278 in Cognitive and Multiagent Interaction: From Cognitive Modeling to Social Simulation, edited by R. Sun. Mahwah, N.J.: Lawrence Erlbaum Associates.

# THE NANO/BIO INTERFACE

# Introduction

Tejal Desai
*University of California, San Francisco*

Hiroshi Matsui
*City University of New York, Hunter College*

The performance of natural systems in various aspects of engineering is often superior to the performance of man-made technologies. Hence, biomimetics in nanoscale are being actively investigated to solve a variety of engineering problems. Biology can provide tools for controlling material synthesis, physical properties, sensing, and mechanical properties at the molecular level. Harnessing biomolecular processes, such as self-assembly, catalytic activity, and molecular recognition, can greatly enhance purely synthetic systems. Therefore, the integration of these fields is a natural evolution in engineering.

The speakers in this session update progress in this field, focusing on the interface between biotechnology and nanotechnology. The first two speakers will present approaches to using biotechnology to solve nanotechnology problems. The third and fourth speakers will describe approaches to using nanotechnology to solve biotechnology problems.

# Biological and Biomimetic Polypeptide Materials

TIMOTHY J. DEMING
*University of California, Los Angeles*

Fibrous structures act as load-bearing components in vivo in many natural proteins besides enzymes, which are soluble globular molecules. Natural evolutionary processes have produced structural proteins that surpass the performance of man-made materials (e.g., mammalian elastin in the cardiovascular system that lasts half a century without loss of function and spider webs composed of silk threads that are tougher than any synthetic fiber) (Kaplan et al., 1994; Mobley, 1994; Viney et al., 1992). These biological polypeptides are all complex copolymers that derive their phenomenal properties from precisely controlled sequences and compositions of the constituent amino acid monomers, which, in turn, lead to precisely controlled self-assembled nanostructures. Recently, there has been a great deal of interest in the development of synthetic routes for preparing natural polymers, as well as de novo designed polypeptide sequences for applications in biotechnology (e.g., artificial tissues and implants), biomineralization (e.g., resilient, lightweight, ordered inorganic composites), and analysis (e.g., biosensors and medical diagnostics) (Capello and Ferrari, 1994).

To be successful, these materials must be “processible,” or better yet, capable of self-assembly into precisely defined structures. Peptide polymers have many advantages over conventional synthetic polymers because they are able to assemble hierarchically into stable, ordered conformations (Voet and Voet, 1990). This ability depends in part on the exact structures of protein chains (i.e., defined molecular weight, monodispersity, stereoregularity, and sequence and composition control at the monomer level). The inherent conformational rigidity of polypeptide chains also contributes to their ability to self-assemble. Depending on the amino acid side-chain substituents, polypeptides are able to adopt a multitude of conformationally stable, regular secondary structures (e.g., helices, sheets, and turns), tertiary structures (e.g., the β-strand-helix-β-strand unit found in β-barrels), and quaternary assemblies (e.g., collagen microfibrils) (Voet and Voet, 1990). Conformational rigidity and precise chain structures work together to produce hierarchically ordered, three-dimensional materials from linear macromolecules.

Beyond laboratory replication of natural structural biopolymers, the synthesis of polypeptides that can self-assemble into nonnatural structures is an attractive challenge for polymer chemists. Synthetic-peptide-based polymers are not new materials; homopolymers of polypeptides have been available for many decades, but their use as structural materials has been limited (Fasman, 1967, 1989). New methods in biological and chemical synthesis have made possible the preparation of increasingly complex polymer sequences of controlled molecular weight that display properties far superior to ill-defined homopolymers.

## SYNTHESIS OF BIOLOGICAL POLYPEPTIDES

The advent of recombinant DNA methodologies has provided a basis for the production of polypeptides with exact sequences that can be controlled by design of a suitable DNA template. The most common technique for biosynthesis of protein polymers has been to design an artificial peptide sequence that can be repeated to form a larger polymer and then to construct the DNA polymer that encodes this protein sequence (McGrath et al., 1990; Tirrell et al., 1991). The DNA sequences are then cloned and expressed, typically in yeast or bacterial hosts, to produce the designed polypeptides, which are then either secreted or isolated by lysing the microorganism cells. This methodology has been used for both small- and large-scale production of biomimetic structural proteins, such as silks and elastins, as well as de novo sequences designed to fold into ordered nanostructures (Kaplan et al., 1994; Mobley, 1994; Viney et al., 1992).

The steps required for synthesizing polypeptides using genetic engineering are shown in Figure 1. Once a sequence of amino acids has been chosen, it is encoded into a complementary DNA sequence, which must be chemically synthesized. Solid-phase synthesis has made significant advances in recent years, and it is now possible to synthesize polynucleotides that can encode chains of hundreds of amino acids (McBride and Caruthers, 1983). By cloning these DNA sequences into circular plasmid DNA, the sequences can be incorporated into host cells (e.g., bacteria). In these cells, the plasmids are amplified through replication and can then be sequenced to check for mutations or deletions before final cloning into an expression plasmid. The expression plasmid contains a promoter sequence that drives transcription. A strong promoter, such as the one in the pET

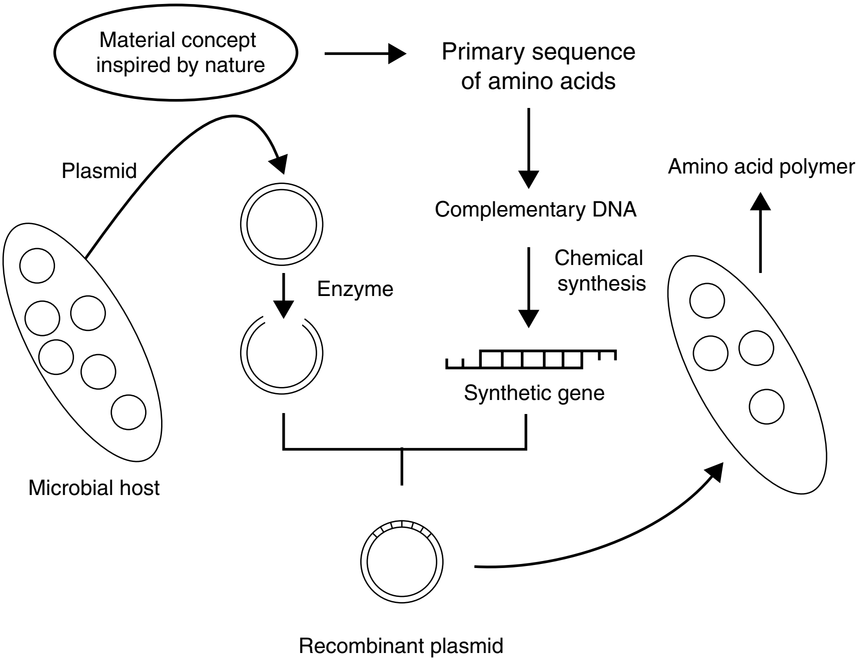


FIGURE 1 Genetic engineering of polypeptides. Source: van Hest and Tirrell, 2001. Reprinted with permission from The Royal Society of Chemistry.

expression vectors developed by Studier et al. (1990), results in high levels of DNA transcription into messenger RNA, which can result in high levels of polypeptide production.

Once the material has been formed, the artificial protein must be isolated from cellular by-products and other proteins. In some cases, the polypeptides form insoluble aggregates in the cells that can be purified by simply extracting all the cellular debris into suitable solvents and recovering the aggregates by filtration. In cases where polypeptides are soluble after cell lysis, they can typically be purified using affinity chromatography by incorporating an affinity marker into the polypeptide sequence (Dong et al., 1994; Zhang et al., 1992).

Recombinant DNA methods have been used to prepare a variety of polypeptide materials. Tirrell's group has prepared lamellar crystals of controlled thickness and surface chemistry (Creel et al., 1991; Krejchi et al., 1994; McGrath et al., 1992), smectic liquid crystals from monodispersed, rod-like α-helical polypeptides (Zhang et al., 1992), and hydrogels from helix-coil-helix triblock copolypeptides (Petka et al., 1998). In related work, van Hest has prepared polypeptide β-sheet fibrils with the recombinant polypeptide chemically coupled to synthetic poly(ethylene glycol) chains to improve solubility (Smeenk et al., 2005). The genetically engineered polypeptide yields well-ordered fibrils of controlled thickness and width.

Chaikof and coworkers have used recombinant DNA methods to prepare protein-based thermoplastic elastomers (Nagapudi et al., 2005). They prepared triblock copolymer sequences that mimic the natural protein elastin; the sequences were modified in such a way that the outer segments were plastic, and the central segment was elastomeric, similar to purely synthetic thermoplastic elastomers, such as styrene-isoprene-styrene rubber. Overall, with genetic engineering we can prepare polypeptides with the complexity of natural proteins. Hence, it is possible to prepare materials with polymer properties that can be manipulated with exquisite control.

## SYNTHESIS OF CHEMICAL POLYPEPTIDES

To circumvent tedious protein purification, large investments of time in gene synthesis, and difficult artificial amino acid incorporation, it would be advantageous if we could synthesize complex copolypeptides chemically from inexpensive amino acid monomers. However, there is a large gap between biologically produced materials and the polypeptides that can be produced synthetically. The most common techniques used for polypeptide synthesis are solid-phase synthesis and polymerization of α-aminoacid-N-carboxyanhydrides (NCAs) (Kricheldorf, 1987, 1990). Solid-phase synthesis, which originated in the pioneering work of Merrifield, involves the stepwise addition of N-protected amino acids to a peptide chain anchored to a polymeric support (Figure 2) (Bodanzsky and Bodanzsky, 1994; Wunsch, 1971). The products of this method are sequence-specific peptides that can be isolated as pure materials. Like solid-phase synthesis of oligonucleotides, the number of peptide residues that can be correctly added to the chains depends on the efficiency of each individual step. However, the capability of strict sequential control in short sequences has made it possible to prepare peptide-based materials.

Solid-phase synthesis has been used to prepare a variety of polypeptide materials. Ghadiri and coworkers used the natural peptide gramacidin-A as a model to develop cyclic peptides that self-assemble into hollow peptide nanotubes (Ghadiri et al., 1993; Khazanovich et al., 1994). The key component of their assembly is the pattern of alternating stereochemistry in the amino acid sequence, which leads to a barrel-like π-helical structure. Numerous other groups have prepared short peptide sequences that assemble into well-defined nanofibers, either through β-sheet or α-helical coiled-coil assembly (Aggeli et al., 1997; Lashuel et al., 2000; Niece et al., 2003; Schneider et al., 2002; Zimenkov et al., 2004).

Recently, Woolfson's group has shown that well-defined kinks can be placed in these fibrils by careful design of the peptide sequence (Ryadnov and Woolfson,

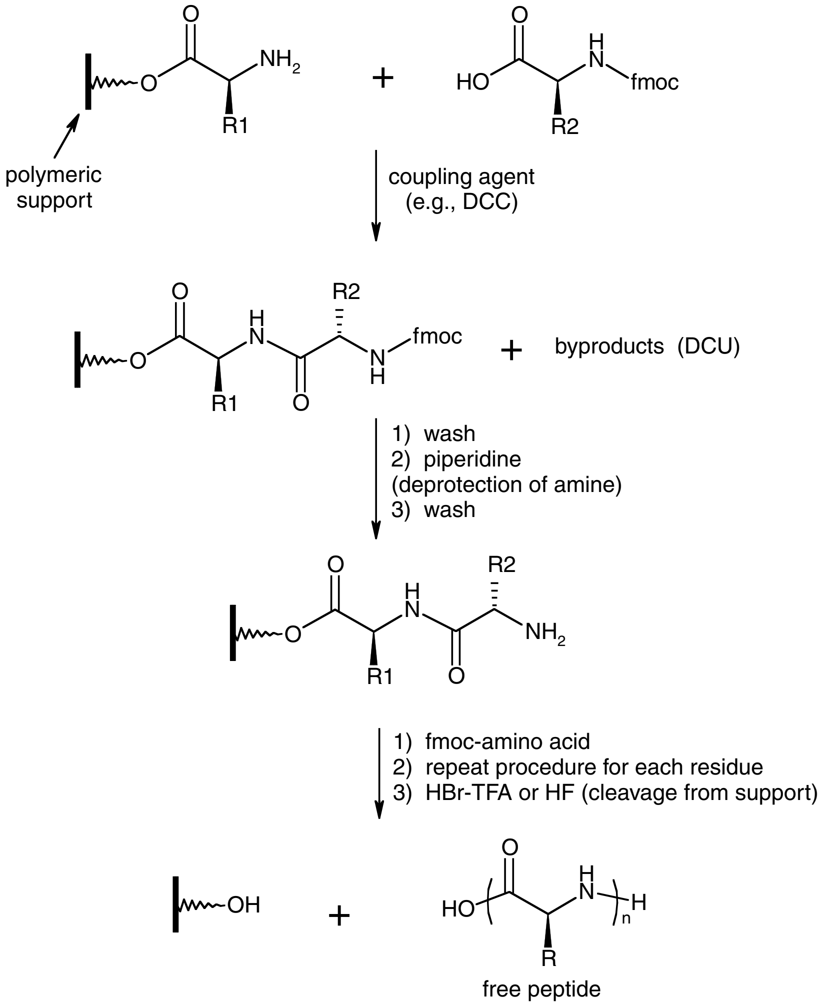
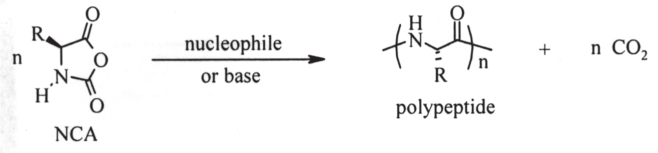


FIGURE 2 Solid-phase peptide synthesis. FMOC = fluorenylmethoxycarbonyl, DCC = dicyclohexylcarbodiimide, DCU = dicyclohexylurea.

2003). Zhang's group has also shown that small, amphiphilic peptides can be designed to form fibrils that create hydrogels that can be used as cell scaffolds (Zhang et al., 2002), as well as membranes that can close off into spherical and tubular vesicles (Vauthey et al., 2002). Solid-phase peptide synthesis has also been used extensively in the preparation of hybrid copolymers, in which a small peptide sequence can have a big effect on overall materials properties (Klok, 2002).

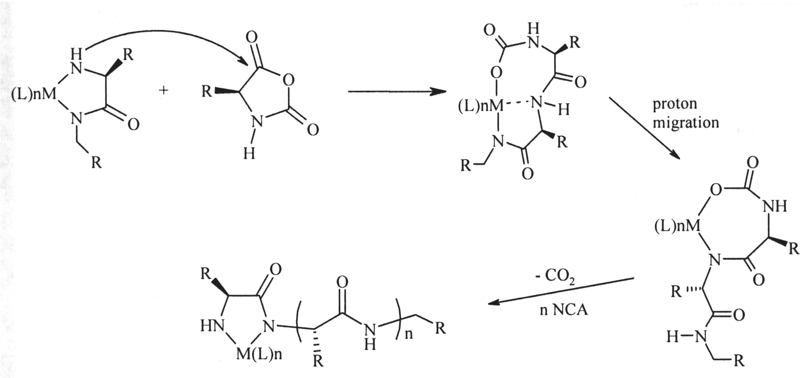


(1)

Equation 1

The most common technique currently used for large-scale preparation of polypeptides is NCAs (eq. 1). However, these materials are almost exclusively homopolymers, random copolymers, or graft copolymers that do not have the sequence specificity and monodispersity of natural proteins (Kricheldorf, 1987, 1990). Therefore, the physical properties of NCAs do not meet the requirements for most applications, even though they have long been available and the monomers are inexpensive and easy to prepare. Recently, a methodology has been developed using transition-metal complexes as active species to control the addition of NCA monomers to polymer chain ends. The use of transition metals to control reactivity has been proven to increase both reaction selectivity and efficiency in organic and polymer synthesis (Boor, 1979; Webster, 1991; Wulff, 1989).

Using this approach, substantial advances in controlled NCA polymerization have been realized in recent years. Highly effective zerovalent nickel and cobalt initiators (i.e., bpyNi(COD) [Deming, 1997] and $(PMe_3)_4Co$ [Deming, 1999]) developed by Deming allow the polymerization of NCAs in a controlled manner without side reactions into high molecular weight polypeptides via an unprecedented activation of the NCAs into covalent propagating species (eq. 2). These cobalt and nickel complexes are able to produce polypeptides with narrow



(2)

Equation 2

chain-length distributions ($M_w/M_n < 1.20$) and controlled molecular weights ($500 < M_n < 500{,}000$) (Deming, 2000). Because this polymerization system is very general, controlled polymerization of a wide range of NCA monomers can be produced as pure enantiomers (D or L configuration) or as racemic mixtures. By adding different NCA monomers, the preparation of block copolypeptides of defined sequence and composition is also feasible (Deming, 2000).

Polypeptide block copolymers prepared via transition-metal-mediated NCA polymerization are well defined, with the sequence and composition of block segments controlled by order and quantity of monomer, respectively, added to initiating species. These block copolypeptides can be prepared with the same level of control as in anionic and controlled radical polymerizations of vinyl monomers, which greatly expands the potential of polypeptide materials. The unique chemistry of these initiators and NCA monomers also allows NCA monomers to be polymerized in any order, which has been a problem in most vinyl copolymerizations. In addition, the robust chain ends allow preparation of copolypeptides with many block domains (e.g., > 4).

The self-assembly of these block copolypeptides has also been investigated (e.g., to direct the biomimetic synthesis of ordered silica structures [Cha et al., 2000], to form polymeric vesicular membranes [Bellomo et al., 2004 (Figure 3); Holowka et al., 2005], and to prepare self-assembled polypeptide hydrogels [Nowak et al., 2002]). Furthermore, poly(L-lysine)-b-poly(L-cysteine) block copolypeptides have been used to generate hollow, organic/inorganic hybrid microspheres composed of a thin inner layer of gold nanoparticles surrounded by a thick layer of silica nanoparticles (Wong et al., 2002). Using the same procedure, hollow spheres were also prepared; these consisted of a thick inner layer of core-shell CdSe/CdS nanoparticles and a thicker outer layer of silica nanoparticles (Cha et al., 2002). The latter spheres are of interest, because they allowed for microcavity lasing without the use of additional mirrors, substrate spheres, or gratings.

## CONCLUSIONS

Many approaches are being investigated for synthesizing new polypeptide materials. Biological approaches have been demonstrated to be extremely powerful for preparing materials with the precision of natural proteins. Chemical techniques are being developed that might be used to prepare polypeptides of any amino acid monomer. The two methodologies complement each other very well. The biological approach is most useful for preparing polypeptides in which monomer sequence and composition must be controlled at the monomer level. The chemical approach is best suited for the preparation of high molecular weight polypeptides in which sequence and composition must only be controlled on length scales of many monomer units (i.e., homopolymer blocks). Future advances in both the biological and chemical arenas are obviously targeted toward

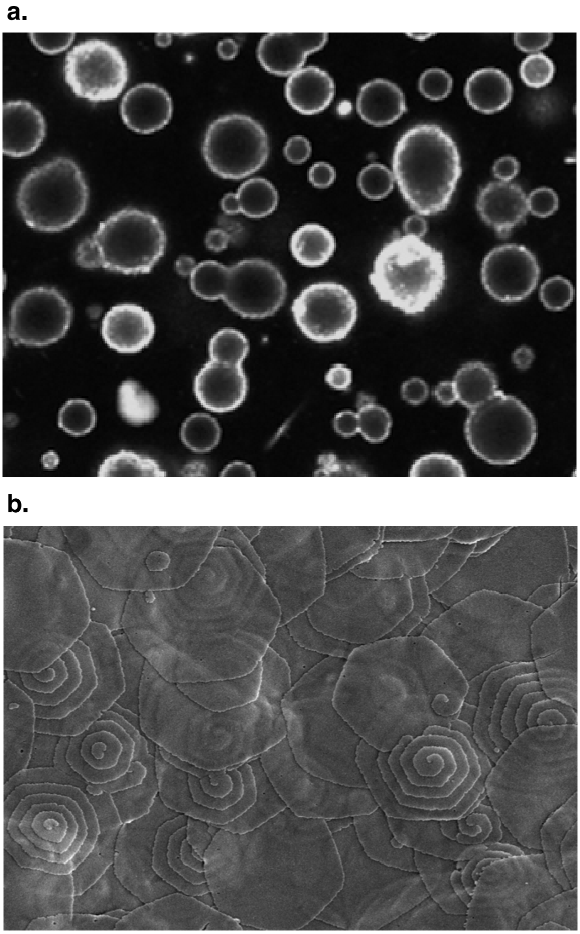


FIGURE 3 (a) Self-assembled vesicles from synthetic diblock copolypeptides. Source: Bellomo et al., 2004. Reprinted with permission from Macmillan Publishers Ltd. (b) Silica-coated plates of α-helical polypeptide single crystals. Source: Bellomo and Deming, 2006. Copyright 2006 American Chemical Society. Reprinted with permission.

conquering the shortcomings of each method—incorporation of artificial amino acids and simplification of preparation/purification for biological synthesis and control of monomer sequence and composition in chemical synthesis. Progress in these areas could expand both methodologies to the point where they might meet and overlap, thus providing scientists with the means of synthesizing any conceivable polypeptide structure.

## REFERENCES

Aggeli, A., M. Bell, N. Boden, J.N. Keen, P.F. Knowles, T.C.B. McLeish, M. Pitkeathly, and S.E. Radford. 1997. Responsive gels formed by the spontaneous self-assembly of peptides into polymeric β-sheet tapes. Nature 386: 259–262.

Bellomo, E.G., and T.J. Deming. 2006. Monoliths of aligned silica-polypeptide hexagonal platelets. Journal of the American Chemical Society 128: 2276–2279.

Bellomo, E., M.D. Wyrsta, L. Pakstis, D.J. Pochan, and T.J. Deming. 2004. Stimuli-responsive polypeptide vesicles by conformation-specific assembly. Nature Materials 3: 244–248.

Bodanzsky, M., and A. Bodanzsky. 1994. Practice of Peptide Synthesis, 2nd ed. New York: Springer-Verlag.

Boor, J. 1979. Zeigler-Natta Catalysts and Polymerizations. New York: Academic Press.

Capello, J., and F. Ferrari. 1994. Microbial Production of Structural Protein Polymers. Pp. 35–92 in Plastics from Microbes. Cincinnati, Ohio: Hanser/Gardner.

Cha, J.N., G.D. Stucky, D.E. Morse, and T.J. Deming. 2000. Biomimetic synthesis of ordered silica structures mediated by block copolypeptides. Nature 403: 289–292.

Cha, J.N., M.H. Bartl, M.S. Wong, A. Popitsch, T.J. Deming, and G.D. Stucky. 2002. Microcavity lasing from block peptide hierarchically assembled quantum dot spherical resonators. Nano Letters 3: 907–911.

Creel, H.S., M.J. Fournier, T.L. Mason, and D.A. Tirrell. 1991. Genetically directed syntheses of new polymeris materials. Efficient expression of a monodisperse copolypeptide containing fourteen tandemmly repeated—$(AlaGly)_4$ProGluGly—elements. Macromolecules 24: 1213–1214.

Deming, T.J. 1997. Facile synthesis of block copolypeptides of defined architecture. Nature 390: 386–389.

Deming, T.J. 1999. Cobalt and iron initiators for the controlled polymerization of α-amino acid-N-carboxyanhydrides. Macromolecules 32: 4500–4502.

Deming, T.J. 2000. Living polymerization of a-amino acid-N-carboxyanhydrides. Journal of Polymer Science, Part A: Polymer Chemistry 38: 3011–3018.

Dong, W., M.J. Fournier, T.L. Mason, and D.A. Tirrell. 1994. Bifunctional hybrid artificial proteins. Polymer Preprints 35(2): 419–420.

Fasman, G.D. 1967. Poly α-Amino Acids. New York: Dekker.

Fasman, G.D. 1989. Prediction of Protein Structure and the Principles of Protein Conformation. New York: Plenum Press.

Ghadiri, M.R., J.R. Granja, R.A. Milligan, D.E. McRee, and N. Khazanovich. 1993. Self-assembling organic nanotubes based on a cyclic peptide architecture. Nature 366: 324–327.

Holowka, E.P., D.J. Pochan, and T.J. Deming. 2005. Charged polypeptide vesicles with controllable diameter. Journal of the American Chemical Society 127: 12423–12428.

Kaplan, D., W.W. Adams, B. Farmer, and C. Viney. 1994. Silk Polymers. American Chemical Society Symposium Series 544. Washington, D.C.: American Chemical Society.

Khazanovich, N., J.R. Granja, D.E. McRee, R.A. Milligan, and M.R. Ghadiri. 1994. Nanoscale tubular ensembles with specified internal diameters. Design of a self-assembled nanotube with a 13-Å Pore. Journal of the American Chemical Society 116: 6011–6012.

Klok, H.-A. 2002. Protein-inspired materials: synthetic concepts and potential applications. Angewandte Chemie International Edition 41: 1509–1513.

Krejchi, M.T., E.D.T. Atkins, A.J. Waddon, M.J. Fournier, T.L. Mason, and D.A. Tirrell. 1994. Chemical sequence control of beta-sheet assembly in macromolecular crystals of periodic polypeptides. Science 265: 1427–1432.

Kricheldorf, H.R. 1987. α-Aminoacid-N-Carboxyanhydrides and Related Materials. New York: Springer-Verlag.

Kricheldorf, H.R. 1990. Polypeptides. Pp. 1–132 in Models of Biopolymers by Ring-Opening Polymerization, edited by S. Penczek. Boca Raton, Fla.: CRC Press.

Lashuel, H.A., S.R. LaBrenz, L. Woo, L.C. Serpell, and J.W. Kelly. 2000. Protofilaments, filaments, ribbons, and fibrils from peptidomimetic self-assembly: implications for amyloid fibril formation and materials science. Journal of the American Chemical Society 122: 5262–5277.

McBride, L.J., and M.H. Caruthers. 1983. An investigation of several deoxynucleoside phosphoramidites useful for synthesizing deoxyoligonucleotides. Tetrahedron Letters 24: 245–248.

McGrath, K.P., D.A. Tirrell, M. Kawai, M.J. Fournier, and T.L. Mason. 1990. Chemical and biosynthetic approaches to the production or novel polypeptide materials. Biotechnology Progress 6: 188–192.

McGrath, K.P., M.J. Fournier, T.L. Mason, and D.A. Tirrell. 1992. Genetically directed syntheses of new polymeric materials. Expression of artificial genes encoding proteins with repeating—(AlaGly)3ProGluGly—elements. Journal of the American Chemical Society 114: 727–733.

Mobley, D. P. 1994. Plastics from Microbes. Cincinnati, Ohio: Hanser/Gardner.

Nagapudi, K., W.T. Brinkman, J. Leisen, B.S. Thomas, E.R. Wright, C. Haller, X. Wu, R.P. Apkarian, V.P. Conticello, and E.L. Chaikof. 2005. Protein-based thermoplastic elastomers. Macromolecules 38: 345–354.

Niece, K.L., J.D. Hartgerink, J.J.J.M. Donners, and S.I. Stupp. 2003. Self-assembly combining two bioactive peptide-amphiphile molecules into nanofibers by electrostatic attraction. Journal of the American Chemical Society 125: 7146–7147.

Nowak, A.P., V. Breedveld, L. Pakstis, B. Ozbas, D.J. Pine, D. Pochan, and T.J. Deming. 2002. Peptides are a gel's best friend. Nature 417: 424–428.

Petka, W.A., J.L. Harden, K.P. McGrath, D. Wirtz, and D.A. Tirrell. 1998. Reverse hydrogels from self-assembling artificial proteins. Science 281: 389–392.

Ryadnov, M.G., and D.N. Woolfson. 2003. Engineering the morphology of a self-assembling protetin fibre. Nature Materials 2: 329–332.

Schneider, J.P., D.J. Pochan, B. Ozbas, K. Rajagopal, L.M. Pakstis, and J. Gill. 2002. Responsive hydrogels from the intramolecular folding and self-assembly of a designed peptide. Journal of the American Chemical Society 124: 15030–15037.

Smeenk, J.M., M.B.J. Otten, J. Thies, D.A. Tirrell, H.G. Stunnenberg, and J.C.M. van Hest. 2005. Controlled assembly of macromolecular b-sheet fibrils. Angewandte Chemie International Edition 44: 1968–1971.

Studier, W.F., A.H. Rosenberg, J.J. Dunn, and J.W. Dubendorf. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods in Enzymology 185: 60–89.

Tirrell, D.A., M.J. Fournier, and T.L. Mason. 1991. Genetic engineering of polymeric materials. Materials Research Society Bulletin 16: 23–28.

van Hest, J.C.M., and D. Tirrell. 2001. Protein-based materials, toward a new level of structural control. Chemical Communications 1897–1904.

Vauthey, S., S. Santoso, H.Y. Gong, N. Watson, and S.G. Zhang. 2002. Molecular self-assembly of surfactant-like peptides to form nanotubes and nanovesicles. Proceedings of the National Academy of Sciences 99: 5355–5360.

Viney, C., S.T. Case, and J.H. Waite. 1992. Biomolecular Materials. Materials Research Society Proceedings 292. Pittsburgh, Penn.: Materials Research Society.

Voet, D., and J.G. Voet. 1990. Biochemistry. New York: John Wiley and Sons.

Webster, O. 1991. Living polymers. Science 251: 887–893.

Wong, M.S., J.N. Cha, K.-S. Choi, T.J. Deming, and G.D. Stucky. 2002. Assembly of nanoparticles into hollow spheres using block copolypeptides. Nano Letters 2: 583–587.

Wulff, G. 1989. Main-chain chirality and optical activity in polymers consisting of C-C chains. Angewandte Chemie International Edition 28: 21–37.

Wunsch, E. 1971. Synthesis of naturally occurring polypeptides, problems of current research. Angewandte Chemie International Edition 10: 786–795.

Zhang, G., M.J. Fournier, T.L. Mason, and D.A. Tirrell. 1992. Biological synthesis of monodisperse derivatives of poly(alpha-L-glutamic acid): model rodlike polymers. Macromolecules 25: 3601–3603.

Zhang, S., D.N. Marini, W. Hwang, and S. Santoso. 2002. Design of nanostructured biological materials through self-assembly of peptides and proteins. Current Opinion in Chemical Biology 6: 865–871.

Zimenkov, Y., V.P. Conticello, L. Guo, and P. Thiyagarajan. 2004. Rational design of a nanoscale helical scaffold derived from self-assembly of a dimeric coiled coil motif. Tetrahedron 60: 7237–7246.

# Applications of Biomimetics

Morley O. Stone
*Air Force Research Laboratory*
*Wright-Patterson Air Force Base, Ohio*

At first glance, imitating nature via biomimetics seems to be a straightforward proposition. For example, if you are a roboticist, add legs to the platform instead of wheels. Unfortunately, as is often the case, the devil is in the details. After a short synopsis of examples of biomimetic material synthesis, sensing, and robotics, I will attempt to identify some lessons learned, some surprising and unanticipated insights, and some potential pitfalls in biomimetics. (For recent perspectives on combining biology with other disciplines, see Naik and Stone, 2005.)

## "JUST DON'T ADD WATER"

Often, biologists and engineers speak completely different languages. Perhaps the most graphic example of this is a comparison of the world of electrical engineers and sensor designers with the world of biologists. Manipulation of biological macromolecules (i.e., nucleic acids and/or proteins) involves the use of buffered solutions (usually pH ~ 7), controlled salinity, and regulated temperatures. Incorporating these biological salt solutions into electronics and sensor architectures seems like an oxymoron. However, the conversion of biological materials away from solution to solid-state processing has been a major objective in our laboratory.

