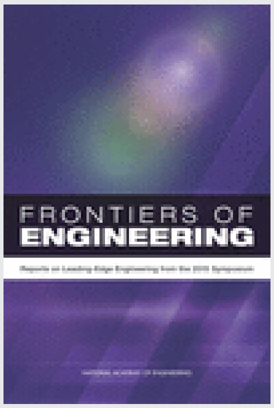

Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2015 Symposium

DETAILS

174 pages | 6 x 9 | PAPERBACK
ISBN 978-0-309-37953-3 | DOI: 10.17226/21825

AUTHORS

National Academy of Engineering

FRONTIERS OF ENGINEERING

Reports on Leading-Edge Engineering from the 2015 Symposium

NATIONAL ACADEMY OF ENGINEERING

THE NATIONAL ACADEMIES PRESS
Washington, DC
www.nap.edu

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

Funding for the activity that led to this publication was provided by The Grainger Foundation, Defense Advanced Research Projects Agency, Department of Defense ASD(R&E) Research Directorate—STEM Development Office, Air Force Office of Scientific Research, Microsoft Research, Cummins Inc., and individual donors. This material is also based upon work supported by the National Science Foundation under Grant No. 1505123. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. In addition, the content of this publication does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

International Standard Book Number-13: 978-0-309-37953-3
International Standard Book Number-10: 0-309-37953-9

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

Printed in the United States of America

Suggestion citation: National Academies of Sciences, Engineering, and Medicine. 2015. *Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2015 Symposium.* Washington, DC: The National Academies Press.

The National Academies of

SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Ralph J. Cicerone is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. C. D. Mote, Jr., is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.national-academies.org**.

ORGANIZING COMMITTEE

ROBERT D. BRAUN (Chair), David and Andrew Lewis Professor of Space Technology, Daniel Guggenheim School of Aerospace Engineering, Georgia Institute of Technology

AMIR AGHAKOUCHAK, Assistant Professor, Department of Civil and Environmental Engineering, University of California, Irvine

DAVID BRUMLEY, Associate Professor, Departments of Electrical and Computer Engineering and Computer Science, Carnegie Mellon University, and CEO, ForAllSecure

JENNIFER DIONNE, Assistant Professor, Department of Materials Science and Engineering, Stanford University

JAMES DONE, Project Scientist II and Willis Research Fellow, National Center for Atmospheric Research Earth System Laboratory

DANIELA OLIVEIRA, Associate Professor, Department of Electrical and Computer Engineering, University of Florida

SARA SEAGER, Class of 1941 Professor of Physics and Planetary Science, Departments of Earth, Atmospheric, and Planetary Sciences and Physics, Massachusetts Institute of Technology

LUKE SWEATLOCK, Research Scientist, Northrop Grumman Aerospace Systems

MITCHELL L. R. WALKER, Associate Professor, Daniel Guggenheim School of Aerospace Engineering, Georgia Institute of Technology

Staff

JANET HUNZIKER, Senior Program Officer
SHERRI HUNTER, Program Coordinator

Preface

This volume presents papers on the topics covered at the National Academy of Engineering's 2015 US Frontiers of Engineering Symposium. Every year the symposium brings together 100 outstanding young leaders in engineering to share their cutting-edge research and innovations in selected areas. The 2015 symposium was held September 9–11 at the Arnold and Mabel Beckman Center in Irvine, California. The intent of this book is to convey the excitement of this unique meeting and to highlight innovative developments in engineering research and technical work.

GOALS OF THE FRONTIERS OF ENGINEERING PROGRAM

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

At the annual 2½-day US Frontiers of Engineering Symposium, 100 of this country's best and brightest engineers—ages 30 to 45, from academia, industry, and government and a variety of engineering disciplines—learn from their peers about pioneering work in different areas of engineering. The number of participants is limited to 100 to maximize opportunities for interactions and exchanges among the attendees, who are chosen through a competitive nomination and selection process. The symposium is designed to foster contacts and learning among promising individuals who would not meet in the usual round of professional meetings. This networking may lead to collaborative work, facilitate the transfer of new techniques and approaches, and produce insights and applications that bolster US innovative capacity.

The four topics and the speakers for each year's meeting are selected by an organizing committee of engineers in the same 30- to 45-year-old cohort as the participants. Speakers describe the challenges they face and communicate the excitement of their work to a technically sophisticated but nonspecialist audience. They provide a brief overview of their field of inquiry; define the frontiers of that field; describe experiments, prototypes, and design studies (completed or in progress) as well as new tools and methods, limitations, and controversies; and assess the long-term significance of their work.

THE 2015 SYMPOSIUM

The topics covered at the 2015 symposium were (1) cybersecurity and privacy, (2) engineering the search for Earth-like exoplanets, (3) optical and mechanical metamaterials, and (4) forecasting natural disasters.

The first session on cybersecurity and privacy focused on how to engineer a system that is as secure as possible given practical construction constraints. A well-engineered system incorporates both layered protection and mechanisms for detecting and mitigating the effects of successful attacks. In addition, the user is just as important to security and privacy as the technology; user-centric designs help the user make good security and privacy decisions. The first talk described the various security and abstraction levels of modern systems as well as security consequences at the application, operating system, and hardware layers. The next speaker described the users' role and how to design interfaces that help them make better security decisions, particularly with regard to mobile platforms. This was followed by a talk on security within medical devices, which have diverse characteristics and pose different challenges to a security engineer. The session concluded with a talk on the US government's views on the challenges and frontiers in engineering cybersecurity.

In the past two decades, astronomers have found thousands of planets orbiting stars other than the sun. There is particular interest in finding exoplanets, as they are called, that orbit their stars' habitable zones where it is possible for liquid water, and therefore life, to exist. The session on Engineering the Search for Earth-like Exoplanets focused on the new generation of space-based telescopes necessary to find and study exoplanets. The first speaker described the James Webb Space Telescope, an international, NASA-led mission to be launched in 2018. The next presentation focused on starlight suppression, or technologies for direct imaging of exoplanets, and covered two main techniques: (1) the internal occulter, or coronagraph, that blocks light inside the telescope and works with wavefront sensing and control to create a stable optical system and (2) the external occulter, or starshade, which is a specially shaped screen tens of meters in diameter that formation flies tens of thousands of kilometers from its telescope and blocks out the star light so that only planet light enters the telescope. The third speaker addressed construction of large structures in space, focusing on large deployables and space-based assembly and construction. The fourth talk discussed leading-edge developments in sensing controls for formation flying and satellite proximity operations, particularly with regard to autonomy for small satellites.

The topic of the third session was optical and mechanical metamaterials—composites that have led to a reassessment of conventionally accepted boundaries on material performance as well as discovery of many useful properties. The first presentation described research in (1) the design of advanced materials that derive extraordinary strength from 3D architecture and microstructure and (2) recoverable mechanical deformation in compliant nanomaterials, leading to impacts on ultra-lightweight batteries and biomedical devices, among others. This was followed by a talk on the development of engineering materials with remarkably light weight and ultra-high stiffness. The next presenter focused on metamaterial-based device engineering, in particular, the connection between microscopic structural properties of metamaterials, such as symmetry and shape, and their macroscopic response. This work leads to useful devices that would not be possible with conventional materials, such as one-way antennas, “invisibility cloaks,” and acoustic circulators. The session concluded with a talk on optical and infrared metamaterials and their applications in devices that could revolutionize optical technologies in communications, photovoltaics, and thermal radiation management.

The final session, Forecasting Natural Disasters, focused on improvements in the prediction of track and intensity of natural hazards, for example, tropical cyclones and flash floods, that have resulted from a better understanding of atmospheric systems, advances in observational technologies, and increased computational power. The session opened with a talk about a physically based probabilistic tropical cyclone risk assessment and management framework that integrates science, engineering, and economic analysis to support coastal resiliency. This was followed by a presentation on behavioral research that could inform how scientific information about natural disasters is presented to the public in order to motivate individuals and organizations to take actions that would reduce risk and prevent economic losses. The third speaker described the Google Earth Engine platform (earthengine.google.org), an experimental application programming interface for massively parallel geospatial analysis on global datasets such as Landsat satellite imagery, elevation data, and weather. Applications based on these data can map, measure, and monitor Earth’s changes in unprecedented detail for the benefit of people and the environment.

In addition to the plenary sessions, the attendees had many opportunities for informal interaction. On the first afternoon, they gathered in small groups for “get-acquainted” sessions during which they presented short descriptions of their work and answered questions from their colleagues. This helped them get to know more about each other relatively early in the program.

On the second afternoon, attendees met in small groups to discuss issues such as industry-academic tech transfer, public understanding of artificial intelligence, technology risk management, effective interdisciplinary collaboration, and public advocacy for engineering, among others. Some of the groups have evolved into communities of interest and continued communicating after the meeting.

Every year a distinguished engineer addresses the participants at dinner on the first evening of the symposium. The 2015 speaker, NAE member and vice president Corale Brierley, principal of Brierley Consultancy LLC, gave the first evening's dinner speech, titled "The Black Swan." She described times in her career when valuable lessons were learned through unexpected experiences. She reminded the group that sometimes the benefit of a new idea is not understood or appreciated until well after the discovery, moving forward entails getting outside one's comfort zone, and failure is only a temporary change in direction to set one straight for the next success.

The NAE is deeply grateful to the following for their support of the 2015 US Frontiers of Engineering symposium:

- The Grainger Foundation
- Defense Advanced Research Projects Agency
- Department of Defense ASD(R&E)–STEM Development Office
- Air Force Office of Scientific Research
- National Science Foundation (this material is based on work supported by the NSF under grant EFMA-1505123)
- Microsoft Research
- Cummins Inc.
- Individual contributors

We also thank the members of the Symposium Organizing Committee (p. iv), chaired by Dr. Robert D. Braun, for planning and organizing the event.

Contents

Cybersecurity and Privacy

Cybersecurity and Privacy

David Brumley
Carnegie Mellon University

Daniela Oliveira
University of Florida

How can systems be engineered to be both secure and respectful of user privacy? Societal dependence on computers makes this question not only extremely relevant, but also nuanced. A series of well-understood steps is involved in engineering highly secure, privacy-respecting systems.

First, an engineer rigorously states the security and privacy goals of the system. Typical goals include the confidentiality of system data and system integrity and availability.

Second, the engineer defines what type of threats the system should be resilient to. For example, will an adversary attempt to infect the system through software vulnerabilities in applications? Or try to compromise the integrity of the operating system, which manages how applications access hardware resources? Worse still, is the adversary targeting the hardware, the lowest level of abstraction? Attacks on hardware render all security solutions at the operating system and application levels useless. Alternatively, the attacker may discover side channels, such as the system's electromagnetic radiation, to find cryptographic keys. The attacker can also leverage weaknesses in network protocols that were designed in the 1960s and still used today to compromise system availability.

Third, the engineer proves that the system design achieves the security goals in the presence of the adversary. And the last step is implementation of the system and formal verification that the implementation is correct.

Rigorous models and proofs, however, are performance expensive and problem specific. You get what you pay for, and highly secure systems are not cheap.

Furthermore, the Internet era exposes the challenge of protecting people's privacy, such as personal information, life habits, social networks, health conditions, and personal beliefs. Who owns and can profit from people's data? How can people delete or hide information from the Internet? Or should they? Isn't that rewriting history?

In practice the question is often not how to build a secure system, but how to engineer a system that is as secure as possible given practical construction constraints. New systems are almost always built on top of existing hardware, operating systems, software, and network protocols that provide fixed capabilities and have both known and unknown weaknesses. A well-engineered system follows a defense in depth strategy that incorporates layered protection and mechanisms for detecting and mitigating the effects of successful attacks. For example, a web server handling credit card numbers may use a network firewall to restrict access to only authorized computers, an intrusion detection system for detecting suspicious behaviors, and a secure communication protocol with its clients to encrypt the credit card numbers.

The best results come when security and privacy are engineered into the design from the beginning. Experience shows that retrofitting security and privacy measures into existing systems is difficult and often results in relatively weak security guarantees.

The user is often just as important to security and privacy as the technology. Users make decisions about what to share, what links to click, and what software to install. Recent research shows that existing systems often have unintuitive security and privacy mechanisms, and thus ultimately make the user the weakest link. Research has also shown that user-centric designs help the user make good security and privacy decisions.

In this session, Bryan Payne (Netflix) started with a talk explaining various security and abstraction levels of modern systems and security consequences at each layer. Franziska Roesner (University of Washington) then described the role of users and how interfaces can be designed to help them make better security decisions, with a focus on mobile platforms. Next, Kevin Fu (University of Michigan) addressed security in medical devices, which have different characteristics and pose different challenges to a security engineer. Tomas Vagoun (National Coordination Office for Networking and Information Technology R&D) concluded the session with a talk on the US government's view of challenges and frontiers in engineering cybersecurity.

Security at Different Layers of Abstractions: Application, Operating Systems, and Hardware

Bryan D. Payne
Netflix

In the field of computer security, it is important to take a threat-centric view of the world to understand who the malicious actors are, what they want, and how they achieve their goals. Computer systems are then evaluated in light of the perceived threats to understand the best defensive measures.

THREATS ACROSS COMPUTATIONAL ABSTRACTIONS

Researchers have extensively studied adversaries over the past two decades, and events in the news such as the Snowden revelations (Greenwald 2014) and attacks by the Equation Group (Kaspersky Lab 2015) have revealed much about threats posed by the most advanced adversaries.

Threat actors typically use the easiest path to achieve their goals. In some cases this involves manipulating an individual, a practice known as *social engineering*. Alternatively, on the technical side, a threat actor may choose to attack the victim's hardware, operating system, or applications. Each option has tradeoffs covering a broad spectrum of sophistication, cost, likelihood of detection, feasibility, and more.

Application-level attacks are often the easiest target for attackers because they cover a broad space across servers, desktops, laptops, and mobile devices. Phishing attacks (Garera et al. 2007) convince a victim to visit a malicious web page that exploits a vulnerability in the user's web browser or email software. The Heartbleed bug (http://heartbleed.com), a recent attack on the OpenSSL cryptographic software that mostly affected web servers, resulted in the disclosure of private information.

Operating system (OS) attacks can be more difficult to execute because modern operating systems have reasonably good default security practices. Even so, they remain an attractive target for attackers because software running at the operating system's access level can view data from anything on the system, including all applications, files, and memory state. Typically these attacks install *rootkits*, malicious software designed to provide OS-level access for an attacker while hiding from the system's users and applications. Because it can be difficult to compromise operating systems remotely, rootkits are usually installed after a successful application-level attack or by compromising the supply chain.

Hardware attacks are the most advanced. They may involve hardware replacement with a malicious version that appears the same to users but allows an attacker to access the system, or a firmware attack that can change programmable portions of hardware to the attacker's benefit. Physical hardware replacement requires a supply chain attack or an onsite operative. Malicious firmware, however, can be installed from a rootkit to improve stealth and retain access across fresh installations of operating systems.

PLENTIFUL TARGETS

Upon hearing "hardware, operating systems, and applications," many people constrain their thinking to the obvious computers in their lives, such as their laptop and smartphone. But computers are everywhere.

Computers called switches and routers manage network traffic on the Internet and mobile phone networks. Computers are in automobiles (Miller and Valasek 2015), medical devices (FDA 2015), smart watches, TVs, refrigerators—even some light bulbs (McMillan 2014). Each of these represents an attack option for malicious actors, whether as a useful target by itself or as a stepping stone to help attackers reach their ultimate goal(s).