The key to overcoming this seemingly insurmountable incompatibility is the use of "bridging" materials systems, such as polymer host materials that capture and maintain biological functionality (Brott et al., 2004). Many polymer systems, such as poly(vinyl alcohol), qualify as hydrogels because they incorporate and maintain an enormous amount of water. While the biological side of this equation can be satisfied via the incorporation of water, polymer systems can be spin-coated, lithographically patterned, made conductive, and undergo a host of other treatments that electrical engineers routinely use. Thus, polymers represent a truly bridging material system in making biological macromolecules mesh with synthetic technology.

Another recent example highlights the potential of biological materials that have been integrated into a common electrical construct, such as a light-emitting diode (LED). To accomplish this, however, there must be a paradigm shift in materials thinking—namely, what would happen if DNA were processed in gram and kilogram quantities, instead of the traditional microgram quantities.

The fishing industry in Japan, which processes tons of seafood yearly, also throws away tons of DNA from fish gametes. Researchers at the Chitose Institute of Science and Technology in Japan, in partnership with our laboratory, have processed this discarded DNA into a surfactant complex and scaled the process up to a multigram scale (Wang et al., 2001). In this form and at this scale, DNA can be spin-coated into traditional electronics architectures. Recently, a DNA electron-blocking layer spin-deposited on the hole-injection side of the electron-hole recombination layer greatly enhanced LED efficiency and performance (Hagen et al., 2006) (Figure 1).

In another approach, we have attempted to use biology indirectly in advanced material synthesis and devices. Similar to the refrain from a commercial for a popular chemical company, biology isn't in the final material, but it makes the final material better. Researchers around the world racing to harness the incredible electronic, thermal, and mechanical properties inherent in single-

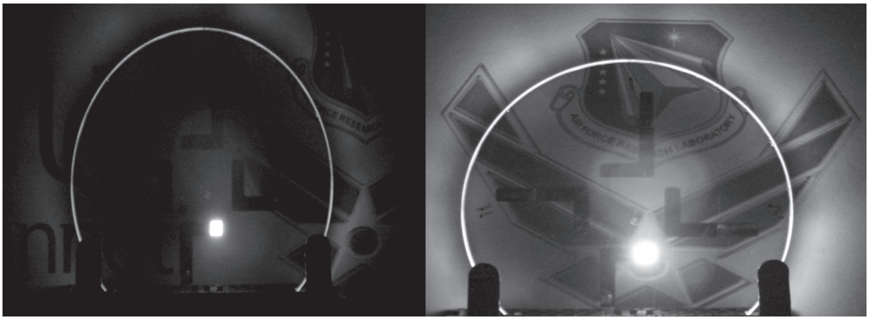


FIGURE 1 Photographs of Alq3 green emitting bio-LED and baseline OLED devices in operation.

walled carbon nanotubes (SWNTs) have run into a formidable obstacle. After a typical synthesis run, there is a large variety (both in length and diameter) of carbon nanotubes. This variety contributes to a number of chiralities, which dictate the metallic or semiconducting nature of the SWNT. Much of this size heterogeneity arises from the heterogeneity of the starting metallic nanoparticles used to catalyze the growth of SWNTs.

Ferritins and ferritin-like proteins sequester iron (in the form of iron oxide) in precisely defined cavities ranging from 8 nm to 4 nm for human and bacterial forms, respectively. We recently engineered a bacterial form called DPS to produce uniform, monodisperse iron oxide particles (Kramer et al., 2005). We reasoned that a monodisperse starting pool of nanoparticles would lead to a more monodisperse population of SWNTs. After the bacterial protein was used to produce the iron oxide particles, the biological shell was removed via sintering in a reduced atmosphere and subjected to gas-phase carbon-nanotube growth. The resulting SWNTs adopted the monodisperse character of the starting catalyst particles. Thus, biology was not in the final product but was used to make a technologically promising material better.

## MATERIALS SCIENCE AND ENGINEERING OVERLAP BIOLOGY

From a materials science and engineering perspective, favorable electronic and structural properties usually emerge when the synthesis process can be controlled at finer and finer levels. Hence the frenzy and hype over nanotechnology. As illustrated in the carbon nanotube example, biology can provide tools for controlling and/or synthesizing materials at the molecular level.

An example of this control is unicellular algae, called diatoms, which make exquisite cellular structures out of silica. Thousands of species of diatoms exist in salt and fresh water. Each diatom species makes unique silica structures and patterns—from hinges to intricate arrays of holes and spines. Silica synthesis occurs at ambient temperature and pH and has a complexity greater than anything we can make synthetically using sol-gel techniques.

Kröger and colleagues (1999) provided insight into the silica-deposition process of diatoms, which has led to a complete rethinking of the molecular evolution of this process (Naik et al., 2002) and how it can be used in practical applications, such as enzymatic encapsulation (Luckarift et al., 2004). Based on work by Kröger and others, the field of biomineralization has expanded the range of materials synthesized via a biological route to encompass not only oxides, but also metals and semiconductors (Slocik et al., 2005).

Specific peptides can now be used for the nucleation and growth of inorganic nanomaterials. When one considers that peptides specific for inorganic binding and nucleation can be combined (i.e., genetically fused) with peptides that bind another moiety, endless possibilities begin to emerge. For example, biological macromolecules might be incorporated directly into electronic structures/devices. One can imagine literally growing a field-effect transistor (FET) metal-oxide-metal architecture via a biological route, rather than relying on standard top-down photolithographic processes.

There might also be new approaches in optics and catalysis where there are now significant challenges in assembling and interconnecting the building blocks of a nanoscale device. One might be able, for example, to address or measure responses electronically at the molecular level. The very real scale gap between the size of the molecule and the limits of lithography is shrinking. Bio-based approaches are being pursued to develop bottom-up self-assembly techniques that provide specificity and versatility. Peptides that recognize inorganic surfaces can be used as templates to organize and/or synthesize inorganic materials. By combining different functional peptides, we can create multifunctional polypeptides that can be used to synthesize and assemble hybrid materials. Recently, we demonstrated that by growing bimetallic systems using a bio-based approach, we can enhance catalytic activity of bimetallic materials (Slocik and Naik, 2006).

In nature, the programmed assembly of amino acids in a polypeptide sequence gives rise to protein molecules that exhibit multifunctional properties. Similarly, using protein engineering, inorganic binding peptide segments can be fused to create multifunctional polypeptides to assemble and/or synthesize hybrid materials. By exploiting the interaction between peptides and inorganic materials, a peptide that contains two separate domains (each responsible for binding to a specific metal species) can be used to assemble hybrid materials (Figure 2). Thus, we can control/program synthesis of bimetallic structures. The bimetallic nanoparticles made by using the designer peptides were found to be efficient catalysts in the hydrogenation of 3-buten-1-ol.

This method represents a generalized approach to achieving hybrid structures by programming the amino acid sequence presented by the peptide interface. The peptide interface may be used to conjugate nanoparticle surfaces to polymers, organic molecules, or other biomolecules. However, to fully appreciate the potential of peptides and other biological building blocks as molecular templates, we will need a better understanding of the interaction between

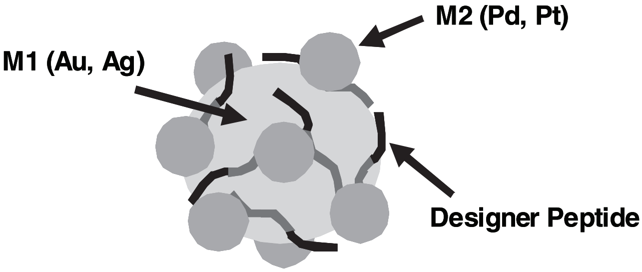


FIGURE 2 Assembly of hybrid materials using designer peptides.

biomolecules and inorganic surfaces. Using computational modeling, we should get a much clearer understanding of the mechanism of peptide-inorganic interactions. In the future, one should be able to create hybrid materials using protein engineering by dialing in the sequence domains to direct their synthesis and assembly.

## BIO-INSPIRED ROBOTICS: APPLIED BIOLOGY AND ENGINEERING

The combination of biological principles, mechanical engineering, and robotics has opened entirely new areas and possibilities. Starting with the question of why legs matter, the field of biodynotics is exploding to encompass why materials properties matter, why mechanics and architecture matter, and how biological insights can lead to completely new capabilities. For example, entirely new lessons and robotic capabilities have emerged, such as dynamic compliance, molecular adhesion, conformal grasping, and dynamic stability, to name just a few of the concepts that have been implemented into robotic platforms.

The first contributions of biology to robotics were based on the insight that a sprawled posture used opposing forces to achieve self-stabilization (Dickenson et al., 2000; Full and Tu, 1991). Much of this early work was focused on understanding the mechanics involved in legged locomotion. The spring-loaded inverted-pendulum model has been accepted as an accurate model of biological locomotion independent of the number of legs or the biological platform (i.e., horse or human or cockroach).

Recently, the Cutkosky laboratory at Stanford, where pneumatically driven hexapod running robots were developed, has been challenged to build a wall-climbing platform capable of emulating gecko-like behavior (Clark et al., 2001). From a materials science perspective, the challenge has been to produce synthetic, self-cleaning hair arrays with a diameter of 200 nm at a density of 1–2e9 hairs/cm$^2$.

The field of robotics would also benefit immensely from the development of tunable (dynamic) modulus materials. Today, compliance is usually tuned mechanically, which entails high costs in weight and power and produces less than satisfactory performance. There are numerous biological models of tunable modulus materials (e.g., the sea cucumber), and extrapolating these lessons to robotics could have a huge impact.

## CONCLUSION

In our research, we are framing future investments in an area we call biotronics, a term that encompasses both bioelectronics and biophotonics. As the LED and FET examples described above suggest, this area is ripe for revolutionary breakthroughs. New capabilities, like tunable dielectrics, could revolutionize sensing and electronic readout. We believe that an integrated package of sensing and readout will emerge.

Biology may also enable us to fabricate materials, structures, and devices from the bottom up. Many believe that we will have to turn to biology for commercially viable nanomanufacturing. Catalysis and self-assembly have been mastered by biological systems like enzymes and viruses, respectively. These lessons are being applied to traditional solid-state electronics, and engineers are beginning to realize the possibilities.

To continue advancements in biomimetics, we must include these principles in undergraduate and graduate training programs. Many other countries are also awakening to this realization. Thus, the future technical base of our country will depend on how well the science and engineering departments in our universities encourage this interdisciplinary training.

## REFERENCES

Brott, L.L, S.M. Rozenzhak, R.R. Naik, S.R. Davidson, R.E. Perrin, and M.O. Stone. 2004. A poly(vinyl alcohol)/carbon-black composite film: a platform for biological macromolecule incorporation. Advanced Materials 16: 592–596.

Clark, J.E, J.G. Cham, S.A. Bailey, E.M. Froehlich, P.K. Nahata, R.J. Full, and M.R. Cutkosky. 2001. Biomimetic Design and Fabrication of a Hexapedal Running Robot. Pp. 3643–3649 in Proceedings of the IEEE International Conference on Robotics and Automation, Vol. 4. Piscataway, N.J.: IEEE.

Dickenson, M.H., C.T. Farley, R.J. Full, M.A.R. Koehl, R. Kram, and S. Lehman. 2000. How animals move: an integrative view. Science 288: 100–106.

Full, R.J., and M.S. Tu. 1991. Mechanics of a rapid running insect: two-, four-, and six-legged locomotion. Journal of Experimental Biology 156: 215–231.

Hagen, J.A., W. Li, A.J. Steckl, and J.G. Grote. 2006. Enhanced emission efficiency in organic light-emitting diodes using deoxyribonucleic acid complex as an electron blocking layer. Applied Physics Letters 88: 171109–171111.

Kramer, R.M., L.A. Sowards, M.J. Pender, M.O. Stone, and R.R. Naik. 2005. Constrained iron catalysts for single-walled carbon nanotube growth. Langmuir 21: 8466–8470.

Kröger, N., R. Deutzmann, and M. Sumper. 1999. Polycationic peptides from diatom biosilica that direct silica nanosphere formation. Science 286: 1129–1132.

Luckarift, H.R., J.C. Spain, R.R. Naik, and M.O. Stone. 2004. Enzyme immobilization in a biomimetic silica support. Nature Biotechnology 22: 211–213.

Naik, R.R., and M.O. Stone. 2005. Integrating biomimetics. Materials Today 8: 18–26.

Naik, R.R., L.L. Brott, S.J. Clarson, and M.O. Stone. 2002. Silica-precipitating peptides isolated from a combinatorial phage display peptide library. Journal of Nanoscience and Nanotechnology 2: 1–6.

Slocik, J.M., and R.R. Naik. 2006. Biologically programmed synthesis of bimetallic nanostructures. Advanced Materials 18: 1988–1992.

Slocik, J.M., M.O. Stone, and R.R. Naik. 2005. Synthesis of gold nanoparticles using multifunctional peptides. Small 1(11): 1048–1052.

Wang, L., J. Yoshida, and N. Ogata. 2001. Self-assembled supramolecular films derived from marine deoxyribo-nucleic acid (DNA)-cationic surfactant complexes: large-scale preparation and optical and thermal properties. Chemistry of Materials 13: 1273–1281.

# Optical Imaging for In Vivo Assessment of Tissue Pathology

Rebekah A. Drezek, Naomi J. Halas, and Jennifer West
*Rice University*
*Houston, Texas*

For hundreds of years, optical imaging at both macroscopic and microscopic levels has been used as a tool to aid clinicians in establishing a diagnosis. Pathologists routinely use a simple compound microscope to examine stained and sectioned tissue at the microscopic level to determine a definitive diagnosis of cancer. At a macroscopic level, clinicians often rely on observed colors as indicators of physiologic status, for instance, yellow skin is associated with jaundice, blue or purple hues with cyanosis, and red with inflammation. In each of these examples, the human eye gathers *qualitative* information about a patient's status based on either the gross visual appearance of tissue or a microscopic evaluation of stained tissue sections or cytologic samples.

Despite the clear importance of these qualitative optical approaches in current medical practice, these strategies are only sensitive to a highly limited subset of the wide array of optical events that occur when light interacts with biologic tissue. In fact, there is a compelling need for more *quantitative* optical imaging strategies that can probe tissue physiology in vivo in real time with high resolution at relatively low cost. In this talk, I describe emerging technologies for quantitative optical imaging and the use of these technologies to diagnose and monitor cancer. Particular emphasis is placed on how advances in nanobiotechnology are leading to new approaches to in vivo medical diagnostics. Because another talk in this session will consider luminescence-based nanomaterials (i.e., quantum dots), the discussion here is focused on nanomaterials, particularly gold-based materials, that provide a scatter-based or absorption-based optical signal. The general biocompatibility of gold, coupled with extensive prior medical applications of gold colloid, suggests a more straightforward regulatory path toward ultimate clinical use than for many other nanomaterials currently under development.

## THE ROLE OF NANOTECHNOLOGY IN OPTICAL IMAGING OF CANCER

For more than 50 years, cancer was the second leading cause of death in the United States, accounting for more than 25 percent of deaths in the population. However, in the past two years, death from cancer has exceeded deaths from heart attacks, and cancer has become the primary cause of deaths in the United States. Early detection is recognized as a highly effective approach to reducing the morbidity and mortality associated with cancer. When diagnosed at an early stage when the cancer is still localized and risk for metastasis is low, most cancers are highly treatable and prognoses are favorable. However, if cancer is not diagnosed until metastasis to distant sites has already occurred, five-year survival is poor for a wide variety of organ sites (Table 1) (American Cancer Society, 2006). Thus, there is a significant clinical need for novel methods of early detection and treatment with improved sensitivity, specificity, and cost effectiveness.

In recent years, a number of groups have demonstrated that photonics-based technologies can be valuable in addressing this need. Optical technologies promise high-resolution, noninvasive functional imaging of tissue at competitive costs. However, in many cases, these technologies are limited by the inherently weak optical signals of endogenous chromophores and the subtle spectral differences between normal and diseased tissue.

In the past several years, there has been increasing interest in combining emerging optical technologies with novel exogenous contrast agents designed to probe the molecular-specific signatures of cancer to improve the detection limits and clinical effectiveness of optical imaging. For instance, Sokolov et al. (2003) recently demonstrated the use of gold colloid conjugated to antibodies to the epidermal growth factor receptor as a scattering contrast agent for biomolecular

TABLE 1 Cancer Survival at Five Years as a Function of Stage at Diagnosis

| Organ | Localized | Regional | Distant |
|---|---|---|---|
| Prostate | ~100% | >85% | 30% |
| Oral | >80% | 50% | 25% |
| Breast | >90% | 80% | 25% |

optical imaging of cervical cancer cells and tissue specimens. In addition, optical imaging applications of nanocrystal bioconjugates have been described by multiple groups, including Bruchez et al. (1998), Chan and Nie (1998), and Akerman et al. (2002). More recently, interest has developed in the creation of nanotechnology-based platform technologies that can couple molecular-specific early detection strategies with appropriate therapeutic intervention and monitoring capabilities.

## METAL NANOSHELLS

Metal nanoshells are a new type of nanoparticle composed of a dielectric core, such as silica, coated with an ultrathin metallic layer, typically gold. Gold nanoshells have physical properties similar to gold colloid, particularly a strong optical absorption due to gold's collective electronic response to light. The optical absorption of gold colloid yields a brilliant red color, which has been used effectively in consumer-related medical products, such as home pregnancy tests. In contrast, the optical response of gold nanoshells depends dramatically on the relative size of the nanoparticle core and the thickness of the gold shell.

By varying the relative thicknesses of the core and shell, the color of gold nanoshells can be varied across a broad range of the optical spectrum that spans the visible and near-infrared spectral regions (Brongersma, 2003; Oldenburg et al., 1998). Gold nanoshells can be made either to absorb or scatter light preferentially by varying the size of the particle relative to the wavelength of the light at their optical resonances. Figure 1 shows a Mie scattering plot of the nanoshell plasmon resonance wavelength shift as a function of nanoshell composition for a 60 nm core gold/silica nanoshell. In this figure, the core and shell of the nanoparticles are shown to relative scale directly beneath their corresponding optical resonances. Figure 2 shows a plot of the core/shell ratio versus resonance wavelength for a silica core/gold shell nanoparticle (Oldenburg et al., 1998). The extremely agile "tunability" of optical resonance is a property unique to nanoshells—in no other molecular or nanoparticle structure can the resonance of the optical absorption properties be "designed" as systematically.

Halas and colleagues have completed a comprehensive investigation of the optical properties of metal nanoshells (Averitt et al., 1997) and achieved quantitative agreement between Mie scattering theory and the experimentally observed optical-resonant properties. Based on this success, it is now possible to design gold nanoshells predictively with the desired optical-resonant properties and then to fabricate the nanoshell with the dimensions and nanoscale tolerances necessary to achieve these properties (Oldenburg et al., 1998). The synthetic protocol developed for the fabrication of gold nanoshells is very simple in concept:

1. Grow or obtain silica nanoparticles dispersed in solution.
2. Attach very small (1–2 nm) metal "seed" colloids to the surface of the

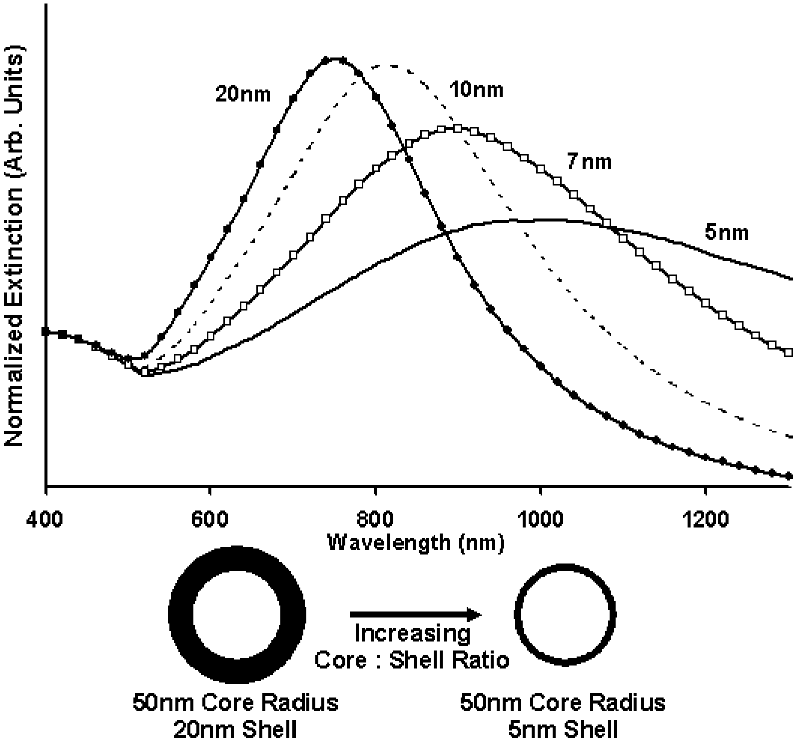


FIGURE 1 Optical resonances of gold-shell, silica-core nanoshells as a function of the core/shell ratio. Respective spectra correspond to the nanoparticles shown below them. Source: Loo et al., 2004. Reprinted with permission.

nanoparticles via molecular linkages; the seed colloids cover the dielectric nanoparticle surfaces with a discontinuous metal colloid layer.

3. Grow additional metal onto the “seed” metal colloid adsorbates via chemical reduction in solution.

This approach has been used successfully to grow both gold and silver metallic shells onto silica nanoparticles. Various stages in the growth of a gold metallic shell onto a functionalized silica nanoparticle are shown in Figure 3. Based on the core/shell ratios that can be achieved with this protocol, gold nanoshells with optical resonances extending from the visible region to approximately 3 μm in the infrared region can currently be fabricated. This spectral region includes the 800–1,300 nm “water window” of the near infrared, a region of high physiological transmissivity that has been demonstrated as the spectral region best suited for optical bioimaging and biosensing. The optical

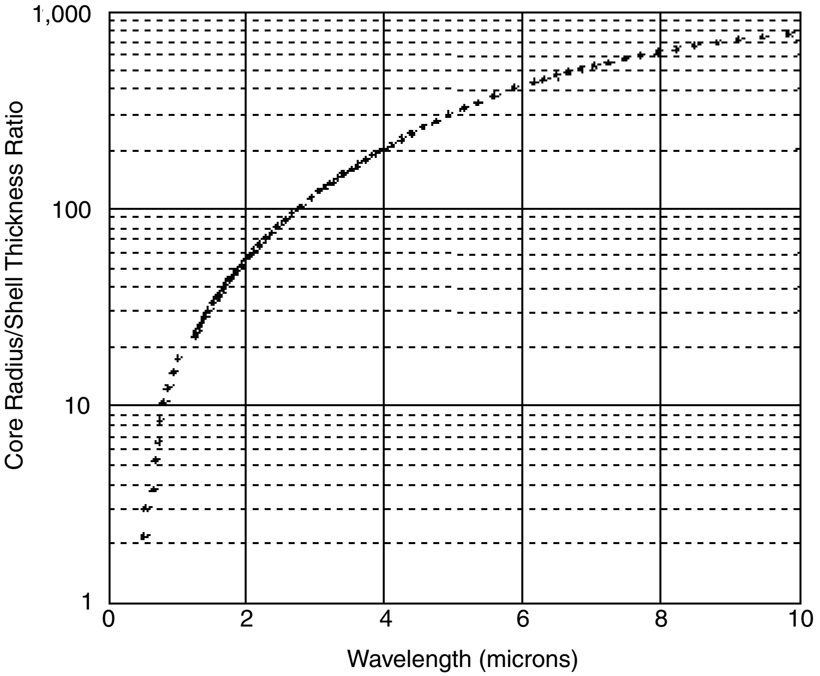


FIGURE 2 Core/shell ratio as a function of resonance wavelength for gold/silica nanoshells. Source: Loo et al., 2004. Reprinted with permission.

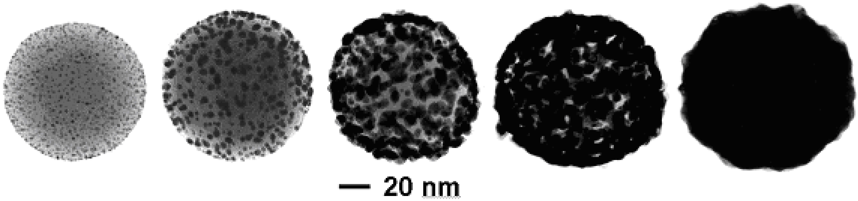


FIGURE 3 Transmission electron microscope images of gold/silica nanoshells during shell growth. Source: Loo et al., 2004. Reprinted with permission.

properties of gold nanoshells, coupled with their biocompatibility and ease of bioconjugation, render them highly suitable for targeted bioimaging and therapeutic applications. By controlling the physical parameters of the nanoshells, it is possible to engineer nanoshells that primarily scatter light, which is desirable for many imaging applications, or alternatively, to design nanoshells that are strong absorbers, which is desirable for photothermal-based therapy applications.

Because the same chemical reaction is used to grow the metal layer of gold nanoshells as is used to synthesize gold colloid, the surfaces of gold nanoshells are virtually chemically identical to the surfaces of the gold nanoparticles universally used in bioconjugate applications. Gold colloid was first used in biological applications in 1971 when Faulk and Taylor (1971) invented the immunogold staining procedure. Since then, the labeling of targeting molecules, especially proteins, with gold nanoparticles has revolutionized the visualization of cellular or tissue components by electron microscopy. The optical and electron-beam contrast qualities of gold colloid have provided excellent detection qualities for immunoblotting, flow cytometry, hybridization assays, and other techniques. Conjugation protocols exist for the labeling of a broad range of biomolecules with gold colloid, such as protein A, avidin, streptavidin, glucose oxidase, horseradish peroxidase, and IgG.

The vast prior history of gold-colloid-based materials has greatly facilitated the development of biomedical applications of newer gold-based nanoparticles. Figure 4 shows one example of the type of medical application enabled by using this class of material. The figure shows an in vitro proof-of-principle example of

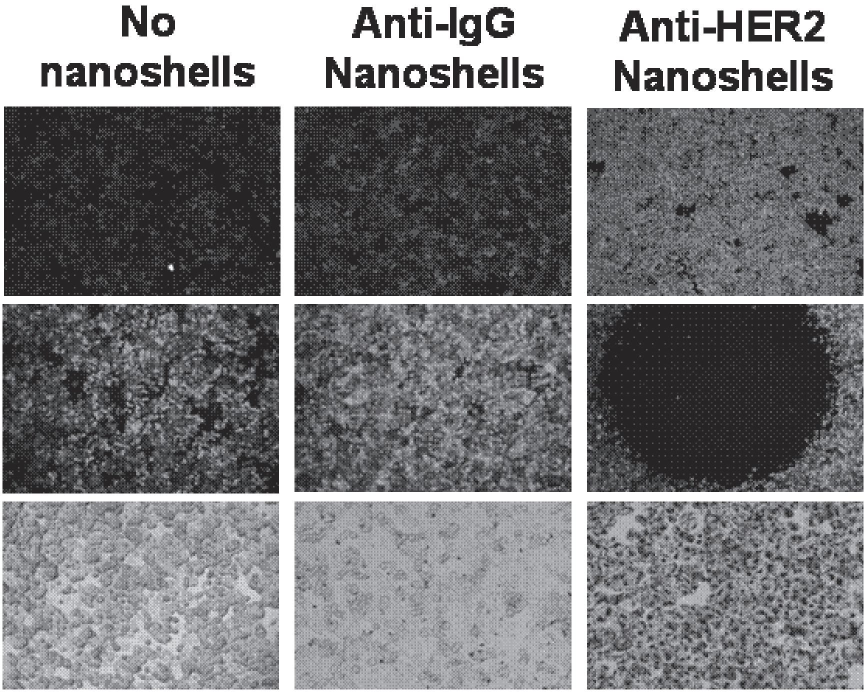


FIGURE 4 Dual imaging/therapy nanoshell bioconjugates. Source: Loo et al., 2005. Copyright 2005 American Chemical Society. Reprinted with permission.

gold nanoshells designed to simultaneously scatter (for imaging) and absorb (for photothermal therapy) near infrared light. Here, scattering and absorbing near infrared nanoshells are conjugated to an antibody for a common breast cancer surface marker. This enables both "lighting up" and, if desired, destroying cells that express this marker without harming other cells. In Figure 4, the top row shows scatter-based imaging of carcinoma cells. By increasing the laser power, it is possible to destroy cells selectively as shown in the middle row, which shows the viability of cells after laser irradiation. The left column shows a no-nanoshell (cells-only) control, and the middle column shows a nonspecific antibody control. The right column indicates successful imaging of cells followed by photothermal destruction (black circle = laser irradiation spot) based on the presence of a chosen marker.

Although in vitro demonstrations can be completed using simple microscopes, in vivo use of this type of nanomaterial requires coupling the development of appropriate materials with the development of optical devices that enable imaging of these materials in tissue. By careful design of these optical systems, it is possible to generate multiple order of magnitude improvements in optical contrast using nanomaterial imaging agents, which could potentially lead to the detection of much smaller lesions. In addition to examples from our own group, work is being done by other laboratories using a variety of other gold-based nanomaterials. In all cases, the move from in vitro cell-based demonstrations to in vivo clinical use is enabled by rapid developments in photonics-based strategies for real-time, low-cost in vivo imaging.

## SUMMARY

Numerous research groups throughout the country are leveraging emerging techniques in optical imaging and nanotechnology to develop powerful new approaches for detecting molecular-specific signatures of precancers and early cancers. These groups are developing several classes of ultrabright contrast agents that strongly scatter and/or absorb at tunable wavelengths throughout the visible and near-infrared spectral bands, as well as methods of targeting these agents to molecular markers of neoplasia. They are demonstrating the efficacy of these agents in biological samples of progressively increasing complexity. These initial efforts will certainly be expanded in future studies.

Ultimately, the use of ultrabright contrast agents will extend the detection limits of optical technologies, increasing their sensitivity and specificity and promoting improved screening and detection of early lesions. We believe there is tremendous potential for synergy between the rapidly developing fields of biophotonics and nanotechnology. Combining the tools of these fields—together with the latest advances in understanding of the molecular origins of cancer—will offer a fundamentally new approach to the detection of cancer, a disease responsible for more than one-quarter of all deaths in the United States today.

## REFERENCES

Akerman, M.E., W. Chan, P. Laakkonen, S.N. Bhatia, and E. Ruoslahti. 2002. Nanocrystal targeting in vivo. Proceedings of the National Academy of Sciences 99: 12617–12621.

American Cancer Society. 2006. Cancer Facts and Figures 2006. Available online at: *http://www.cancer.org/docroot/STT/content/STT_1x_Cancer_Facts__Figures_2006.asp*.

Averitt, R.D., D. Sarkar, and N.J. Halas. 1997. Plasmon resonance shifts of Au-coated Au2S nanoshells: insights into multicomponent nanoparticles growth. Physiology Review Letters 78: 4217–4220.

Brongersma, M.L. 2003. Nanoshells: gifts in a gold wrapper. Nature Materials 2: 296–297.

Bruchez, M., M. Moronne, P. Gin, S. Weiss, and A.P. Alivisatos. 1998. Semiconductor labels as fluorescent biological labels. Science 281: 2013–2016.

Chan, W.C.W., and S. Nie. 1998. Quantum dot bioconjugates for ultrasensitive nonisotopic detection. Science 281: 2016–2018.

Faulk, W.T., and G. Taylor. 1971. An immunocolloid method for the electron microscope. Immunochemistry 8: 1081–1083.

Loo, C., L. Hirsch, J. Barton, N. Halas, J. West, and R. Drezek. 2004. Nanoshell-enabled photonics-based cancer imaging and therapy. Technology in Cancer Research and Treatment 3: 33–40.

Loo, C., A. Lowery, N. Halas, J. West, and R. Drezek. 2005. Immunotargeted nanoshells for integrated imaging and therapy of cancer. Nano Letters 5: 709–711.

Oldenburg, S.J., R.D. Averitt, S.L. Westcott, and N.J. Halas. 1998. Nanoengineering of optical resonances. Chemical Physics Letters 288: 243–247.

Sokolov, K., M. Follen, J. Aaron, I. Pavlova, A. Malpica, R. Lotan, and R. Richards-Kortum. 2003. Real-time vital optical imaging of precancer using anti-epidermal growth factor receptor antibodies conjugated to gold nanoparticles. Cancer Research 63: 1999–2004.

# Commercialization and Future Developments in Bionanotechnology

Marcel P. Bruchez
*Carnegie Mellon University*
*Pittsburgh, Pennsylvania*

Perfection in nanotechnology has long been achieved by biological systems. An enzyme represents a nearly perfect robot, stamping out molecular patterns from unique templates designed to execute individual tasks with nearly perfect efficiency. Evolution has driven these efficient designs to enable life forms to thrive in harsh environments. Evolutionary improvements have developed over a period of at least 3.5 billion years, with impressive results. The fact that we are here at all is a testament to the power and vast potential of nanotechnology.

Recently, we have made the first blunt-fingered attempts to extend the capabilities of biological systems by harnessing innovations in materials chemistry and electronics coupled with biologically defined specificity for both magnetic and fluorescent probes. In these cases, we have succeeded in introducing relatively limited new functionalities to existing biological systems. But we have barely tapped the potential of engineering of these systems, and from here on, our efforts will undoubtedly expand dramatically.

At the present time, we are guided by empirical observations and not by a detailed understanding of the interactions of biological systems with the materials and devices we are preparing. Thus, not only are we blunt-fingered, but we are also nearly blind. Before we can realize substantial commercial rewards and benefits in health and medicine, we will have to expand our efforts dramatically to develop new characterization methods and basic specifications and predictors of biological performance.

## DEFINITION OF BIONANOTECHNOLOGY

At the present time, there is no consensus definition of bionanotechnology. To take advantage of the enthusiasm of funding agencies, a number of old (and important) areas, such as colloid science, molecular biology, and implantable materials surface science, have been relabeled "nanotechnology." In fact, all of these fields, coupled with biological systems, should be included in bionanotechnology. In general, the idea of bionanotechnology is the engineering of interfaces between molecules or materials and biological systems. Looking ahead, the key areas for commercialization will be bringing engineered systems into biological contact and biological function.

The version of bionanotechnology popularized in the media has been largely oversold. The general idea, which was popular 20 years ago as the "magic-bullet" theory of biotechnology and has been adopted as the bionanotechnology target, can be described as the "dump truck" model of technology. In this conception, the technology components consist of a targeting moiety, either biological or nanotechnological, and one or more cargoes, which are envisioned as small machines capable of specific destructive or corrective action.

In reality, designing targeting molecules that are selective for diseased tissues and capable of delivering cargoes larger than a typical antibody has proven extraordinarily difficult, and molecular targeting of nanoscale devices greater than 5 nm outside the vascular space may prove to be prohibitively difficult. However, with no guiding principles for the effective biological direction of nonbiological molecules, this is still an open question.

In this paper, I describe three recent examples of commercialized bionanotechnology, beginning with the one that is the best characterized system. The three are antibody-directed enzyme prodrug therapy (ADEPT), superparamagnetic iron oxide particles for enhancing contrast on magnetic resonance images (MRIs), and quantum-dot technology for biological detection. Each of these shows the potential power and some of the challenges of integrating technologies at the molecular level.

## ANTIBODY-DIRECTED ENZYME PRODRUG THERAPY

Perhaps the most salient and relevant example of a bionanotechnology currently being commercialized is the ADEPT method being investigated by Genencor and Seattle Genetics (Figure 1). In this method, an antibody-enzyme fusion is first prepared and isolated. This molecule can be designed with precise chemical (biological) composition, precise linkage geometry, and complete definition and characterization using standard molecular-biology techniques and biochemical methods. The antibody, linked to the enzyme, can be used to target the particular antibody-enzyme conjugate to the site of interest. In this way, a small antibody fragment is used to target a molecular machine (an enzyme) to a par-

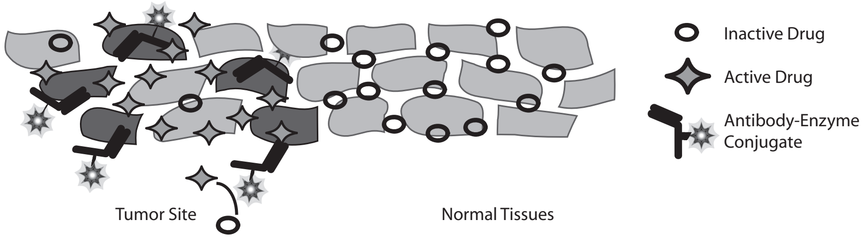


FIGURE 1 ADEPT uses antibodies to target particular cells with an enzyme that then converts prodrug molecules to an active drug at the target site.

ticular site of interest, and the machine is then used to generate a specific molecule at that site. In the clinic, a prodrug (a drug molecule modified to an inactive state that can be converted to an active state in situ) is administered to the patient. After the antibody-enzyme construct reaches its target site, the prodrug is administered and is converted by the enzyme to an active state. The local concentration of the active drug can be driven very high, even though the overall concentration remains very low. Thus, the therapy is both safer and more effective than a high dose of the toxic compound.

ADEPT is a highly characterized, highly effective example of bionanotechnology in action. However, even after 15 years of active research, these targeted prodrug strategies are still in the research or early clinical trial stage and not in general practice. This is a reflection of the complexities of developing biospecific performing technologies, which is likely to be a general problem for the development of nanotechnologies with high clinical impact.

## SUPERPARAMAGNETIC IRON OXIDE PARTICLES

A second, more recognizable example of bionanotechnology in clinical use is Ferridex and Combidex superparamagnetic particles, marketed by Advanced Magnetics, which are being commercialized for enhancing MRI signals (Figure 2). The particles are modified with dextran (a polymerized sugar molecule) to create a biocompatible coating, which dramatically reduces nonspecific interactions in the body and increases the contrast of the instrument wherever the particles are present. When administered intravenously, they can easily be measured in a standard clinical MRI instrument. These materials are currently approved for imaging cancers of the liver and spleen.

Recently, the Combidex agent was rejected by the Food and Drug Administration (FDA) because of safety concerns and a lack of efficacy data. Questions regarding safety had arisen when at least one patient died in a clinical trial investigating the use of the Combidex agent for sentinel-node detection (finding

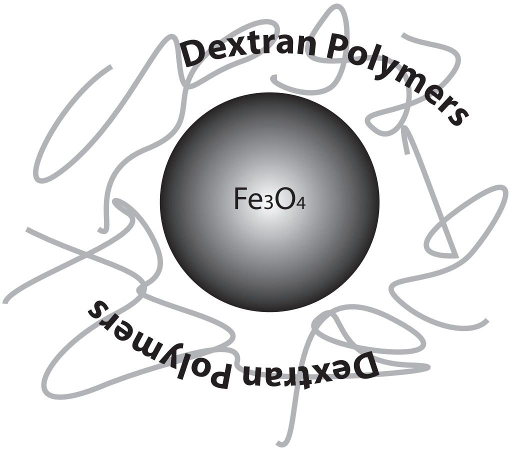


FIGURE 2 Combidex particles, which enhance contrast on MRIs, consist of superparamagnetic iron oxide nanoparticles covered with dextran. Because these particles are excluded from tumor-bearing regions of lymph nodes, they can help identify tumor-bearing nodes noninvasively.

near-nodes that are most likely to contain cancerous cells), which is critical to the grading, staging, and appropriate treatment of cancers. The FDA did recommend, however, that with further appropriately designed trials, the compound may be approvable for specific indications.

## QUANTUM-DOT TECHNOLOGY

I have been extensively involved in work on the third example—the use of fluorescent quantum dots for biological detection in research and, ultimately, clinical applications (Figure 3). Semiconductor nanocrystals (i.e., quantum dots), specifically designed to have intense monochromatic emission spectra, are coupled to biological targeting molecules, such as antibodies and nucleic acids. The conjugates can then be used to detect the presence of particular analytes in biological samples. Although these particles dramatically increase experimental information and sensitivity, the clinical community has been slow to adopt them because of subtle protocol differences between these materials and the typical fluorescent dyes and enzymatic methods used in detection. Many of the protocol differences are thought to arise from distinct size differences between typical probes and nanotechnology-based probes. Such idiosyncrasies are likely to be ubiquitous in nanotechnology-enabled product commercialization.

The technology for the use of quantum dots in biology was initially published in September 1998 in two simultaneous papers in *Science* (Bruchez et al., 1998; Chan and Nie, 1998). Although these articles generated significant enthusiasm in the scientific community, biologically useful products were not launched until November 2002. In the meantime, Quantum Dot Corporation was accused of hoarding the technology, stalling progress, and many other things.

In fact, the reasons for the delay were hardly nefarious—we have no rational framework for "optimizing" these materials. Therefore, although we were working very hard to make products that could be used successfully by the average biologist, every time we made an improvement to any aspect of the system, the entire process had to be revalidated. This empirical approach to product development resulted in a very long development time.

This delay was in addition to unavoidable delays in process development. A nanoparticle designed for a particular application is a complex multilayer structure, shown schematically in Figure 3. Scale-up of the initial chemistries used to make these nanoparticles (as published in *Science*) was exceptionally dangerous; procedures involved pyrophoric precursors, flammable solvents, and rapid additions and releases of explosive gases. To develop safe, scalable procedures, our scientists had to develop innovative techniques in all aspects of nanoparticle chemistry. Again, every innovation had to be validated through to the utility of

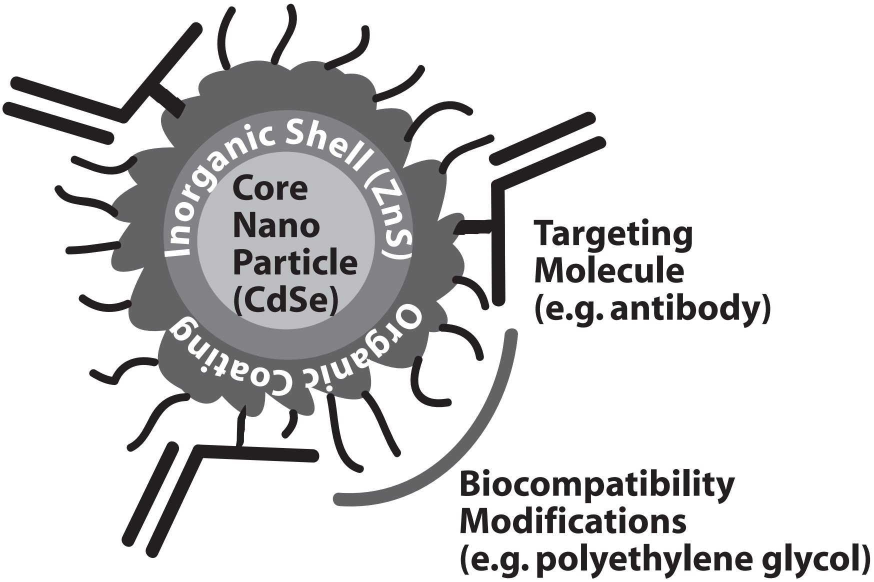


FIGURE 3 Quantum-dot conjugates have bright emission and multicolor capability, allowing researchers to view many targets in a single sample and simplifying detection strategies. The overall structure of these conjugates is designed and optimized for performance in biological applications (e.g., stability, brightness, and specificity).

the final material, making the development cycle exceptionally onerous. Just imagine if, to validate a change in one hose of a car, you had to build an entire car with only that change included.

The lack of specifiability of our modules was a key challenge to commercialization. Specification will require detailed basic investigations of the properties and chemistry of nanoparticle materials in biological systems. In addition, we will have to establish analytical tools and quantitative descriptors to detail the distribution of properties present in a population of nanoparticles. This is categorically different from specification for organic molecules and proteins, in which properties can be effectively described by an average. In nanomaterials, performance properties may be dominated by a relatively small population of particles, so averaging cannot always be used.

## THE CHALLENGE OF CHARACTERIZATION

Dramatically different tools are necessary for characterization of the three examples I have described. ADEPT, a fully biological system, can be characterized structurally, chemically, and on the basis of activity to ensure that each component of the system is capable of acting independently and that this behavior is preserved as the system components are brought together. Nevertheless, for reasons related to biological complexity, the use of ADEPT in the clinic has not yet proven beneficial. This problem gives some indication of the challenges ahead for nanotechnology solutions.

The second example, Combidex technology, is a homogeneous-particle technology covered with a natural material, dextran, that minimizes the complexity of the system. In this case, the particle size and shape (which can be characterized by electron microscopy) dictate its magnetic properties. The interaction of dextran on the surface dictates the in vivo behavior of these materials. Although the components can be characterized in great detail, the interaction of the dextran with the surface (e.g., the number of surface iron atoms that are actually covered) may be crucial to the fate of these materials in clinical use, an obstacle that was not predicted.

The interaction of molecules with surfaces in complex environments represents a critical area for analytical development. At the moment, many studies are carried out by x-ray photoelectron spectroscopy (XPS), a vacuum technique that does not show many solution interactions in the normal biological environment. Micro-rheology techniques might be valuable in addressing this issue.

The final example, quantum dots, an entirely engineered material, presents many characterization challenges. First, the particle itself is a complex structure, and the best available tools for characterizing these materials are capital intensive and often inaccessible. Essentially, methods such as energy-dependent XPS require a synchrotron source. Other methods, such as Z-contrast scanning transmission-electron microscopy, require unique instrumentation that is available only in a few laboratories in the world. In addition, these tools are suited either for measuring average properties or measuring single-particle properties, but not both. Bridging the gap to a statistical method that shows single-particle properties in a large population of particles would allow for discrimination of population properties from single-particle properties.

Moving out in the structure, the surface is coated with ligands. Thus, surface interactions will cause the same problems as have arisen for Combidex—routine tools do not give a detailed molecular picture of interactions at the surface. The problem is further complicated as particles are modified for biological applications, for instance by coupling polyethylene glycol molecules to the surface.

The characterization of chemistry on the surface of these particles has not matured to the level of typical organic chemical reactions. In fact, much of the characterization is still inferential (i.e., we analyze what does not react with the particle to determine what does react with it). The tools we have today neither discriminate between adsorbed and reacted materials nor determine whether the chemistry is homogeneous or heterogeneous from particle to particle. These distinctions will be critical for the development of nanoparticle tools with biomedical applications.

## OUTLOOK

Where, then, will bionanotechnology take us? As the examples I have described show, advances have progressed from ADEPT, a completely characterized system with a defined molecular structure (still encountering difficulty in clinical acceptance), to a system in which components are well characterized (Combidex), to quantum dots, a system we still cannot fundamentally characterize. Chemists have tools like mass spectrometry and nuclear magnetic resonance spectroscopy to guide them. Engineers have testing and measurement systems for validating systems as small as a few hundred nanometers. In the middle range, however, nanoengineers (or nanochemists) still do not have the fundamental tools to determine how well they have done their jobs or in what direction they should look for improvements.

Devices are synthesized on molar ($\sim 10^{26}$) scales, but the characterization tools designed for molecules do not work effectively for bionanotechnology systems. Clearly, the device characterization methods (typically single "device" characterization on enough devices to ensure a reliable measurement of production-run statistics) are inappropriate, especially when a dose is $10^{13}$ devices and a minor population component can dominate the bad effects (for instance pore-clogging).

Thus, we have an acute and growing need for specifiability in the design of bionanotechnology tools. To achieve engineerable systems, a concerted effort must be made to conduct a basic scientific investigation of the impact of materials properties on the biological behavior of bionanotechnology systems, combined with a physical scientific investigation of new methods to characterize the detailed physical and population properties of nanometer-scale materials and components. Specifiability will make predictability, falsifiability, and rapid progress in commercial bionanotechnology feasible.

## FURTHER READING

### ADEPT Technology

Alderson, R.F., B.E. Toki, M. Roberge, W. Geng, J. Basler, R. Chin, et al. 2006. Characterization of a CC49-based single-chain fragment-beta-lactamase fusion protein for antibody-directed enzyme prodrug therapy (ADEPT). Bioconjugate Chemistry 17(2): 410–418.

Bagshawe, K.D., S.K. Sharma, and R.H.J. Begent. 2004. Antibody-directed enzyme prodrug therapy (ADEPT) for cancer. Expert Opinion on Biological Therapy 4(11): 1777–1789.

Wu, A.M., and P.D. Senter. 2005. Arming antibodies: prospects and challenges for immunoconjugates. Nature Biotechnology 23(9): 1137–1146.

### Combidex Technology

Harisinghani, M.G., and R. Weissleder. 2004. Sensitive, noninvasive detection of lymph node metastases. PLOS Medicine 1(3): 202–209.

Harisinghani, M.G., J. Barentsz, P.F. Hahn, W.M. Deserno, S. Tabatabaei, C.H. van de Kaa, et al. 2003. Noninvasive detection of clinically occult lymph-node metastases in prostate cancer. New England Journal of Medicine 348(25): 2491–2495.

### Quantum-Dot Technology

Alivisatos, A.P. 2004. The use of nanocrystals in biological detection. Nature Biotechnology 22(1): 47–52.

Bruchez, M.P., P. Gin, M. Morrone, S. Weiss, and A.P. Alivasotos. 1998. Semiconductor nanocrystals as fluorescent biological labels. Science 281(5385): 2013–2016.

Chan, W.C.W., and S. Nie. 1998. Quantum dot bioconjugates for ultrasensitive nonisotopic detection. Science 281(5385): 2016–2018.

Michalet, X., F.F. Pinaud, L.A. Bentolila, J.M. Tsay, S. Doose, J.J. Li, et al. 2005. Quantum dots for live cells, in vivo imaging, and diagnostics. Science 307(5709): 538–544.

# ENGINEERING PERSONAL MOBILITY FOR THE 21ST CENTURY

# Introduction

Apoorv Agarwal
*Ford Motor Company*
*Dearborn, Michigan*

William F. Schneider
*University of Notre Dame*
*Notre Dame, Indiana*

Humans have historically spent roughly the same share of their time and income traveling daily, but modern technology, especially the automobile, has greatly increased both the range and convenience of personal travel. Personal mobility enabled by automobiles is closely related to a sense of personal freedom. Current approaches to providing this mobility have had a major impact on the landscape of our cities and suburbs, on the environment as a whole, and on energy consumption. Providing the same levels of personal mobility in the future in a cost-effective, energy-efficient, and environmentally sustainable manner in both the developing and developed worlds is one of the great challenges for the 21st century.

The speakers in this session explore the history and evolution of personal mobility, including its availability and the expectations it raises, the energy and environmental challenges of current forms of personal mobility, and prospective technologies that could transform personal mobility for us and future generations.

# Long-Term Trends in Global Passenger Mobility

ANDREAS SCHÄFER*
*University of Cambridge*
*Cambridge, United Kingdom*

Anticipating changes in travel demand on aggregate levels is critical for industries making decisions about meeting the demand for vehicles and fuel and for governments planning infrastructure expansions, predicting transport-sector (greenhouse-gas) emissions, and evaluating mitigation policies. In this paper, I show that only a few variables are necessary to explain past levels and project internally consistent future levels of aggregate, world-regional travel demand and transport modes. I then highlight the enormous challenges that must be met to reduce greenhouse-gas emissions, especially from passenger aircraft, the fastest growing transport mode.

## DETERMINANTS OF AGGREGATE TRAVEL DEMAND AND TRANSPORT MODES

Growth in per capita income and population are the two single most important factors in passenger mobility. During the past 50 years, global average per capita income has increased slightly more than threefold, and world population has more than doubled. This combined growth, by a factor of 7.4, has translated into a nearly proportional increase in passenger mobility. The nearly direct rela-

---

*The fundamental ideas underlying the model described in this paper were developed jointly with David G. Victor, Stanford University. See Schäfer and Victor (2000).

tionship can be explained by so-called travel budgets, roughly constant averages of expenditure shares of money and time.

Although the amount of time spent traveling is highly variable on an individual level, large groups of people spend about 5 percent of their daily time traveling. The stability of the aggregate "travel-time budget," first hypothesized in similar form for urban travelers by the late Yacov Zahavi (1981), is illustrated in Figure 1 for a wide range of income levels. On average, residents in African villages, the Palestinian Territories, and the suburbs of Lima spend between 60 and 90 minutes per day traveling, the same as for people living in the automobile-dependent societies of Japan, Western Europe, and the United States.

A similar transition, from variability at disaggregate levels to stability at aggregate levels, can be observed for travel-expenditure shares (i.e., the percentage of income dedicated to travel). Zahavi observed that households that rely exclusively on nonmotorized modes of transport and public transportation spend only about 3 to 5 percent of their income on travel; that percentage rises to 10 to 15 percent for people who own at least one motor vehicle. Figure 2 shows that the aggregate "travel-money budget," here defined as total travel expenditures divided by the gross domestic product (GDP), follows a similar pattern, rising from about 5 percent of GDP at motorization rates close to zero cars per 1,000 capita (nearly all U.S. households in 1909 and the least developed countries

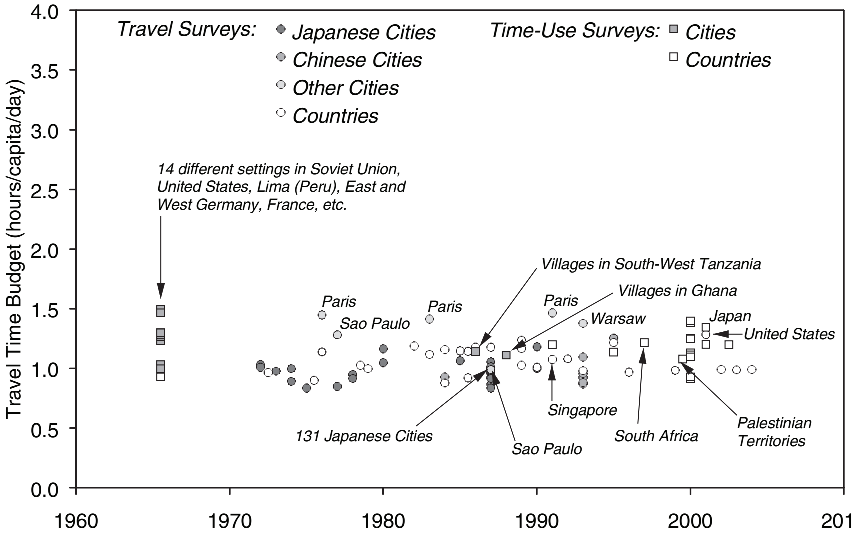


FIGURE 1 Average daily travel time as a function of per capita GDP.

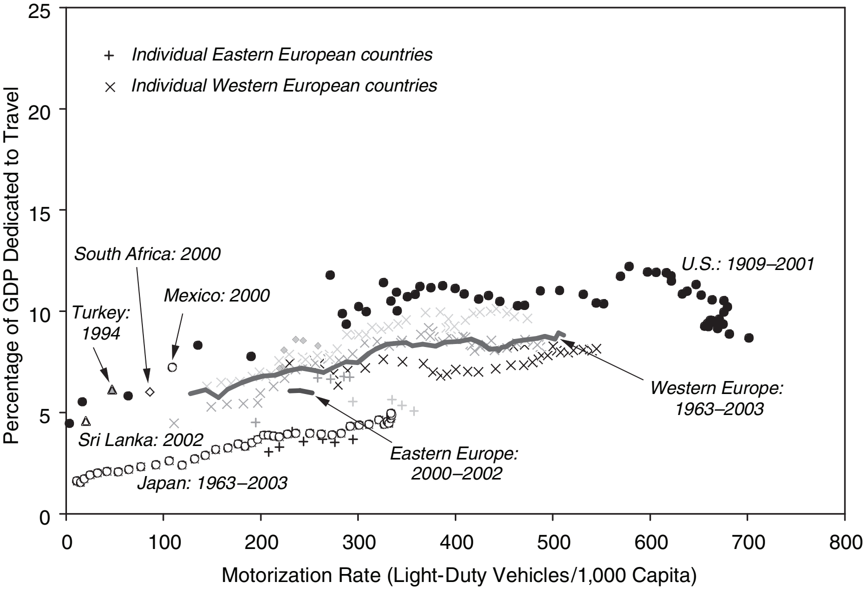


FIGURE 2 Travel expenditures as a fraction of income for annual travel distance.

today) to around 10 percent of GDP at about 300 cars per 1,000 capita (industrialized countries today), an ownership level of about one car per household of three to four people. In addition, travel demand and choice of transport mode depend on average door-to-door speed and travel costs to the consumer. The drastic decrease in the cost of air travel in the past decades has contributed to the rising share of that transport mode.

## THE PAST FIVE DECADES IN WORLD TRAVEL DEMAND

To study the historical development and project the future development of global travel demand, we estimated a unique data set of passenger mobility for 11 world regions, covering passenger-kilometers (km) traveled (PKT) using four major modes of transport (light-duty vehicles, buses, railways, and aircraft) and spanning 51 years (1950 to 2000).[1] The overall relationship between GDP per capita and per person PKT is shown in Figure 3. The saturating travel-money budget, described above, helps explain how rising GDP translates into rising travel demand.

---

[1] This data set is an update of an older data set for 1960 to 1990 by the same author. See Schäfer (1998).

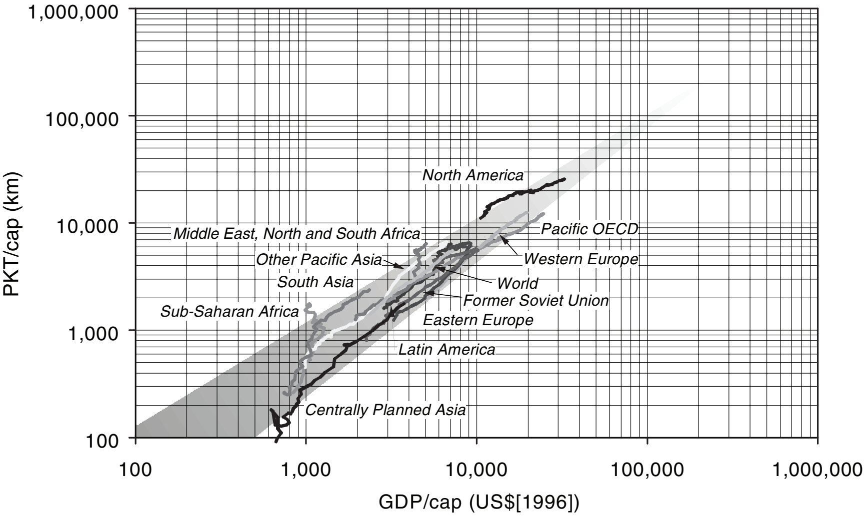


FIGURE 3 Passenger-km per capita by per capita GDP for 11 world regions and the entire world from 1950 to 2000.

Over the past five decades, Earth's inhabitants have increased their travel demand from an average of 1,400 to 5,500 km, using a combination of automobiles, buses, railways, and aircraft. Since the world population grew nearly 2.5-fold during the same period, world PKT increased by one order of magnitude, from nearly 3.6 to some 33 trillion PKT. The biggest increase in PKT, by a factor of more than 20, occurred in the developing world, where the combined growth in per capita GDP and population was largest.

However, the "mobility gap" between developing and industrialized regions remains substantial. In 2000, residents in North America, the Pacific Organisation for Economic Co-operation and Development (Japan, Australia, and New Zealand), and Western Europe traveled 17,000 PKT per capita on average, five times as much as people in the developing world. These differences are even larger on a world-regional scale. Residents of North America, the region with the highest level of mobility, traveled 25,600 km per year, while people in sub-Saharan Africa (not including South Africa) traveled just 1,700 km.

GDP is the most important, but not the only, determinant of per-person PKT. As Figure 3 shows, the average travel per person differs significantly at different income levels, mainly because of different costs for transportation, but also because of the size of purchasing power parity (PPP) adjustments in the socioeconomic data set (Heston et al., 2002).

While the travel-money budget translates rising per capita GDP into rising

PKT per capita, the fixed travel-time budget requires that the increasing travel demand be satisfied in the same amount of time. Since each transport mode operates within a known range of speeds, the increasing per-person travel demand can only be satisfied by shifting toward increasingly rapid transport.

Figure 4 shows the continuous shifts toward faster modes of transport, from low-speed public transportation (the aggregate of buses and low-speed railways), to light-duty vehicles (automobiles and personal trucks, but, for simplicity, referred to as automobiles), to high-speed modes of transportation (aircraft and high-speed rail), again for a 51-year historical time horizon. Three distinct phases of dominance by a single transport mode can be seen. Low-speed public transportation is dominant for mobility levels of up to 1,000 km/cap; light-duty vehicles between 1,000 and 10,000 km/cap; and high-speed transport modes at even higher levels of mobility.

Similar to differences in total mobility, differences in travel costs and in urban land-use characteristics can lead to different levels in shares for transport modes at a given level of PKT per capita. However, the impact of policy measures on choice of transport mode is limited—at least on the aggregate levels shown. In Eastern Europe and the former Soviet Union, access to private automobiles was severely restricted until the transition toward a market economy in the early 1990s. Nevertheless, the modal trajectories have evolved largely within the shaded envelopes.

## THE NEXT FIVE DECADES IN WORLD TRAVEL DEMAND

If travel-expenditure shares remain approximately stable, future increases in per capita GDP will continue to cause a rise in PKT. At the same time, the fixed travel-time budget will continue to push travelers toward faster modes of transport. The highest level of travel demand would be achieved if travelers used the fastest mode of transport for their entire daily travel-time budget 365 days a year. Assuming that aircraft gradually increase their current average "door-to-door" speed from about 260 km/h (including transfers to and from the airport) to 660 km/h, the current average airport-to-airport speed for domestic flights in the United States, and a travel-time budget of 1.2 h/d for 365 d/y, the annual per-person traffic volume would result in approximately 289,000 km. At that high mobility level, most travel would be international. Prices would adjust, and so would income levels.

Hence, regional differences in per capita traffic volume at a given GDP per capita, mainly resulting from differences in land use and prices, would decline, and the 11 trajectories would ultimately converge into a single point in the far future. Given historical development, it is assumed that the GDP per capita value of that "target point" would correspond to US$(2000) 289,000. (Sensitivity analyses show that the exact location of the target point has almost no impact on the levels projected for 2050.)