The challenge for security professionals is to understand the broad scope of threats while building systems that provide meaningful security. For example, if one is concerned about network traffic monitoring on the Internet, then a reasonable countermeasure is to encrypt that traffic. Similarly, if one is worried about attackers using guessed or stolen passwords to access an online service, then a reasonable countermeasure is to use two-factor authentication (e.g., a code is sent to the user's phone and must be provided along with the password for access). But when everything is interconnected and the threat landscape is enormous, it is not always obvious what protections are most important to implement.

In addition to rich interconnections, modern computing systems from companies such as Facebook, Google, and Netflix operate at an extremely large scale, with some managing millions of computers (Ballmer 2013). Such a scale shifts the security challenges considerably. Securing each piece of hardware and each operating system requires dedicated teams that must work closely to ensure that security problems do not arise on the boundaries. Even finding the application data that require protection can be challenging. In addition, retrofitting security to an existing large ecosystem requires that the security not disrupt reliability or performance.

CASE STUDY: PROTECTING NETWORK TRAFFIC IN THE CLOUD

When using a web browser, most people know to look for a lock icon indicating a secure connection before making a credit card transaction or other sensitive operation. This security is enabled by a network protocol called *transport layer security* (TLS). It works because there is a preestablished trust relationship between the individual computer and a third party that has verified the authenticity of the visited website.

In addition to protecting the network traffic between a browser and a website, TLS is often used to protect the internal network traffic of a cloud or data center that supports the website's operation. In these cases the technology is the same but the deployment strategies are different to reflect both the scale and the threats across many abstraction layers. The use of TLS in a cloud or data center mitigates threats both from malicious network switches "eavesdropping on" or altering network traffic and from the software virtualization layers used to enable cloud computing.

But deploying TLS at scale for a cloud application presents many challenges. Principal among these is the need to create trust relationships—like those used to make credit card transactions with a web browser—that are reliable and maintainable. This kind of trust relationship is based on a *public key infrastructure* (PKI), which comprises the hardware, software, and policies that manage the encryption that ensures security. Unfortunately, in practice, it can be very difficult to deploy and maintain a PKI; for example, critical security credentials may sometimes be lost or compromised. Traditionally this problem has been addressed using revocation lists, but these lists are difficult to maintain and do not scale well.

The use of short-lived credentials (Topalovic et al. 2012) in lieu of revocation lists has been growing in popularity. This approach can increase software complexity through the automation required to create and deploy the updated credentials (Clark 2014), raising the bar for an attacker and resulting in a more maintainable PKI.

The establishment of PKI and use of TLS to protect network traffic mitigate certain classes of attacks, as described above, but if malicious actors can access other levels in the computing ecosystem they can compromise the very data that TLS is designed to protect.

Security practitioners must estimate whether a particular threat actor is sufficiently capable and motivated to break the assumptions and compromise the system. This is why security is often considered to be both an art and a science.

CHALLENGES AND FUTURE DIRECTIONS

In practice, hardware-level security is a domain that only nation-state actors can afford to engage in both offensively and defensively. In general, the best strategy is to assume that the hardware is secure and leverage some of its security features. Building from this foundation, there are many opportunities to improve security at the OS and application layers.

Services that operate in the cloud today generally need to take a leap of faith and assume that the underlying infrastructure is secure. Often this is a very reasonable assumption because the major cloud providers have thus far demonstrated that they are better at infrastructure security than all but the largest data center operators. However, rather than merely trusting, it would be *much safer to "trust and then verify" that cloud providers are in fact secure*.

Another major challenge in security is simply that it is too difficult to achieve. Security design, including cryptographic engineering and designing to mitigate the impact of malicious adversaries, requires expertise that the average developer cannot possibly be expected to have. Complexity must be reduced at all levels, by *reducing software engineering complexity*, making security *more science and less art*, and providing *secure building blocks for creating complex software*.

These technical challenges are huge and critically important. The only way to solve them is with many talented computer security professionals. Unfortunately, the community has a *massive talent shortage*. As with the other challenges above, solving this will require close partnerships between academia and industry.

REFERENCES

Ballmer S. 2013. Keynote speech. Microsoft Worldwide Partner Conference, Houston, July 8. Available at http://news.microsoft.com/2013/07/08/steve-ballmer-worldwide-partner-conference-2013-keynote/.

Clark R. 2014. SSL everywhere with ephemeral PKI. OpenStack Summit, Paris, November 3–7. Available at https://www.openstack.org/summit/openstack-paris-summit-2014/session-videos/presentation/ssl-everywhere-with-ephemeral-pki.

FDA [US Food and Drug Administration]. 2015. Symbiq Infusion System by Hospira: FDA Safety Communication – Cybersecurity Vulnerabilities. Safety Alert for Human Medical Products, July 31. Silver Spring, MD. Available at http://www.fda.gov/Safety/MedWatch/SafetyInformation/SafetyAlertsforHumanMedicalProducts/ucm456832.htm.

Garera S, Provos N, Chew M, Rubin AV. 2007. A framework for detection and measurement of phishing attacks. Proceedings of the ACM Workshop on Recurring Malcode (WORM), Alexandria, VA, October 29–November 2.

Greenwald G. 2014. No Place to Hide: Edward Snowden, the NSA, and the US Surveillance State. New York: Metropolitan Books.

Kaspersky Lab. 2015. Equation Group: Questions and Answers (Version 1.5). February. Moscow. Available at https://securelist.com/files/2015/02/Equation_group_questions_and_answers.pdf.

McMillan P. 2014. Attacking the Internet of Things using time. DEF CON 22, Las Vegas, August 7–10. Available at https://www.youtube.com/watch?v=uzGXxWuDwxc.

Miller C, Valasek C. 2015. Remote exploitation of an unaltered passenger vehicle. Black Hat USA. August. Available at http://illmatics.com/Remote%20Car%20Hacking.pdf.

Topalovic E, Saeta B, Huang L-S, Jackson C, Boneh D. 2012. Towards short-lived certificates. Proceedings of the Web 2.0 Security & Privacy Workshop (W2SP), San Francisco, May 24. Available at http://www.w2spconf.com/2012/papers/w2sp12-final9.pdf.

Computer Security and Privacy: Where Human Factors Meet Engineering

Franziska Roesner
University of Washington

As the world becomes more computerized and interconnected, computer security and privacy will continue to increase in importance. In this paper I review several computer security and privacy challenges faced by end users of the technologies we build, and considerations for designing and building technologies that better match user expectations with respect to security and privacy. I conclude with a brief discussion of open challenges in computer security and privacy where engineering meets human behavior.

INTRODUCTION

Over the past several decades new technologies have brought benefits to almost all aspects of daily life for people all over the world, transforming how they work, communicate, and interact with each other.

Unfortunately, however, new technologies also bring new and serious security and privacy risks. For example, smartphone malware is on the rise, often tricking unsuspecting users by appearing as compromised versions of familiar, legitimate applications (Ballano 2011); by recent reports, more than 350,000 variants of Android malware have been identified (e.g., Sophos 2014). These malicious applications incur direct and indirect costs by stealing financial and other information or by secretly sending costly premium SMS messages. Privacy is also a growing concern on the web, as advertisers and others secretly track user browsing behaviors, giving rise to ongoing efforts at "do not track" technology and legislation.[1]

[1] Congress introduced, but did not enact, the Do-Not-Track Online Act of 2013, S.418. Available at https://www.congress.gov/bill/113th-congress/senate-bill/418.

Such concerns cast a shadow on modern and emerging computing platforms that otherwise provide great benefits. The need to address these and other computer-related security and privacy risks will only increase in importance as the world becomes more computerized and interconnected. This paper focuses specifically on computer security and privacy at the intersection of human factors and the engineering of new computer systems.

DESIGNING WITH A "SECURITY MINDSET"

It is vital to approach the engineering design process with a "security mindset" that attempts to anticipate and mitigate unexpected and potentially dangerous ways technologies might be (mis)used. Research on computer security and privacy aims to systematize such efforts by (1) studying and developing cryptographic techniques, (2) analyzing or attempting to attack deployed technologies, (3) measuring deployed technical ecosystems (e.g., the web), (4) studying human factors, and (5) designing and building new technologies. Some of these (e.g., cryptography or usable security) are academic subdisciplines unto themselves, but all work together and inform each other to improve the security and privacy properties of existing and emerging technologies.

Many computer security and privacy challenges arise when the expectations of end users do not match the actual security and privacy properties and behaviors of the technologies they use—for example, when installed applications secretly send premium SMS messages or leak a user's location to advertisers, or when invisible trackers observe a user's behavior on the web.

There are two general approaches to try to mitigate these discrepancies. One involves trying to help users change their mental models about the technologies they use to be more accurate (e.g., to help users think twice before installing suspicious-looking applications), by educating them about the risks and/or by carefully designing the user interfaces (UIs) of app stores. Recent work by Bravo-Lillo and colleagues (2013) on designing security-decision UIs to make them more difficult for users to ignore is a nice example of this approach.

The alternative is to *(re)design technologies themselves* so that they better match the security and privacy properties that users intuitively expect, by "maintaining agreement between a system's security state and the user's mental model" (Yee 2004, p. 48).

Both approaches are valuable and complementary. This paper explores the second: designing security and privacy properties in computer systems in a way that goes beyond taking human factors into account to actively remove from users the burden of explicitly managing their security and/or privacy at all. To illustrate the power of this approach, I review research on user-driven access control and then highlight other recent examples.

RETHINKING SMARTPHONE PERMISSIONS WITH USER-DRIVEN ACCESS CONTROL

Consider smartphones (such as iOS, Android, or Windows Phone), other modern operating systems (such as recent versions of Windows and Mac OS X), and browsers. All of these platforms both allow users to install arbitrary applications and limit the capabilities of those applications in an attempt to protect users from potentially malicious or "buggy" applications. Thus, by default, applications cannot access sensitive resources or devices such as the file system, microphone, camera, or GPS. However, to carry out their intended functionality, many applications need access to these resources.

Thus an open question in modern computing platforms in recent years has been: How should untrusted applications be granted permissions to access sensitive resources?

Challenges

Most current platforms explicitly ask users to make the decision about access. For example, iOS prompts users to decide whether an application may access a sensitive feature such as location, and Android asks users to agree to an install-time manifest of permissions requested by an application[2] (Figure 1).

Unfortunately, these permission-granting approaches place too much burden on users. Install-time manifests are often ignored or not understood by users (Felt et al. 2012a), and permission prompts are disruptive to the user's experience, teaching users to ignore and click through them (Motiee et al. 2010). Users thus unintentionally grant applications too many permissions and become vulnerable to applications that use the permissions in malicious or questionable ways (e.g., secretly sending SMS messages or leaking location information).

In addition to outright malware (Ballano 2011), studies have shown that even legitimate smartphone applications commonly leak or misuse private user data, such as by sending it to advertisers (e.g., Enck et al. 2010).

Toward a New Approach

It may be possible to reduce the frequency and extent of an application's illegitimate permissions by better communicating application risks to users (e.g., Kelley et al. 2013) or redesigning permission prompts (e.g., Bravo-Lillo et al. 2013). However, my colleagues and I found in a user survey that people have existing expectations about how applications use permissions; many believe, for example, that an application cannot (or at least will not) access a sensitive resource

[2] Android M will use runtime prompts similar to iOS instead of its traditional install-time manifest.

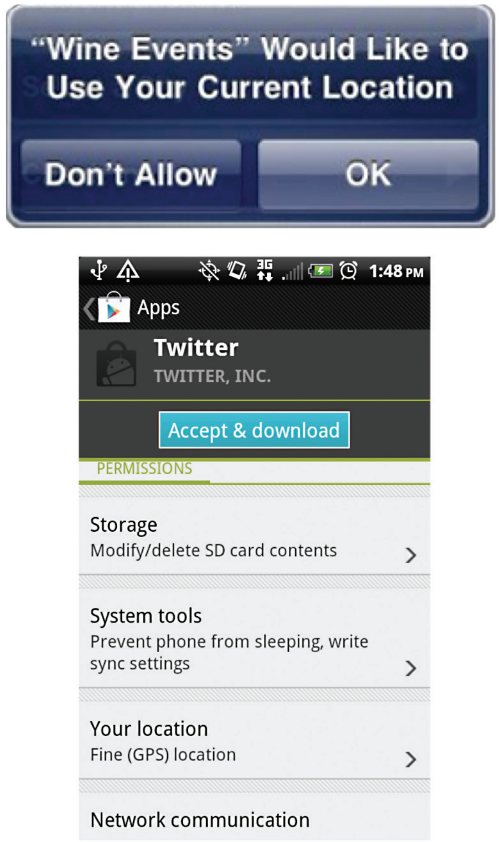

FIGURE 1 Existing permission-granting mechanisms that require the user to make explicit decisions: runtime prompts (as in iOS, above) and install-time manifests (as in Android, below).

such as the camera unless it is related to the user's activities within the application (Roesner et al. 2012b). In reality, however, after being granted the permission to access the camera (or another sensitive resource) once, Android and iOS applications can continue to access it in the background without the user's knowledge.

Access Based on User Intent

This finding speaks for an alternate approach: modifying the system to better match user expectations about permission granting. To that end, we developed *user-driven access control* (Roesner et al. 2012b) as a new model for granting permissions in modern operating systems.

Rather than asking the user to make explicit permission decisions, user-driven access control grants permissions automatically based on existing user actions within applications. The underlying insight is that *users already implicitly indicate the intent to grant a permission through the way they naturally interact with an application.* For example, a user who clicks on the "video call" button in a video chat application implicitly indicates the intent to allow the application to access the camera and microphone until the call is terminated.

If the operating system could interpret the user's permission-granting intent based on this action, then it would not need to additionally prompt the user to make an explicit decision about permissions and it could limit the application's access to a time intended by the user. The challenge, however, is that the operating system cannot by default interpret the user's actions in the custom user interfaces (e.g., application-specific buttons) of all possible applications.

Access Control Gadgets

To allow the operating system to interpret permission-granting intent, we developed *access control gadgets* (ACGs), special, system-controlled user interface elements that grant permissions to the embedding application. For example, in a video chat application, the "video call" button is replaced by a system-controlled ACG. Figure 2 shows additional examples of permission-related UI elements that can be easily replaced by ACGs to enable user-driven access control.

The general principle of user-driven access control has been introduced before (as discussed below), but ACGs make it practical and generalizable to

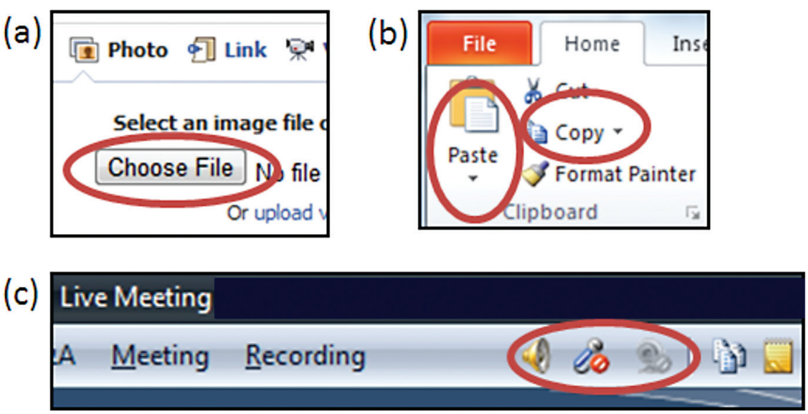

FIGURE 2 Access control gadgets (ACGs) are system-controlled user interface elements that allow the operating system to capture a user's implicit intent to grant a permission (e.g., to allow an application to access the camera).

multiple sensitive resources and permissions, including those that involve the device's location and clipboard, files, camera, and microphone. User-driven access control is powerful because it improves users' security and privacy by changing the system to better match their expectations, rather than the other way around. That is, users' experience interacting with their applications is unchanged while the underlying permissions granted to applications match the users' expectations.

OTHER EXAMPLES

User-driven access control follows philosophically from Yee (2004) and a number of other works. CapDesk (Miller 2006) and Polaris (Stiegler et al. 2006) were experimental desktop computer systems that applied a similar approach to file system access, giving applications minimal privileges but allowing users to grant applications permission to access individual files via a "powerbox" user interface (essentially a secure file picking dialog). Shirley and Evans (2008) proposed a system (prototyped for file resources) that attempts to infer a user's access control intent from the history of user behavior. And BLADE (Lu et al. 2010) attempts to infer the authenticity of browser-based file downloads using similar techniques. Related ideas have recently appeared in mainstream commercial systems, including Mac OS X and Windows 8,[3] whose file picking designs also share the underlying user-driven access control philosophy.

Note that automatically managing permissions based on a user's interactions with applications may not always be the most appropriate solution. Felt and colleagues (2012b) recommended combining a user-driven access control approach for some permissions (e.g., file access) with other approaches (e.g., prompts or post facto auditing) that work more naturally for other permissions or contexts. The approach of designing systems to better and more seamlessly meet users' security and privacy expectations is much more general than the challenges surrounding application permissions discussed above.

Following are a few of the systems that are intentionally designed to better and more automatically match users' security and privacy expectations. Such efforts are often well complemented by work that attempts to better communicate with or educate users by changing the designs of user interfaces.

Personal Communications

To secure communications between two parties, available tools such as PGP have long faced usability challenges (Whitten and Tygar 1999). Efforts

[3] Apple posted information in 2011 on the "App sandbox and the Mac app store" (available at https://developer.apple.com/videos/wwdc/2011/), and Windows information on "Accessing files with file pickers" is available at http://msdn.microsoft.com/en-us/library/windows/apps/hh465174.aspx.

that remove the burden from users while providing stronger security and privacy include Vanish (Geambasu et al. 2009), which supports messages that automatically "disappear" after a period of time; ShadowCrypt (He et al. 2014), which replaces existing user input elements on websites with ones that transparently encrypt and later decrypt user input; and Gyrus (Jang et al. 2014), which ensures that only content that a user has intended to enter into a user input element is what is actually sent over the network ("what you see is what you send").

Commercially, communication platforms such as email and chat are increasingly moving toward providing transparent end-to-end encryption (e.g., Somogyi 2014), though more work remains to be done (Unger et al. 2015). For example, the security of journalists' communications with sensitive sources has come into question in recent years and requires a technical effort that provides low-friction security and privacy properties to these communications (McGregor et al. 2015).

User Authentication

Another security challenge faced by many systems is that of user authentication, which is typically handled with passwords or similar approaches, all of which have known usability and/or security issues (Bonneau et al. 2012). One approach to secure a user's accounts is two-factor authentication, in which a user must provide a second identifying factor in addition to a password (e.g., a code provided by an app on the user's phone). Two-factor authentication provides improved security and is seeing commercial uptake, but it decreases usability; efforts such as PhoneAuth (Czeskis et al. 2012) aim to balance these factors by using the user's phone as a second factor only opportunistically when it happens to be available.

Online Tracking

As a final example, user expectations about security and privacy do not match the reality of today's systems with respect to privacy on the web. People's browsing behaviors are invisibly tracked by third-party advertisers, website analytics engines, and social media sites.

In earlier work we discovered that social media trackers, such as Facebook's "Like" or Twitter's "tweet" button, represent a significant fraction of trackers on popular websites (Roesner et al. 2012a). To mitigate the associated privacy concerns, we applied a user-driven access control design philosophy to develop ShareMeNot, which allows tracking *only when* the user clicks the associated social media button. ShareMeNot's techniques have been integrated into the Electronic Frontier Foundation's Privacy Badger tool (https://www.eff.org/privacybadger), which automatically detects and selectively blocks trackers without requiring explicit user input.

CHALLENGES FOR THE FUTURE

New technologies are improving and transforming people's lives, but they also bring with them new and serious security and privacy concerns. Balancing the desired functionality provided by increasingly sophisticated technologies with security, privacy, and usability remains an important challenge.

This paper has illustrated efforts across several contexts to achieve this balance by designing computer systems that remove the burden from the user. However, more work remains to be done in all of these and other domains, particularly in emerging areas that rely on ubiquitous sensors, such as the augmented reality technologies of Google Glass and Microsoft HoloLens.

By understanding and anticipating these challenges early enough, and by applying the right insights and design philosophies, it will be possible to improve the security, privacy, and usability of emerging technologies before they become widespread.

REFERENCES

Ballano M. 2011. Android threats getting steamy. Symantec Official Blog, February 28. Available at http://www.symantec.com/connect/blogs/android-threats-getting-steamy.

Bonneau J, Herley C, van Oorschot PC, Stajano F. 2012. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. Proceedings of the 2012 IEEE Symposium on Security and Privacy. Washington: IEEE Society.

Bravo-Lillo C, Cranor LF, Downs J, Komanduri S, Reeder RW, Schechter S, Sleeper M. 2013. Your attention please: Designing security-decision UIs to make genuine risks harder to ignore. Proceedings of the Symposium on Usable Privacy and Security (SOUPS), July 24–26, Newcastle, UK.

Czeskis A, Dietz M, Kohno T, Wallach D, Balfanz D. 2012. Strengthening user authentication through opportunistic cryptographic identity assertions. Proceedings of the 19th ACM Conference on Computer and Communications Security, October 16–18, Raleigh, NC.

Enck W, Gilbert P, Chun B, Cox LP, Jung J, McDaniel P, Sheth AN. 2010. TaintDroid: An information-flow tracking system for realtime privacy monitoring on smartphones. Proceedings of the 9th USENIX Conference on Operating System Design and Implementation. Berkeley: USENIX Association.

Felt AP, Ha E, Egelman S, Haney A, Chin E, Wagner D. 2012a. Android permissions: User attention, comprehension, and behavior. Proceedings of the Symposium on Usable Privacy and Security (SOUPS), July 11–13, Washington, DC.

Felt AP, Egelman S, Finifter M, Akhawe D, Wagner D. 2012b. How to ask for permission. Proceedings of the 7th Workshop on Hot Topics in Security (HotSec), August 7, Bellevue, WA.

Geambasu R, Kohno T, Levy A, Levy HM. 2009. Vanish: Increasing data privacy with self-destructing data. Proceedings of the 18th USENIX Security Symposium. Berkeley: USENIX Association.

He W, Akhawe D, Jain S, Shi E, Song D. 2014. ShadowCrypt: Encrypted web applications for everyone. Proceedings of the ACM Conference on Computer and Communications Security. New York: Association for Computing Machinery.

Jang Y, Chung SP, Payne BD, Lee W. 2014. Gyrus: A framework for user-intent monitoring of text-based networked applications. Proceedings of the Network and Distributed System Security Symposium (NDSS), February 23–26, San Diego.

Kelley PG, Cranor LF, Sadeh N. 2013. Privacy as part of the app decision-making process. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery.

Lu L, Yesneswaran V, Porras P, Lee W. 2010. BLADE: An attack-agnostic approach for preventing drive-by malware infections. Proceedings of the 17th ACM Conference on Computer and Communications Security. New York: Association for Computing Machinery.

McGregor SE, Charters P, Holliday T, Roesner F. 2015. Investigating the computer security practices and needs of journalists. Proceedings of the 24th USENIX Security Symposium. Berkeley: USENIX Association.

Miller MS. 2006. Robust composition: Towards a unified approach to access control and concurrency control. PhD thesis, Johns Hopkins University.

Motiee S, Hawkey K, Beznosov K. 2010. Do Windows users follow the principle of least privilege? Investigating user account control practices. Proceedings of the 6th Symposium on Usable Privacy and Security (SOUPS). New York: Association for Computing Machinery.

Roesner F, Kohno T, Wetherall D. 2012a. Detecting and defending against third-party tracking on the web. Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation, April 25–27, San Jose.

Roesner F, Kohno T, Moshchuk A, Parno B, Wang HJ, Cowan C. 2012b. User-driven access control: Rethinking permission granting in modern operating systems. Proceedings of the IEEE Symposium on Security and Privacy, May 20–23, San Francisco.

Shirley J, Evans D. 2008. The user is not the enemy: Fighting malware by tracking user intentions. Proceedings of the Workshop on New Security Paradigms. New York: Association for Computing Machinery.

Somogyi S. 2014. An update to end-to-end. Google Online Security Blog, December 16. Available at http://googleonlinesecurity.blogspot.com/2014/12/an-update-to-end-to-end.html.

Sophos. 2014. Our top 10 predictions for security threats in 2015 and beyond. Blog, December 11. Available at https://blogs.sophos.com/2014/12/11/our-top-10-predictions-for-security-threats-in-2015-and-beyond/.

Stiegler M, Karp AH, Yee K-P, Close T, Miller MS. 2006. Polaris: Virus-safe computing for Windows XP. Communications of the ACM 49:83–88.

Unger N, Dechand S, Bonneau J, Fahl S, Perl H, Goldberg I, Smith M. 2015. SoK: Security messaging. Proceedings of the IEEE Symposium on Security and Privacy, May 17–21, San Jose.

Whitten A, Tygar JD. 1999. Why Johnny can't encrypt: A usability evaluation of PGP 5.0. Proceedings of the USENIX Security Symposium 8:14.

Yee K-P. 2004. Aligning security and usability. IEEE Security and Privacy 2(5):48–55.

On the Technical Debt of Medical Device Security

Kevin Fu
University of Michigan

Cybersecurity shortfalls in medical devices trace to decisions made during early engineering and design. The industry is now paying the cybersecurity "technical debt" for this shortsightedness.

INTRODUCTION

Computer networking, wireless communication, wireless power, the Internet, and a host of other engineering innovations, combined with electronic health records and the reengineering of clinical workflow, have enabled innovative therapeutics and diagnostics—but at the cost of information security and privacy, or *cybersecurity*.

Complexity breeds technological insecurity. In the past few decades, medical devices have evolved from simple analog components to complex digital systems with an amalgam of software, circuits, and advanced power sources that are much more difficult to validate and verify. Whereas a classic stethoscope depended on well-understood analog components, modern devices such as linear accelerators, pacemakers, drug infusion pumps, and patient monitors depend critically on computer technology.

When a medical device is compromised, its behavior becomes unpredictable: the device may deliver incorrect diagnostic information to clinicians, be unavailable to deliver patient care during repair, and, in extreme cases, cause patient harm. Lack of security may even introduce an unconventional dimension of risk to safety and effectiveness: intentional harm.

Much cybersecurity risk is attributable to legacy medical devices dependent on Windows XP and other unmaintainable operating systems that no longer

FIGURE 1 An old drug compounder running on a consumer operating system, Windows XP Embedded, was infected with malware, according to a US Food and Drug Administration Manufacturer and User Facility Device Experience (MAUDE) adverse event report in 2010. A former engineer from the company later explained that the malware was accidentally spread to other compounders during the repair. © Kevin Fu.

receive security patches (Figure 1). But proprietary embedded systems are no less vulnerable.

Complexity introduced at the design stage is the root cause of many cybersecurity problems, not hackers. Complexity increases the *attack surface*, the points of unintended access to a computer system. By uncovering the implications of the flaws baked in from early engineering choices, hackers are merely the “collectors” and messengers of this cybersecurity technical debt.

BRIEF HISTORY OF MEDICAL DEVICE SECURITY

Research: Case Studies

There is a rich history of efforts to ensure trustworthy medical device software (Fu 2011). The classic and eye-opening Therac-25 study showed how a linear accelerator caused a number of injuries and deaths from massive radiation overdoses in the late 1980s and early 1990s (Leveson and Turner 1993). While project mismanagement, complacency, and overconfidence in unrealistic probabilities played a role, the most interesting root cause was the adoption of poorly designed software instead of well-understood analog components to safely control the radiation delivery.

More recently, research on ways to improve medical device security led to an interdisciplinary paper on the security of an implantable cardiac defibrillator (Halperin et al. 2008). The study took several years because of the interdisciplinary nature of the problem and clinical challenges such as attending live surgery to fully understand the threat model. In the paper my colleagues and I demonstrated that it was possible to wirelessly disable the device's life-saving shocks and induce ventricular fibrillation (a deadly heart rhythm).[1]

We articulated the engineering principle that a secure medical device should not be able to run an operation that causes it to induce the hazardous state (in this instance, ventricular fibrillation) it is designed to prevent. The paper also includes a number of defensive approaches primarily centered on the concept of zero-power security, requiring the provision of wireless power to ensure that the implant can protect the availability of its precious battery power. And, importantly, our research showed that, despite the security risks, patients predisposed to health risks who are prescribed a wireless medical device are far safer accepting the device than not.

The Role of Hackers

A few years later, the hacker community began to replicate academic experiments on medical devices. Barnaby Jack famously replicated our pacemaker/defibrillator experiment in a manner more appealing to the general public. Although formal peer-reviewed proceedings were rare, the hackers gave captivating talks and pointed demonstrations that attracted attention to the subject. The hacker community began to find new security flaws in medical devices such as insulin pumps and infusion pumps (e.g., demonstrations by Billy Rios, Barnaby Jack, Jay Radcliffe, Scott Erven, and others). The hacker community uncovered security vulnerabilities that have led to unprecedented FDA actions.

National Facilities for Medical Device Security

To promote deeper intellectual inquiry into medical device security, I created the Open Medical Device Research Library (OMDRL) to collect and share hard-to-find implants with security researchers. Unfortunately, the demand did not justify the high cost of biohazard decontamination, and computer science staff were uncomfortable with managing biohazard facilities, so the library was short lived. However, researchers from the Massachusetts Institute of Technology did engage with the OMDRL to invent a novel radio frequency (RF) jamming protocol that blocks legacy implanted cardiac devices from transmitting insecure "plaintext" messages and overlays an encrypted version (Gollakota et al. 2011).

[1] The device has not been sold for several years, and the manufacturer established a rigorous training program for security engineering.

Device manufacturers have difficulty testing beyond the component level because of (1) the diverse array of configurations and interoperating medical devices and (2) uncertain risk to patients during live testing. For this reason, the OMDRL is adapting from a library to a testbed at the University of Michigan.

The notional Bring Your Own Hospital (BYOH) testbed, as part of the Archimedes Center for Medical Device Security (secure-medicine.org), will enable security testing and experimentation on systems of medical devices with automated and highly configurable threat simulators to better prepare manufacturers and hospitals to cope with the changing threat landscape. Efforts will include control studies to compare the effectiveness of different hospital information security policies, and emergency preparedness "fire drills" to train manufacturers and clinicians on how to respond to cyberattacks and malware infections that affect the timely delivery of care. The first experiment involves mapping out the infection vectors created by reuse of USB drives to understand how fast infections can spread in clinical facilities and determine the most effective ways to control an outbreak.

RECENT FEDERAL AND OTHER MEASURES

In July 2015, a few days after the National Highway Traffic Safety Administration issued the first recall of an automobile solely because of a cybersecurity risk (Kessler 2015), Hospira became the first medical device company to receive an FDA (2015) safety communication because of a cybersecurity risk. Although not legally a recall, the FDA notice had a similar effect: the agency strongly discouraged healthcare facilities from purchasing the company's infusion pump because of a cybersecurity vulnerability that could let hackers induce over- or underinfusion of drugs and thus potentially cause patient harm.