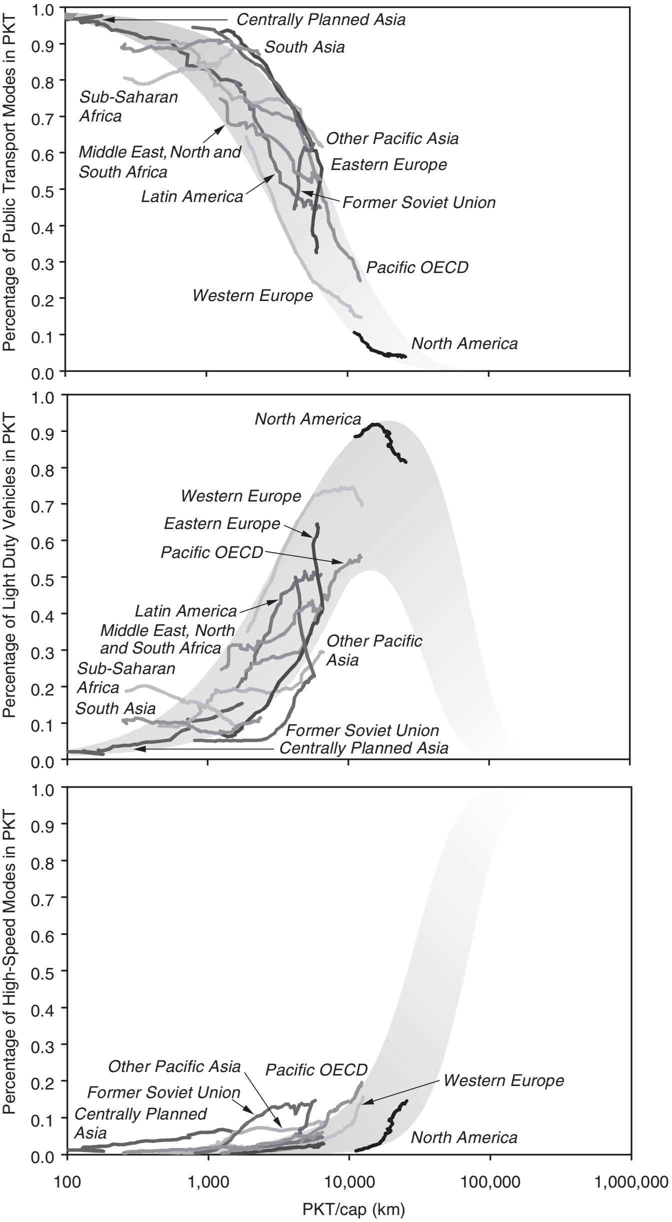


FIGURE 4 Three stages in the evolution of motorized mobility (1950–2000): the declining share of low-speed public transport modes (top), the growth and relative decline of the automobile (middle), and the rise of high-speed transportation (bottom).

This imaginary world of high-speed transportation helps in projecting future levels of PKT by approximating each of the 11 world-regional trajectories in Figure 3 by one and the same type of regression equation and by constraining one parameter of that equation so that the simulated trajectory must pass through the target point.[2] Future levels are then determined by the predicted levels of GDP/cap and population. The world-regional GDP per capita projections used here are derived from recent reference runs of the MIT Joint Program on the Science and Policy of Global Change Systems model, after slightly reducing the growth rates of industrialized regions and the reforming economies of Eastern Europe and the former Soviet Union and slightly increasing those of developing countries to match the mean values of past projections more closely (Nakicenovic, 2000; Paltsev et al., 2005). Overall, PPP adjusted gross world product per capita is projected to nearly double from US$(2000) 7,500 in 2000 to US$(2000) 14,200 in 2050. In addition to the 50 percent growth in world population, as suggested by the medium variant of the United Nations population projections (2004), gross world product would rise by a factor of nearly three.

Based on these changes in socioeconomic conditions, the stable relationship between growth in GDP and traffic volume implies that world-travel demand will increase approximately in proportion to the projected level of income, from 33 trillion passenger-km in 2000 to 105 trillion in 2050. Because of their projected higher growth in income and population, developing regions will contribute a rising share, ultimately accounting for 60 percent of world passenger-traffic volume in 2050, up from about 50 percent in 2000. (Higher growth rates of GDP in developing regions will further increase their absolute and relative importance in traffic volume.)

Given fixed travel time, future shares of low-speed public transportation modes, light-duty vehicles, and high-speed transportation systems must remain largely within the shaded envelopes in Figure 4. (The target point condition requires that high-speed transportation account for the entire traffic volume in a hypothetical world where the target point can be reached.) The precise shift in shares of transport modes, necessary to satisfy the projected travel demand through 2050, can be derived in a number of ways, but perhaps most convincingly by estimating the parameters of the functional form of statistical consumer-choice models. However, in this application, those models would require time-series data (ideally for 1950 to 2000) on speeds and travel costs for each transport

---

[2] The general form of the regression equation is $\mathrm{PKT/cap} = \chi \cdot (\mathrm{GDP/cap})^{\alpha} \cdot (P_T)^{\beta}$, with parameters $\chi$ being a constant, $\alpha$ the income elasticity, and $\beta$ the price elasticity. However, because long-term historical data of travel costs ($P_T$) are available for very few countries, this factor is dropped. Thus, $\chi$ includes the averages of travel-money budget and price of travel in a particular region. Imposing the target point condition leads to an estimate of $\chi$.

mode. These data can be derived for the United States and, to a limited extent, for a few European countries and Japan, but they are not available for most countries in the world.

Without these data, we can perform simplified projections by determining plausible future shares in each modal envelope at the projected level of per capita PKT, depending on whether a particular region is an early adopter or a latecomer to the diffusion of automobiles. Latecomers achieve lower shares of light-duty vehicle travel, here assumed to develop along the lower boundary of the automobile envelope in Figure 4, as they have already "leapfrogged" into high-speed travel and thus develop along the higher boundary of the high-speed transportation envelope in the same figure. (For a general introduction to diffusion theory, see Grübler, 1990.)

By contrast, future shares of high-speed transportation in the three industrialized regions and two reforming economies are estimated as the mean value of the upper and lower boundary of the envelope at the projected level of per capita GDP. The projection for the industrialized regions are then retrospectively compared to estimates from more sophisticated statistical-choice models, for which more complete speed and cost data are available. For example, the estimate of a detailed statistical-choice model for North America yields a 2050 share of 55 percent for high-speed transportation, which compares to 56 percent using the simplified approach. The use of statistical-choice models also allows us to conduct sensitivity tests (e.g., with regard to the stability of the travel-time budget). Should the travel-time budget increase from 1.2 to 1.5 hours per person per day (a 25 percent increase), the projected 2050 share of high-speed transportation would decline from 55 percent to 44 percent (a 20 percent decline). Although the decline in the 2050 share of high-speed transportation is significant, a 44 percent share still corresponds to three times the share for that transport mode in 2000.

In the industrialized world, light-duty vehicles and high-speed transportation modes will account for nearly the entire traffic volume in 2050 and for roughly equal shares. By contrast, in reforming economies and developing regions, automobiles will supply most of the PKT, followed by low-speed public transportation. In both meta-regions, however, high-speed transportation is also on the rise, accounting for nearly 20 percent of the 2050 passenger-traffic volume. Globally, the traffic shares of automobiles and low-speed public transport modes will decline by about 6 and 12 percentage points, respectively, below the 2000 level by 2050. At the same time, the relative importance of high-speed modes will increase from 10 percent to about 30 percent. Figure 5 summarizes the global development in PKT by major mode of transport for 1950, 2000, and 2050 (projected).

The simplified model necessarily has a number of limitations. Perhaps most important, the mode shifts in Figure 4 represent only the aggregate of two markets and do not capture substitutions in the urban and intercity transport seg-

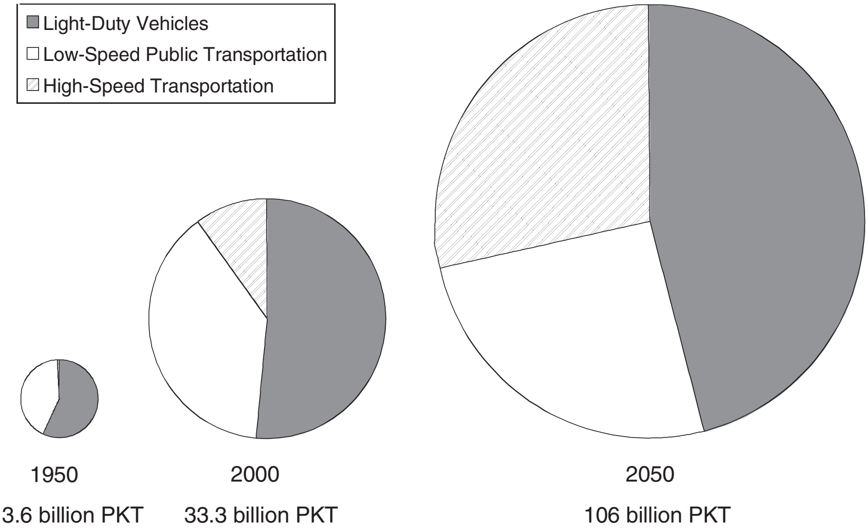


FIGURE 5 Global passenger-km traveled, by major mode of transport, in 1950, 2000, and 2050 (projected). Size of pies corresponds to PKT, which has been multiplied by nearly 10 times through 2000 and is likely to be multiplied by a factor of 30 by 2050. For comparison, GDP has grown by factors of 7 and 20, respectively.

ments. In the urban transportation segment, light-duty vehicles become more important to the cost of low-speed public transportation. By contrast, in intercity transport, automobiles are displaced by high-speed transportation modes. By disaggregating the data set into these two transportation markets and estimating (the functional form of) a nested, discrete-choice model, we could project plausible levels of shares for transport modes over time periods of more than 50 years. (Whether these projections will ultimately be achieved in reality is a different subject, which raises questions about whether such models are more valuable as tools for understanding interactions between humans and technology than for making exact predictions of the future.)

Another limitation is that the projection of future passenger mobility was performed in an unconstrained world. However, a separate analysis of potentially limiting factors, including the resource base of oil products, the need for higher aircraft speeds, the potential substitution of travel by telecommunication, increasing travel congestion, and so on, suggests that none of these constraints is likely to be binding in the next five decades. At some point in the future, however, the finite characteristics of our planet will necessarily have an impact on transportation systems.

## IMPLICATIONS

The growth in travel demand and the shift toward faster transport modes have a number of implications. Two of the most important are for the amount of travel time spent in different transport modes and the impact on energy use and greenhouse-gas emissions.

Figure 6 shows the daily per-person travel distance by mode of transport (left) and the associated daily travel time (right) for 1950, 2000, and 2050 (projected) in North America (essentially the United States and Canada). Over the past 50 years, the daily travel distance has more than doubled, from 30 km to 70 km, while per-person travel time has likely remained stable (no time-use data are available for 1950). Over the next five decades, based on our projection of per capita GDP, daily mobility will double again, to 140 km, with high-speed transportation accounting for 56 percent. However, despite the growing demand for high-speed transport, travelers will continue to spend most of their travel time on the road. Although automobile travel time will decrease only slightly, the main change in travel-time allocation will be a net substitution of high-speed transportation for low-speed public transportation. A traveler in 2050 will spend an average of 12 minutes per day in the air or on high-speed railways (compared to two minutes today). If the per-person travel-time budget increases to 1.5 hours per day, the average daily high-speed travel time will decrease to about 9 minutes.

Although total travel time may not be affected by the increase in travel demand, energy use and greenhouse-gas emissions will change. All factors being

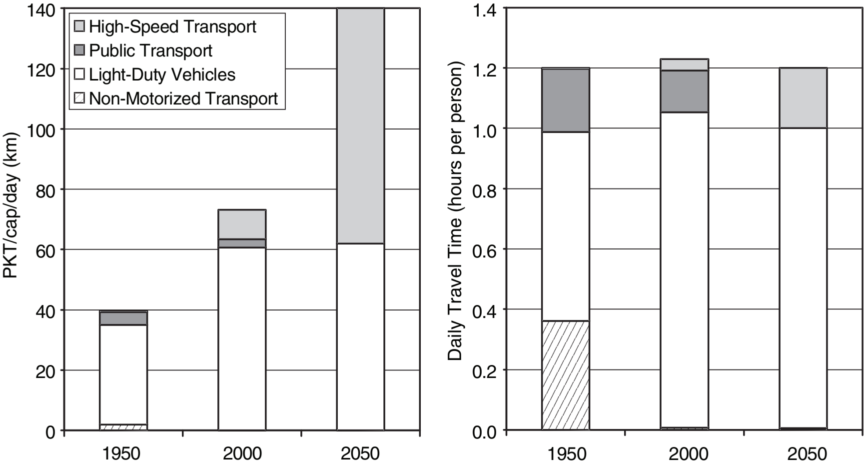


FIGURE 6 Daily per-person travel distance by mode of transport (left) and associated daily travel time (right) for 1950, 2000, and 2050 (projected) in North America.

equal, rising travel demand will cause a proportional increase in passenger-travel energy use. Given that (synthetic) oil products are likely to continue dominating the future fuel supply of the transportation system, over the next five decades, directly released greenhouse-gas emissions will also rise, roughly in proportion to travel demand. (The increase in life-cycle greenhouse-gas emissions could be greater if there is a significant shift toward unconventional oil, such as extra-heavy oil, oil sands, shale oil, or synthetic oil from natural gas or, especially, coal.) Changes in passenger-travel energy intensity (i.e., energy use per PKT) will also influence levels of passenger-travel energy use and greenhouse-gas emissions.

In the absence of more fuel-efficient transport modes, three trends will determine future levels of energy intensity. First, any increase in travel speed will cause an increase in energy intensity. Based on current and average U.S. data, a complete shift from low-speed public transportation to light-duty vehicles in urban travel would increase energy intensity by 25 percent. For intercity travel, a complete shift from low-speed public transportation to light-duty vehicles would increase energy intensity by almost 60 percent. A shift toward air travel would increase it by another 40 percent. In the United States, most of these changes have already taken place. If the ongoing shift from automobile intercity travel toward aircraft continues, intercity passenger-travel energy intensity will increase by 15 to 20 percent by 2050.

Second, the change to air travel for intercity transport will also increase the relative importance of urban automobile travel, which is more energy intense than intercity automobile travel because of varying engine loads and low occupancy rates. Thus, this shift will cause an increase in average automobile-travel energy intensity. However, even though the energy intensity of urban travel is twice that of intercity travel, it will probably only increase 10 percent or less, because nearly all PKT by automobiles already occurs over relatively short distances.

Third, the substitution of air transportation for intercity automobile travel mainly occurs at trip distances of less than 1,000 km, distances at which aircraft energy use is mainly determined by the energy-intensive takeoff and climb stages. Aircraft energy intensity at such stage lengths can be twice as high as for trips of more than 1,000 km (Babikian et al., 2002).

In North America, the strong growth in air travel suggests that the combined effect of these three trends is determined mainly by the change in aircraft energy intensity resulting from the relative growth in different market segments. However, because the average energy intensity of total air travel is lower than for automobiles and low-speed public transport modes operating in urban areas, the overall effect of these changes is likely to be negligible. Thus, total energy use and greenhouse-gas emissions will rise roughly in proportion to the growth in PKT (i.e., by 2050, 130 percent over the 2000 level, based on our assumptions of GDP growth). In Western Europe, the combination of these trends may result in

TABLE 1 Projected Percentage Change in Aircraft Energy Use by 2050

| | Low Estimate | High Estimate |
|---|---|---|
| Aircraft technology | –25 | –45 |
| PAX load factor | –10 | –10 |
| Direct flights | 0 | –11 |
| Shift to high-speed rail | –1 | –1 |
| Totals | –33 | –56 |

Note: Estimates for high-speed rail are based on 50 percent market share in 10 U.S. high-density corridors, with a cumulative great-circle distance of 16,700 km. Sources: Adapted from Lee et al., 2001; Jamin et al., 2004.

passenger-travel energy intensity rising by 2050 by as much as 20 percent above the level for 2000.

The situation is fundamentally different, especially in developing countries, where the substitution of automobiles for low-speed public transportation is just beginning. Combined with a future decline in vehicle occupancy rates (mainly a consequence of increasing female participation in the labor force), the impact of these trends on passenger-travel energy intensity may be 50 to 100 percent. Compensating for this increase in energy intensity in developed countries already requires sophisticated fuel-saving technology.

In the passenger-transport sector, air travel accounts for the fastest growth in energy use and greenhouse-gas emissions. In addition to the projected ninefold increase in global air-travel demand, future levels of air-travel greenhouse-gas emissions will depend on which technologies are used, assuming that (synthetic) oil products continue to fuel air transportation. Table 1 shows the major opportunities for reductions in aircraft energy use for a given travel demand based on recent studies. Even if aircraft energy use is reduced by 33 to 56 percent by 2050, the 2000 level of carbon dioxide emissions would still be multiplied by a factor of four to six. Thus, controlling greenhouse-gas emissions from transportation will remain a significant challenge for generations to come.

## REFERENCES

Babikian, R., S.P. Lukachko, and I.A. Waitz. 2002. Historical fuel efficiency characteristics of regional aircraft from technological, operational, and cost perspectives. Journal of Air Transport Management 8(6): 389–400.

Grübler, A. 1990. The Rise and Fall of Infrastructures. Heidelberg, Germany: Physica-Verlag.

Heston, A., R. Summers, and B. Aten. 2002. Penn World Table Version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP), October 2002. Available online at: *http://pwt.econ.upenn.edu/php_site/pwt_index.php*.

Jamin, S., A. Schäfer, M.E. Ben-Akiva, and I.A. Waitz. 2004. Aviation emissions and abatement policies in the United States: a city pair analysis. Transportation Research D 9(4): 294–314.

Lee, J.J., S.P. Lukachko, I.A. Waitz, and A. Schäfer. 2001. Historical and future trends in aircraft performance, cost, and emissions. Annual Review of Energy and the Environment 26: 167–200.

Nakicenovic, N., ed. 2000. Special Report on Emissions Scenarios. Intergovernmental Panel on Climate Change. Cambridge, U.K.: Cambridge University Press.

Paltsev, S., J.M. Reilly, H.D. Jacoby, R.S. Eckaus, J. McFarland, M. Sarofim, M. Asadoorian, and M. Babiker. 2005. The MIT Emissions Prediction and Policy Analysis (EPPA) Model: Version 4. MIT Joint Program on the Science and Policy of Global Change, Report 125. Cambridge, Mass.: MIT Press. Available online at: *http://web.mit.edu/globalchange/www/reports.html#pubs*.

Schäfer, A. 1998. The global demand for motorized mobility. Transportation Research A 32(6): 455–477.

Schäfer, A., and D.G. Victor. 2000. The future mobility of the world population. Transportation Research A 34(3): 171–205.

United Nations. 2004. World Population Prospects: The 2004 Revision Population Database. United Nations Population Division. Available online at: *http://esa.un.org/unpp/*.

Zahavi, Y. 1981. The UMOT-Urban Interactions. DOT-RSPA-DPB 10/7. Washington, D.C.: U.S. Department of Transportation.

## FURTHER READING

Marchetti, C. 1994. Anthropological invariants in travel behavior. Technological Forecasting and Social Change 47: 75–88.

Zahavi, Y., and A. Talvitie. 1980. Regularities in travel time and money expenditures. Transportation Research Record 750: 13–19.

# Energy and Environmental Impacts of Personal Mobility

MATTHEW J. BARTH
*University of California, Riverside*

Personal mobility is critical for a progressive society. The freedom to travel anywhere at anytime with few restrictions is a precondition for a vibrant economy. However, mobility is often restricted by limitations in the transportation infrastructure. For example, when many people use the infrastructure at the same time, congestion invariably occurs. One can look at this phenomenon as a resource-management problem. If resources (i.e., the transportation infrastructure) are limited and demand is high, congestion is likely to occur. Two ways of solving this problem are (1) by providing additional resources and/or (2) reducing demand.

In the United States, the transportation system has primarily been developed around the automobile, and the majority of personal trips are made by driving cars to various destinations. Statistics show that only a small percentage of the U.S. travel demand is satisfied by public transit. Instead, we have invested billions of dollars into building a large network of roadways that allow people to drive automobiles almost anywhere. Since the major buildup of roads from the 1950s through the 1990s, it has become significantly more difficult to construct new roadways because of higher population densities and subsequent land-use restrictions. Instead, transportation officials are now investigating intelligent transportation systems and other means to increase the capacity of existing roadways through computer, communications, and control technologies (DOT, 2001). The theory is that by improving overall capacity, congestion on the roadways would be reduced.

Nevertheless, studies have shown that roadway congestion continues to get worse. For example, the Texas Transportation Institute (TTI) conducts an Annual Mobility Study that includes estimates of traffic congestion in many large cities and the impact on society (Schrank and Lomax, 2005). The study defines congestion as "slow speeds caused by heavy traffic and/or narrow roadways due to construction, incidents, or too few lanes for the demand." Because traffic volume has increased faster than road capacity, congestion has gotten progressively worse, despite the push toward alternative modes of transportation, new technologies, innovative land-use patterns, and demand-management techniques.

Some of the major concerns raised by roadway congestion are impacts on energy consumption and air quality. The TTI Annual Mobility Study estimates that billions of gallons of fuel are wasted every year because of congestion (Schrank and Lomax, 2005). In addition, heavy congestion often leads to greater mobile-source emissions. One way to estimate the energy and emissions impacts of congestion is to examine velocity patterns of vehicles operating under different levels of congestion. Roadway congestion is often categorized by the "level of service" (LOS) (TRB, 1994). For freeways (i.e., uninterrupted flow), LOS can be represented as a ratio of traffic flow to roadway capacity. There are several different LOS values that are represented by the letters A through F. For each LOS, a typical vehicle-velocity trajectory will have different characteristics.

Examples of these velocity trajectories are shown in Figure 1 (EPA, 1997). Under LOS A, vehicles typically travel near the highway's free-flow speed, with few acceleration/deceleration perturbations. As LOS conditions get progressively worse (i.e., LOS B, C, D, E, and F), vehicles travel at lower average speeds with more acceleration/deceleration events. For each representative vehicle-velocity trajectory (such as those shown in Figure 1), it is possible to estimate both fuel consumption and pollutant emissions. For automobiles, we are most often concerned about emissions of carbon monoxide (CO), hydrocarbons (HCs), oxides of nitrogen (NOx), and particulate matter.

Figure 2 shows examples of automobile fuel consumption and emission rates that correspond to the average speeds of the representative velocity trajectories shown in Figure 1. Fuel consumption and emission rates are normalized by distance traveled, given in units of grams per unit mile. As expected, when speeds are very low, vehicles do not travel very far; therefore, grams per mile emission rates are quite high. In fact, when a car is not moving, we get an infinite-distance normalized emission rate. Conversely, when vehicles travel at higher speeds, they experience higher engine load requirements and, therefore, have higher fuel consumption and emission rates. As a result, this type of speed-based emission-factor curve has a distinctive parabolic shape, with high emission rates on both ends and a minimum rate at moderate speeds of around 45 to 50 mph.

Figure 2 shows a fuel-consumption and emissions curve for a vehicle (an average "composite" vehicle representing the 2005 vehicle fleet in southern Cali-

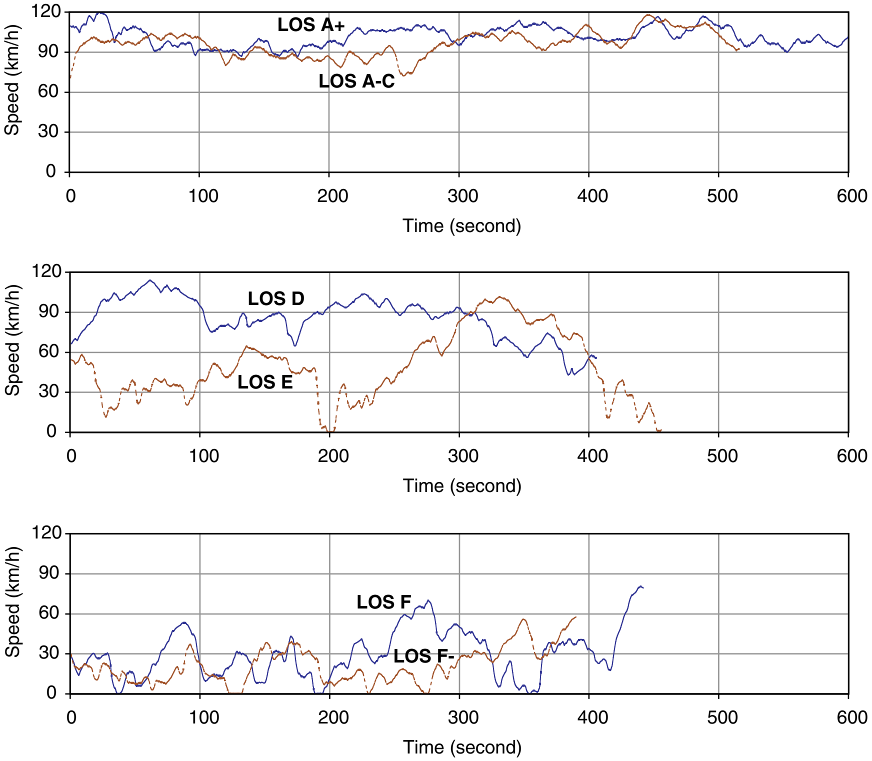


FIGURE 1 Sample vehicle-velocity trajectories for different LOS on a freeway (based on EPA facility-cycle development). Source: EPA, 1997.

fornia) traveling at a perfectly constant, steady-state speed. Of course, vehicles moving in traffic must do some amount of "stop-and-go" driving, and the associated accelerations and decelerations lead to higher fuel consumption and emissions. The constant, steady-state speed line in Figure 2 shows the lower bound of fuel consumption and emissions for any vehicle traveling at that speed.

Several important results can be derived from this information:

- In general, whenever congestion brings the average vehicle speed below 45 mph (for a freeway scenario), there is a negative net impact on fuel consumption and emissions. Vehicles spend more time on the road, which results in lower fuel economy and higher total emissions. Therefore, in this scenario, reducing congestion will improve fuel consumption and reduce emissions.
- If congestion brings average speeds down from a free-flow speed of about

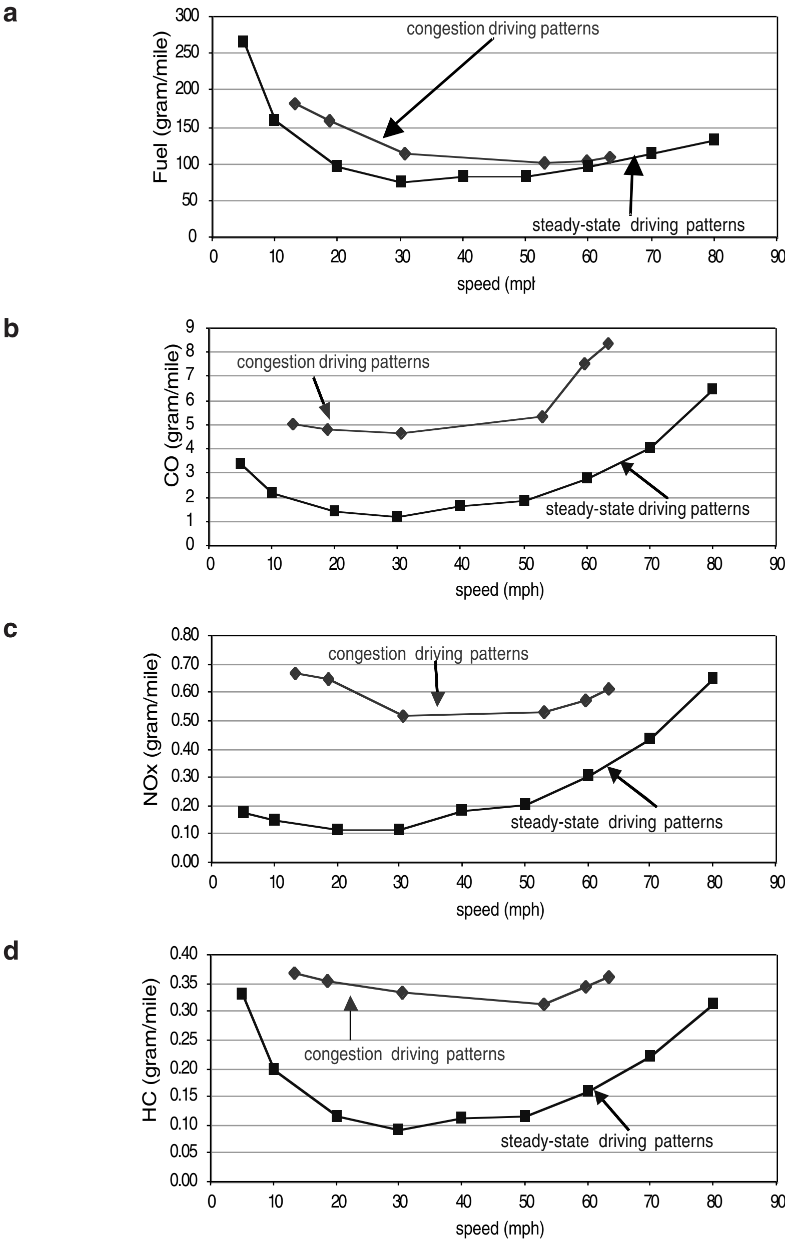


FIGURE 2 Fuel consumption and emissions based on average speed for typical passenger vehicles. a. fuel consumption. b. carbon monoxide. c. hydrocarbons. d. oxides of nitrogen.

65 mph to a slower speed of 45 to 50 mph, congestion can actually *improve* fuel consumption and lower emissions. If relieving congestion increases average traffic speed to the free-flow level, fuel consumption and emissions levels will go up.

- If the real-world, stop-and-go velocity pattern of vehicles could somehow be smoothed out so an average speed could be maintained, significant fuel consumption savings and emissions reductions could be achieved.

A similar analysis can be performed for roadway travel on arterials and residential roads (i.e., interrupted flow patterns). These analyses are a bit more complicated, but they too show that any measure that keeps traffic flowing smoothly for longer periods of time (e.g., operational measures, such as synchronization of traffic signals) improves overall fuel economy and lowers emissions. It is important to note that fuel/emissions congestion effects are much more pronounced for heavy-duty trucks, which tend to have much lower power-to-weight ratios than cars.

Three general areas can be addressed for decreasing congestion and improving mobility and accessibility: supply, demand, and land use. ***Supply management*** techniques include adding resources and capacity to the transportation infrastructure. Examples include building additional roads and adding lanes to existing roads to increase roadway capacity; building bike paths or lanes and walkways to promote these alternative modes of transportation; improving transit facilities and services, as well as intermodal facilities and services, to encourage people to use mass transit more often; improving overall system operations (e.g., responding quickly to roadway incidents); and implementing intelligent transportation system techniques to improve travel efficiency.

***Demand management*** could involve pricing mechanisms to limit the use of resources; providing a much greater range of alternative modes of transportation; encouraging alternative work locations and flexible work schedules; and encouraging or even requiring employers to provide travel-support programs. ***Land-use management*** would require better urban design, mixed-use land development, increased housing and industry density, innovative planning and zoning, and growth management.

It is important that these different areas should be addressed together, rather than separately. If supply alone is increased, this will likely induce additional demand. On the other hand, providing demand management without increasing supply could limit economic growth. Therefore, the best way to approach the problem of traffic congestion is to address all three areas together.

Within these three general areas are several specific programs that can reduce congestion and also help reduce energy consumption and emissions. The following list is not exhaustive, but it provides examples of some things that are being done today and some that could be done in the future.

***Intelligent Speed Adaptation*** (ISA) typically consists of an onboard system that monitors the location and speed of a vehicle, compares it to a defined set speed, and takes corrective action, such as advising the driver and/or limiting the top speed of the vehicle. Researchers in Europe are actively investigating ISA systems and are currently evaluating their effects on safety, congestion, and environmental impacts (Servin et al., 2006).

***Carsharing*** is a new mobility strategy that offers an alternative to individual vehicle ownership by providing a fleet of vehicles that can be shared throughout the day by different users. Carsharing improves overall transportation efficiency by reducing the number of vehicles required to meet total travel demand (e.g., Barth and Shaheen, 2002).