In addition, FDA (2014) premarket guidance on cybersecurity calls for a technical cybersecurity risk analysis in all applications for premarket clearance to sell medical devices in the United States. In addition, the FDA is expected to release a postmarket guidance document on coordinated vulnerability disclosure, incident reporting, and continuous surveillance of emerging cybersecurity risks. The preparation of this document is more complicated because it involves a number of unusual bedfellows, ranging from the vulnerability research community to the Department of Homeland Security to the US Computer Emergency Response Team.

Complementing these federal measures, the Association for the Advancement of Medical Instrumentation (AAMI) sets the major standards for medical device safety. The AAMI medical device security working group consists of both healthcare providers and medical device engineers who have written a technical information report (TIR 57) (currently under ballot) that provides much-needed advice to engineers on how to think methodically about cybersecurity across the product development lifecycle of a medical device.

FIGURE 2 My former students and I cofounded Virta Labs, Inc. to detect malware and anomalies on medical devices using a smart power outlet rather than installed software. Machine learning and side channels in reactive AC power enable detection in a manner that does not cause alarm fatigue (Clark et al. 2013).

ANALOG CYBERSECURITY

Many of the vulnerabilities and solutions (Figure 2) for medical device security involve analog cybersecurity. Cybersecurity risks that begin in the analog world can infect the digital world by exploiting semipermeable digital abstractions. Analog cybersecurity is the focus of research on side channels and fault injection attacks that transcend traditional boundaries of computation. Security problems tend to occur in boundary conditions where different abstractions meet. In particular, the analog-digital abstraction poses subtle security weaknesses for cyberphysical systems such as medical devices and the Internet of Things.

Researchers have demonstrated how an adversary can violate widely held computing abstractions as fundamental as the value of a bit. For example, ionizing radiation and computing faults cause smartcards and processors to divulge cryptographic secrets (Boneh et al. 2001; Pellegrini et al. 2010). Intentional electromagnetic interference causes sensors to deliver incorrect digital values to closed-loop feedback systems such as pacemakers (Kune et al. 2013). Acoustic and mechanical vibrations cause drones to malfunction by hitting the resonant frequency of a microelectromechanical systems (MEMS) gyroscope (Son et al. 2015). The row hammer attack[2] enabled malicious flipping of bits in computer

[2] Project Zero: Exploiting the DRAM rowhammer bug to gain kernel privileges, March 9, 2015. Available at http://googleprojectzero.blogspot.com/2015/03/exploiting-dram-rowhammer-bug-to-gain.html.

memory to adjacent physical rows of memory (Kim et al. 2014). The GSMem paper (Guri et al. 2015) shows how computer memory can emit RF signals in cellular frequencies.

Such analog cybersecurity weaknesses will likely remain a significant challenge for automated systems such as medical devices. Traditional research and education in cybersecurity focus on software flaws and solutions. I believe that threats to the digital abstraction represent the next frontier of security engineering. The Internet of Things and cyberphysical systems such as medical devices, automobiles, and aircraft have awakened interest in analog threats affecting digital vulnerabilities that have physical consequences.

CONCLUSION

Medical devices help patients lead more normal and healthy lives. The innovation of such devices results from a complex interplay of medicine, computer engineering, computer science, human factors, and other disciplines, and this complexity breeds design-induced cybersecurity risks.

The greatest near-term risk is old malware that accidentally breaks into clinical systems running old operating systems, causing damage to the integrity and availability of medical devices and in turn interrupting clinical workflow and patient care. While targeted malware will likely become a problem in the future, medical devices are currently challenged by basic cybersecurity hygiene, such as hospitals spreading malware with USB drives or vendors infecting their own products by accident.

To enhance the trustworthiness of emerging medical devices and patients' confidence in them, manufacturers need to address cybersecurity risks during the initial engineering and design, and maintain postmarket surveillance throughout the product lifecycle.

ACKNOWLEDGMENTS

This work is supported in part by the Archimedes Center for Medical Device Security and the National Science Foundation under the Trustworthy Health and Wellness project (THAW.org; award CNS-1330142). Any opinions, findings, and conclusions expressed in this material are those of the author and do not necessarily reflect the views of the NSF. For the full-length technical report, contact the author (kevinfu@umich.edu).

REFERENCES

Boneh D, Demillo RA, Lipton RJ. 2001. On the importance of eliminating errors in cryptographic computations. Journal of Cryptology 14:101–119.

Clark SS, Ransford B, Rahmati A, Guineau S, Sorber J, Xu W, Fu K. 2013. WattsUpDoc: Power side channels to nonintrusively discover untargeted malware on embedded medical devices. Proceedings of the USENIX Workshop on Health Information Technologies, August 12, Washington.

FDA [US Food and Drug Administration]. 2014. Content of Premarket Submissions for Management of Cybersecurity in Medical Devices: Guidance for Industry and Food and Drug Administration Staff. Silver Spring, MD: Center for Devices and Radiological Health and Center for Biologics Evaluation and Research. Available at http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm356190.pdf.

FDA. 2015. Cybersecurity vulnerabilities of Hospira Symbiq infusion system. FDA Safety Communication, July 31. Silver Spring, MD. Available at http://www.fda.gov/MedicalDevices/Safety/AlertsandNotices/ucm456815.htm.

Fu K. 2011. Appendix D: Trustworthy medical device software. In: Public Health Effectiveness of the FDA 510(k) Clearance Process: Measuring Postmarket Performance and Other Select Topics: Workshop Report. Washington, DC: National Academies Press.

Gollakota S, Hassanieh H, Ransford B, Katabi D, Fu K. 2011. They can hear your heartbeats: Non-invasive security for implanted medical devices. ACM SIGCOMM Computer Communication Review 41(4):2–13.

Guri M, Kachlon A, Hasson O, Kedma G, Mirsky Y, Elovici Y. 2015. SMem: Data exfiltration from air-gapped computers over GSM frequencies. Proceedings of the 24th USENIX Security Symposium, August 12–14, Washington, DC. pp. 849–864.

Halperin D, Heydt-Benjamin TS, Ransford B, Clark SS, Defend B, Morgan W, Fu K, Kohno T, Maisel WH. 2008. Pacemakers and implantable cardiac defibrillators: Software radio attacks and zero-power defenses. Proceedings of the 29th Annual IEEE Symposium on Security and Privacy, May 18–22, Oakland. pp. 129–142.

Kessler AM. 2015. Fiat Chrysler issues recall over hacking. New York Times, July 24. Available at http://www.nytimes.com/2015/07/25/business/fiat-chrysler-recalls-1-4-million-vehicles-to-fix-hacking-issue.html.

Kim Y, Daly R, Kim J, Fallin C, Lee JH, Lee D, Wilkerson C, Lai K, Mutlu O. 2014. Flipping bits in memory without accessing them: An experimental study of DRAM disturbance errors. Proceeding of the 41st Annual International Symposium on Computer Architecture (ISCA '14), June 14–18, Minneapolis. Washington, DC: IEEE Press. pp. 361–372.

Kune DF, Backes J, Clark SS, Kramer DB, Reynolds MR, Fu K, Kim Y, Xu W. 2013. Ghost talk: Mitigating EMI signal injection attacks against analog sensors. Proceedings of the 34th Annual IEEE Symposium on Security and Privacy, May. Washington, DC: IEEE Computer Society. pp. 145–159.

Leveson N, Turner C. 1993. An investigation of the Therac-25 accidents. IEEE Computer 26(7):18–41.

Pellegrini A, Bertacco V, Austin T. 2010. Fault-based attack of RSA authentication. Proceedings of the Conference on Design, Automation, and Test in Europe (DATE), March 8–12, Dresden. pp. 855–860.

Son Y, Shin H, Kim D, Park Y, Noh J, Choi K, Choi J, Kim Y. 2015. Rocking drones with intentional sound noise on gyroscopic sensors. Proceedings of the 24th USENIX Security Symposium, August 12–14, Washington, DC. pp. 881–896.

Challenges of Engineering Cybersecurity: A Government Perspective

TOMAS VAGOUN
National Coordination Office for Networking and Information Technology R&D

The US government is a principal source of funding for basic research in cybersecurity, and as such is in a position to direct research on fundamental issues in cybersecurity toward novel and game-changing solutions. Among the federal strategic cybersecurity research themes, Moving Target Defense and Science of Security are great examples of engineering- and science-based efforts to significantly improve the security of information technology (IT) systems.

CALL FOR GAME-CHANGING CYBERSECURITY RESEARCH

The nation's security, economic progress, and social fabric are now inseparably dependent on cyberspace. But the digital infrastructure and its foundations are not secure. Cybervulnerabilities can be exploited by criminals for illicit financial gains, by state-sponsored mercenaries to compromise national security interests, and by terrorist groups to cause large-scale disruptions in critical national infrastructures.

The status quo is unacceptable. Recognizing this problem, the federal government has been a champion of high-risk, high-payoff cybersecurity research. Its strategy, set forth in *Trustworthy Cyberspace: Strategic Plan for the Federal Cybersecurity Research and Development Program* (NSTC 2011), directs federal agencies and challenges the research community at large to pursue game-changing advances in cybersecurity.

MOVING TARGET DEFENSE

In the current environment, cyberattackers win by taking advantage of the relatively static nature of systems. They can plan at their leisure, relatively safe in the assumption that key IT assets will look the same for a long time. They can map out likely responses and stockpile a set of exploits that escalate in sophistication as better defenses are deployed. They can afford to invest significant resources in their attacks because they expect to persist for a long time and reuse the attacks across many targets.

To reverse this asymmetry, it is essential to decrease both the predictability of systems and the return on investment for developing and executing attacks. A cyberterrain that is made to appear chaotic to attackers will force them to do reconnaissance and launch exploits anew for every desired penetration—ideally, they will enjoy no amortization of development costs.

The federal cybersecurity R&D community has proposed the development of such capabilities under the rubric of Moving Target Defense (MTD). This strategy calls for the development of technologies such as nonpersistent execution environments, randomized execution of code, randomized network and host identities, randomizing compilers, dynamic address spaces, and automated patch synthesis and installation.

Many natural systems are far more complex than cybersystems but nonetheless extremely robust, resilient, and effective. The biological immune system, for example, functions remarkably well in distributed, complex, and ever-changing environments, even when subject to a continuous barrage of attacks. Immune systems exhibit a wealth of interesting mechanisms that could be the inspiration for new methods relevant to MTD objectives, such as distributed processing, pathogenic pattern recognition, multilayered protection, decentralized control, diversity, and signaling. Designing and developing computing systems that implement such capabilities could bring about game-changing advances in cybersecurity.

DARPA CRASH PROGRAM

Announced in 2010 and ending in 2015, the Clean-Slate Design of Resilient, Adaptive, Secure Hosts (CRASH) Program of the Defense Advanced Research Projects Agency (DARPA) promotes novel ways of thinking about enhancing computing system security, taking inspiration from immune systems. The objective is to design systems that can adapt to continue rendering useful services after a successful attack, learn from previous attacks, and repair themselves after the attack.

The program's multipronged approach looks at hardware, programming languages, operating systems, and theorems.

Hardware was designed to enforce operating rules by tagging every individual piece of data with its type, size, and ownership to enforce access and use restrictions on data at the hardware level.

Newly developed programming languages are explicit about information flows and access control rights. These languages allow programmers to state exactly what rules apply to each module of code, and the operating systems enforce these rules dynamically when the program runs.

Similarly, a new type of operating system has been developed based on a large number of cooperative but mutually independent modules. Each module is designed with a specific purpose and the lowest level of access privileges needed. The modules are also designed to be suspicious of each other, checking one another's results to make sure they conform to the rules and policies that govern them. This creates a system where more than one component would have to be specifically compromised for an attacker to succeed.

When these self-monitoring systems detect a violation, they invoke built-in system services that attempt to diagnose the problem, using replay and reasoning techniques to isolate and characterize it; recover from the problem by having multiple redundant methods to achieve any given goal; synthesize filters to detect the same type of attack in the future and prevent it from succeeding; and automatically generate a patch to fix the underlying vulnerability.

The DARPA CRASH program successfully demonstrated that it is possible to develop significantly more secure computing systems that incorporate game-changing ideas that address core deficiencies of today's cyberspace, as summarized in Table 1.

SCIENCE OF SECURITY

Prioritized by the federal cybersecurity R&D strategy and supported by research funding from a number of federal agencies, MTD has become an active area of R&D. At least 40 moving target techniques have been proposed, at all levels of a computing system—hardware, operating system, applications, network, and system of systems (for examples see Okhravi et al. 2013).

While the techniques propose innovative approaches to increasing agility, diversity, and redundancy of computing systems, and hence increase attackers' workload and decrease their return on investment, MTD techniques are subject to the same limitations as others: lack of knowledge about how to systematically assess the efficacy of security techniques, how to measure security benefits, how to compare different techniques, or how to provably determine the security characteristics of the techniques.

MTD techniques can make systems appear chaotic and unpredictable to attackers, but they do so at the cost of increased complexity. What are the best ways to assess whether the benefits outweigh the costs? Some approaches have been proposed—for example, incorporating MTD into formal security models such as the Hierarchical Attack Representation Model (HARM; Hong and Kim 2015)—but it remains to be seen whether they provide the foundations necessary for formally assessing MTD.

TABLE 1 Cybersecurity Improvements Developed Under the DARPA Clean-Slate Design of Resilient, Adaptive, Secure Hosts (CRASH) Program Based on Aspects of Biological Immunity.

Cybersecurity problem	Biological approach	DARPA CRASH innovation
Systems are easily penetrated	Innate immunity • Fast-reacting defenses to known pathogens	New hardware and OS that eliminate common technical vulnerabilities. Examples: • CHERI (Capability Hardware Enhanced RISC Instructions): hardware-supported, in-process memory protection and sandboxing (Watson et al. 2015) • TESLA (Temporally Enforced Security Logic Assertions): compiler-generated runtime instrumentation for continuous validation of security properties (Anderson et al. 2014)
Cleanup and repair are slow, unpredictable, and costly	Adaptive immunity • Slower-reacting defenses to unknown pathogens • Learning and adaptation	Adaptive software that determines causes of vulnerabilities and dynamically repairs flaws. Example: • GenProg: genetic programming for automated software repairs (Le Goues et al. 2012)
Computing homogeneity • Large pool of targets, large return on investment for attackers • No enterprise-wide survivability	Diversity • Sustains population survival	Techniques that increase entropy, make systems unique, and raise work factor for attackers: instruction set randomization, address space randomization, functional redundancy. Example: • Advanced Adaptive Application (A3) Environment (Pal et al. 2014)

The inability to assess the strengths and weaknesses of security measures, MTD or otherwise, in a systematic, measurable, and repeatable manner, points to a fundamental weakness: There is no foundation to ground the development of secure systems in a rigorous and scientific approach that would facilitate the discovery of laws, hypothesis testing, repeatable experiments, standardized metrics, and common terminology. The lack of scientific foundations is a critical problem and barrier to achieving effective and sustained improvements in cybersecurity. Nurturing the development of a science of security is therefore another key objective of the federal cybersecurity R&D strategy.

The most focused science-of-security research initiative funded by the federal government is the set of Science of Security Lablets, funded by the National Security Agency and launched in 2012. Four universities—Carnegie Mellon University, University of Illinois at Urbana-Champaign, North Carolina State

University, University of Maryland—were selected to lead research and education projects specifically aimed at investigating scientific foundations of cybersecurity. The projects are initially targeting five areas of interest: resilient architectures, scalability and composability, secure collaboration, metrics, and human behavior.

The growing emphasis on the science of security is strengthening foundations of security across many areas, including MTD. Efforts to develop and evaluate MTD techniques from a theoretical basis are growing, including, for example, a project that assesses how MTD techniques increase a system's entropy and decrease the predictability of its behavior (Zhuang et al. 2014).