Public transit is seldom used in the United States because it has typically been inflexible and unreliable. However, new *Enhanced Transit Systems*, such as Bus Rapid Transit (BRT) and other systems that provide intermodal linkages of standard transit routes, are becoming available (Levinson et al., 2003).

***Smart Parking*** is a strategy that can lead to significant savings in fuel consumption and reductions in emissions. Smart parking uses advanced technologies to direct drivers to available parking spaces at transit stations (and other locations). This encourages the use of mass transit, reduces driver frustration, and reduces congestion on highways and arterial streets (Rodier et al., 2005).

***Transit-Oriented Developments*** (TODs) promote the use of mass transit by integrating multiple transit options in high-density developments that include residential, commercial, and retail areas. TODs have been demonstrated to increase the use of mass transit and pedestrian traffic and reduce the use of private vehicles (Cervero et al., 2004).

***Innovative Modes of Transportation*** can be used in addition to automobiles to satisfy travel demand. New travel modes can include the Segway™ human transporter, electric bicycles, and neighborhood electric vehicles.

## SUMMARY

Roadway congestion and associated environmental impacts will continue to get worse unless a number of alternatives are introduced. Although a certain amount of congestion can have a positive impact on fuel consumption and vehicle emissions by slowing traffic, severe congestion has the opposite effect. A number of transportation innovations can be implemented to improve overall personal mobility with minimal energy and environmental impacts.

## REFERENCES

Barth, M., and S. Shaheen. 2002. Shared-use vehicle systems: a framework for classifying carsharing, station cars, and combined approaches. Transportation Research Record No. 1791: 105–112.

Cervero, R., R. Arrington, J. Smith-Heimer, and R. Dunphy. 2004. Transit-Oriented Development in the United States: Experiences, Challenges, and Prospects. Washington, D.C.: Transit Cooperative Research Program.

DOT (U.S. Department of Transportation). 2001. The National ITS Architecture: A Framework for Integrated Transportation into the 21st Century. CD book from U.S. Department of Transportation, ITS Joint Program Office. Available online at: *www.itsa.org/public/archdocs/national.html*.

EPA (Environmental Protection Agency). 1997. Development of Speed Correction Cycles. Technical Document #M6.SPD.001, June 1997, prepared by Sierra Research. Washington, D.C.: EPA.

Levinson, H., S. Zinnerman, J. Clinger, S. Rutherford, and R. Smith. 2003. Bus Rapid Transit Systems. Transit Cooperative Research Programs Report #90, Transportation Research Board, The National Academies. Washington, D.C.: Transportation Research Board.

Rodier, C., S. Shaheen, and A. Eaken. 2005. Transit-based smart parking in the San Francisco Bay Area, California: assessment of user demand and behavioral effects. Transportation Research Record 1927: 167–173.

Schrank, D., and T. Lomax. 2005. The 2005 Urban Mobility Report. Texas Transportation Institute, Texas A&M University System. Available online at: *http://mobility/tamu.edu*.

Servin, O., K. Boriboonsomsin, and M. Barth. 2006. An Energy and Emissions Impact Evaluation of Intelligent Speed Adaptation. Pp. 1257–1262 in Proceedings of the 2006 IEEE Intelligent Transportation Systems Conference, Toronto, Canada. Piscataway, N.J.: IEEE.

TRB (Transportation Research Board). 1994. Highway Capacity Manual. Special Report 209. Washington, D.C.: National Academy Press.

# New Mobility: The Next Generation of Sustainable Urban Transportation

Susan Zielinski
*University of Michigan*
*Ann Arbor, Michigan*

In a classic 1950s photograph, a scientific-looking man in a light suit is dwarfed by a mammoth mainframe computer he's programming. It is unlikely that the idea of a "nanopod" would have entered his mind, let alone mesh networking, GIS, or "Googling." He wouldn't have conceived of the connectivity that a mere half-century later has brought these elements together, transformed the world, and evolved into one of the fastest growing, most pervasive global industries.

Today, we are on the cusp of a comparable transformation for cities called New Mobility. Accelerated by the emergence of new fuel and vehicle technologies; new information technologies; flexible and differentiated transportation modes, services, and products; innovative land use and urban design; and new business models, collaborative partnerships are being initiated in a variety of ways to address the growing challenges of urban transportation and to provide a basis for a vital New Mobility industry (MTE and ICF, 2002).

## CONNECTIVITY

An early and very successful example of integrated innovation in New Mobility is the Hong Kong Octopus system, which links multiple transit services, ferries, parking, service stations, access control, and retail outlets and rewards via an affordable, contactless, stored-value smart card. The entire system is de-

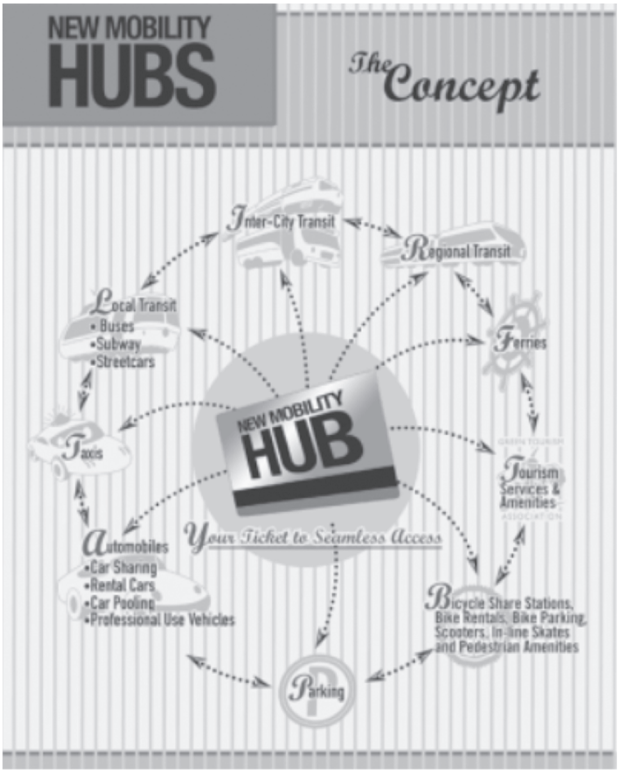


FIGURE 1 The New Mobility hub concept. Source: MTE, 2004.

signed and engineered to support seamless, sustainable door-to-door trips (Octopus, 2006).

A more recent innovation, referred to as New Mobility hub networks, began in Bremen, Germany, and is evolving and spreading to a number of other European cities, as well as to Toronto, Canada (Figure 1). New Mobility hubs connect a variety of sustainable modes of transportation and services through a network of physical locations or "mobile points" throughout a city or region, physically and electronically linking the elements necessary for a seamless, integrated, sustainable door-to-door urban trip (MTE, 2004). Hubs are practical for cities in the developed or developing world because they can be customized to fit local needs, resources, and aspirations. Hubs can link and support a variety of diverse elements:

- multiple transportation operators, modes, and services
- taxis and car-sharing of a variety of vehicle types and sizes
- "slugging" (Slug-Lines.com, 2006)
- free or fee-for-use bicycle sharing (Bikeshare/CBN, 2006)
- walkable, bikable, and transit-oriented spatial design and development (Kelbaugh, 1997)
- cafes and meeting places
- wi-fi amenities
- electronic fare-payment options and pricing mechanisms for all transportation modes and services
- satellite-enhanced, real-time, urban traveler information for all modes of transportation provided at on-street kiosks and by PDA.

## FACTORS DRIVING THE DEVELOPMENT OF NEW MOBILITY

The evolution of New Mobility is inspired by emerging innovations and propelled by pressing needs, not the least of which is rapid urbanization. Although a few cities are shrinking, especially in the developed world, by 2030 more than 60 percent of the world population and more than 80 percent of the North American population will live in urban regions (UN, 1996). With increasing motorization, traffic volume and congestion are already resulting in lost productivity and competitiveness, as well as health and other costs related to smog, poor air quality, traffic accidents, noise, and, more recently, climate change (WBCSD, 2001).

At the same time, sprawling, car-based, urban-development patterns can mean either isolation or chauffeur dependence for rapidly aging populations, as well as for children, youths, and the disabled (AARP, 2005; Hillman and Adams, 1995; O'Brien, 2001; WBCSD, 2001). In developing nations, aspirations toward progress and status often translate into car ownership, even as the risks and costs of securing the energy to fuel these aspirations rise (Gakenheimer, 1999; Sperling and Clausen, 2002; WBCSD, 2001).

## ENGINEERING FOR NEW MOBILITY

The factors described above have created not only compelling challenges for engineering, but also opportunities for social and business innovation. New Mobility solution building is supported by new ways of thinking about sustainable urban transportation, as well as emerging tools and approaches for understanding, implementation, and commercialization. In this article, I focus on three frontiers of thinking and practice for New Mobility: complexity, accessibility, and new business models.

## COMPLEXITY

### Tools for Understanding

A variety of tools and approaches have been developed to support the analysis and modeling of complex urban transportation systems. At least three types of complementary systems analysis (top-down, bottom-up, and simulations) can be applied to transportation and accessibility. Top-down analyses generally start with self-generated variables or hypotheses and develop a causal-loop diagram using software that highlights patterns, dynamics, and possible intervention points. Once a basic analysis is complete, more in-depth data gathering and modeling can be done. Some of the most extensive transportation-related work of this kind has been undertaken by Professor Joseph Sussman at MIT (Dodder et al., 2002; Sussman, 2002; Sussman and Hall, 2004). Figure 2 shows a passenger-transportation subsystem for Mexico City.

Bottom-up, or agent-based, models are computer-based models that use empirical and theoretical data to represent interactions among a range of components, environments, and processes in a system, revealing their influence on the overall behavior of the system (Axelrod and Cohen, 2000; Miller and Roorda, 2006; Miller and Salvini, 2005; Zellner et al., 2003). Ethnographic research can also be applied to transportation as a bottom-up research tool. By giving subjects documentation tools (e.g., cameras) over a fixed period of time, patterns of behavior can be observed without interference by researchers.

Simulations and scenario-building software can draw from and build upon both top-down and bottom-up analyses. Simulations graphically depict and manipulate transportation and other urban dynamics to inform decision making and identify opportunities for innovation. MetroQuest (2006) is a good example of an effective urban-transportation simulation tool.

### Sophisticated Solution Building

Complex transportation challenges call for sophisticated solutions. “Single-fix” approaches (e.g., alternative fuels alone, pricing mechanisms alone, or policy changes alone) cannot address the serious urban challenges and conditions noted above. Informed by complex systems analysis, systems-based solution building involves “connecting the dots,” that is, enhancing or transforming existing conditions with customized, integrated innovations in products, services, technologies, financing, social conditions, marketing, and policies and regulations (ECMT, 2006; MTE and ICF, 2002; Newman and Kenworthy, 1999). Sophisticated solution building usually involves multisector interdisciplinary collaboration.

A good example of systems-based solution building is the New Mobility hub network described above. Hub networks can catalyze engineering and busi-

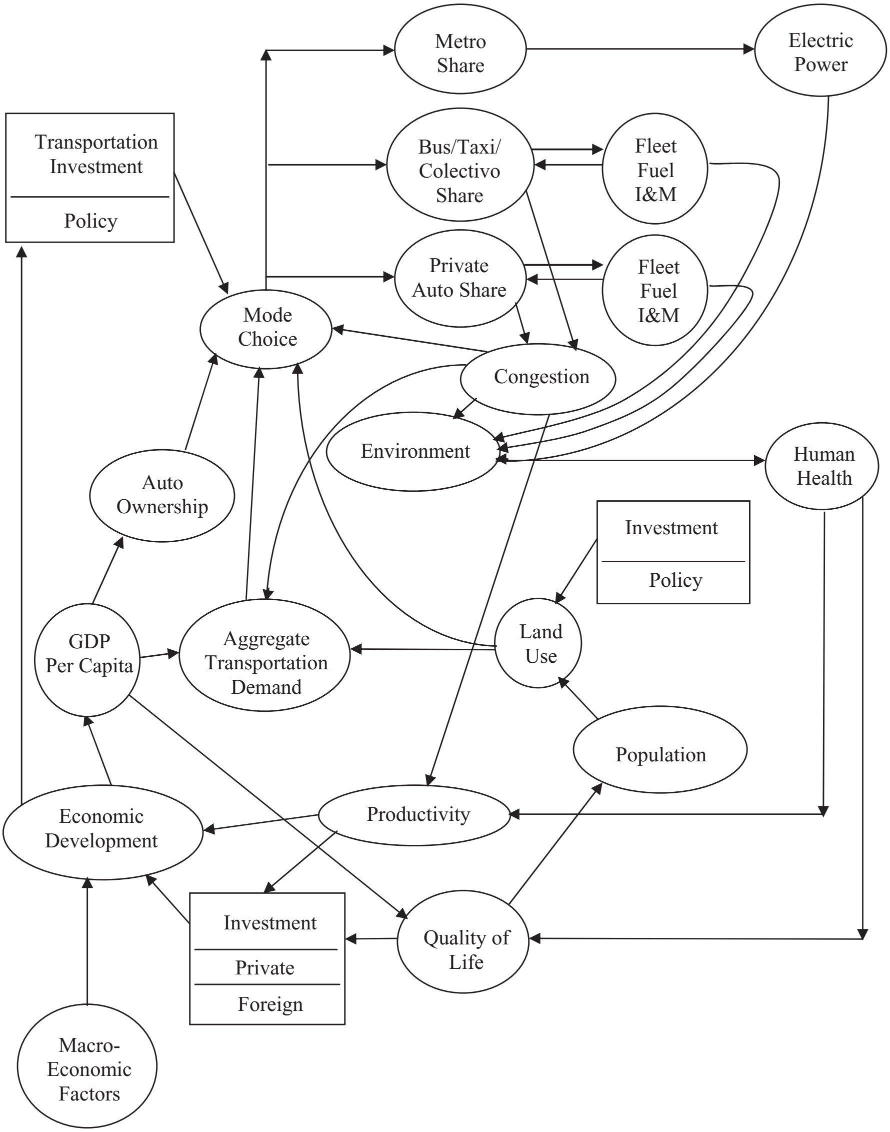


FIGURE 2 Part of a larger analysis showing a passenger-transportation subsystem for Mexico City. Source: Dodder et al., 2002.

ness opportunities related not only to the design and implementation of individual product and service innovations, but also to the engineering of physical and digital connections between them.

## ACCESSIBILITY

Over the past 50 years, measures of regional and economic success have become increasingly linked to (motorized) mobility and speed of travel (TTI, 2005). This association originated in the West and has been widely adopted in cities of the developing world. However, transportation is only a means to an end, or a derived demand, so measures and applications of accessibility do not focus on how fast or how far one can travel in a certain period of time. Instead, they focus on how much can be accomplished in a given time frame and budget or how well needs can be met with available resources. For example, on a typical day in Los Angeles, you may drive long distances at high speeds to fit in three meetings. In Bremen, Germany, a more accessible place, you may be able to fit in five meetings and a leisurely lunch, covering only half the distance at half the speed and for half the price (Levine and Garb, 2002; Thomson, 1977; Zielinski, 1995).

Accessibility can be achieved in at least three ways: wise land use and design, telecommunication technologies that reduce the need for travel, and seamless multimodal transportation. Among other benefits, connected accessibility options can help address the demographic, equity, and affordability needs of seniors, children, the poor, and the disabled. At the same time, integrated accessibility can help build more adaptable and resilient networks to meet the challenges of climate change and emergency situations in cities. Dynamic and flexible accessibility and communications systems can support quick responses to unforeseen urban events.

The University of Michigan's SMART/CARSS project (2006) is currently developing an accessibility index to compare and rate accessibility in metropolitan regions as a basis for urban policy reform and innovation (see Box 1).

## NEW BUSINESS MODELS

In a 2002 study by Moving the Economy, the current value and future potential of New Mobility markets were measured in billions of dollars (MTE and ICF, 2002). New Mobility innovations and opportunities go beyond the sectoral bounds of the traditional transportation industry. They encompass aspects of telecommunications; wireless technologies; geomatics; e-business and new media; tourism and retail; the movement of goods; supply chain management (Zielinski and Miller, 2004); the design of products, services, and technologies; real estate development; financial services; and more.

New Mobility innovations not only improve local competitiveness and qual-

### BOX 1
### University of Michigan SMART/CARSS Project

SMART (Sustainable Mobility and Accessibility Research and Transformation), an interdisciplinary initiative at the University of Michigan in Ann Arbor, is grounded in complexity theory and practice. The goal of the project is to move beyond purely technical and mobility-based approaches to urban transportation to address challenges and opportunities raised by the complex interactions of social, economic, environmental, and policy factors. A project of CARRS (Center for Advancing Research and Solutions for Society), SMART brings together experts on issues, theoretical approaches, and practical and policy applications to tackle the complexity, sophistication, impacts, and opportunities related to urban transportation and accessibility, particularly for growing urban populations worldwide. SMART works collaboratively across disciplines and sectors to:

- catalyze systemic and fundamental transformations of urban mobility/accessibility systems that are consistent with a sustainable human future
- harness emerging science on complex adaptive systems to meet future mobility and accessibility needs in an ecologically and socially sustainable way and identify “tipping points” to guide the evolution of such systems
- inform and develop integrated New Mobility innovation and business models
- provide diverse academic opportunities related to sustainable urban mobility and accessibility
- contribute to a growing multidisciplinary, multistakeholder, global network of applied learning in sustainable mobility and accessibility.

ity of life (Litman and Laube, 2002; Newman and Kenworthy, 1999), they also provide promising export and economic development opportunities for both mature and “base-of-the-pyramid” markets (Hart, 2005; Prahalad, 2004). Because urban transportation represents an increasingly urgent challenge worldwide, and because urban mobility and accessibility solutions can, in most cases, be adapted and transferred, regions, nations, and enterprises that support New Mobility (supply-side) innovation, as well as industry clustering and the development of new business models, stand to gain significantly from transportation export markets in the coming years (MTE and ICF, 2002).

## ENGINEERING AND BEYOND

New Mobility has the potential to revitalize cities and economies worldwide and can open up a wealth of engineering and business opportunities. But obstacles will have to be overcome, not all of them related to engineering. For example, increased motorization and the high social status it represents in developing countries, along with seemingly unstoppable urban sprawl in the West, are challenges that must be addressed on psychological and cultural levels, as well as infrastructural and economic levels. Progress toward a positive, integrated, and sustainable future for urban transportation will require more than moving people and goods. It will also involve the complex task of moving hearts and minds.

## ACKNOWLEDGMENTS

Thomas Gladwin and Jonathan Levine, University of Michigan, and Moira Zellner, University of Illinois, Chicago (all members of SMART/CARSS), made helpful contributions to this paper. Background research was provided by Sathyanarayanan Jayagopi, a student in the master's program, University of Michigan Institute for Global Sustainable Enterprise.

## REFERENCES

AARP (Association for the Advancement of Retired People). 2005. Universal Village: Livable Communities in the 21st Century. Available online at: *http://www.aarp.org/globalaging*.

Axelrod, R., and R. Cohen. 2000. Harnessing Complexity: Organizational Implications of a Scientific Frontier. New York: Basic Books.

Bikeshare/CBN (Community Bicycle Network). 2006. Available online at: *http://communitybicyclenetwork.org/index.php?q=bikeshare*.

Dodder, R., J. Sussman, and J. McConnell. 2002. The Concept of CLIOS Analysis: Illustrated by the Mexico City Case. Working Paper Series. Cambridge, Mass.: Engineering Systems Division, MIT. Available online at: *http://www.google.com/search?hl=en&q=sussman+%26+ Dodder+The+Concept+of+CLIOS+analysis&btnG=Google+Search*.

ECMT (European Conference of Transport Ministers). 2006. Implementing Sustainable Urban Travel Policies: Applying the 2001 Key Messages. Council of Ministers of Transport, Dublin, May 17–18. Available online at: *http://www.cemt.org/council/2006/cm200603fe.pdf*.

Gakenheimer, R. 1999. Urban mobility in the developing world. Transportation Research Part A (33): 671–689.

Hart, S. 2005. Capitalism at the Crossroads: The Unlimited Business Opportunities in Solving the World's Most Difficult Problems. Philadelphia, Pa.: Wharton School Publishing.

Hillman, M., and J. Adams. 1995. Children's Freedom and Safety. Pp. 141–151 in Beyond the Car: Essays in Auto Culture, edited by S. Zielinski and G. Laird. Toronto: Steel Rail Publishing.

Kelbaugh, D. 1997. Common Place: Toward Neighbourhood and Regional Design. Seattle: University of Washington Press.

Levine, J., and Y. Garb. 2002. Congestion pricing's conditional promise: promotion of accessibility or mobility. Transportation Policy 9(3): 179–188.

Litman, T., and Laube, F. 2002. Automobile Dependency and Economic Development. Available online at: *http://www.vtpi.org/ecodev.pdf*.

MetroQuest. 2006. Available online at: *http://www.envisiontools.com*.

Miller, E.J., and M.J. Roorda. 2006. Prototype Model of Household Activity Travel Scheduling. Transportation Research Record 1831, Paper 03.3272. Washington, D.C.: Transportation Research Board of the National Academies.

Miller, E.J., and P. Salvini. 2005. ILUTE: An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems. Pp. 217–234 in Networks and Spatial Economics 5. The Netherlands: Springer Science and Business Media, Inc.

MTE (Moving the Economy). 2004. Bremen and Toronto New Mobility Hub Case Studies and Day in the Life Scenario. Available online at: *http://www.movingtheeconomy.ca/content/csPDF/BremenVideoSummaryAug2.pdf* and *http://www.movingtheeconomy.ca/content/mte_hubAbout.html* and *http://www.movingtheeconomy.ca/content/ditl.html*.

MTE and ICF (ICF Consulting). 2002. Building a New Mobility Industry Cluster in the Toronto Region. Available online at: *http://www.movingtheeconomy.ca*.

Newman, P., and P. Kenworthy. 1999. Sustainability and Cities. Washington, D.C.: Island Press.

O'Brien, C. 2001. Trips to School: Children's Experiences and Aspirations. York Centre for Applied Sustainability. Available online at: *http://plasma.ycas.yorku.ca/documents/ontario_walkability_study_rep.pdf*.

Octopus. 2006. Available online at: *http://lnweb18.worldbank.org/External/lac/lac.nsf/Sectors/Transport/D5A576A039A802C0852568B2007988AD?OpenDocument* and *http://en.wikipedia.org/wiki/Octopus_card*.

Prahalad, C.K. 2004. Fortune at the Bottom of the Pyramid. Philadelphia, Pa.: Wharton School Publishing.

Slug-Lines.com. 2006. Available online at: *http: //www.slug-lines.com/slugging/About-slugging.asp*.

SMART/CARSS. 2006. Available online at: *http://www.isr.umich.edu/carss*.

Sperling, D., and E. Clausen. 2002. The Developing World's Motorization Challenge. Available online at: *http://www.issues.org/19.1/sperling.htm*.

Sussman, J.M. 2002. Collected Views on Complexity in Systems. Pp. 1–25 in Proceedings of the Engineering Systems Division Internal Symposium. Cambridge, Mass.: Engineering Systems Division, MIT.

Sussman, J.M., and R.P. Hall. 2004. Sustainable Transportation: A Strategy for Systems Change. Working Paper Series. Cambridge, Mass.: Engineering Systems Division, MIT.

Thomson, J.M. 1977. Great Cities and Their Traffic. London: Peregrine.

TTI (Texas Transportation Institute). 2005. Urban Mobility Report: 2005. Available online at: *http://tti.tamu.edu/documents/mobility_report_2005_wappx.pdf*.

UN (United Nations). 1996. Urban and Rural Areas. Department of Economic and Social Affairs, Population Division. Available online at: *http://www.un.org/esa/population/pubsarchive/ura/uracht1.htm*.

WBCSD (World Business Council on Sustainable Development). 2001. Mobility 2001: World Mobility at the End of the Twentieth Century and Its Sustainability. Available online at: *http://www.wbcsd.org/web/projects/mobility/english_full_report.pdf*.

Zellner, M., R. Riolo, W. Rand, S.E. Page, D.G. Brown, and L.E. Fernandez. 2003. Interaction Between Zoning Regulations and Residential Preferences as a Driver of Urban Form. Available online at: *http://www.caup.umich.edu/acadpgm/urp/utesymposium/publication/zellner.pdf*.

Zielinski, S. 1995. Access over Excess. Pp. 131–155 in Change of Plans, edited by M. Eichler. Toronto: Garamond Press.

Zielinski, S., and G. Miller. 2004. Integration Technologies for Sustainable Urban Goods Movement. Moving the Economy and Canadian Urban Institute. Available online at: *http://www.tc.gc.ca/pol/en/Report/UrbanGoods/Report.htm*.

## BIBLIOGRAPHY

Burwell, D., and T. Litman. 2003. Issues in Sustainable Transportation. Available online at: *http://vtpi.org/sus_iss.pdf*.

Jacobs, J. 1985. Cities and the Wealth of Nations: Principles of Economic Life. New York: Random House.

Kennedy, C., E.J. Miller, A. Shalaby, H. Maclean, and J. Coleman. 2005. The four pillars of sustainable urban transportation. Transport Reviews 25(4): 393–414.

Levine, J. 1998. Rethinking accessibility and jobs-housing balance. Journal of the American Planning Association 64: 133–150.

Levine, J. 2005. Zoned Out: Regulation, Markets, and Choices in Transportation and Metropolitan Land Use. Washington, D.C.: Resources for the Future Press.

Sterman, J. 2000. Business Dynamics: Systems Thinking and Modeling for a Complex World. New York: Irwin/McGraw Hill.

Sussman, J. 2000. Introduction to Transportation Studies. Boston: Artech House.

# SUPPLY CHAIN MANAGEMENT APPLICATIONS WITH ECONOMIC AND PUBLIC IMPACT

# Introduction

Jennifer K. Ryan
*University College Dublin*
*Dublin, Ireland*

Julie L. Swann
*Georgia Institute of Technology*
*Atlanta, Georgia*

A supply chain is a network that includes all facilities, materials, and activities necessary to bring a product or service to the end-user, or consumer. Supply chain management (SCM) is the process of planning, implementing, and controlling the operations of a supply chain, with the goal of satisfying customer requirements as efficiently as possible. For large multinational corporations that manufacture complex products, such as automobiles, electronics, or aircraft, supply chains are highly complex systems, and the management of these systems is a large-scale problem involving many interrelated components, including facility location and network design, production planning and scheduling, inventory control, transportation and vehicle routing, information systems, and so on.

SCM is further complicated because most supply chains operate in highly variable and uncertain environments with facilities or stages in the supply chain that may be independently owned and/or operated. Because of these complexities, SCM relies heavily on methods developed by operations research, such as optimization and stochastic processes, as well as an understanding of engineering, economic, and business processes.

Over the past two decades, effective SCM has become a significant source of competitive advantage for private companies in both manufacturing industries (e.g., Dell Computer) and service industries (e.g., Wal-Mart). Recently, researchers and practitioners have begun to focus on the public impact of SCM, exploring the relationship between SCM and health care, housing policy, the environment, and national security.

The goal of this session is to provide an introduction to the problems and methods of SCM, focusing on the matching of supply and demand in complex, variable, and/or uncertain environments. To illustrate the widespread applicability of SCM, our speakers consider problems in a variety of settings, including manufacturing, the military, security, and public policy. These papers will provide examples of some of the work being done in SCM. However, potential applications of SCM methods could also include many other areas, such as water-resource management, certain nanoenvironments, network design of electrical systems, and so on.

# Supply Chain Applications of Fast Implosion*

Brenda L. Dietrich
*IBM T.J. Watson Research Center*
*Yorktown Heights, New York*

The chapter Supply Chain Applications of Fast Implosion by Brenda L. Dietrich is available only in the printed book.

---

*An earlier version of this paper was published in Dietrich et al. (2005). This version is included here with kind permission of Springer Science and Business Media.

The chapter Supply Chain Applications of Fast Implosion by Brenda L. Dietrich is available only in the printed book.

The chapter Supply Chain Applications of Fast Implosion by Brenda L. Dietrich is available only in the printed book.

The chapter Supply Chain Applications of Fast Implosion by Brenda L. Dietrich is available only in the printed book.

The chapter Supply Chain Applications of Fast Implosion by Brenda L. Dietrich is available only in the printed book.

The chapter Supply Chain Applications of Fast Implosion by Brenda L. Dietrich is available only in the printed book.

The chapter Supply Chain Applications of Fast Implosion by Brenda L. Dietrich is available only in the printed book.

The chapter Supply Chain Applications of Fast Implosion by Brenda L. Dietrich is available only in the printed book.

The chapter Supply Chain Applications of Fast Implosion by Brenda L. Dietrich is available only in the printed book.

# From Factory to Foxhole: Improving the Army's Supply Chain

Mark Y.D. Wang
*RAND Corporation*
*Santa Monica, California*

At first glance, the end-to-end supply chain by which repair parts are procured and moved to support U.S. Army troops worldwide looks similar to commercial supply chains. Both have suppliers, wholesale distribution centers, retail suppliers, customers, and transportation carriers that move parts from point to point. The main differences between military and commercial supply chains relate, not surprisingly, to the challenges the military faces and the way those challenges must be met.

In 1999, a team of RAND analysts was awarded Al Gore's Hammer Award for support of the Army's Velocity Management Initiative, which dramatically improved ordering and shipping times (OST) for repair parts. Current efforts are now focused further upstream in the supply chain to improve the Army's purchasing and supply management (PSM), integrate supplier management to increase stock availability, and lower total cost. As shown in Figure 1, these initiatives span the entire Army supply chain from factory to foxhole.

## MILITARY VERSUS TRADITIONAL SUPPLY CHAINS

Traditional commercial supply chains focus on physical efficiency, with the emphasis on operating at the lowest possible cost, minimizing investment in inventory, and maximizing capacity utilization. Supply chains that support just-in-time manufacturing (e.g., the Toyota production system) smooth the flow of material from supplier to manufacturing line (Liker, 2003). Management of

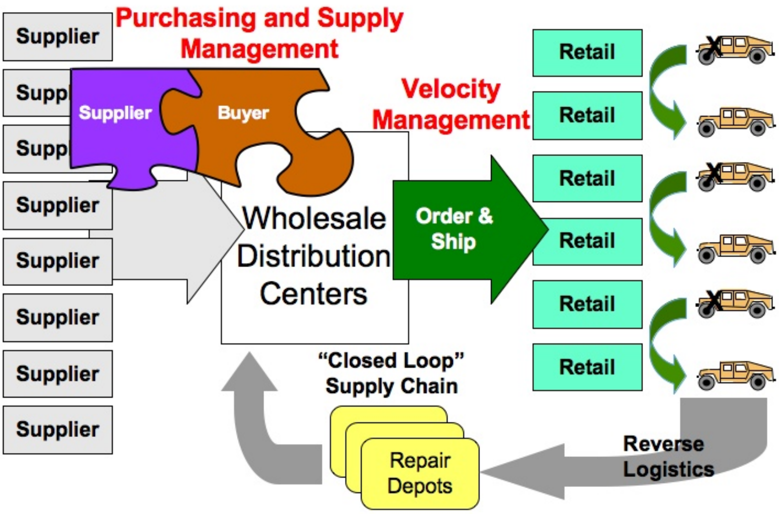


FIGURE 1 The Army's factory-to-foxhole supply chain and how velocity management and purchasing and supply management can improve support for warfighters. Source: Wang, 2000.

physically efficient supply chains may include active management of demand, (e.g., "everyday low prices") to minimize surges and spikes and address inaccuracies in forecasting.