SUMMARY

Dependence on cyberinfrastructure is far too great to hope that incremental enhancements will bring about substantial security improvements. In the absence of market-driven solutions, the federal government has initiated high-risk/high-payoff R&D programs that focus on game-changing advances in security. The government's strategy of MTD techniques and the development of the field of science of security show promising results in both areas.

REFERENCES

Anderson J, Watson R, Chisnall D, Gudka K, Davis B, Marinos I. 2014. TESLA: Temporally Enhanced System Logic Assertions. Proceedings of the 2014 European Conference on Computer Systems (EuroSys 2014), April 14–16, Amsterdam, Article No. 19.

Hong JB, Kim DS. 2015. Assessing the effectiveness of moving target defenses using security models. IEEE Transactions on Dependable and Secure Computing 99.

Le Goues C, Dewey-Vogt M, Forrest S, Weimer W. 2012. A systematic study of automated program repair: Fixing 55 out of 105 bugs for $8 each. Proceedings of the 2012 International Conference on Software Engineering (ICSE), June 2–9, Zurich, pp. 3–13.

NSTC [National Science and Technology Council]. 2011. Trustworthy Cyberspace: Strategic Plan for the Federal Cybersecurity Research and Development Program. Washington, DC. Available at http://www.nitrd.gov/subcommittee/csia/fed_cybersecurity_rd_strategic_plan_2011.pdf.

Okhravi H, Rabe MA, Mayberry TJ, Leonard WG, Hobson TR, Bigelow D, Streilein WW. 2013. Survey of cyber moving targets. Technical Report 1166, ESC-EN-HA-TR-2012-109, MIT Lincoln Laboratory. Available at http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA591804.

Pal P, Schantz R, Paulos A, Benyo B. 2014. Managed execution environment as a Moving-Target Defense infrastructure. IEEE Security and Privacy 12(2):51–59.

Watson R, Woodruff J, Neumann PG, Moore SW, Anderson J, Chisnall D, Dave N, Davis B, Gudka K, Laurie B, Murdoch SJ, Norton R, Roe M, Son S, Vadera M. 2015. CHERI: A hybrid capability-system architecture for scalable software compartmentalization. Proceedings of the 2015 IEEE Symposium on Security and Privacy, May 18–20, San Jose, pp. 20–37.

Zhuang R, DeLoach SA, Ou X. 2014. Towards a theory of moving target defense. Proceedings of the First ACM Workshop on Moving Target Defense, November 3, Scottsdale, pp. 31–40.

Engineering the Search for Earth-like Exoplanets

Engineering the Search for Earth-like Exoplanets

Sara Seager
Massachusetts Institute of Technology

Mitchell Walker
Georgia Institute of Technology

Every star in the sky is a sun, and if Earth's Sun has planets then it seems logical that other stars should have planets also. And they do: in the past two decades, astronomers have found thousands of "exoplanets" orbiting stars other than the Sun. Surprisingly, no solar system copies have yet been found; instead, an incredible diversity of exoplanets and planetary systems has been uncovered. Extrapolating from these discoveries, every star in the Milky Way galaxy should have at least one planet. With hundreds of billions of stars in the Milky Way and upwards of hundreds of billions of galaxies in the universe, the chance for one of those planets to be Earth-like should be a near certainty.

But currently available techniques to find and study small rocky planets have measurement capabilities that limit the detectable planet size (or mass) and orbit. In other words, similar-sized planets such as Venus (with a scorching surface hot enough to melt lead) and Earth (with a clement surface and temperatures supportive of a liquid water ocean and suitable for life) would look the same to current observational capabilities. The ability to observe rocky exoplanet atmospheres and detect biosignature gases—gases produced by life that accumulate in the atmosphere to detectable levels—is a prime goal of the search for other Earths.

A new generation of space-based telescope is needed to find and identify an Earth-like exoplanet (even though such a telescope will have only the nearest stars within reach). It must operate above the blurring effects of Earth's atmosphere. Moreover, the signal of an Earth-like planet orbiting a nearby sun-like star is so dim that less than one visible-light photon would strike the telescope's primary mirror each second for a mirror 10 m in diameter (the larger the aperture, the more stars are accessible and the higher the chance of finding an exo-Earth). But many billions of photons from that planet's host star will flood the telescope in that same second, requiring precise separation, suppression, and/or shadowing of the star if the Earth-like exoplanet is to be detected and studied.

Ultraprecise starlight suppression and the deployment or construction of large optical telescopes in space depend on advanced engineering to enable the search for Earth-like exoplanets. The presenters in this session described aspects of this advanced engineering. Amy Lo (Northrop Grumman) set the stage for large space telescopes by describing the James Webb Space Telescope (JWST), an international, NASA-led mission to be launched in 2018. JWST has nearly four times the collecting area of the Hubble Space Telescope and is cryogenically cooled to detect infrared wavelengths. She provided the industry perspective for large civilian space missions. Next, Dmitry Savransky (Cornell University) explained the two main techniques for starlight suppression: the internal occulter, or coronagraph, that blocks light inside the telescope and works with wavefront sensing and control to create a stable optical system; and the external occulter, or starshade, a specially shaped screen tens of meters in diameter that formation flies tens of thousands of kilometers from the telescope, blocking out the star light so that only planet light enters the telescope. Jeremy Banik then addressed the construction of large structures in space, from large deployables to space-based assembly and construction. He leads the Large Deployable Structures Technology Thrust Area at the Air Force Research Laboratory Space Vehicles Directorate at Kirtland Air Force Base. The final speaker, Jonathan Black (Virginia Tech), presented the cutting edge in sensing controls for formation flying and satellite proximity operations (primarily to enable autonomy for small satellites) and discussed challenges in the advances needed to apply those technologies to space-based planet-finding missions.

Engineering the James Webb Space Telescope

Amy Lo
Northrop Grumman Aerospace Systems

NASA's James Webb Space Telescope (JWST) is the premier space telescope of its time. Set to launch in October 2018, it is designed to look at "first light" to image the formation of stars and galaxies. Building on the successes of the Hubble Space Telescope (HST), the JWST will look farther back in time—13.5 billion years, to the beginning of the universe—to dimmer, redder targets that were some of the very first objects to form in the universe.

The JWST is an engineering marvel unlike any space telescope before. The 6.5 meter aperture is accompanied by a 14 m × 22 m sunshield to passively cool the entire telescope portion of the observatory to cryogenic temperatures. Transported on an Ariane 5 launch vehicle, the JWST must fold up like an origami and travel 1.5 million km from Earth, to the Second Lagrange Point. At this distance, the JWST will not be serviceable like the Hubble and must deploy and operate flawlessly. The engineering required for this time machine is itself opening new pathways of technological capacity.

This paper describes the engineering needed to meet the JWST science goals, focusing on the precision design and testing required for the sunshield, which provides crucial protection to the telescope and scientific instruments. In addition, some of the challenges facing the alignment of the sunshield structure are discussed. The concluding section provides the latest status of the observatory and remaining steps to the launch in 2018.

CURRENT VS. NEWLY ENGINEERED CAPABILITIES

Wavelengths that are inaccessible from ground-based astronomical observations are accessed by launching telescope observatories into space. NASA has launched a series of such observatories and significantly advanced understanding of the universe.

One of the greatest currently active space-based observatories is the Hubble Space Telescope. It has revolutionized just about every aspect of astronomy, making its greatest contributions in observations of the early universe. Figure 1 shows the HST view of the extreme deep field (XDF) of a small region in the Fornax constellation. This is one of the most sensitive images taken in the visible wavelength, and captures some of the oldest galaxies ever imaged—formed when the universe was just 450 million years old (it is now ~13.7 billion years old) (Illingworth et al. 2013).

Using data from the XDF, astronomers are able to probe the structure and organization of the beginning of the universe. But the galaxies that appear in the

FIGURE 1 Hubble extreme deep field (XDF), showing some of the oldest galaxies ever imaged. They appear as dim, red, fuzzy blobs. Figure in color at http://www.nap.edu/catalog/21825. Credit: NASA; European Space Agency (ESA); G. Illingworth, D. Magee, and P. Oesch, University of California, Santa Cruz; R. Bouwens, Leiden University; and the Hubble Ultra Deep Field 2009 (HUDF09) team.

XDF are not the oldest galaxies ever formed: because of their distance and age, such galaxies are too dim to be captured by the HST and redder than the longest HST wavelength. The oldest, or the first, galaxies are thought to have formed around 200 million years after the Big Bang, or another 250 million years earlier than Hubble can see.

To enhance understanding beyond what can be learned from the HST, NASA has set out science goals for the JWST in four areas: first light and reionization (e.g., the formation of structures in the universe); assembly of galaxies (e.g., how galaxies are formed and what happens when small and large galaxies merge); the birth of stars and protoplanetary systems (e.g., the formation of stars and planetary systems); and planets and origins of life (e.g., planet formation, orbits, and habitable zones). Details of the JWST science goals are available from NASA (http://jwst.nasa.gov/science.html) and other sources (e.g., Gardner et al. 2006). The focus of this paper is on the engineering needed to support achievement of the first light goals.

As the successor to the HST, the James Webb Space Telescope is designed to have higher resolution, with optics seven times the surface area of the HST, making it 100 times more powerful. It will be able to probe the infrared portion of the electromagnetic spectrum to see these first light objects. Additionally, compared to the HST, the JWST has capabilities to 28 microns, versus the HST's maximum wavelength of 2.5 microns. With these enhanced technical capacities, the JWST will be able to image objects as old as 13.5 billion years, or ~200 million years after the Big Bang.

ENGINEERING THE JAMES WEBB SPACE TELESCOPE

The four science goals may be distilled to the following JWST primary imaging requirements :

- targets anywhere in the sky
- faint targets (requiring high sensitivity with low background)
- small targets (requiring high resolution with low jitter)
- infrared objects (requiring a cool telescope to reduce background)

The mission architecture of JWST developed to meet these requirements necessitated a significant amount of technology development. NASA initiated Phase A technology development for the JWST in the late 1990s, and in 2007 ten JWST critical developments were brought to Technology Readiness Level (TRL) 6 (Figure 2), a criterion for program confirmation (Gardner et al. 2006).

Among the technologies developed for the JWST, this paper will focus on the sunshield (Figure 3).

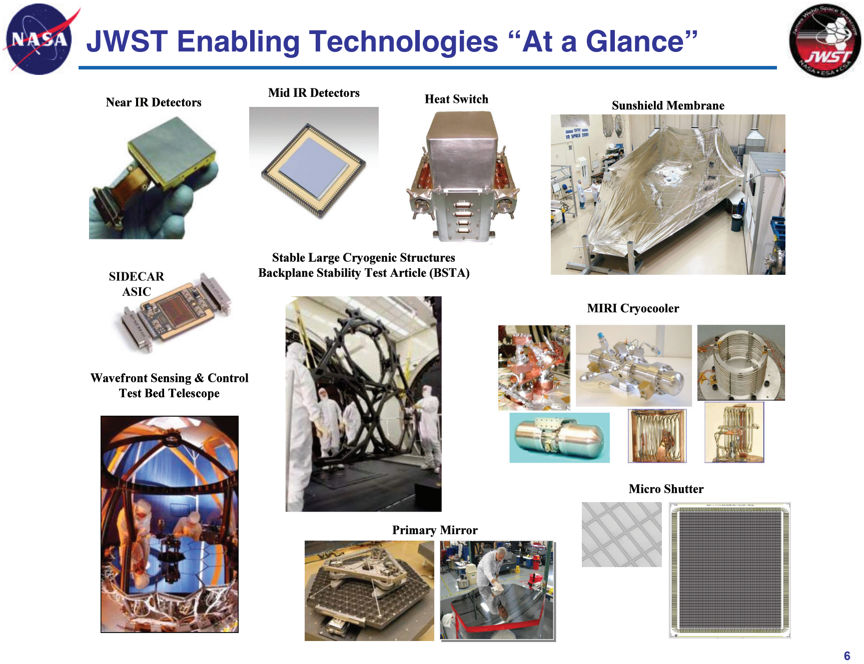

FIGURE 2 James Webb Space Telescope (JWST) technology development items identified at the start of the flight program. All of the items completed Technology Readiness Level (TRL) 6 development by 2007. IR = infrared; MIRI = mid-infrared instrument; SIDECAR ASIC = System for Image Digitization, Enhancement, Control and Retrieval Application Specific Integrated Circuit. Credit: NASA/Goddard Space Flight Center.

THE JWST SUNSHIELD

Measuring 14 m × 22 m (the size of a tennis court) when deployed, the sunshield has to protect the sensitive telescope and instruments from visible and thermal radiation from the Sun, Earth, Moon, and the spacecraft itself.

Design, Function, and Performance

Unlike the HST barrel assembly forming a cylinder centered on the primary optics, the open JWST sunshield design leaves the telescope exposed to space to facilitate the passive cooling of the optics to cryogenic temperatures of 40 to 50 Kelvin. This design also provides shadow over the JWST's pitch angles of +5° to −45° and roll angles of +5° to −5°.

The sunshield performs three major functions: (1) it shields the telescope from direct sunlight, earthlight, and moonlight, enabling the rejection of incident

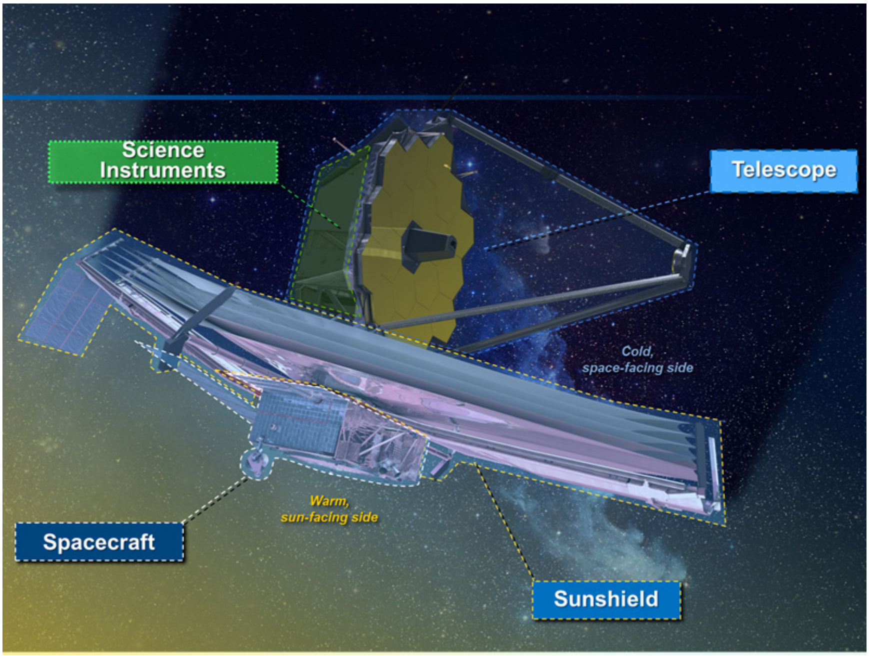

FIGURE 3 Main elements of the James Webb Space Telescope, with the tennis court–sized sunshield shading the telescope and science instruments from intense solar radiation. Credit: Northrop Grumman Corporation.

solar radiation such that only ~1 part in 300,000 is transmitted; (2) the upper layers prevent stray light from entering the telescope; (3) the lower layers absorb and prevent the bouncing of light into the telescope. These features keep the telescope cold and limit background noise effects on the science images.