In contrast, military supply chains focus on responsiveness and surge capabilities. The Army must be able to deploy quickly anywhere in the world, and its supply chain must be able to adapt and respond to unpredictable demands and rapidly changing environments. In preparation for Operation Iraqi Freedom, the equivalent of more than "150 Wal-Mart superstores" was moved to Kuwait to support 250,000 soldiers, sailors, airmen, and marines (Walden, 2003).

The nature of commodities, functional or innovative, dictates whether supply chains must be physically efficient or demand responsive (Fischer, 1997). Thus, industries that produce innovative products with very uncertain forecasts (e.g., high-tech, high-fashion, or even toy/entertainment industries) rely on demand-responsive supply chains (Sheffi, 2005). The nature of the military *mission* requires a demand-responsive supply chain. In addition, the characteristics of Army repair parts add to the challenge.

Repair parts are not only highly specialized and weapon-system-specific, but are also often produced by sole-source suppliers who have no commercial market to fall back on. Many parts, such as engines and transmissions, are "reparable" and must be overhauled as a source of future supply (Diener, 2004). Thus, the military not only has to manage a forward logistics pipeline, but must also manage an equally big reverse logistics, or "retrograde," pipeline in a "closed-loop" supply chain (Blumberg, 2004). For every engine, transmission, or rotor blade replaced in the field, a carcass must be moved back to an Army repair depot or national maintenance program location. When you take into account the commodity characteristics and a mission that must respond to unpredictable surges and spikes in demand, the differences between the Army's supply chain and the supply chains of commercial companies become readily apparent.

## VELOCITY MANAGEMENT TO SPEED UP FLOW

The purpose of the Army's Velocity Management Initiative, begun in 1995, was to improve the responsiveness, reliability, and efficiency of a logistics system based on massive stockpiles of supplies and weapon systems, many of them prepositioned "just in case" (Dumond et al., 2001). Although this was a world-class system for supporting a Cold War army, it has become increasingly less effective and unaffordable for the current force-projection army.

To measure the Army's logistics performance, the velocity-management team developed a percentile bar-chart presentation of OST that takes into account not only times for peak distribution, but also times for the tail end of the distribution. Figure 2 shows the time distribution of OST for moving in-stock

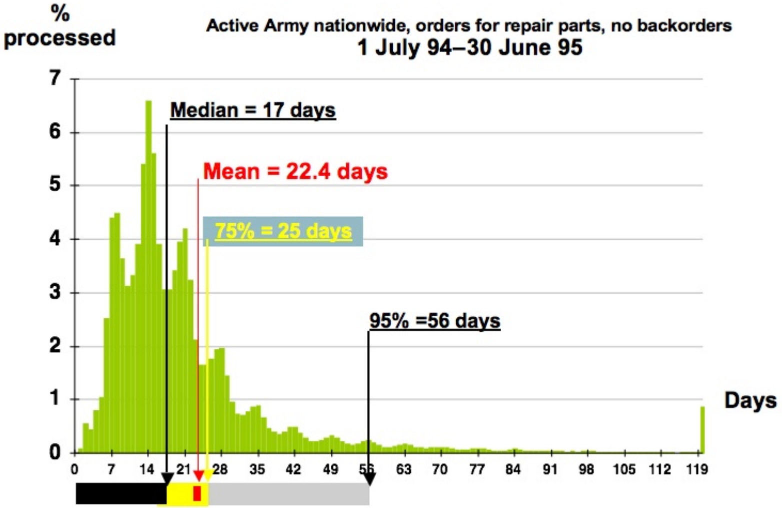


FIGURE 2 In 1994–1995, lengthy OST times were combined with long, variable distribution times. Source: Wang, 2000.

materiel from wholesale defense distribution centers to the Army's retail supply locations. The horizontal axis shows OST measured in days, and the vertical axis shows the percentage of total requisitions. On the lower horizontal bar, the black region represents the time it took to receive half the requisitions for repair parts (17 days during the baseline period). The light (intermediate) and gray (final) regions show the time it took to receive 75 percent and 95 percent of the requisitions, respectively. The square marker shows the mean time (22.4 days during the baseline period). As this figure shows, the difference between the average time and the 95th percentile varied by a factor of two or three. Thus, soldiers waiting for repair parts could not plan repair schedules or maintain the combat readiness of their weapons systems. They simply had to wait, frustrated customers of an unreliable and unresponsive distribution system.

The velocity-management team used a define-measure-improve methodology to "walk the process," following the flow of requisitions and materiel. An excellent example of a win-win solution was the optimization of truck deliveries, which was accomplished by replacing a mix of delivery modes with a reliable, high-volume, high-performing distribution system based on scheduled deliveries. The Army now has premium-level service that is faster, better, and cheaper. Other improvements include better coordinated requisition processing and financial reviews, the adoption of simple rules to "clear the floor" daily, and the establishment of a high-level governance structure to measure performance and ensure continuous improvement.

As Figure 3 shows, through velocity management, the Army dramatically

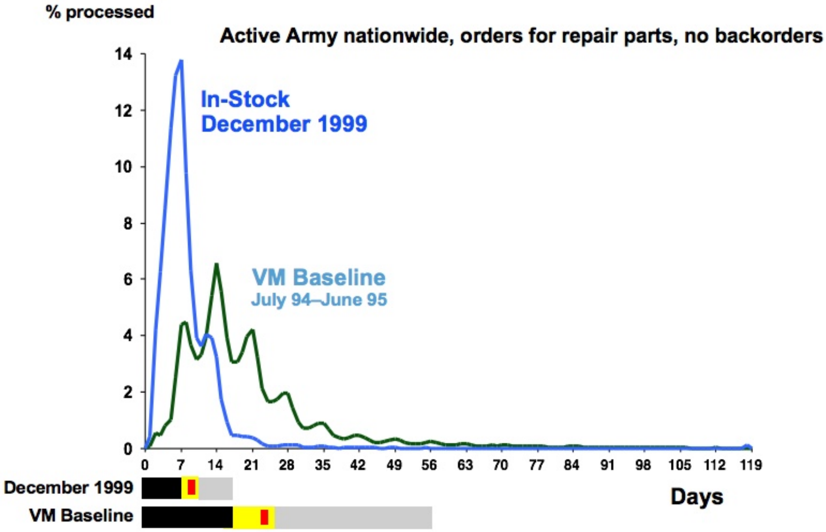


FIGURE 3 Army OST dropped dramatically during the implementation of velocity management. Source: Wang, 2000.

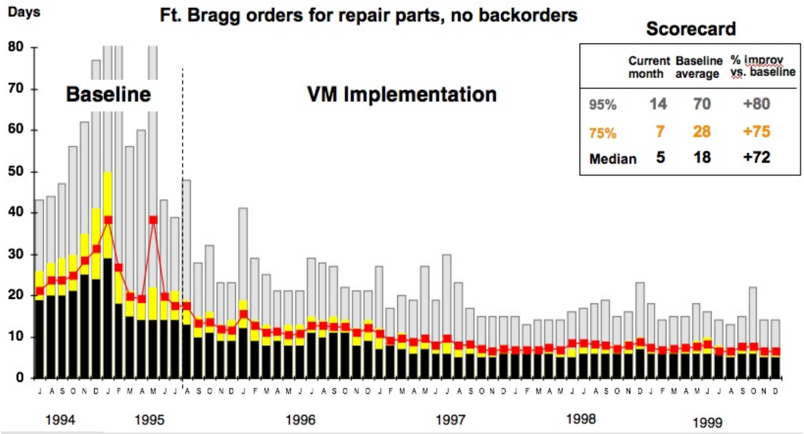


FIGURE 4 Improvements in OST have been most dramatic at major FORSCOM installations, such as Fort Bragg. Source: Wang, 2000.

streamlined its supply process, cutting OST for repair parts by nearly two-thirds nationwide (Wang, 2000). The greatest improvements, which cut OST by more than 75 percent, were achieved at the major forces command (FORSCOM) installations and other installations in the active Army (Figure 4). Today, Army customers nationwide and worldwide routinely receive the same quick, dependable service expected from a high-performing commercial supply chain.

## IMPROVING PURCHASING AND SUPPLY MANAGEMENT

Distribution improvements achieved through velocity management were focused on *moving* in-stock parts. More recent efforts have been focused on improving the Army's PSM processes to ensure that parts are *kept* in stock. During Operation Iraqi Freedom, when both the operating tempo and demand for repair parts were consistently high, requisition backorders of Army-managed items at the national wholesale level skyrocketed, reaching 35 percent for the active Army (Peltz et al., 2005). Backorder rates are a key performance metric because they indicate longer customer waiting times for parts, longer repair-cycle times, and, ultimately, adverse impacts on the availability of weapons systems and unit readiness (Folkeson and Brauner, 2005).

Many factors were contributors to the Army's stock-availability challenges. Besides the contingency surge, they included financial delays and the underfunding of war-reserve inventory prior to the war. The implementation of best PSM practices, such as reducing lead times and total costs, could greatly improve future supply performance. In the commercial world, there has been a paradigm shift from managing items and contracts to managing suppliers and supplier capacity. This has greatly reduced the "bullwhip effect"—wide variations in demand caused by a lack of coordination and information that cascade back through a supply chain (Lee et al., 1997). Best PSM practices call for collaborative planning, forecasting, and replenishment by buyers and suppliers, which leads to better supplier management and more integrated supplier relationships. As the Army's supply chain becomes more responsive to demand, it continues to move toward these PSM goals.

RAND is currently performing high-level analyses of the Army's spending for goods and services, more than $300 billion in FY05, to identify opportunities for improving purchasing (e.g., aggregating requirements when there are many contracts or many suppliers for the same commodity). Another important step toward rationalizing the Army's supply base will be the development of improved supply strategies. As long-term agreements are made with the best suppliers, overall supplier performance will improve, and the Army and suppliers can work together to integrate business processes. Army Materiel Command, the headquarters organization responsible for PSM, is planning to conduct several pilot tests of PSM principles in the coming year.

## SUMMARY

The Army's supply chain faces unique challenges because it must operate in and provide support for highly unpredictable contingencies. As a result, it must be demand-responsive, that is, able to surge and adapt as conditions and demand change. Dramatic reductions in the Army's OST have accelerated flow and streamlined the Army's supply chain. The current challenge is to leverage the distribution improvements achieved by velocity management with higher and more robust wholesale stock availability. Efforts are under way to improve the Army's PSM by adopting best practices in commercial PSM to improve the management of suppliers and supplier capacity.

## REFERENCES

Blumberg, D. 2004. Introduction to Reverse Logistics and Closed-Loop Supply Chain Processes. Boca Raton, Fla.: CRC.

Diener, D. 2004. Value Recovery from the Reverse Logistics Pipeline. RAND MG-238-A. Santa Monica, Calif.: RAND Corporation.

Dumond, J., M.K. Brauner, R. Eden, J. Folkeson, and K. Girardini. 2001. Velocity Management: The Business Paradigm That Has Transformed U.S. Army Logistics. RAND MR-1108-A. Santa Monica, Calif.: RAND Corporation.

Fischer, M. 1997. What is the right supply chain for your product? Harvard Business Review 75(2): 105–116.

Folkeson, J., and M.K. Brauner. 2005. Improving the Army's Management of Reparable Spare Parts. RAND MG-205-A. Santa Monica, Calif.: RAND Corporation.

Lee, H., V. Padmanabhan, and S. Whang. 1997. The bullwhip effect in supply chains. Sloan Management Review 38: 93–102.

Liker, J. 2003. The Toyota Way: Fourteen Management Principles from the World's Greatest Manufacturer. New York: McGraw-Hill.

Peltz, E., H.J. Halliday, M.L. Robins, and K.J. Girardini. 2005. Sustainment of Army Forces in Operation Iraqi Freedom: Major Findings and Recommendations. MG-342-A. Santa Monica, Calif.: RAND Corporation.

Sheffi, Y. 2005. The Resilient Enterprise: Overcoming Vulnerability for Competitive Advantage. Cambridge, Mass.: MIT Press.

Walden, J. 2003. The Forklifts Have Nothing to Do: Lessons in Supply Chain Leadership. Lincoln, Nebr.: iUniverse.

Wang, M. 2000. Accelerated Logistics: Streamlining the Army's Supply Chain. RAND MR-1140-A. Santa Monica, Calif.: RAND Corporation.

# Managing Disruptions to Supply Chains

LAWRENCE V. SNYDER
*Lehigh University*
*Bethlehem, Pennsylvania*

ZUO-JUN MAX SHEN
*University of California, Berkeley*

For as long as there have been supply chains, there have been disruptions, and no supply chain, logistics system, or infrastructure network is immune to them. Nevertheless, supply chain disruptions have only recently begun to receive significant attention from practitioners and researchers. One reason for this growing interest is the spate of recent high-profile disruptions, such as 9/11, the West Coast port lockout of 2002, and hurricanes Katrina and Rita in 2005.

Another reason is the focus in recent decades on the philosophy of “lean” supply chains, which calls for slimmed-down systems with little redundancy or slack. Although lean supply chains are efficient when the environment behaves as predicted, they are extremely fragile, and disruptions can leave them virtually paralyzed. Evidently, there is some value to having slack in a system.

A third reason for the growing attention paid to disruptions is that firms are much less vertically integrated than they were in the past, and their supply chains are increasingly global. A few decades ago, many firms manufactured products virtually from scratch. For example, IBM used to talk, only slightly hyperbolically, about sand and steel entering one end of the factory and computers exiting the other. In contrast, today’s firms tend to assemble final products from increasingly complex components procured from suppliers rather than produced in-house. These suppliers are located throughout the globe, many in regions that are unstable politically or economically or subject to wars and natural disasters. In his recent book *End of the Line*, Barry Lynn (2006) argues that this globalization has led to extremely fragile supply chains.

Supply chain disruptions can have significant physical costs (e.g., damage to facilities, inventory, electronic networks) and subsequent losses due to downtime. A recent study (Kembel, 2000) estimates the cost of downtime (in terms of lost revenue) for several online industries that cannot function if their computers are down. For example, the cost of one hour of downtime for eBay is estimated at $225,000, for Amazon.com, $180,000, and for brokerage companies, $6,450,000. Note that these numbers do not include the cost of paying employees who cannot work because of an outage (Patterson, 2002) or the cost of losing customers' goodwill. Moreover, a company that experiences a supply chain disruption can expect to face significant declines in sales growth, stock returns, and shareholder wealth for two years or more following the incident (Hendricks and Singhal, 2003, 2005a, 2005b).

The huge costs of disruptions show that business continuity is vital to business success, and many companies are actively pursuing strategies to ensure operational continuity and quick recovery from disruptions. For example, Wal-Mart operates an Emergency Operations Center that responds to a variety of events, including hurricanes, earthquakes, and violent criminal attacks. This facility receives a call from at least one store with a crisis virtually every day (Leonard, 2005). Other firms have outsourced their business continuity and recovery operations. IBM and SunGard, the two main players in this field, provide secure data, systems, networks, and support to keep businesses running smoothly during and after disruptions.

Supply chains are multilocation entities, and disruptions are almost never local—they tend to cascade through the system, with upstream disruptions causing downstream "stockouts." In 1998, for example, strikes at two General Motors parts plants led to shutdowns of more than 100 other parts plants, which caused closures of 26 assembly plants and led to vacant dealer lots for months (Brack, 1998). Another, scarier, example relates to port security (Finnegan 2006):

> National-security analysts estimate that if a terrorist attack closed New York Harbor in winter, New England and upstate New York would run out of heating fuel within ten days. Even temporarily hampering the port's operations would have immeasurable cascading effects.

Nevertheless, very little research has been done on disruptions in multi-location systems. Current research is focused mostly on single-location systems and the local effects of disruptions. The research discussed below is a step toward filling this gap.

## UNDERLYING CONCEPTS

Supply uncertainty (SU) and demand uncertainty (DU) have several similarities. In both cases, the problem boils down to having too little supply to meet demand, and it may be irrelevant whether the mismatch occurs because of too much demand or too little supply. Moreover, firms have used similar strategies—holding extra inventory, using multiple suppliers, or improving their forecasts—to protect against both SU and DU.

These similarities offer both good and bad news. The good news is that supply chains under DU have been studied for decades, and we know a lot about them. The bad news is that much of the conventional wisdom about DU is exactly wrong for SU. Thus, we need research on supply chains under SU to determine how they behave and to develop strategies for coping with disruptions in supply.

## RELATED LITERATURE

In the early 1990s, researchers began to embed supply disruptions into classical inventory models, assuming that a firm's supplier might experience a disruption when the firm wished to place an order. (See Nahmias [2005] for an introduction to inventory theory and Zipkin [2000] for a more advanced treatment.) Examples include models based on the economic order quantity model (Berk and Arreola-Risa, 1994; Parlar and Berkin, 1991), the $(R,Q)$ model (Gupta, 1996; Parlar, 1997), and the $(s,S)$ model (Arreola-Risa and DeCroix, 1998). All of these models are generally less tractable than their reliable supply counterparts, although they can still be solved easily using relatively simple algorithms.

More recent literature has addressed higher level, strategic decisions made by firms in the face of disruptions. For example, Tomlin (2006) explores strategies for coping with disruptions, including inventory, dual sourcing, and acceptance (i.e., simply accepting the risk of disruption and not protecting against it), and shows that the optimal strategy changes as the disruption characteristics change (e.g., disruptions become longer or more frequent). Tomlin and Snyder (2006) examine how strategies change when a firm has advance warning of an impending disruption. Lewis, Erera, and White (2005) consider the effects of border closures on lead times and costs. Chopra, Reinhardt, and Mohan (2005) evaluate the error that results from "bundling" disruptions and yield uncertainty (another form of SU) when making inventory decisions.

Only a very small body of literature is focused on disruptions in multilocation supply chains. Hopp and Yin (2006) investigated optimal locations for capacity and inventory buffers in a multilocation supply chain and concluded that as potential disruptions become more severe, buffer points should be located closer to the source of disruptions. Kim, Lu, and Kvam (2005) evaluated the effects of yield uncertainty in a three-tier supply chain. They addressed the consequences of the decision maker's risk aversion, an important factor when modeling infrequent but high-impact events.

A growing literature addresses disruptions in the context of facility location. Here, the objective is to choose locations for warehouses and other facilities that minimize transportation costs to customers and, at the same time, account for possible closures of facilities that would necessitate rerouting of the product. Although these are multilocation models, they focus primarily on the local effects of disruptions (see Snyder et al. [2006] for a review). We discuss these models in greater detail below.

## SUPPLY UNCERTAINTY VS. DEMAND UNCERTAINTY

In the sections that follow, we discuss the differences between SU and DU in multi-echelon supply chains. (An *echelon* is a "tier" of a supply chain, such as a factory, warehouse, retailer, etc.) We consider several studies, each of which examines two possible answers to a question of supply chain design or management. Each study demonstrates that one answer is optimal for SU while the opposite answer is optimal for DU. Some of these results may be proven theoretically. Others are demonstrated using simulation by Snyder and Shen (2006).[1]

### Centralization vs. Decentralization

Consider a system with one warehouse that serves *N* retailers (Figure 1). Under DU, it is well known that if the holding costs are equal at the two echelons and transportation times are negligible, then the optimal strategy is to hold inventory at the warehouse (a *centralized* system) rather than at the individual retailers (a *decentralized* system). This is because of the *risk-pooling* effect, which says that the total mean cost is smaller in the centralized system because cost is proportional to the standard deviation of demand. The standard deviation, in turn, is proportional to the square root of *N* in the centralized system but is linear in the decentralized system (Eppen, 1979). Although the assumptions of equal holding costs and negligible lead times are unrealistic, the risk-pooling effect and the insights that arise from it are applied widely in supply chain planning and management.

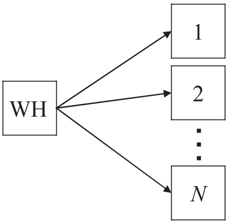


FIGURE 1 One-warehouse, multiretailer system.

---

[1] Although we use terminology suggestive of private-sector supply chains (e.g., "firms" and "retailers"), the results discussed in this paper are also applicable to noncommercial supply networks (e.g., military, health care, and humanitarian networks).

Now consider the same system under SU (with deterministic demand). In this case, if inventory sites are subject to disruptions, it may be preferable to hold inventory at the retailers rather than at the warehouse. Under this decentralized strategy, a disruption would affect only a fraction of the retailers; under a centralized strategy, a disruption would affect the whole supply chain. In fact, the *mean* costs of the two strategies are the same, but the decentralized strategy results in a smaller *variance* of cost. This is referred to as the *risk-diversification* effect, which says that disruptions are equally frequent in either system, but they are less severe in the decentralized system (Snyder and Shen, 2006).

## Inventory Placement

In a serial system (Figure 2), a common question is which stages should hold inventory. Under DU, the tendency is to push inventory as far upstream as possible (to the left in Figure 2), because the cost of holding inventory tends to increase as one moves downstream in a supply chain. Under SU, however, the tendency is reversed. It is preferable to hold inventory downstream, where it can protect against disruptions elsewhere in the supply chain. For example, this might mean that a manufacturing firm should hold inventory of raw materials under DU but of finished goods under SU.

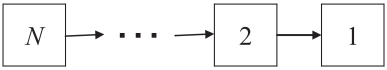


FIGURE 2 Serial system.

## Hub-and-Spoke vs. Point-to-Point Networks

Figure 3 shows two possible networks for a firm with a single factory that wants to distribute its product to multiple retailers. The network in Figure 3a is a

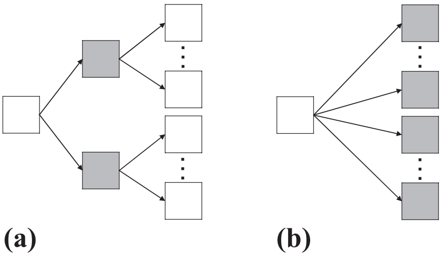


FIGURE 3 a. Hub-and-spoke network. b. Point-to-point network. Sites that hold inventory are shaded.

*hub-and-spoke* network, with intermediate warehouses that hold inventory and distribute it to retailers. The network in Figure 3b is a *point-to-point* network in which the warehouses are bypassed and retailers hold the inventory. Many firms operate hub-and-spoke networks because of economies of scale and other savings from consolidating inventory locations. Even absent economies of scale, however, the hub-and-spoke network is optimal under DU because of the risk-pooling effect (there are fewer inventory-stocking locations, hence a smaller total inventory requirement). Under SU, however, the point-to-point network is preferable because of the risk-diversification effect (increasing the number of stocking locations reduces the severity of disruptions).

A relevant analogy comes from the airline industry. Large U.S. carriers have primarily adopted a hub-and-spoke model because of the economies of scale it offers regarding airport infrastructure and the scheduling of flight connections. However, when a disruption (e.g., a thunderstorm) occurs at a hub, it can affect the carrier's entire domestic flight network. In contrast, smaller carriers have tended to adopt point-to-point networks that allow flight schedules to be somewhat more flexible and reactive.

### Supplier Redundancy

Consider a single firm with a single supplier trying to determine the value of adding backup suppliers. Suppose that each supplier has sufficient capacity to meet the mean demand plus a few standard deviations. Under DU, backup suppliers have little value because they would fill in only when demand exceeds capacity, which happens infrequently. Under SU, however, backup suppliers play a vital role because they provide capacity both to meet demand *during* a disruption to the primary supplier and to ramp up supply *after* a disruption.

### Facility Location

Classical facility-location models choose locations for plants, warehouses, and other facilities to minimize transportation cost or achieve some other measure of proximity to both suppliers and customers (Daskin, 1995; Drezner and Hamacher, 2002), typically ignoring both DU and SU. A recent model finds that under DU the optimal number of facilities decreases because of the risk-pooling effect and economies of scale from consolidation (Shen et al., 2003). Conversely, when facilities face potential disruptions (i.e., under SU), the optimal number of facilities increases because of the risk-diversification effect (Snyder and Daskin, 2005). A model currently under development incorporates both DU and SU, thus balancing these competing tendencies (Jeon et al., 2006).

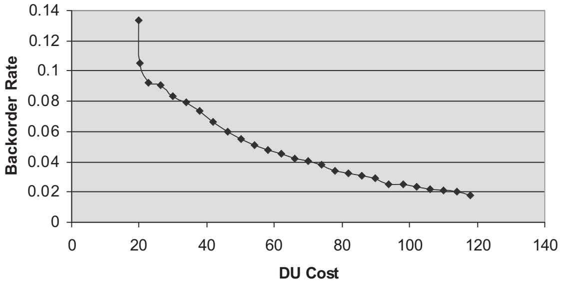


FIGURE 4 Trade-off curve.

## Cost of Reliability

A firm that is used to planning primarily for DU may recognize the importance of planning for SU but may be reluctant to do so if it requires a large upfront investment in inventory or infrastructure. Fortunately, a small amount of extra inventory goes a long way toward protecting against disruptions. Figure 4 shows the trade-offs between the vulnerability of a system to disruptions (on the *y*-axis, measured by the percentage of demands that cannot be met immediately) and the cost under DU (on the *x*-axis, measured in the cost the firm is used to considering).

Each point in Figure 4 represents a possible solution, with the left-most point representing the optimal solution if there are no disruptions. This solution is cheap but very vulnerable to disruptions. The left-hand portion of the curve is steep, suggesting that large improvements in reliability are possible with small increases in cost. For example, the second point shows 21 percent fewer stockouts but is only 2 percent more expensive. This trend is fairly common and has been identified in other contexts, including facility location with disruptions (Snyder and Daskin, 2005).

## CONCLUSIONS

Studies of SU and DU in multi-echelon supply chains show that the two types of uncertainty require different strategies in terms of centralization, inventory placement, and supply chain structure. In fact, the optimal strategy for dealing with SU is, in many cases, the exact opposite of the optimal strategy for DU. However, we are not suggesting that firms are currently doing everything wrong. Rather, we are arguing that although DU leads to certain tendencies in supply chain management (e.g., centralization), SU suggests opposite strategies that should also be considered when making supply chain decisions. Fortunately, it can be relatively inexpensive to shift the balance enough to account for SU, in the sense that the trade-offs between the two types of uncertainty are favorable.

In virtually all practical settings, both DU and SU are present, and the optimal strategy must account for interactions between them. For example, since upstream inventory is cost effective under DU but downstream inventory is most helpful under SU, a firm may wish to adopt a hybrid strategy that combines the advantages of both. For example, many firms hold inventory of both raw materials (upstream) and finished goods (downstream), with raw material inventory accounting for the bulk of the firm's inventory holdings but finished goods inventory acting as a key buffer against uncertainty. Alternately, a hybrid strategy may involve holding inventory near the middle of the supply chain (rather than at both ends). For example, Dell holds inventory of sophisticated components, assembling them into finished goods only after orders are placed.

It is our hope that researchers will continue investigating the causes and effects of supply chain disruptions, as well as strategies for coping with them. One important area for future research is the development of analytical tools for understanding the interdependence of risks faced by a supply chain. A single event (e.g., an economic downturn or a bird-flu pandemic) might cause multiple types of disruptions (e.g., a shortage of raw materials and absenteeism among the firm's own workforce), and these risks may be subtly related. In other words, the supply chain's total risk may not be a simple sum of its parts.

Another promising avenue for future research is to develop strategies for designing resilient supply chains. How can a supply chain's infrastructure be designed so that buffers are located in the right places and in the right quantities to protect against disruptions and other forms of uncertainty? What forms should these buffers take (e.g., inventory, capacity, redundant supply)?

Many of the analytical models for designing and managing supply chains under uncertainty assume that the decision maker has some knowledge of the risk of disruption, for example, the probability that a disruption will occur or the expected duration of a disruption. In practice, these parameters can be very hard to estimate. Therefore, we suggest, as a third area for future research, the development of models that are insensitive to errors in these parameters.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge support from National Science Foundation grants #DMI-0522725 and #DMI-0348209.

## REFERENCES

Arreola-Risa, A., and G.A. DeCroix. 1998. Inventory management under random supply disruptions and partial backorders. Naval Research Logistics 45: 687–703.

Berk, E., and A. Arreola-Risa. 1994. Note on "Future supply uncertainty in EOQ models." Naval Research Logistics 41: 129–132.

Brack, K. 1998. Ripple effect from GM strike build. Industrial Distribution 87(8): 19.

Chopra, S., G. Reinhardt, and U. Mohan. 2005. The Importance of Decoupling Recurrent and Disruption Risks in a Supply Chain. Working paper, Northwestern University, Evanston, Illinois.

Daskin, M.S. 1995. Network and Discrete Location: Models, Algorithms, and Applications. New York: John Wiley and Sons.

Drezner, Z., and H.W. Hamacher, eds. 2002. Facility Location: Applications and Theory. New York: Springer-Verlag.

Eppen, G.D. 1979. Effects of centralization on expected costs in a multi-location newsboy problem. Management Science 25(5): 498–501.

Finnegan, W. 2006. Watching the waterfront. New Yorker, June 19, 2006, pp. 52–63.

Gupta, D. 1996. The $(Q,r)$ inventory system with an unreliable supplier. INFOR 34(2): 59–76.

Hendricks, K.B., and V.R. Singhal. 2003. The effect of supply chain glitches on shareholder wealth. Journal of Operations Management 21(5): 501–522.

Hendricks, K.B., and V.R. Singhal. 2005a. Association between supply chain glitches and operating performance. Management Science 51(5): 695–711.

Hendricks, K.B., and V.R. Singhal. 2005b. An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm. Production and Operations Management 14(1): 35–52.

Hopp, W.J., and Z. Yin. 2006. Protecting Supply Chain Networks Against Catastrophic Failures. Working paper, Northwestern University, Evanston, Illinois.

Jeon, H.-M., L.V. Snyder, and Z.-J.M. Shen. 2006. A Location-Inventory Model with Supply Disruptions. Working paper, Lehigh University, Bethlehem, Pennsylvania.

Kembel, R. 2000. The Fibre Channel Consultant: A Comprehensive Introduction. Tucson, Ariz.: Northwest Learning Associates.

Kim, H., J.-C. Lu, and P.H. Kvam. 2005. Ordering Quantity Decisions Considering Uncertainty in Supply-Chain Logistics Operations. Working paper, Georgia Institute of Technology, Atlanta.

Leonard, D. 2005. “The Only Lifeline Was the Wal-Mart.” Fortune 152(7): 74–80.

Lewis, B.M., A.L. Erera, and C.C. White. 2005. An Inventory Control Model with Possible Border Disruptions. Working paper, Georgia Institute of Technology, Atlanta.

Lynn, B.C. 2006. End of the Line: The Rise and Coming Fall of the Global Corporation. New York: Doubleday.

Nahmias, S. 2005. Production and Operations Analysis. New York: McGraw-Hill/Irwin.

Parlar, M. 1997. Continuous-review inventory problem with random supply interruptions. European Journal of Operational Research 99: 366–385.

Parlar, M., and D. Berkin. 1991. Future supply uncertainty in EOQ models. Naval Research Logistics 38: 107–121.

Patterson, D.A. 2002. A Simple Way to Estimate the Cost of Downtime. Pp. 185–188 in Proceedings of LISA ’02: 16th Systems Administration Conference. Berkeley, Calif.: USENIX Association.

Shen, Z.-J.M., C.R. Coullard, and M.S. Daskin. 2003. A joint location-inventory model. Transportation Science 37(1): 40–55.

Snyder, L.V., and M.S. Daskin. 2005. Reliability models for facility location: the expected failure cost case. Transportation Science 39(3): 400–416.

Snyder, L.V., and Z.-J.M. Shen. 2006. Disruptions in Multi-Echelon Supply Chains: A Simulation Study. Submitted for publication.

Snyder, L.V., M.P. Scaparra, M.S. Daskin, and R.L. Church. 2006. Planning for Disruptions in Supply Chain Networks. Pp. 234–257 in TutORials in Operations Research, edited by H. Greenberg. Baltimore, Md.: INFORMS.