To meet these performance requirements, the sunshield needs to be carefully aligned to ensure that the deployed structure maintains its edges within a few centimeters of the nominal location. Detailed performance analyses were necessary, starting with the on-orbit environment, where perturbations to the deployed sunshield were assessed and controlled. Environmental factors that affect the sunshield on orbit are thermal distortion, composite dry-out, the elastic response of the sunshield structure under tension, and the inelastic response (or creep) of the sunshield structure under tension. These effects were carefully modeled and quantified.

Measurement and Testing

Rigorous processes are used to control the tolerances of relevant sunshield parts; manufacturing and as-built structure dimensions are examined and precisely measured. Exposure of the observatory to the harsh launch vehicle environment can induce otherwise solidly attached parts to shift slightly at joints that are not bonded; these launch shifts are calculated based on design tolerances and allocated.

The need for deployment also affects the ultimate alignment of the sunshield, as the precision of the deployment is dependent on the details of the mechanical part performing the deployment. An extensive error budget has been constructed to account for these effects: environmental distortion, manufacturing and installation error, launch shift, and deployment repeatability.

A thorough set of alignment tests were baselined to measure and test the JWST sunshield in order to verify and/or demonstrate the values in the alignment error budget and quantify its on-orbit performance. Starting at the unit level and extending to the end of the observatory integration and test program, every major piece of the sunshield is measured, tested, and measured again to ensure that it has been properly characterized and its on-orbit performance is well understood.

FINAL STEPS TOWARD LAUNCH

As of the writing of this paper, most of the major components of the JWST sunshield are being manufactured; some subassemblies are undergoing testing and a high-fidelity pathfinder has recently completed deployment. The alignment program is well under way, measurements of the sunshield are being compiled, and pretest analysis is being done.

For the rest of the observatory, the telescope structure is complete and was shipped to Goddard Space Flight Center (GSFC) in August 2015. All mirrors have finished fabrication and are being installed at GSFC. The instruments in support of the four science goals have been integrated into their holding structure and are undergoing extensive testing. The spacecraft bus structure is complete and is in testing, as are many of its subsystems, preparing for integration into the spacecraft.

In 2017 the completed telescope portion, along with mirrors and instruments, will return to Northrop Grumman's Space Park facility in Redondo Beach, California, to be integrated with the spacecraft and sunshield. The observatory will then undergo final testing and be shipped to French Guiana for launch, planned for October 2018.

REFERENCES

Gardner JP, Mather JC, Clampin M, Doyon R, Greenhouse MA, Hammel HB, Hutchings JB, Jakobsen P, Lilly SJ, Long KS, and 13 others. 2006. The James Webb Space Telescope. Space Science Reviews 123(4):485–606.

Illingworth GD, Magee D, Oesch PA, Bouwens RJ, Labbé I, Stiavelli M, van Dokkum PG, Franz M, Trenti M, Carollo CM, Gonzalez V. 2013. The HST extreme deep field (XDF): Combining all ACS and WFC3/IR data on the HUDF region into the deepest field ever. Astrophysical Journal Supplement 209(1):6–19.

Starlight Suppression: Technologies for Direct Imaging of Exoplanets

Dmitry Savransky
Cornell University

Of the nearly 2,000 planets confirmed to exist outside the solar system, only a handful were detected directly rather than inferred from their interaction with the stars they orbit. The vast majority were discovered by sifting years of observations of thousands of stars for periodic changes in the stars' colors or fluxes, which indicate an orbiting planet.

Imaging allows for planetary detection with just one observation and confirmation with a few observations taken only months apart. More importantly, it enables spectroscopic characterization of exoplanets, often at spectral resolutions significantly exceeding those possible with any other detection method. With spectroscopy it is possible to probe the atmospheric and, potentially, surface composition of exoplanets and to validate models of planet formation and evolution.

Imaging is therefore a crucial component of exoplanet detection and characterization. Exoplanet imagers in ground observatories are already producing exciting discoveries (Macintosh et al. 2015), and the next generation of space instrumentation has the potential to detect Earth-like planets and indications of the presence of life.

This short paper reviews the challenges of exoplanet imaging, techniques used to overcome them, and the status of technology development for exoplanet imagers in space.

IMAGING OF EXOPLANETS

A telescope operates by collecting light from an astronomical source using a finite-sized aperture—either the entrance pupil of a refractive system or the primary mirror of a reflective one—and bringing the light to a focus on an imaging detector, such as a charge-coupled device (CCD). Diffraction effects limit the spatial resolving capabilities of telescopes, with angular resolution inversely proportional to the size of the aperture. For a circular pupil the minimum angular resolution (α_r) of the system will be greater than approximately $1.22\lambda/D$, where λ is the wavelength of light and D the diameter of the aperture. This is most easily understood by considering the point spread function (PSF) of the telescope—the impulse response generated by imaging a point source.

For a circular aperture, the PSF is an Airy disk, shown in Figure 1—a bright central spot of radius α_r surrounded by a series of annuli that decrease exponentially in brightness with angular separation. Between these annuli are small, dark regions called *nulls*.

Planets are significantly fainter than their host stars, with contrasts of 10^6 for the very brightest, self-luminous, young Jovian planets, and 10^{10} for Earth-sized planets in reflected visible light. Even if it were possible to observe a planet when it was precisely located on one of the deeper nulls of a perfect, diffraction-limited telescope, it would still not be possible to image it, because no detector has the dynamic range required to capture both the signals from the planet and the PSF of the star in the same image.

CORONAGRAPHY

Fortunately, this problem was partially solved in the early 20th century by solar astronomers studying the sun's corona, which is 1 million times fainter than the sun itself. Previously, the corona could be studied only during full solar eclipses. In the 1930s, however, Bernard Lyot demonstrated the first solar coronagraph—a system designed specifically to block bright, on-axis sources and thus enable the study of faint, off-axis ones (Lyot 1939).

Figure 2 shows a schematic view of a Lyot (pron. Lee-*o'*) coronagraph, along with images taken at the pupil of a coronagraphic system. The pupil is conjugate with the entrance pupil of the whole system, so that the first image, where no coronagraph elements are in place, is equivalent to the intensity distribution seen by the entrance aperture of the whole system. The central dark spot in the first image is due to the secondary mirror that partially obscures the primary aperture in this system. On-axis starlight, together with off-axis planet light, enters the telescope and is brought to a focus where the on-axis source is blocked by a small, hard-edged focal plane mask (FPM). The remaining light is propagated to the next pupil plane (middle image). Most of the starlight is blocked by the FPM, but some will diffract around the mask's edges, creating a pattern of rings along with a bright central spot. The remaining light is blocked by introducing another hard-edged mask, called a *Lyot stop*, into the pupil, leaving very little residual light, as shown in the image on the right. The Lyot stop can also include additional features to handle diffraction about other mechanical elements, such as struts that hold up the secondary mirror.

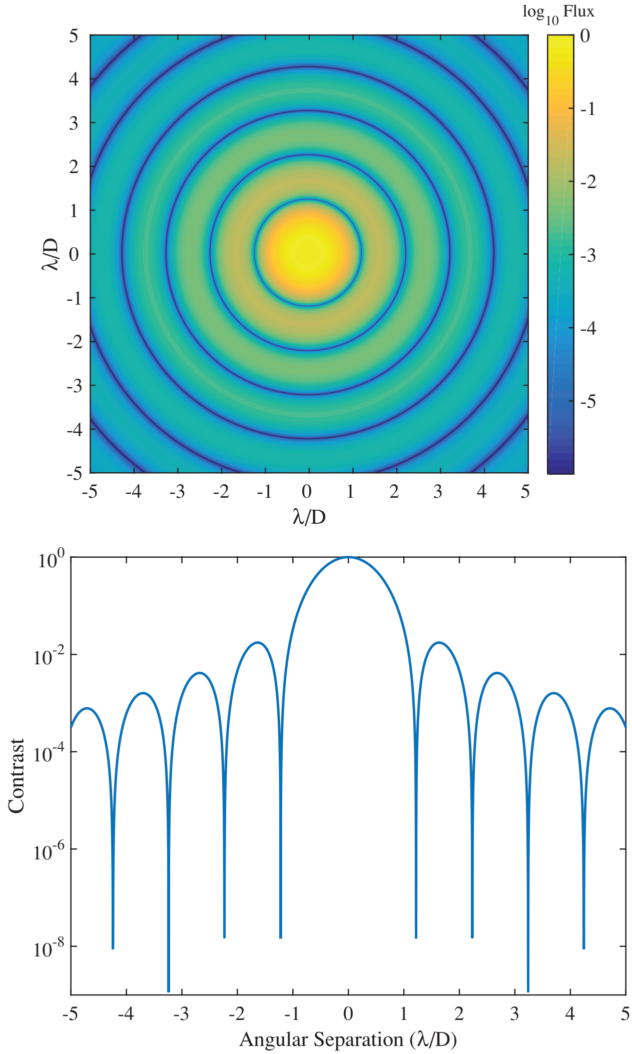

FIGURE 1 Numerical simulation of an Airy disk—the impulse response of a circular aperture. *Top:* Image of a point source, with color indicating intensity in log scale. Figure in color at http://www.nap.edu/catalog/21825. *Bottom:* The (radially symmetric) contrast profile of the image. λ = wavelength; D = aperture diameter.

To deal with residual diffracted light in the classical Lyot coronagraph, an additional pupil plane can be introduced before the FPM with a partially transmissive mask to apodize the beam and minimize diffraction effects downstream (Soummer 2005). An alternate approach is to introduce a specially shaped hard-

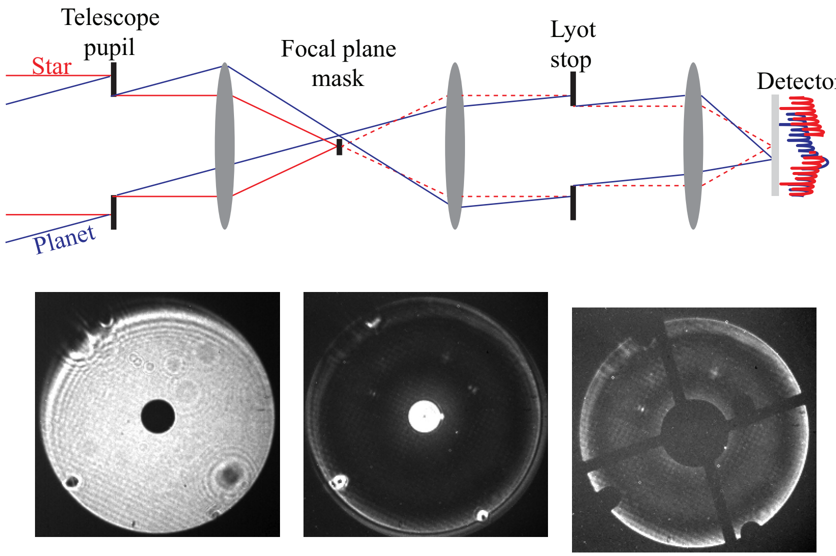

FIGURE 2 *Top:* Schematic of a Lyot coronagraph. Based on Sivaramakrishnan et al. (2001). *Bottom:* Images taken at the pupil plane of the Lyot stop showing (*left*) the unobscured entrance pupil, (*middle*) only the focal plane mask, and (*right*) both the focal plane mask and Lyot stop. Each image is individually stretched, and the final image has less than 1 percent of the light in the first image.

edged pupil ahead of the FPM to change the PSF so it is no longer radially symmetric, leaving high-contrast regions in the downstream focal plane (Kasdin et al. 2005).

Other strategies involve replacing the hard-edged FPM of the original Lyot coronagraph with a phase-shift mask to produce destructive interference of the on-axis light (i.e., the starlight cancels itself out, while planet light is mostly unaffected; Roddier and Roddier 1997). One can also achieve the beam apodization by using pupil-mapping mirrors to change the geometrical redistribution of the light (Guyon et al. 2005).

All of these approaches have various pros and cons, but all share the same basic limitations. Coronagraphs are still limited by the diffraction limit of the telescope. A coronagraph's design is highly specific to a particular telescope design, and most coronagraphs remove some of the planet light along with the starlight. Many coronagraph designs are also highly sensitive to misalignment, vibration, and optical surface errors.

Coronagraphs being evaluated for use in space all rely on active wavefront control via deformable mirrors, which have only recently begun to be demonstrated for use in space (Cahoy et al. 2014).

STARSHADES

An alternate approach, first suggested by Lyman Spitzer (1962), involves blocking the starlight before it enters the telescope. This method requires a space telescope to fly in formation, over a baseline of tens of thousands of kilometers, with an occulting spacecraft, or *starshade*. The starshade must be tens of meters in diameter and specifically shaped, as diffraction effects would cause light to be scattered back into the shadow cast by a simple flat plate. Fortunately, by Babinet's principle, the occulter is complementary to a pinhole camera, allowing starshades to be designed in much the same way as shaped pupil masks for internal coronagraphs, via numerical optimization (Vanderbei et al. 2007). This allows for constraints to be placed on minimum feature sizes to ensure that the produced designs are manufacturable. The resulting optimized shapes are radially symmetric, with a circular central core surrounded by petal-like extensions (Figure 3).

The main advantage of starshades is that they can achieve high contrasts with any conventional telescope design and without any active wavefront control. Furthermore, the minimum angular separation of a detectable planet for a starshade is determined solely by geometry—the size of the starshade and its distance from the telescope—and is the same at all wavelengths.

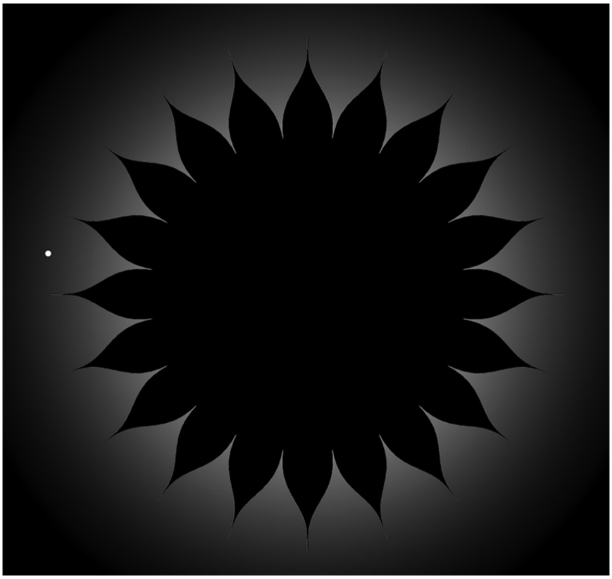

FIGURE 3 Schematic view of a starshade as seen by an occulted telescope. The starshade creates a shadow region of high contrast at the telescope aperture, allowing for exoplanets to be detected. Starshade design from Vanderbei et al. (2007).

On the other hand, while a telescope with an internal coronagraph merely needs to pivot in order to observe a new target, a starshade must be repositioned over large distances, with weeks of slew time between observations, leading to fewer stars observed and making scheduling optimization more difficult (Savransky et al. 2010). Starshade contrasts are also highly sensitive to shape, positioning, and alignment errors, leading to µm order manufacturing tolerances, mm deployment tolerances, and meter-scale alignment tolerances throughout the course of an observation (Shaklan et al. 2010). The requirement for precise alignment for extended periods also makes formation flying more difficult in geocentric orbits, so that most starshade mission concepts plan for operations in orbit about the second Earth-Sun Lagrange point (Kolemen et al. 2012). Finally, unlike coronagraphs, it is impossible to fully test starshades on the ground, so proxy experiments must be developed to build confidence in this technique (Sirbu et al. 2014).