Tomlin, B.T. 2006. On the value of mitigation and contingency strategies for managing supply chain disruption risks. Management Science 52(5): 639–657.

Tomlin, B.T., and L.V. Snyder. 2006. Inventory Management with Advanced Warning of Disruptions. Working paper, Lehigh University, Bethlehem, Pennsylvania.

Zipkin, P.H. 2000. Foundations of Inventory Management. New York: McGraw-Hill/Irwin.

# Engineering Methods for Planning Affordable Housing and Sustainable Communities

MICHAEL P. JOHNSON
*Carnegie Mellon University*
*Pittsburgh, Pennsylvania*

Housing is a key component of the U.S. economy. In 2001, housing comprised more than one-third of the nation's tangible assets, and, in the form of home building and remodeling, housing consumption and related spending represented more than 21 percent of the U.S. gross domestic product. Since 2001, home sales, prices, equity, and debt have all increased substantially, enabling millions of Americans to purchase goods and services (Joint Center for Housing Studies of Harvard University, 2006).

Decent, affordable housing (generally defined as housing that consumes less than 30 percent of a family's income) often enables families to enjoy stability, good health, employment, education, and recreation. Decent, affordable housing also contributes to the physical, economic, environmental, and social health—the *sustainability*—of communities (Millennial Housing Commission, 2002). These impacts are especially important for lower income households and other underserved populations.

Despite the general strength of the U.S. housing market, the benefits of housing and stable, vibrant communities are not distributed equally. Examples of inequalities include: residential segregation, differences in homeownership rates by race, sprawl-type development patterns, and shortages of affordable housing. In the wake of Hurricane Katrina, for example, the challenges of securing basic shelter and rebuilding homes and communities have fallen disproportionately on minority and low-income populations (de Souza Briggs, 2006; Joint Center for Housing Studies of Harvard University, 2006; Millennial Housing Commission, 2002). These and similar circumstances justify social intervention by government and nongovernmental organizations.

The purpose of this paper is to highlight new, creative research in a variety of disciplines—especially decision sciences—that can help determine when, where, what type, and by what means affordable housing and sustainable communities might be built, redeveloped, and maintained. As a prelude to the subject, it is useful to link housing planning and supply chain management, the theme of this Frontiers of Engineering session.

A *supply chain* is a network of facilities and modes of transportation that uses production and logistics processes to transform inputs into finished goods and services, thereby integrating supply and demand management. A central feature of supply chain management is temporal planning—strategic, tactical, operational, and technical (e.g., the location of facilities at which operations are performed). Housing and community development (a social enterprise) are not literally examples of supply chain management. However, facility location—here, the location of housing—is central to both, and the temporal scope of housing and community development planning spans strategic, tactical, and operational time horizons. Finally, effective housing and community development planning, like supply chain management, is an attempt to match supply and demand for goods and services—in this case, affordable shelter and sustainable communities.

Initiatives to make affordable housing and sustainable communities more accessible must address the needs of stakeholders (e.g., employers, housing developers, citizens, government agencies); policy objectives (minimize housing costs and environmental impacts, "deconcentrating" poverty); and actions (the creation of new housing alternatives, protection of current alternatives, changes in attitudes and preferences) (cf. de Souza Briggs, 2005).

Engineering and related disciplines can influence all of these dimensions of housing policy. Civil, environmental, and mechanical engineering, for example, can generate methods of implementing housing initiatives with more efficient and effective construction. Urban and regional planning, especially land-use and transportation planning, in contrast, focus on social efficiency and equitable development outcomes, given current or best-practice construction technologies. Decision sciences (e.g., operations research and management science) represent a link between engineering and planning methods; they generate specific, actionable strategies for optimizing social efficiency, effectiveness, and equity. Decision sciences may take as given current or best practices in construction technologies or planning methods, or both, or neither.

The remainder of this paper is focused on research results in engineering construction methods and urban and regional planning methods related to the development of affordable housing and a discussion of the unique contributions of decision sciences. We also identify a number of promising areas for continued research.

## ENGINEERING-BASED METHODS FOR HOUSING CONSTRUCTION

Traditional engineering is well suited to the efficient development of cost-effective housing. Improvements in construction technologies can result in increased affordability, energy efficiency, and structural integrity and decreased negative environmental impacts. Recent European research addressing "sustainable" development from an engineering perspective, focused mostly on minimizing negative environmental impacts, has shown that, even when construction techniques are modified to decrease the ecological impacts associated with "flows" of energy, construction materials, and water, the resulting innovations are often contradicted by increased resource usage by housing occupants and ineffective national policies (e.g., Priemus, 2005). Ultimately, Priemus argues, the policy with the greatest impact on sustainability may be a policy that discourages, or even decreases, the construction of new housing.

Other engineering approaches have focused on best practices for reducing energy consumption through energy-conserving materials, such as windows, insulation, and appliances; alternative energy sources, such as solar power; improved construction methods for foundations and walls; and more efficient heating and air-conditioning systems (Steven Winter Associates Inc., 2001). Building-design strategies are based on advanced computer simulations comparing energy savings from novel designs with actual outcomes, as well as architectural choices, such as site selection and building orientation for maximum passive solar exposure, and compact floor plans. A specially designed house that incorporated these technologies used 46 percent less energy than the average U.S. house (Balcomb et al., 1999).

These technologies are also available for the rehabilitation of existing housing in low-income areas through retrofitting, improved gas metering, and increased cooperation between stakeholders. Estimated cost savings in energy for a low-income family are on the order of one month's rent per year (Katrakis et al., 1994).

Engineering methods also influence construction processes. Examples include concurrent engineering to help meet customer requirements for industrialized housing (Armacost et al., 1994) and knowledge management to improve coordination between the owners, designers, and developers of affordable housing (Ibrahim and Nissen, 2003).

## URBAN PLANNING FOR AFFORDABLE HOUSING AND COMMUNITY DEVELOPMENT

American planners and analysts have been dealing, with limited success, with the problems of affordable housing and community design for more than 80 years (von Hoffman, 1996). In central cities, planners in the 1930s and 1940s embraced the idea of vertical towers grouped in communities distinct from surrounding neighborhoods. These enclaves often resulted in social dysfunction and physical decay, which have only been remedied in a substantial way in the past decade under the Federal HOPE VI Program. In contrast, post-World War II suburbs were designed to be affordable, accessible to central cities via freeways, and uniform in appearance.

In recent years, dense, transit-friendly, mixed-use developments in central cities or nearby suburbs, often on land previously used for residential or industrial purposes, have converged with the redevelopment of distressed inner-city neighborhoods into mixed-income, joint ventures (Bohl, 2000). Although U.S. consumers still overwhelmingly prefer the traditional suburban model of detached, single-family, owner-occupied housing, market demand is increasing for housing units and communities that appear to be more sustainable socially and environmentally (Myers and Gearin, 2001).

The impact of assisted housing development has been limited in recent years because of stagnant federal funding for subsidized and affordable housing. Planning researchers are turning to decision models and geographic information systems to generate alternative strategies for optimizing social objectives (Ayeni, 1997). However, very little work in this area, or in traditional urban planning, is being done on decision-support models designed specifically for planning affordable housing.

## DECISION-SCIENCE METHODS FOR AFFORDABLE HOUSING POLICY AND PLANNING

Decision models can help planners improve access to affordable housing and sustainable communities by simultaneously, and explicitly, addressing space, opportunity, design, and choice alternatives. *Space* and *opportunity* are factors in decisions about the physical location of housing units and their proximity to community amenities, which are important to improved quality of life. *Design* decisions are important to the development of policies that enable families to participate in housing programs, as well as in establishing development priorities and configuring mixed land-use and mixed-housing communities. *Choice* decisions are essential to individuals choosing housing and neighborhood destinations that best meet their needs and preferences. In contrast to engineering construction and planning methods, decision models for housing development are quantitative, stylized, prescriptive, forward-looking, and multiobjective.

One type of strategic decision we address is choosing and evaluating housing and community development policies. A solution to this problem consists of program *types* (e.g., housing subsidies) and *intensities* (e.g., funding levels or number of program participants). Caulkins et al. (2005) developed a model to predict long-term population outcomes associated with stylized, large-scale programs in which low-income families use housing subsidies to relocate to low-poverty neighborhoods. The purpose of the model is to identify the circumstances under which a large-scale housing program might preserve the health of destination communities. The authors model changes in the stock of middle-class families in a typical region as a result of (1) normal demographic changes, (2) a large-scale housing mobility program resulting in low-income families that "assimilate" to the middle class, and (3) middle-class "flight" in response to in-movers. Figure 1 shows that, for base-case values of structural parameters, equilibrium would be maintained over the long term (near $X = 1$) in a generic metropolitan area with a low-intensity housing-mobility program; in the long term, the size of middle-class communities would decrease only slightly.

Given support, in a strategic sense, for a particular housing policy, a tactical decision must be made about the amount and type(s) of housing to be provided in a specific region over a specific period of time. Addressing this decision requires specifying program *locations* (municipalities, neighborhoods, or land parcels) and *configurations* (different numbers of different-sized rental- or owner-occupied housing units). Gabriel et al. (2006) developed a multiobjective optimization model for identifying land parcels for development that balances the needs of planners, developers, environmentalists, and government.

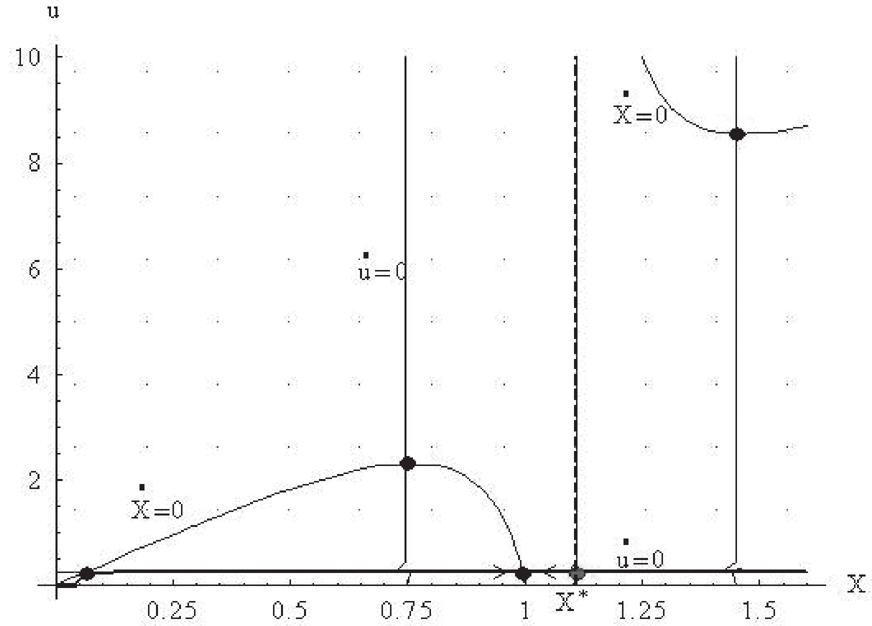


FIGURE 1 Dynamic optimization model solution for a housing mobility program—base-case parameters. Source: Caulkins et al., 2005. Reprinted with permission.

Johnson (2006) solves two complementary optimization models specifically for affordable housing: (1) a longer range model for identifying regional investment levels that maximize social benefits and (2) a shorter range model for identifying specific locations and development sizes that balances social benefits and equity. Figure 2 shows Pareto frontiers associated with solutions to the multiobjective optimization problem for owner-occupied and renter-occupied housing using data for Allegheny County, Pennsylvania. The curves show that a range of policy alternatives can support a "most-preferred" solution.

The last decision problem considered here, operational in scope, is a client's choice of a most-preferred housing program, neighborhood, or housing unit, within defined, affordable, housing-policy priorities. Solving this problem requires specifying detailed characteristics (*attributes*) of housing units and neighborhoods, decision models by which participants can rank potential destinations (*alternatives*), and information systems to help standardize and automate the process (*decision support systems*).

Johnson (2005) developed a prototype spatial decision-support system (SDSS) for tenant-based subsidized housing that addresses qualitative concerns (which attributes of housing units and neighborhoods are important to the client) and quantitative concerns (how a client can rank a "short list" of alternatives to

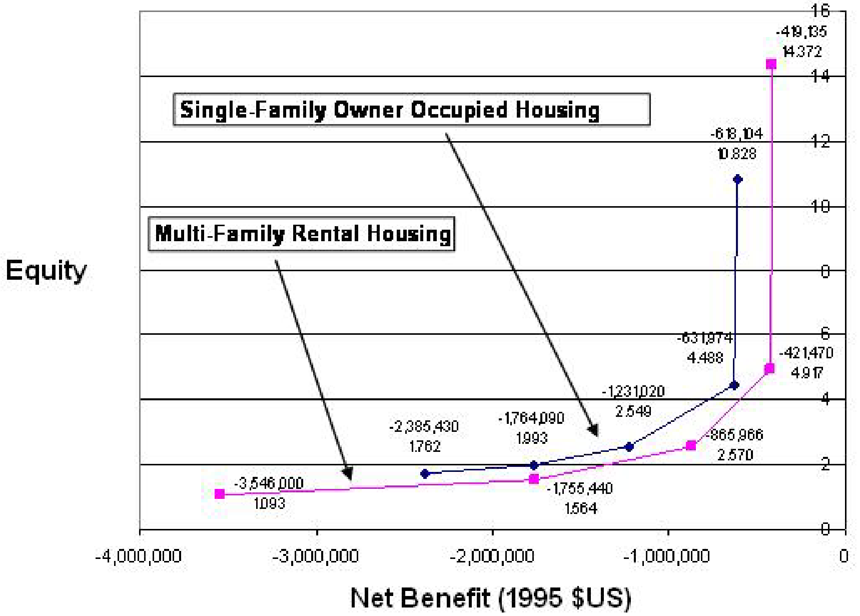


FIGURE 2 Pareto frontiers for a case study of an affordable-housing location problem. Source: Johnson, 2006. Reprinted with permission from Pion Limited, London.

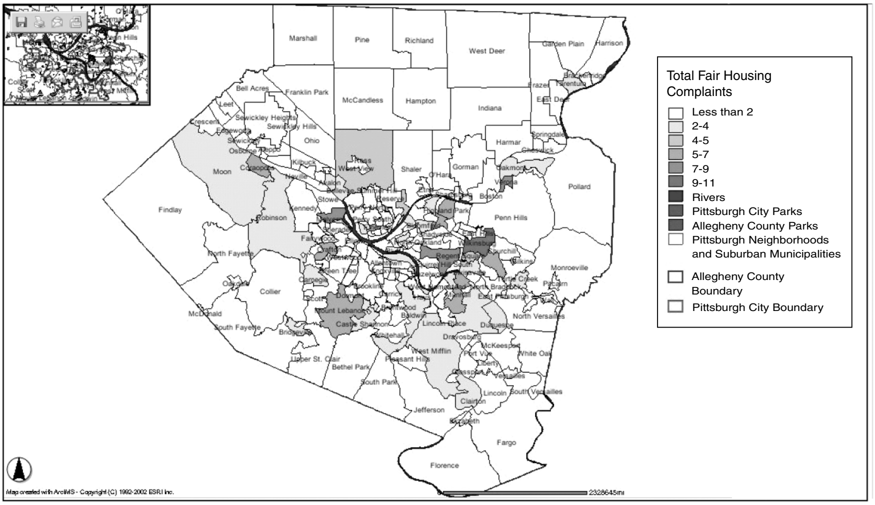


FIGURE 3 Spatial-data interface for counseling SDSS. Source: Johnson, 2005. Reprinted with permission from Elsevier. (Figure can be viewed in color at *http://www.andrew.cmu.edu/user/johnson2/SearchPittsburghNeighborhoods.jpg*.)

maximize satisfaction and minimize the burden of the housing search). The SDSS uses geographic information systems to illustrate neighborhood characteristics, a relational database to store information on specific housing units, and a multi-criteria decision model to help clients make relocation decisions. Figure 3 illustrates the spatial-data interface with fair housing data for Allegheny County, Pennsylvania.

## RESEARCH NOW AND IN THE FUTURE

A number of analytical methods can be used to make affordable housing and sustainable communities more accessible. In one stream of current research, civil, environmental, and mechanical engineering methods are being used to design housing units that improve on current practices in terms of energy efficiency, cost, structural quality, and efficiency of construction processes. In another stream of current research, urban and regional planning are being used to help stakeholders define development strategies that reflect best knowledge of social science-based program evaluation, land-use and transportation planning standards, and community-level partnerships. Decision sciences can provide opportunities to design housing- and community-development policies that improve on current practices in construction-oriented engineering and planning in terms of social outcomes, multistakeholder negotiations, and housing program client choices.

Because affordable housing and sustainable community development are not currently top priorities for market-rate housing providers, government support for the engineering of residential housing may be necessary to increase environmental sustainability and reduce user costs. However, housing policies that optimize various social criteria must also address technological aspects of housing and be based on best practices in urban and regional planning to be considered sustainable and affordable.

The decision-sciences research described in this paper suggests a number of promising areas for future research. The most important is to provide evidence that implementation of the decision models described above result in improved outcomes for communities and individuals. Other areas for research include: (1) choices of housing design and construction strategies that balance housing-unit- and community-level sustainability measures; (2) the development of dynamic models for designing strategic housing policies to address place-based housing strategies (i.e., new construction and rehabilitation of existing housing units); and (3) the design of realistic and tractable decision models to guide developers of affordable housing who must routinely choose a handful of sites to develop from many alternatives, with limited funding, to maximize the probability of neighborhood revitalization.

As long as urban sprawl, environmental degradation, and geographical barriers to affordable housing and opportunity remain policy problems, researchers have an opportunity to devise novel and creative solutions at the nexus of engineering, planning, and decision sciences.

## ACKNOWLEDGMENTS

My thanks to Jeannie Kim and Vincent Chiou for assisting in this research and to Julie Swann and Jennifer Ryan for encouraging me to participate in the 2006 Frontiers of Engineering Symposium.

## REFERENCES

Armacost, R.L., J. Paul, M.A. Mullens, and W.W. Swart. 1994. An AHP framework for prioritizing customer requirements in QFD: an industrialized housing application. IIE Transactions 26(4): 72–80.

Ayeni, B. 1997. The Design of Spatial Decision Support Systems in Urban and Regional Planning. Pp. 3–22 in Decision Support Systems in Urban Planning, edited by H. Timmermans. London: E & F N Spon.

Balcomb, J.D., C.E. Hancock, and G. Barker. 1999. Design, Construction, and Performance of the Grand Canyon House. National Renewable Energy Laboratory, U.S. Department of Energy. Available online at: *http://www.nrel.gov/docsfy00osti/24767.pdf.*

Bohl, C.C. 2000. New urbanism and the city: potential applications and implications for distressed inner-city neighborhoods. Housing Policy Debate 11(4): 761–801.

Caulkins, J.P., G. Feichtinger, D. Grass, M.P. Johnson, G. Tragler, and Y. Yegorov. 2005. Placing the poor while keeping the rich in their place: separating strategies for optimally managing residential mobility and assimilation. Demographic Research 13(1): 1–34. Available online at: *http://www.demographic-research.org/Volumes/Vol13/1/default.htm.*

de Souza Briggs, X. 2005. Politics and Policy: Changing the Geography of Opportunity. Pp. 310–341 in The Geography of Opportunity: Race and Housing Choice in Metropolitan America, edited by X. de Souza Briggs. Washington, D.C.: Brookings Institution Press.

Gabriel, S.A., J.A. Faria, and G.E. Moglen. 2006. A multiobjective optimization approach to smart growth in land development. Socio-Economic Planning Sciences 40: 212–248.

Ibrahim, R., and M. Nissen. 2003. Emerging Technology to Model Dynamic Knowledge Creation and Flow Among Construction Industry Stakeholders During the Critical Feasibility-Entitlements Phase. In Information Technology 2003: Towards a Vision for Information Technology in Civil Engineering, edited by I. Flood. Reston, Va.: American Society of Civil Engineers. Available on CD-ROM.

Johnson, M.P. 2005. Spatial decision support for assisted housing mobility counseling. Decision Support Systems 41(1): 296–312.

Johnson, M.P. 2006. Planning Models for Affordable Housing Development. Forthcoming in Environment and Planning B: Planning and Design.

Joint Center for Housing Studies of Harvard University. 2006. The State of the Nation's Housing 2006. Available online at: *http://www.jchs.harvard.edu/publications/markets/son2006/son2006.pdf*.

Katrakis, J.T., P.A. Knight, and J.D. Cavallo. 1994. Energy-Efficient Rehabilitation of Multifamily Buildings in the Midwest. Argonne National Laboratory, Decision and Information Sciences Division. Available online at: *http://www.eere.energy.gov/buildings/info/documents/pdfs/multigu.pdf*.

Millenial Housing Commission. 2002. Meeting Our Nation's Housing Challenges: A Report of the Bipartisan Millenial Housing Commission Appointed by the Congress of the United States. Available online at: *http://govinfo.library.unt.edu/mhc/MHCReport.pdf*.

Myers, D., and E. Gearin. 2001. Current preferences and future demand for denser residential environments. Housing Policy Debate 12(4): 633–659.

Priemus, H. 2005. How to make housing sustainable?: the Dutch experience. Environment and Planning B 32(1): 5–19.

Steven Winter Associates Inc. 2001. Building America Field Project: Results for the Consortium for Advanced Residential Buildings (CARB), January to October 2001. National Renewable Energy Laboratory, U.S. Department of Energy. Available online at: *http://www.nrel.gov/docs/fy03osti/31380.pdf.*

von Hoffman, A. 1996. High ambitions: the past and future of American low-income housing policy. Housing Policy Debate 7(3): 423–446.

# DINNER SPEECH

# The Changing Face of Industrial Research

W. Dale Compton
*Purdue University*
*West Lafayette, Indiana*

Giving this talk is a great pleasure—not only because I get to meet all of you and be part of this meeting, but also because this venue brings back many memories for me.

You are meeting at the Ford Research Laboratories during tumultuous times for the U.S. automotive industry. Although I can't offer insights into the current state of the industry or Ford, I believe I can offer some insight into the factors that have led to the creation, and then the near demise, of some of this country's great industrial research establishments.

Major changes have occurred in recent years. Bell Labs is a distant memory. The size and focus of IBM research, GE research, Westinghouse research, and Xerox research have all been reduced, and research activities in many other companies have undergone similar changes. The focus of the industrial research that remains has changed—except for the pharmaceutical and biomedical industries, industrial research no longer includes basic research.

Why and how did these changes come about? Let's begin by examining the philosophical justifications that made industrial research popular some 50 years ago. This philosophy laid the groundwork for the great tide of industrial research that ultimately shaped the research posture today.

The story begins with World War II, when scientists and engineers from all over the country were called upon to leave their universities and come together to create the technologies and systems that were crucial to winning the war. The key developments included radar, the proximity fuse, the atomic bomb, and many others.

In November 1944, following the cessation of hostilities, President Franklin Roosevelt wrote to Dr. Vannevar Bush requesting his recommendations for post-war activities. In Roosevelt's words, "New frontiers of the mind are before us, and if they are pioneered with the same vision, boldness, and drive with which we waged this war, we can create a fuller and more fruitful employment and a fuller and more fruitful life." Dr. Bush, who had been the director of the Office of Scientific Research and Development during the war, responded with a report, *Science—The Endless Frontier*.

A number of very important actions were taken in response to the recommendations in that report. First, in 1945, the Office of Naval Research was founded, the first government agency responsible for funding research that did not necessarily address immediate requirements for the military. This was the start of federal funding for basic research in our universities. In January 1946, the Research Grants Office was created at NIH to administer projects of the Office of Scientific Research and Development and to operate a program of extramural research grants and fellowship awards. A few years later, in 1950, the National Science Foundation (NSF) was created. NSF is still the principal source of funding for basic research in the physical sciences, engineering, and social sciences. Other federal departments soon followed suit and established their own extramural funding activities. Many of you have participated in one or more of these federal programs. In fact, NSF is providing partial support for this conference.

A number of actions were taken by industry soon after this. Principally, industrial research was initiated—primarily in the physical sciences and engineering. Back in 1951, Ford created the Ford Research Laboratories, as Gerhard Schmidt[1] mentioned earlier. During the 50th anniversary celebration for the laboratories, Ford characterized those 50 years in a very useful way. The period from 1951 to 1970 was called the "golden era" of research, 1970 to 1985 the era of redirection and regulation, and 1985 to 2001 the era of relevance. If we replaced *regulation* in the middle era with *deregulation*, as in the case of telecommunications, airlines, and energy, this characterization would be rather accurate for industry as a whole.

## THE GOLDEN ERA

What was the emphasis during the golden era of research, the first 25 years after World War II? At that time, people looked at what scientists and engineers had done during the war and concluded that they could do the same for industry. The principal justification for supporting research was that good research would lead to good products and good profits. Please note the absence of any mention of the topics to be explored. The second justification was people. A research

---

[1] Vice President, Research and Advanced Engineering, Ford Motor Company.

organization staffed by outstanding individuals would be available to consult on internal company problems and would serve as the eyes and ears of the company in the global world of science and technology.

Based on these two justifications, a wide variety of corporate laboratories flourished. Great progress was made in science, publications were abundant, and opportunities for employment abounded. Again, notice that the research output was not directly tied to the profits of the companies.

The quality of science in physics, chemistry, and metallurgy at Ford from 1951 to 1970 was outstanding. If one were to compare the physics department at Ford with academic physics departments throughout the country, I think it is safe to say Ford would have been in the top ten—maybe even the top five.

Al Overhauser at Ford was the discoverer of the Overhauser effect, which has been of enormous importance for solid-state physics. The first SQUID—any of you in magnetic measurements will recognize the SQUID as the most sensitive detector of magnetic fields that has ever been developed—was also invented at Ford. The first frequency doubler for the laser was demonstrated here. The scientists at Ford were widely known and recognized as outstanding. Scientists in the Ford chemistry and metallurgy departments were similarly talented.

## THE ERA OF REDIRECTION AND REGULATION

Now we enter the era of redirection and regulation—1970 to 1986. Allow me to share briefly some of my findings on joining Ford in 1970. First, the people were outstanding. But, although they had great eyes and ears in the technical world, they had no effective communication with the operations sector of the company. In fact, they had little credibility with operations. Programs in the physics area were related to general relativity and solving math problems that were fun but had hardly any relevance to the company's operations. More important, I found no systems effort at all in the laboratory.

During this era, we made serious efforts to refocus the laboratory on problems relevant to the company while keeping the research as long range and basic as possible. At the time, the company was beset by external demands that required new directions for research. The Clean Air Act became law and emissions regulations, fuel-economy regulations, and regulations limiting emissions from manufacturing plants were passed.

It was not hard to find research areas that could provide a fundamental understanding that would be relevant to meeting these demands, but it took time for people to reorient their thinking and embrace problems that were relevant to the new needs of the company. In due course, however, a number of new activities were initiated.

An atmospheric-science program was started, and a major effort in systems engineering was undertaken. These soon began to provide dividends. Ford was the first company to develop a simulation model for hybrid vehicles and then to demonstrate the viability of a hybrid. Ford developed the first flexible-fuel vehicles that could operate on 100 percent methanol, 100 percent gasoline, or any mixture of the two—all automatically. We developed the first electronic engine controls in 1972. In 1975, they went into production as the first fully programmable electronic, adaptive control system in a production vehicle.

That production program led to the largest off-campus reeducation of engineers up to that time. Most of the mechanical engineers involved in the development of the engine and the controls needed up-to-date training on digital electronics, so many of them pursued master's degree programs in electrical engineering at Wayne State University. General courses were taught on the Wayne State campus. Courses based on proprietary information were taught at Ford.

Catalytic converters were critical to controlling emissions from vehicles. Ford sponsored the development of the monolithic catalyst structure, which later became the model for the industry. The key active ingredients in the catalyst were platinum and rodium, and reducing the amount of platinum became a long-standing goal. Haren Gandhi,[2] who is sitting here in the front row, participated in that effort. You may have noticed, on the outside wall in the large hall, the Medal of Technology that President Bush presented to Haren for his efforts to improve the efficiency of catalysts, and, hence, reduce their cost by reducing the amount of platinum they required.

Despite the success of the catalyst, it was also an unfortunate example of how hard it is to communicate across "silos" in a large company. While Haren was successfully reducing the amount of platinum in each catalyst, and thus reducing the cost of each unit, the financial sector of the company held a very large forward position in platinum. Because there was no good mechanism for discussions between the company's research and financial sectors, Ford lost a lot of money when the need for less platinum was announced and the value of platinum plummeted.

Research in many other areas was also ongoing—sensors, stamping dies, new paints, new high-strength alloys, and so on. Basic research also continued, at a significantly lower level but at a high enough level to be effective. In other words, there was still an effective mass concentrated in areas that would logically support the large, more applied programs that, in turn, supported the operating divisions. I am sure Gerhard will understand that maintaining basic research programs was possible only because we were able to hide them. The programs directly related to the company's operating divisions were large enough and important enough that management was not interested in asking questions about the other programs.

The new regulatory environment had a profound impact on the company,

---

[2] Ford Technical Fellow and Manager, Chemical Engineering Department, Ford Motor Company.

which found itself working toward technical objectives with which it had no experience and little knowledge. Communication by researchers with bright people in operations who were willing to look at the needs of the company was possible, but not easy. There were still many detractors—people at high levels in the company who wanted to reduce the research budget by as much as 30 percent in one year. Fortunately, clearer heads prevailed, and the research laboratories were given enough time to make the necessary reorientation. If they had been forced to do so in a very short time, it is unlikely that they would have continued to exist.

## THE ERA OF RELEVANCE

Now to the era of relevance, namely post-1986. In this era, essentially all programs must contribute either to the products or the processes used by the company that sponsors them. This is true not just at Ford, but throughout industry. As companies have reduced the size, or even eliminated, their research laboratories or, at least, eliminated long-term research in their laboratories, many have increased their cooperation with universities, particularly in the bioscience and engineering sectors. As a result, little basic research is being done today by industry, although many industry segments continue to support some basic research, mostly in universities rather than in their own laboratories.

But there are some problems with this arrangement. For example, graduates interested in pursuing careers in research have fewer opportunities. In addition, questions have arisen about ownership of intellectual property and, most important, about the funding philosophy of industry and federal agencies supporting research in universities. For example, how long a view can researchers take? And how many risks?

## LESSONS LEARNED

Let me share with you my personal biases about what we have learned in the last 60 years. Good research can be done on relevant problems, but those problems frequently lead to questions that can only be answered by fundamental research, which takes time. The reason the management of research in industry is so very difficult is that researchers must identify and be working on problems well before the operations sectors even realize they have a problem. Those problems cannot be solved immediately. Research must be out in front, and that takes tremendous foresight, which, in turn, requires that the research sector be in close touch with operations. That's the only way these problems can be anticipated and understood. Only after that, can a company decide what can be done to solve them and which problems will only be solvable through basic research.

A research organization in industry is in a very fragile position. If it is too close to operations, the pressures for short-term results may increase to the point that long-term research is crowded out. If operations are held at arm's length, however, the research sector can lose its credibility and the support of the people who would benefit the most from its results.

Those of you who work in universities may think none of this applies to you. I hope that this is not the case, because choosing the right problem to work on at the right time is as critical to your success as to the success in industry. In fact, it is critical that you get funding to pursue that research.

## THE PROBLEMS AHEAD

Just as we learned that research in industry can only prosper in the long term when the research sector maintains contact with its customer, namely the company, so must we identify and tackle the really important problems confronting not only the company, but also the country and world markets. Some of these problems are technical, and some are not. Just for illustration, I will give you examples of each.