CURRENT AND FUTURE EXOPLANET IMAGERS

Multiple exoplanet imagers, such as the Gemini Planet Imager (Macintosh et al. 2014) and SPHERE (Sauvage et al. 2013), are currently operating at some of the largest ground-based observatories. These instruments couple advanced coronagraphs with extreme adaptive optics systems (Tyson 2010), which correct for the effects of the turbulent atmosphere, to produce the highest levels of contrast ever demonstrated from the ground. Still, these systems will be able to detect only the very youngest, self-luminous giant planets on relatively large orbits—akin to Jupiter in the first 100 million years of its existence.

Advances in adaptive optics and coronagraphy, and the construction of the next generation of extremely large telescopes, will allow for the detection of smaller, older planets, but the best chance of directly imaging and getting spectra of an Earth-like planet lies in space.

NASA is developing a coronagraphic instrument for the Wide Field Infrared Space Telescope (WFIRST; Spergel et al. 2015), the next major astrophysics mission to be launched after the James Webb Space Telescope. The WFIRST coronagraph will be capable of detecting a wide variety of exoplanets, ranging from those like Neptune to “super-Earths,” a class of potentially rocky planets up to twice the radius of Earth that do not exist in our solar system.

Figure 4 shows a simulation of the population of planets around nearby stars, based on statistics from indirect surveys (Fressin et al. 2013; Howard et al. 2010), along with the expected contrast of one design for the WFIRST coronagraph with varying assumptions of telescope stability (Krist 2014).

The NASA Astrophysics Division has sponsored two science and technology definition teams to study $1 billion mission concepts based on a starshade (Exo-S)

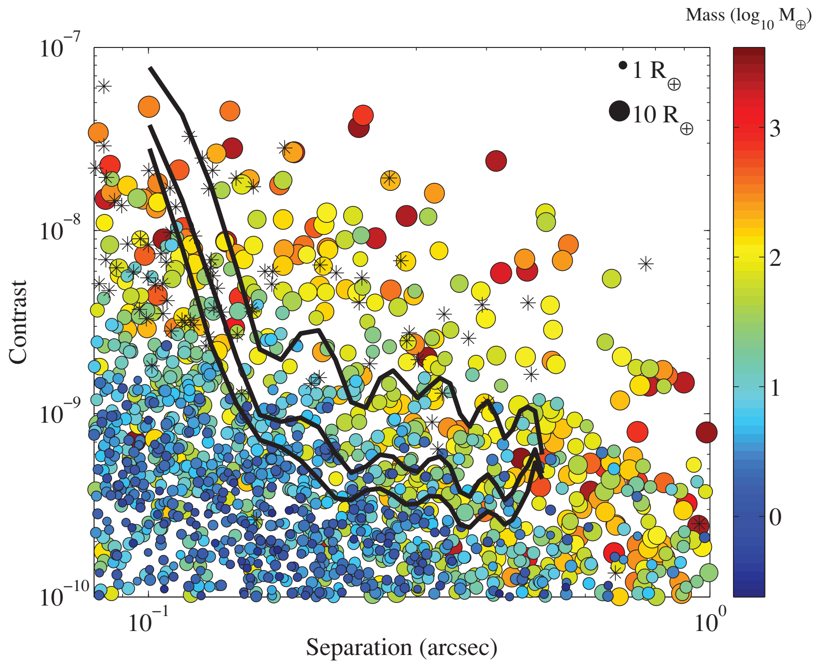

FIGURE 4 Colored points represent a simulated population of exoplanets based on prior surveys; asterisks represent contrast estimates for known, indirectly detected exoplanets at their most favorable viewing geometries; and the black lines are the predicted contrasts for one design of the Wide Field Infrared Space Telescope (WFIRST) coronagraph at various levels of telescope stability. Figure in color at http://www.nap.edu/catalog/21825. M⊕ = mass of Earth; R⊕ = radius of Earth. Based on Savransky (2013).

and coronagraph (Exo-C).[1] These and other concept studies are helping to identify the remaining engineering challenges in direct imaging and guide technology development programs.

These efforts may lead to the operation of a space-based direct imaging instrument in the next decade, producing exciting new science and helping validate the technologies needed to discover Earth-like planets and perhaps even alien life.

---

[1] For information on WFIRST, see http://wfirst.gsfc.nasa.gov/, and for information on Exo-S and Exo-C, see https://exep.jpl.nasa.gov/stdt/.

REFERENCES

Cahoy KL, Marinan AD, Novak B, Kerr C, Nguyen T, Webber M, Falkenburg G, Barg A. 2014. Wavefront control in space with MEMS deformable mirrors for exoplanet direct imaging. Journal of Micro/Nanolithography, MEMS, and MOEMS 13(1):011105-1–011105-14.

Fressin F, Torres G, Charbonneau D, Bryson ST, Christiansen J, Dressing CD, Jenkins JM, Walkowicz LM, Batalha NM. 2013. The false positive rate of Kepler and the occurrence of planets. Astrophysical Journal 766(2):81.

Guyon O, Pluzhnik EA, Galicher R, Martinache F, Ridgway ST, Woodruff RA. 2005. Exoplanet imaging with a phase-induced amplitude apodization coronagraph. I. Principle. Astrophysical Journal 622(1):744.

Howard AW, Marcy GW, Johnson JA, Fischer DA, Wright JT, Isaacson H, Valenti JA, Anderson J, Lin DNC, Ida S. 2010. The occurrence and mass distribution of close-in super-Earths, Neptunes, and Jupiters. Science 330(6004):653–655.

Kasdin NJ, Vanderbei RJ, Littman MG, Spergel DN. 2005. Optimal one-dimensional apodizations and shaped pupils for planet finding coronagraphy. Applied Optics 44(7):1117–1128.

Kolemen E, Kasdin NJ, Gurfil P. 2012. Multiple Poincaré sections method for finding the quasiperiodic orbits of the restricted three body problem. Celestial Mechanics and Dynamical Astronomy 112(1):47–74.

Krist JE. 2014. End-to-end numerical modeling of AFTA coronagraphs. SPIE Astronomical Telescopes and Instrumentation, vol. 9143. Bellingham, WA: International Society for Optics and Photonics.

Lyot B. 1939. The study of the solar corona and prominences without eclipses. George Darwin Lecture. Monthly Notices of the Royal Astronomical Society 99:580.

Macintosh B, Graham JR, Ingraham P, Konopacky Q, Marois C, Perrin M, Poyneer L, Bauman B, Barman T, Burrows AS, and 37 others. 2014. First light of the Gemini Planet Imager. Proceedings of the National Academy of Sciences 111(35):12661–12666.

Macintosh B, Graham JR, Barman T, De Rosa RJ, Konopacky Q, Marley MS, Marois C, Nielsen EL, Pueyo L, Rajan A, and 78 others. 2015. Discovery and spectroscopy of the young jovian planet 51 Eri b with the Gemini Planet Imager. Science 350(6256):64–67.

Roddier F, Roddier C. 1997. Stellar coronagraph with phase mask. Publications of the Astronomical Society of the Pacific 109:815–820.

Sauvage J-F, Beuzit J-L, Roelfsema R, Feldt M, Dohlen K, Mouillet D, Puget P, Wildi F, Abe L, Baruffolo A, and 52 others. 2013. Sphere: Complete laboratory performance and prediction for on-sky first light. SPIE Optical Engineering and Applications, vol. 8864. Bellingham, WA: International Society for Optics and Photonics.

Savransky D. 2013. Space mission design for exoplanet imaging. SPIE Optical Engineering and Applications, vol. 8864. Bellingham, WA: International Society for Optics and Photonics.

Savransky D, Kasdin NJ, Cady E. 2010. Analyzing the designs of planet finding missions. Publications of the Astronomical Society of the Pacific 122(890):401–419.

Shaklan SB, Noecker MC, Lo AS, Glassman T, Dumont PJ, Jordan EO, Kasdin NJ Jr, Cash WC, Cady EJ, Lawson PR. 2010. Error budgeting and tolerancing of starshades for exoplanet detection. SPIE Space Telescopes and Instrumentation, vol. 7731. Bellingham, WA: International Society for Optics and Photonics.

Sirbu D, Kasdin NJ, Vanderbei RJ. 2014. Diffractive analysis of limits of an occulter experiment. In: SPIE Astronomical Telescopes and Instrumentation, vol. 9143. Bellingham, WA: International Society for Optics and Photonics.

Sivaramakrishnan A, Koresko CD, Makidon RB, Berkefeld T, Kuchner MJ. 2001. Ground-based coronagraphy with high-order adaptive optics. Astrophysical Journal 552(1):397.

Soummer R. 2005. Apodized pupil Lyot coronagraphs for arbitrary telescope apertures. Astrophysical Journal Letters 618(2):L161–L164.

Spergel D, Gehrels N, Baltay C, Bennett D, Breckinridge J, Donahue M, Dressler A, Gaudi BS, Greene T, Guyon O, and 45 others. 2015. Wide-field infrared survey telescope–astrophysics focused telescope assets (WFIRST-AFTA) 2015 report. WFIRST-AFTA Science Definition Team. Washington: National Aeronautics and Space Administration.

Spitzer L. 1962. The beginnings and future of space astronomy. American Scientist 50(3):473–484.

Tyson R. 2010. Principles of Adaptive Optics. Boca Raton, FL: CRC Press.

Vanderbei RJ, Cady E, Kasdin NJ. 2007. Optimal occulter design for finding extrasolar planets. Astrophysical Journal 665(1):794–798.

Realizing Large Structures in Space

Jeremy Banik
Air Force Research Laboratory

Since the dawn of space access in 1957 the size of spacecraft payloads and solar arrays has steadily grown. Yet the demand for larger antenna arrays and telescope mirrors and higher power on spacecraft continues to outpace availability.

BACKGROUND

Large structures in space have one simple purpose: to support payloads that manipulate the electromagnetic spectrum. Telescopes operate at short wavelengths for astronomy and surveillance missions, from ultraviolet to near infrared. Starshades block visible sunlight for direct imaging of Earth-like planets. Sunshields reject the shortwave spectrum to keep instruments cool. Solar sails reflect photons in the short wave, generating constant propulsion for interplanetary exploration and station-keeping maneuvers.

For practical Earth-based applications, communication and radar antennas transmit and receive microwaves for communications, coarse imaging, and object tracking. The larger the antenna or radar in space, the smaller the antenna on a soldier's back and the greater number of objects tracked. Photovoltaic solar arrays convert photons to electrons for charging satellite batteries. Radio frequency missions require larger antennas to generate higher-resolution radar imaging for Earth science, and to keep up with the higher data throughput demands of modern hand-held devices.

In each of these examples the larger the space structure, the more effective the mission data return. Exoplanet discovery is no exception. Larger optical mirrors and bigger starshade occulters are needed to detect both a larger number of stars, thus increasing Earth-like exoplanet discovery opportunities, and a broad range of the electromagnetic spectrum for exoplanet spectral characterization.

But the design, construction, and deployment of larger antennas and mirrors are fraught with highly specialized challenges. The largest commercial radio frequency antenna on-orbit is 22 meters in diameter, and the largest space telescope mirror is 6.5 meters. These structures must be designed to hold extreme dimensional precision and stability tolerances.

Optical and radio signal quality are directly related to the precision of the surface from which the signal emanates. The larger the structure, the more difficult it is to achieve a given figure precision. Radio frequency missions operate on long wavelengths so precision requirements are not as stringent as for optical missions, but because signal gain scales inversely with the square of wavelength, radio antennas require larger apertures than optical.

Once unfolded in space, these structures face extreme temperature swings that can cause large static and dynamic dimensional changes. Spacecraft in a geosynchronous orbit endure daily temperature swings from −200°C to +200°C over a typical 15-year lifetime. Materials must exhibit a low coefficient of thermal expansion, and assembly interfaces are extensively tested to control the characteristics of such expansion.

CURRENT CAPABILITIES AND CONSTRAINTS

Figure 1 illustrates the indirect relationship between structure size and dimensional precision. As this ratio grows, payload cost escalates. Some of the highest-performing space structures to date are represented on this chart, yet they are restricted to relatively low diameter-to-precision ratios when compared to future needs in the tens to hundreds of meters.

Aside from dimensional precision challenges, large space structures must also be folded compactly for a violent trip to orbit. Once designed, built, and stowed in a 5 meter launch vehicle fairing, these payloads endure 10 to 70 g peak accelerations during the 10-minute trip to low Earth orbit, reaching a velocity of 7 km/s. Once on-orbit, the structure must unfold precisely and reliably.

As an example, an exoplanet starshade must unfurl from 5 meters to 34 meters into a shape that is within 0.1 mm precision, approximately the width of a human hair, across a span similar to an eight-lane interstate overpass. Figure 2 and Table 1 show the relative scale of the largest launch vehicles, the typical large structures that must stow in them, and the respective packaging ratios.

The mechanical approach thus far for realizing large communication antennas, telescopes, and radar antennas has generally been to fold a deep truss structure and self-deploy using multiple pin-clevis joints, motors, torsion springs, and dampers. This approach has led to incremental improvements in size, weight, and power over the past 50 years, but these heritage mechanisms and structural support schemes are reaching size and mass limits. Adding hinges to package

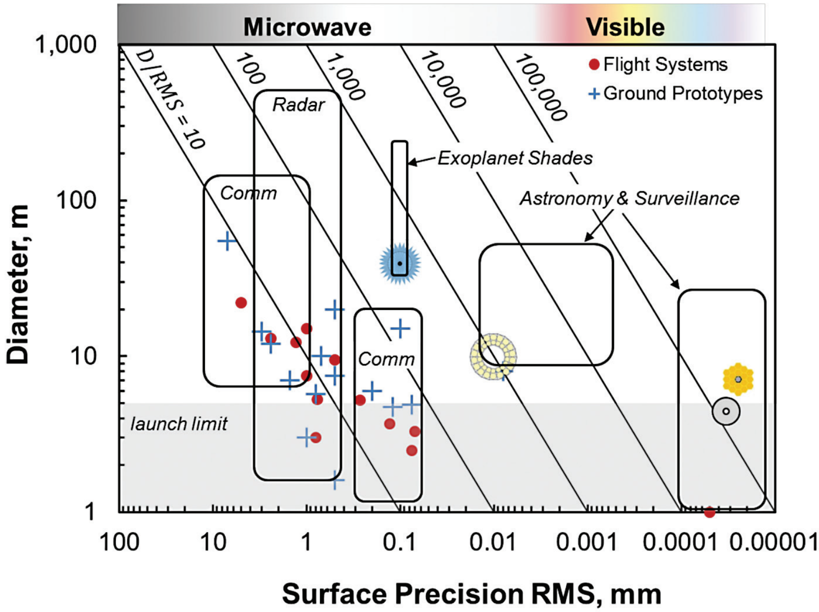

FIGURE 1 Size and surface figure precision are indirectly related in space structure design. Boxes indicate the current art of the possible. Comm = communications; JWST = James Webb Space Telescope; MOIRE = Membrane Optical Imager for Real-Time Exploitation; RMS = root mean square.

larger payloads into these limited launch volumes is causing reliability concerns and cost escalation.

EMERGING APPROACHES

Two structural design techniques show high payoff potential: (1) tension-aligned antennas and optics and (2) high-strain composite mechanisms. These methods have been used successfully for decades, but in very limited form because of the absence of high-strength, high-stiffness carbon fiber composites and robust analytical and test tools.[1]

[1] Two historical space structures that used lower-stiffness glass and aramid fiber composites are the Continuous Longeron Mast of the 1960s and the Wrap-Rib reflector of the 1970s.