First, a nontechnical problem—the lack of understanding of the consequences of political decisions, both local and national, related to technology. A search of the Congressional Research Office database for congressmen with "engineer" in their titles turned up only one. There may be a few more, but only one was found in that search. That is pathetic, but it reflects how difficult it is for technical people to reach out and be part of the political system. We desperately need to think about how to break down that barrier.

We must address the whole issue of technical literacy, for both technical and nontechnical people. I might observe that one of the benefits of a symposium like this is that it increases your technical literacy in subjects that are not in your special area of expertise. At my own institution, Purdue University, there are no courses that teach technology to nontechnical students. This is a travesty.

Now for the technical problems. A lot of things could be used as examples, but I will show my biases with two of them—energy independence and health care delivery. First, energy independence. We have to find alternative fuels. We have to find a way to become less dependent on the petroleum sources in this world. The conversion of cellulose to liquid fuel and coal to liquid are viable sources of liquid-based fuels, but the technology is not yet at a point that would make these viable.

My second example is a newly emerging research area for engineering. The National Academy of Engineering and the Institute of Medicine recently published a study on the subject of engineering and health care delivery. The focus was not just on bioengineering and biomedical engineering, as important as they are, but also on the system by which care is provided to people—such as system optimization, sensors, remote communications, telemedicine, and making every hospital room an intensive care unit. Another question is how we can take advantage of the Internet in the delivery of health care, the long-term role of which we cannot predict.

## MAJOR CHALLENGES

At the beginning of this new century, the technical community is facing two major challenges. The first is ensuring the continuing availability of innovation, which is critical to our national prosperity. The second is supporting the necessary level of research to ensure that innovation continues.

There is a tendency to think we have come full circle since 1944, from research is golden, to research is unnecessary, to a realization that research is critical to future innovation. However, we live in a time of globalization, when competition is fierce, money is limited, and expenditures that are not directly relevant to a company's mission must be justified. In fact, this is a more difficult environment than the environment of the 1970s when we were suffering the effects of the oil embargo. We must find ways to meet these challenges through both technical and nontechnical means.

I close with a quote from the recent National Academies study, *Rising Above the Gathering Storm.* "This nation must prepare with great urgency to preserve its strategic and economic security. Because other nations have, and probably will continue to have, the competitive advantage of a low-wage structure, the United States must compete by optimizing its knowledge-based resources, particularly in science and technology, and by sustaining the most fertile environment for new and revitalized industries and the well-paying jobs they bring."

I leave you with a big question. How will, or can, our institutions respond to these challenges?

# APPENDIXES

# Contributors

**Robert L. Axtell** is an associate professor at the Center for Social Complexity and Computational Social Science at George Mason University in Fairfax, Virginia. Previously, Dr. Axtell was a senior fellow at the Brookings Institution in Washington, D.C., and he also served as a visiting or adjunct professor at the Santa Fe Institute, New School University, Johns Hopkins University, and Georgetown University. He earned a Ph.D. in engineering and public policy from Carnegie Mellon University, where he studied economics, computer science, game theory, operations research, and environmental science. His book, *Growing Artificial Societies: Social Science from the Bottom Up* (MIT Press, 1996), co-authored with J. Epstein, was an early exploration of the potential of multi-agent systems modeling in the social sciences. Dr. Axtell's research has been published in academic journals (e.g., *Science, Proceedings of the National Academy of Sciences, Economic Journal, Computational and Mathematical Organization Theory, Journal of Regulatory Economics*) and reprised in the popular science press (e.g., *Scientific American, Science News, New Scientist, Discover, Technology Review*), newspapers (e.g., *Wall Street Journal, Los Angeles Times, Washington Post*), and magazines (e.g., *Atlantic Monthly, New Yorker*). His latest book, *Artificial Economies of Adaptive Agents: The Multi-Agent Systems Approach to Economics*, was published by MIT Press in 2006. Dr. Axtell has been a consultant to industry and government, through the former BiosGroup, with NuTech Solutions, and most recently with BAE Systems.

**Matthew J. Barth** is director of the College of Engineering Center for Environmental Research and Technology at the University of California (UC), Riverside. His Transportation Systems and Vehicle Technology Research Laboratory has several full-time staff members and provides research experience for undergraduate and graduate students; the research is focused on intelligent transportation systems and air quality. From 1985 to 1986, Dr. Barth was a member of the technical staff in the Advanced Technologies Division of General Research Corporation, Santa Barbara. From 1986 to 1987, he was a visiting research student at the University of Tokyo. After completing his Ph.D., he returned to Japan as a visiting researcher at Osaka University, where he conducted research in systems engineering from 1989 to 1991. When he returned to the United States, he joined the faculty of the UC-Riverside College of Engineering. Dr. Barth is a member of the Institute of Electrical and Electronic Engineers, the Air and Waste Management Association, the Transportation Research Board Transportation and Air Quality Committee and New Technology Committee, and the ITS America Energy and Environment Committee. He has also served on several National Research Council committees. Dr. Barth received his M.S. and Ph.D. in electrical and computer engineering from the University of California, Santa Barbara, in 1986 and 1990, respectively.

**Marcel Bruchez** is program manager at the Technology Center for Networks and Pathways and visiting associate research professor in the Department of Chemistry at Carnegie Mellon University. From 1998 to 2005, at Quantum Dot Corporation in Hayward, California, a company he cofounded, Dr. Bruchez was founding scientist and senior scientist in the chemistry division, principal scientist for labels product development, and director of marketing for labels products. He is the author of 12 manuscripts, 12 issued patents, and 20 published patent applications. Dr. Bruchez's other professional activities include reviewer for *Journal of the American Chemical Society, Advanced Materials, Angewandte Chemie, NanoLetters, Nature Materials, Nature Medicine, Nature Methods, and Nature Biotechnology* (2002–2003). He is also co-editor of *Methods in Molecular Biology: Quantum Dots in Biological Applications* (Humana Press, 2006). Dr. Bruchez was the recipient of the Rank Prize Optoelectronics Award (2005) and MIT TR100 Award (2004). In 2003 he was recognized by *Science* for one of the Top Ten Scientific Innovations of 2003—quantum dots for biological detection. Dr. Bruchez received a Ph.D. in physical chemistry from the University of California, Berkeley (1998).

**W. Dale Compton** is the Lillian M. Gilbreth Distinguished Professor of Industrial Engineering, Emeritus, at Purdue University. His research interests include materials science, automotive engineering, combustion engineering, materials engineering, manufacturing engineering, and management of technology. From 1986 to 1988, as the first National Academy of Engineering (NAE) Senior Fellow, Dr. Compton directed activities related to industrial issues and engineering education. He came to NAE from the Ford Motor Company, where he was vice president of research. Before that, he was professor of physics and director of the

Coordinated Sciences Laboratory at the University of Illinois. Dr. Compton has served as a consultant to numerous government and industrial organizations and is a fellow of the American Physical Society, American Association for the Advancement of Science, Society of Automotive Engineers, and Engineering Society of Detroit. He has received the M. Eugene Merchant Manufacturing Medal from the American Society of Mechanical Engineers and the Society of Manufacturing Engineers, the University of Illinois College of Engineering Alumni Award for Distinguished Service, and the Science Trailblazers Award from the Detroit Science Center and the Michigan Sesquicentennial Commission. Dr. Compton was elected a member of NAE in 1981 and is currently NAE home secretary.

**Timothy J. Deming** is a professor in the Department of Bioengineering at the University of California, Los Angeles. Previously, he held positions in the Materials and Chemistry Departments and the Interdepartmental Program of Biomolecular Science and Engineering at the University of California, Santa Barbara. Dr. Deming has received many awards and honors, including the International Union of Pure and Applied Chemistry (IUPAC) Macromolecular Division, Samsung-IUPAC Young Scientist Award from the World Polymer Congress (2004), Materials Research Society Young Investigator Award (2003), Camille Dreyfus Teacher-Scholar Award (2000), Beckman Young Investigator Award (1998), Alfred P. Sloan Research Fellow (1998), and National Science Foundation CAREER Award (1997). In 2002, he was a Rothschild-Mayent Foundation Fellow at the Institut Curie in Paris. Dr. Deming is currently a member of the editorial advisory boards of *Macromolecules*, *Macromolecular Bioscience*, and *Biopolymers*. He has also served on numerous professional society and faculty committees and worked with middle-school students. Dr. Deming has filed 11 patent applications. He received a Ph.D. in chemistry from the University of California, Berkeley (1993).

**Brenda L. Dietrich** is director of mathematical sciences at the IBM Thomas J. Watson Research Center. Her areas of research include manufacturing scheduling, services resource management, transportation logistics, integer programming, and combinatorial duality. She is a member of the Advisory Board of the Industrial Engineering/Management Science Department of Northwestern University; a member of the Industrial Advisory Board for both the Institute for Mathematics and Its Applications and the Center for Discrete Mathematics and Theoretical Computer Science at Rutgers University; and IBM's delegate to the Massachusetts Institute of Technology Supply Chain 2020 Program. She has participated in numerous conferences of the Institute for Operations Research and the Management Sciences (INFORMS), Math Programming, Society for Industrial and Applied Mathematics, Council of Logistics Management, and Association for Operations Management. She holds a dozen patents, has co-authored numerous publications, and has co-edited *Mathematics of the Internet: E-Auction and Markets* (Springer-Verlag, 2002). Dr. Dietrich has been a member of the INFORMS Roundtable, served on the INFORMS board as vice president for Practice, was chair of the advisory committee for the first two Practice meetings, and is currently the president-elect of INFORMS. In addition, Dr. Dietrich has served on the editorial board of *M&SOM* and is currently on the editorial board of *Logistics Research Quarterly*. She received a B.S. in mathematics from the University of North Carolina and an M.S. and Ph.D. in operations research/industrial engineering from Cornell University.

**Rebekah Anna Drezek,** an associate professor of bioengineering and electrical and computer engineering at Rice University, is affiliated with numerous institutes at Rice, including the Institute for Biosciences and Bioengineering, Computer and Information Technology Institute, Rice Quantum Institute, Center for Biological and Environmental Nanotechnology, and Center for Nanoscale Science and Technology. Among her many awards are the MIT TR100 Award (2004), the Beckman Young Investigator Award (2005), the Coulter Foundation Early Career Translational Research Award (2005), and the American Association for Medical Instrumentation Becton Dickinson Career Achievement Award (2005). She has been an invited speaker or panelist at numerous colloquia, meetings, and workshops, including the American Association for Cancer Research Annual Meeting (2006), the 36th Annual Colloquium on the Physics of Quantum Electronics (2006), and the IEEE International Biomedical Imaging Symposium (2006). Dr. Drezek received her M.S. and Ph.D. in electrical engineering from the University of Texas at Austin, in 1998 and 2001, respectively.

**Michael P. Johnson** is an associate professor of management science and urban affairs at the H. John Heinz III School of Public Policy and Management at Carnegie Mellon University. His research interests are focused on public-sector facility location and service delivery, especially for affordable housing and sustainable community development. He has taught courses on operations research, decision-support systems, cost-benefit analysis, and a capstone project course for public policy master's students. His extensive participation in academic service and community affairs has enriched his research and teaching. Dr. Johnson recently co-edited a volume of tutorials in operations research and is currently president of a professional society section on location analysis. He also founded and co-directed the Carnegie Mellon University/University of Pittsburgh Applied Decision Modeling Seminar Series. He recently evaluated plans by the Pittsburgh Public Schools to open, close, and resize various public schools and is currently a member of a committee to redesign the district's program for gifted students. He has been a member of the board of the Highland Park Community Development Corporation and director of the development of a community plan for the Highland Park neighborhood of Pittsburgh. Dr. Johnson has received master's degrees in operations research from the University of California, Berkeley, and in electrical engineering from the Georgia Institute of Technology. He received a Ph.D. from Northwestern University in 1997. His previous professional experience includes consulting in logistics, operations management, and information systems.

**Risto Miikkulainen,** professor of computer sciences at the University of Texas at Austin, has conducted recent research on methods of evolving neural networks and applying these methods to game playing, robotics, and intelligent control. He is an author of more than 200 articles on neuroevolution, connectionist natural-language processing, and the computational neuroscience of the visual cortex. Dr. Miikkulainen is an editor of the *Machine Learning Journal* and *Journal of Cognitive Systems Research*. He received an M.S. in engineering from the Helsinki University of Technology, Finland (1986), and a Ph.D. in computer science from the University of California, Los Angeles (1990).

**Andreas Schäfer** is a lecturer (associate professor) in the Department of Architecture and a research associate with the Institute for Aviation and the Environment at the University of Cambridge. He is also a research affiliate with the Massachusetts Institute of Technology (MIT). Previously, he spent five years at the International Institute for Applied Systems Analysis in Laxenburg, Austria, and seven years at MIT. Dr. Schäfer has been working for more than 10 years in the area of technology, human behavior, and the environment. His main areas of interest are modeling the demand for energy services, assessing characteristics of greenhouse-gas-emission technologies, and simulating the optimum technology dynamics in a greenhouse-gas-constrained energy system. He has published widely on global travel-demand modeling, transport-system technology assessment, and the introduction of technology. Dr. Schäfer holds an M.Sc. in aeronautical and astronautical engineering and a Ph.D. in energy economics, both from the University of Stuttgart, Germany.

**Alan Schultz,** director of the Navy Center for Applied Research in Artificial Intelligence at the Naval Research Laboratory in Washington, D.C., conducts research on human-robot interaction, evolutionary robotics, learning in robotic systems, and adaptive systems. The recipient of an Alan Berman Research Publication Award, he has published more than 75 articles on machine learning and robotics. Dr. Schultz is currently co-chair of the American Association for Artificial Intelligence (AAAI) Symposium Series and program chair of the 2007 Association for Computing Machinery (ACM)/Institute of Electrical and Electronics Engineers International Conference on Human Robot Interaction. In 2006, he was program co-chair of the 2006 ACM International Conference on Human-Robot Interaction. In 1999 and 2000, he chaired the AAAI Mobile Robot Competition and Exhibitions.

**Lawrence V. Snyder** is the Frank Hook Assistant Professor of Industrial and Systems Engineering at Lehigh University and co-director of Lehigh's Center for Value Chain Research. His research interests include modeling and solving problems in supply chain management, facility location, and logistics, especially under the threat of disruptions or other sources of uncertainty. His research has received awards from INFORMS and the Institute of Industrial Engineers. He has worked as a supply chain engineer and consultant for major producers of both perishable and durable goods. Dr. Snyder received a Ph.D. in industrial engineering and management sciences from Northwestern University.

**Morley O. Stone** is a former program manager in the Defense Sciences Office of the Defense Advanced Research Projects Agency and principal research biologist and biotechnology lead for the Air Force Research Laboratory (AFRL) at Wright-Patterson Air Force Base, where he has been a materials research engineer, research biologist, senior research biologist, and biotechnology group leader. In addition, he is an adjunct faculty member in the Department of Materials Science and Engineering at Ohio State University. His honors and awards include Fellow, AFRL (2005); Carnegie Mellon Alumni Award (2005); Vincent J. Russo Leadership Excellence Award (2003); and MIT TR100 nominee (2003). He is a member of the American Chemical Society, American Association for the Advancement of Science, and Materials Research Society. Dr. Stone received a Ph.D. in biochemistry from Carnegie Mellon University (1997).

**Mark Y.D. Wang** is a senior physical scientist at the RAND Corporation in Santa Monica, California, where he works in the area of purchasing and supply chain management, including inventory and multimodal distribution logistics; business process reengineering; acquisition/purchasing strategy. Recent research projects have included improving contracting at the city of Los Angeles Airport, Port, and Department of Water and Power; reverse logistics; high-tech manufacturing; and technology transfer from federally funded research and development. He was associate director of RAND's National Security Research Division and associate director of RAND's Science and Technology Policy Institute. Dr. Wang received an Sc.D. in physics from the Massachusetts Institute of Technology in 1994.

**Lloyd Watts** is founder, chair, and chief technology officer of Audience Inc., a venture-backed company in Silicon Valley developing high-performance audio-signal processing systems for the telecommunications industry. He has worked at Microtel Pacific Research, Synaptics, Arithmos, and Interval Research. He received a B.Sc. in engineering physics from Queen's University (1984), an M.A.Sc. in electrical engineering from Simon Fraser University (1989), and a Ph.D. in electrical engineering from the California Institute of Technology (1992).

**Susan Zielinski** is managing director of SMART (Sustainable Transportation and Access Research and Transformation) at the University of Michigan in Ann Arbor; SMART is a project of the Center for Advancing Research and Solutions for Society (CARSS). Just before joining SMART/CARSS, Ms. Zielinski spent a year as a Harvard Loeb Fellow working on New Mobility innovation and leadership. Prior to 2004, she co-founded and directed Moving the Economy, an innovative Canada-wide “link tank” that catalyzes and supports New Mobility industry development. For more than 15 years, she was a transportation planner for the city of Toronto, where she worked on developing and leading transportation and air-quality policies and initiatives, with a focus on sustainable transportation/New Mobility. Ms. Zielinski has been an advisor to local, national, and international initiatives, including the National Advisory Committee on Energy Efficiency, Transport Canada’s Sustainable Development Advisory Committee, Gridlock Panel of the Ontario Smart Growth Initiative, Organisation for Economic Co-operation and Development Environmentally Sustainable Transport Project, the jury of the Stockholm Partnerships for Sustainable Cities, the European Conference of Transport Ministers, the Centre for Sustainable Transportation, and the Kyoto Cities Initiative International Advisory Panel. After receiving her undergraduate degree from the University of Toronto and a graduate fellowship to study for a year in France, she received a master’s degree in environmental studies from York University. She is a registered professional planner and a member of the Canadian Institute of Planners.

# Program

## NATIONAL ACADEMY OF ENGINEERING

2006 U.S. Frontiers of Engineering Symposium
September 21–23, 2006

Chair: Julia M. Phillips, Sandia National Laboratory

## THE RISE OF INTELLIGENT SOFTWARE SYSTEMS AND MACHINES

Organizers: M. Brian Blake, Georgetown University, and David Fogel, Natural Selection, Inc.

*Commercializing Auditory Neuroscience*
Lloyd Watts

*Creating Intelligent Agents in Games*
Risto Miikkulainen

*Co-evolution of the Computer and Social Science*
Robert Axtell

*Using Computational Cognitive Models to Build Better Human-Robot Interaction*
Alan Schultz

***

## THE NANO/BIO INTERFACE

Organizers: Tejal Desai, University of California, San Francisco, and
Hiroshi Matsui, CUNY Graduate Center and Hunter College

Part I: Solving nanotechnology problems using biotechnology

*Biological and Biomimetic Polypeptide Materials*
Timothy J. Deming, University of California, Los Angeles

*Biomimetics and the Application to Devices*
Morley O. Stone, Air Force Research Laboratory

Part II: Solving biotechnology problems using nanotechnology

*Optical Imaging for In Vivo Assessment of Tissue Pathology*
Rebekah A. Drezek, Rice University

*Commercialization and Future Developments in Bionanotechnology*
Marcel Bruchez, Carnegie Mellon University

***

## ENGINEERING PERSONAL MOBILITY FOR THE 21ST CENTURY

Organizers: Apoorv Agarwal, Ford Motor Company, and
William Schneider, University of Notre Dame

*Long-Term Trends in Global Passenger Mobility*
Andreas Schäfer, University of Cambridge

*Energy and Environmental Impacts of Personal Mobility*
Matthew J. Barth, University of California, Riverside

*New Mobility: The Next Generation of Sustainable Urban Transportation*
Susan Zielinski, University of Michigan

*Advancing the State of Hybrid Technology*
Andreas Schell, DaimlerChrysler Corp.

***

## SUPPLY CHAIN MANAGEMENT AND APPLICATIONS WITH ECONOMIC AND PUBLIC IMPACT

Organizers: Jennifer Ryan, University College Dublin, and Julie Swann, Georgia Institute of Technology

*Supply Chain Applications of Fast Implosion*
Brenda L. Dietrich, IBM Thomas J. Watson Research Center

*Factory to Foxhole: Improving the Army's Supply Chain*
Mark Wang, RAND Corporation

*Supply Chain Management under the Threat of Disruptions*
Lawrence V. Snyder, Lehigh University

*Engineering-Based Methods for Affordable Housing and Sustainable Community Development*
Michael P. Johnson, Carnegie Mellon University

***

## DINNER SPEECH

*The Changing Face of Industrial Research*
W. Dale Compton

# Participants

## NATIONAL ACADEMY OF ENGINEERING

2006 U.S. Frontiers of Engineering Symposium
September 21–23, 2006

Alexis Abramson
Warren E. Rupp Assistant Professor
Department of Mechanical and Aerospace Engineering
Case Western Reserve University

Stephanie G. Adams
Assistant Dean of Research
AAAS/NSF Science and Engineering Fellow
University of Nebraska–Lincoln

Apoorv Agarwal
EVA Engine PMT Leader
Scientific Research Laboratories
Ford Motor Company

David V. Anderson
Associate Professor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Mark H. Anderson
Associate Scientist
College of Engineering
University of Wisconsin–Madison

Ana I. Antón
Associate Professor
Department of Computer Science
North Carolina State University

Robert Axtell
Associate Professor
Department of Computational and Data Sciences
George Mason University

George D. Bachand
Principal Member of the Technical Staff
Sandia National Laboratories

Qing Bai
Member of Technical Staff
Agilent Technologies, Inc.

Matthew Barth
Director
Center for Environmental Research and Technology
University of California, Riverside

Bradley R. Bebee
Vice President, Systems Software
Information Systems Worldwide Corporation

Chandra R. Bhat
Professor
Department of Civil, Architechtural, and Environmental Engineering
University of Texas at Austin

Stephan Biller
Lab Group Manager
General Motors

M. Brian Blake
Associate Professor of Computer Science
Georgetown University

Richard K. Boger
Application Engineer
ABAQUS, Inc.

Marcel P. Bruchez
Program Manager, Technology Center for Networks and Pathways
Visiting Associate Research Professor, Department of Chemistry
Carnegie Mellon University

David Bruemmer
Principal Research Scientist
Idaho National Laboratory

Bryan Cantrill
Senior Staff Engineer
Sun Microsystems

Kimberly A. Chaffin
Principal Scientist
Medtronic, Inc.

Jane P. Chang
Associate Professor
Department of Chemical and Biomolecular Engineering
University of California, Los Angeles

Jia Chen
Research Staff Member
IBM T.J. Watson Research Center

Aref Chowdhury
Member of Technical Staff
Bell Laboratories, Lucent Technologies

Zissis Dardas
Group Leader
United Technologies Research Center

Jarrett Datcher
Engineer/Scientist
Boeing–Phantom Works

Moses M. David
Lead Senior Research Specialist
Corporate Research Process Research Laboratory

Matthew DeLisa
Assistant Professor
School of Chemical and Biomolecular Engineering
Cornell University

Timothy Deming
Professor
Department of Bioengineering
University of California, Los Angeles

Tejal A. Desai
Professor of Physiology/
Bioengineering
University of California, San
Francisco

Brenda L. Dietrich
Director, Mathematical Sciences
IBM T.J. Watson Research Center

Anne C. Dillon
Senior Research Scientist
National Renewable Energy
Laboratory

Scott W. Doebling
Program Manager
Los Alamos National Laboratory

Rebekah Anna Drezek
Associate Professor
Departments of Electrical and
Computer Engineering and
Bioengineering
Rice University

John Dunagan
Researcher
Systems and Networking Group
Microsoft Research

Richard Elander
Advanced Pretreatment Team Leader
Bioprocess Engineering Group
National Bioenergy Center
National Renewable Energy
Laboratory

Ye Fang
Research Associate
Science and Technology Division
Corning Inc.

Andrei G. Fedorov
Assistant Professor
Woodruff School of Mechanical
Engineering
Georgia Institute of Technology

David B. Fogel
Chief Executive Officer
Natural Selection, Inc.

Suresh V. Garimella
Professor
School of Mechanical Engineering
Purdue University

Michael J. Garvin
Assistant Professor
Department of Civil and
Environmental Engineering
Virginia Polytechnic Institute & State
University

Irene Georgakoudi
Assistant Professor
Department of Biomedical
Engineering
Tufts University

Cindie Giummarra
Senior Engineer
Alcoa Technical Center

Christine S. Grant
Professor
Department of Chemical Engineering
North Carolina State University

Paula Hicks
Principal Scientist
Cargill

Ron Ho
Senior Research Scientist
Sun Microsystems Research Labs

Mani Janakiram
Principal Engineer
Intel Corporation

Michael P. Johnson
Associate Professor
H. John Heinz III School of Public Policy
Carnegie Mellon University

Paul C. Johnson
Professor and Executive Dean
Ira A. Fulton School of Engineering
Arizona State University

Ruben Juanes
Assistant Professor
Department of Civil and Environmental Engineering
Massachussetts Institute of Technology

Jack W. Judy
Associate Professor and Director of the NeuroEngineering Program
Department of Electrical Engineering
University of California, Los Angeles

Krishna Kalyan
Principal Staff Engineer
Motorola

Deepak Khosla
Senior Research Scientist
HRL Laboratories, LLC

Steven Kou
Associate Professor of Industrial Engineering
Columbia University

Olga Anna Kuchar
Research Scientist 4
Pacific Northwest National Laboratory

David A. LaVan
Assistant Professor
Department of Mechanical Engineering
Yale University

Christian Lebiere
Research Faculty
Psychology Department
Carnegie Mellon University

Philip LeDuc
Assistant Professor
Departments of Mechanical Engineering, Biomedical Engineering, and Biological Sciences
Carnegie Mellon University

Lorraine H. Lin
Senior Engineering Specialist
Bechtel National, Inc.

Melissa Lunden
Staff Scientist
Lawrence Berkeley National Laboratory

Teng Ma
Associate Professor
FAMU-FSU College of Engineering
Florida State University

Surya K. Mallapragada
Professor
Department. of Chemical and Biological Engineering
Iowa State University

Rajit Manohar
Associate Professor
School of Electrical and Computer Engineering
Cornell University

Hiroshi Matsui
Associate Professor
Department of Chemistry
CUNY Graduate Center and Hunter College

Lawrence Megan
Manager
Advanced Control and Optimization R&D
Praxair

Matthew M. Mehalik
Visiting Assistant Professor
Department of Industrial Engineering
University of Pittsburgh

Nafaa Mekhilef
Senior Research Scientist
Arkema, Inc.

Sergey Melnik
Researcher
Microsoft Research

Risto Miikkulainen
Professor
Department of Computer Science
University of Texas at Austin

Josh Molho
Staff R&D Engineer
Caliper Life Sciences

Jeffrey Montanye
Research and Development Director
Dow Automotive
Dow Chemical Company

Kumar Muthuraman
Assistant Professor
School of Industrial Engineering
Purdue University

Timothy R. Nolen
Senior Research Associate and Lab Head
Eastman Chemical Company

Eric A. Ott
Senior Engineer
GE Aviation

Julia M. Phillips
Director
Physical and Chemical Sciences Center
Sandia National Laboratories

Jonathan Preston
Technical Fellow
Lockheed Martin Aeronautics Company

William D. Provine
Research Manager
Central Research & Development
DuPont Company

Adam Rasheed
Aerospace Research Engineer
GE Global Research

Carol Rego
Vice President
CDM

Jennifer Ryan
Senior Lecturer
Management Department
University College Dublin

Todd Salamon
Member of Technical Staff
Math and Algorithmic Sciences Research Department
Bell Laboratories, Lucent Technologies

Andreas Schäfer
Associate Professor
Department of Architecture
University of Cambridge

Andreas Schell
Office of North American Operations
DaimlerChrysler

William F. Schneider
Associate Professor
Department of Chemical and Biomolecular Engineering, Concurrent in Chemistry
University of Notre Dame

Alan Schultz
Director
Navy Center for Applied Research in Artificial Intelligence
Naval Research Laboratory

Corey J. Schumacher
Senior Research Aerospace Engineer
Air Force Research Laboratory

Kenneth Shepard
Associate Professor
Department of Electrical Engineering
Columbia University

Abhijit V. Shevade
Member of Engineering Staff
Jet Propulsion Laboratory

Michael Siemer
President
Mydea Technologies

Vijay Singh
Assistant Professor
Department of Agricultural and Biological Engineering
University of Illinois at Urbana–Champaign

Sanjiv K. Sinha
Corporate Director
Environmental Consulting & Technology Inc.

Lawrence Snyder
Professor
Department of Industrial and Systems Engineering
Lehigh University

Hyongsok Soh
Assistant Professor
Biomolecular Science and Engineering & Mechanical Engineering
University of California, Santa Barbara

Morley O. Stone
Principal Research Biologist and Biotechnology Lead
Materials and Manufacturing Directorate
Air Force Research Lab

Julie Swann
Assistant Professor
Industrial and Systems Engineering
Georgia Institute of Technology

Margarita P. Thompson
Senior Research Engineer
Delphi Research Labs

Sammy Tin
Associate Professor
Department of Mechanical, Materials, and Aerospace Engineering
Illinois Institute of Technology

Manuel Torres
Section Manager and Assistant Vice President
Advanced Concepts Business Unit
SAIC

Aditya Tyagi
Senior Water Resources Engineer
CH2M Hill

Michiel Van Nieuwstadt
Technical Leader
Scientific Research Laboratory
Ford Motor Company

Mark Wang
Senior Physical Scientist
RAND Corporation

Michael L. Washington
Industrial Engineer
National Immunization Program
Immunization Services Division
U.S. Centers for Disease Control and Prevention

Lloyd Watts
Founder, CEO, and CTO
Audience Inc.

Paul K. Westerhoff
Associate Professor
Department of Civil and Environmental Engineering
Arizona State University

Colin S. Whelan
Senior Principal Engineer
Raytheon Company

Christopher Wolverton
Technical Leader
Ford Motor Company

Gerard Wong
Assistant Professor
Department of Materials Science and Engineering
University of Illinois at Urbana–Champaign

Robert Wray
Chief Scientist
Soar Technology

Aleksey Yezerets
Technical Advisor and Leader
Catalyst Technology Group
Cummins Inc.

ChengXiang Zhai
Assistant Professor
Department of Computer Science
University of Illinois at Urbana–Champaign

Jingren Zhou
Researcher
Microsoft Research

Susan Zielinski
Managing Director, SMART
Center for Advancing Research and Solutions for Society
University of Michigan

*Guests*

Adnan Akay
Division Director
Division of Civil, Mechanical, and Manufacturing Innovation
Directorate for Engineering
National Science Foundation

Marshall Lih
Senior Advisor
Division of Chemical, Bioengineering, Environmental, and Transport Systems
Directorate for Engineering
National Science Foundation

Mary Lou Maher
Program Director
Human-Centered Computing Cluster
Directorate for Computer and Information Science and Engineering
National Science Foundation

*Dinner Speaker*

Dr. W. Dale Compton
Lillian M. Gilbreth Distinguished Professor of Industrial Engineering, Emeritus
School of Industrial Engineering
Purdue University

*National Academy of Engineering*

Wm. A. Wulf
President

Lance A. Davis
Executive Officer

Janet Hunziker
Senior Program Officer

Gin Bacon
Senior Program Assistant

*Ford Motor Company*

Gerhard Schmidt
Vice President, Research and Advanced Engineering

Haren Gandhi
Ford Technical Fellow and Manager, Chemical Engineering Department

Barb Rutkowski
Administrative Assistant to Haren Gandhi

John Markee
Manager, Business Operations

Jim Bretz
Research Technologist, Business Operations

Rose Gossman
Business Operations