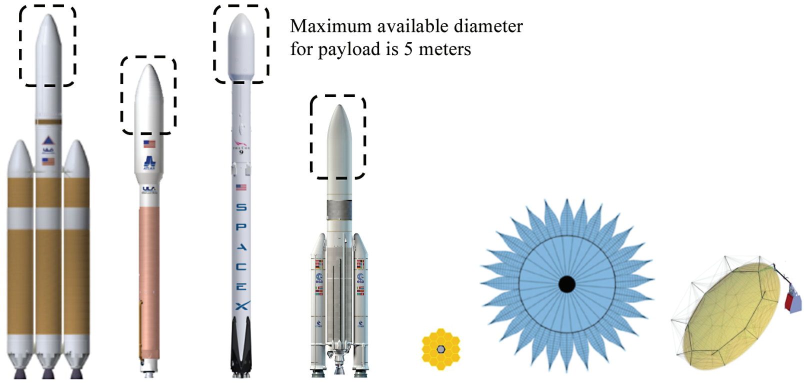

FIGURE 2 Scale of the largest launch vehicles (*left*) relative to representative large space structures (*right*). Images reprinted with permission.

TABLE 1 Packaged Size of Typical Precision Space Mirror, Shade, and Antenna

	Deployed size	Stowed size	Packaging ratio
JWST Primary	6.5 m	4.0 m	1.6:1
Exo-S Starshade	34 m	5.0 m	9:1
SkyTerra-1 Mesh Reflector	22 m	2.4 m	9:1

JWST = James Webb Space Telescope; m = meter.

Tension-Aligned Antennas and Optics

As more sophisticated computerized testing and analysis tools became available and the use of high-strength carbon fibers in aircraft became prevalent, feasible new space architectures began to surface, enabling the ground testing of the Innovative Space-based Radar Antenna truss in 2007 (Lane et al. 2011), the flexible unfurlable and refurlable lightweight (FURL) solar sail in 2010 (Banik and Ardelean 2010), and in 2013 the Membrane Optical Imager for Real-Time Exploitation (MOIRE) brassboard telescope (Domber et al. 2014). Most recently, the roll-out solar array (ROSA; Spence et al. 2015) was manifested for a space-flight experiment to the International Space Station in 2016.

Other advanced concepts currently under development by government and industry all share high-strain composite features for folding functionality and/or tension as the means of structural stability: a low-cost multiarm radial composite

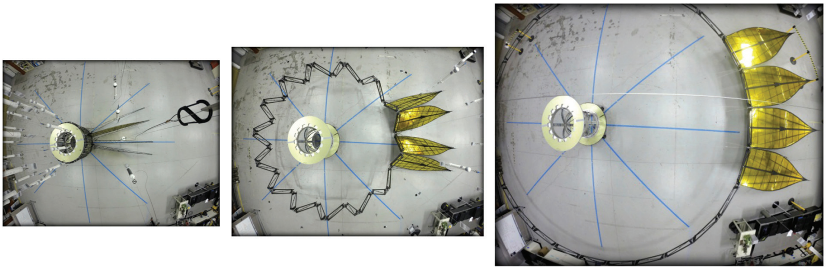

FIGURE 3 Ground deployment sequence of a starshade demonstration model shows a perimeter truss collapsed against a central drum (*left*). As the truss articulates radially (*center*) petals begin to rotate into position. Finally the perimeter truss is fully locked out (*right*), providing full support to the four petals. SOURCE: NASA (2015).

radio-frequency reflector (Footdale and Banik 2016), an isogrid column (Jeon et al. 2016), a starshade occulter (Figure 3; NASA 2015), an extremely high expansion deployable structure (Warren et al. 2007), a triangular rollable and collapsible mast (Banik and Murphey 2010), and a tensioned planar membrane antenna (Warren et al. 2015).

High-Strain Composite Mechanisms

High-strain composites are defined as thin carbon and glass fiber polymer matrix laminate materials used to construct shell structures that undergo large elastic deformations during folding and then release the stored strain energy to enforce deployment.

Architectures constructed from these materials have 7× greater deployment force, 20× greater dimensional stability, and 4× higher stiffness compared to traditional metallic flexure mechanisms (Murphey et al. 2013, 2015; Welsh et al. 2007). Moreover, when compared to traditional pin-clevis-type hinges, the payoff is a reduced mechanism part count, less susceptibility to binding, and more robust deployments. The hinges have greater lateral and torsion compliance during the transition from folded to deployed, a critical transition when there is a high risk of binding (e.g., when asymmetric solar heating and resulting expansion induce side loads on hinges). High-strain composite hinges can operate through this state and achieve a repeatable, dimensionally stable locked-out condition due to the near-zero coefficient of thermal expansion of carbon fibers.

Combined with the kinematic determinacy of a tensioned antenna or optic, these technologies are cracking open the door to a new era of space structures where 50-meter telescopes, 100-meter antennas, and megawatt-class solar arrays are all feasible.

LONG-TERM POSSIBILITIES AND CONSIDERATIONS

Despite the potential of tension-aligned payloads and high-strain composite mechanisms, even these will eventually reach limitations in size scaling, mass efficiency, and dimensional stability. It will be necessary to push beyond these limits to meet long-term civil and military space needs.

As the promise of robotic assembly and in-space additive manufacturing technologies is only beginning to emerge, the best tack is not yet clear. Certainly the unique realities and constraints of space flight must remain front and center in any pursuit of exciting new technologies.

Few industries are more risk-averse than those involved in space flight. NASA and Department of Defense program managers regularly spend hundreds of millions (sometimes billions) of dollars on a single spacecraft to try to ensure mission success. For example, the price tag on the 6.5 meter James Webb Space Telescope has reportedly reached $8.7 billion during the 16 years from inception to launch (Leone 2011). If this paradigm holds, then a 20-meter space telescope will remain in development for 87 years and cost $47.5 billion (Arenberg et al. 2014).

It is a spiraling effort. As more money is spent, additional testing and analysis are required to try to increase the certainty of spacecraft success. Schedules are then drawn out, further adding to the costs.

Spaceflight is a one-shot business. Hundreds of critical systems must work together flawlessly the first time or the mission is lost. Deployable structures are notorious as one of the highest sources of failure. Mission managers therefore expend great effort, time, and resources on testing in space simulation chambers. But even then the effects of gravity and the lack of a truly representative space environment always raise questions about the validity of these tests despite decades of experience and evidence.

Of course, the pursuit of innovation should not be deterred by these challenges. Rather, they should motivate the continued evaluation of new architectures against all measures of success—not only structural performance but also cost factors such as ease of ground testing and validation, simplicity of analysis methods, and reduced quantity of mechanical interfaces and unique parts.

It is difficult to quantify each key cost factor in the early conceptual design phase, but this is precisely when critical design decisions that most affect cost are made. Cost evaluation metrics are therefore essential. Until they are available, rational comparison of competing structural architectures must rely on structural performance metrics along with subjective cost assessments. A common list of such metrics is provided in Table 2; note the highly sophisticated telescope mission cost metric (bottom row).

TABLE 2 Common Metrics for Evaluation of Large Space Structure Performance

Metric	Description	Equation
Packaging ratio	deployed length/stowed length	L_d/L_s
Linear packaging density	deployed size/stowed volume	D/V
Areal packaging density	deployed area/stowed volume	A/V
Beam performance index (Murphey 2006)	strength moment, bending stiffness, linear mass density	$\mu = \frac{(M^2EI)^{1/5}}{w}$
Solar array scaling index (Banik and Carpenter 2015)	acceleration load, frequency, boom quantity, length, area, blanket areal mass density, total mass	$\kappa = (af)^{0.216}\, n^{0.231} L_{pb}\, A^{0.755} \frac{\gamma_b^{0.176}}{m}$
Mass efficiency	diameter/mass	D/m
Surface precision	diameter/root mean square (RMS) figure error	D/RMS
Dimensional stability	coefficient of thermal expansion	α
Telescope mission cost (Arenberg et al. 2014; Stahl et al. 2012)	diameter, wavelength, temperature of operation	$MC = \frac{C\, D^{1.7} \lambda^{-0.3} T^{-0.25}}{0.11 + 0.09 ln(D)}$

CONCLUSION

The challenges to realize large space structures are great, but if they are successfully addressed the opportunities will be well worth the effort and cost. No doubt, one of the most exciting possibilities is discovery of Earth-like planets that might have sustained life—or perhaps still do. . . .

REFERENCES

Arenberg J, Atkinson C, Breckinridge J, Conti A, Feinberg L, Lillie C, MacEwen H, Polidan R, Postman M, Matthews G, Smith E. 2014. A new paradigm for space astrophysics mission design. SPIE Astronomical Telescopes and Instrumentation, June 22–27, Montreal.

Banik J, Ardelean E. 2010. Verification of a retractable solar sail in a thermal-vacuum environment. 51st AIAA Conference on Structures, Structural Dynamics, and Materials, April 12–15, Orlando, FL.

Banik J, Carpenter B. 2015. Analytic representation of space solar array scaling performance. 42nd IEEE Photovoltaic Specialist Conference, June 14–19, New Orleans.

Banik J, Murphey T. 2010. Performance validation of the triangular rollable and collapsible mast. 24th AIAA/USU Conference on Small Satellites, August 9–12, Logan, UT.

Domber J, Atcheson P, Kommers J. 2014. MOIRE: Ground test bed results for a large membrane telescope. 1st AIAA Spacecraft Structures Conference, January 13–17, National Harbor, MD.

Footdale J, Banik J. 2016. Design and deployment testing of the Multi-Arm Radial Composite (MARCO) Reflector Antenna. 3rd AIAA Spacecraft Structures Conference, January 4–8, 2016, San Diego.

Jeon S, Banik J, Peterson M. 2016. Design and validation of a deployable isogrid. 3rd AIAA Spacecraft Structures Conference, January 4–8, 2016, San Diego.

Lane S, Murphey T, Zatman M. 2011. Overview of the Innovative Space-based Radar Antenna Technology Program. AIAA Journal of Spacecraft and Rockets 48(1):135–145.

Leone D. 2011. NASA: James Webb Space Telescope to now cost $8.7 billion. Space News, 27 August. Available at http://www.space.com/12759-james-webb-space-telescope-nasa-cost-increase.html.

Murphey T. 2006. Booms and trusses. In: Recent Advances in Gossamer Spacecraft, ed. Jenkins CHM. Progress in Astronautics and Aeronautics Series, vol. 212, pp. 1–43. Reston, VA: American Institute of Aeronautics and Astronautics.

Murphey T, Peterson M, Grigoriev M. 2013. Four point bending of thin unidirectional composite laminas. 54th AIAA Conference on Structures, Structural Dynamics, and Materials, April 9, Boston.

Murphey T, Francis W, Davis B. 2015. High strain composites. 2nd AIAA Spacecraft Structures Conference, January 5–9, Kissimmee, FL.

NASA JPL [National Aeronautics and Space Administration Jet Propulsion Laboratory]. 2015. Exo-S Starshade Probe-Class: Exoplanet Direct Imaging Mission Concept Final Report. Pasadena. Available at http://exep.jpl.nasa.gov/stdt/Exo-S_Starshade_Probe_Class_Final_Report_150312_URS250118.pdf.

Spence B, White S, Takeda R, LaPointe M, Schmid K, Kiefer S, Carter G, Paskin A, Allmandinger T, Douglass M, LaCorte P. 2015. Flexible blanket solar arrays: Next-generation game-changing technology. 42nd IEEE Photovoltaic Specialist Conference, June 17, New Orleans.

Stahl HP, Henrichs T, Luedtke A, West M. 2012. Update on multivariable parametric cost models for ground and space telescopes. SPIE Proceedings 8442: Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave, September 21.

Warren P, Dobson B, Silver M. 2007. Experimental characterization of an extremely high expansion deployable gossamer structure. 48th AIAA Conference on Structures, Structural Dynamics, and Materials, April 23–26, Honolulu.

Warren P, Steinbeck J, Minelli R, Mueller C. 2015. Large, deployable S-band antenna for a 6U Cubesat. 29th AIAA/USU Conference on Small Satellites, August 8–13, Logan, UT.

Welsh J, Mayes J, Biskner A. 2007. Experimental and numerical failure predictions of biaxially-loaded quasi-isotropic carbon composites. 16th International Conference on Composite Materials (ICCM), July 8–13, Kyoto.

Sensing Controls for Space-based Planet Finding

Jonathan T. Black
Virginia Tech

Detection and characterization of Earth "twins," defined as Earth-sized planets with Earth's geometric albedo of 0.2 in the habitable zone, are the most challenging of the space-based planet-finding missions. Proposed designs to conduct the search for such exoplanets include direct imaging missions that require formation flying of starshades dozens of meters in diameter at distances of 25,000–50,000 km. Position sensing and control under these conditions are substantially more challenging than those involved in orbital rendezvous and docking.

This paper presents the cutting edge in sensing controls for formation flying and satellite proximity operations with a focus on enabling autonomy for small satellites, and discusses current challenges and limitations on advances that are required to apply those technologies to space-based planet-finding missions, recent work in the field, and long-term challenges.

BACKGROUND

NASA recently commissioned two studies to produce exoplanet direct imaging design reference missions (DRMs), one based on a coronagraph observatory (NASA 2015b) and the other on a free-flying starshade dozens of meters in diameter with a telescope up to 4 m in diameter called Exo-S (NASA 2015a). The Exo-S DRMs were based on the Kepler observatory and leveraged satellite technologies with extensive flight heritage; a graphical overview is shown in Figure 1.

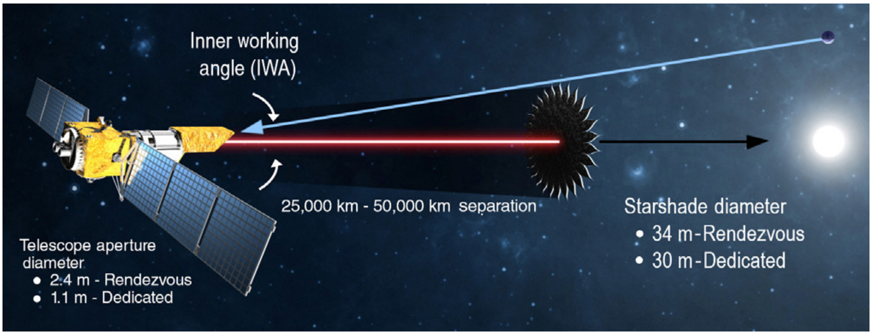

FIGURE 1 Exo-S mission overviews for Rendezvous Mission, in which the starshade works with the Wide-Field Infrared Survey Telescope (WFIRST) Observatory, and Dedicated Mission, in which the starshade is colaunched with a dedicated small observatory. Image courtesy NASA/JPL-Caltech (NASA 2015a).

FORMATION FLYING

Direct imaging requires laterally aligning the telescope to within 1 m of the starshade–star axis, with a separation of 25,000–50,000 km between the starshade and the observatory (Figure 1). While this submeter position control is routine in orbital rendezvous and docking, sensing in this case is altogether more challenging as positions must be sensed three to five times more finely than the control requirement. The lateral offset of the starshade must therefore be sensed to 30 cm at maximum separation, yielding a bearing measurement precision of 6 nrad (nanoradians) (1.25 milliarcseconds, mas) during imaging (NASA 2015a). The sensing requirements during all phases of observatory operation are shown in Figures 2 and 3.

SMALL SATELLITE PROXIMITY OPERATIONS AND FORMATION FLYING

Automated proximity operations (proxops) have enabled new mission capabilities and enhanced space situational awareness, but additional technology development was needed for exoplanet direct imaging. Several successful flight missions, including the Orbital Express (Defense Advanced Research Projects Agency)[1] and the SpaceX Dragon capsule docking with the International Space Station, involve active external illumination and sensing (Ogilvie et al. 2008),

[1] Archived information about the 2007 Orbital Express is available from the DARPA Tactical Technology Office at http://archive.darpa.mil/orbitalexpress/index.html.

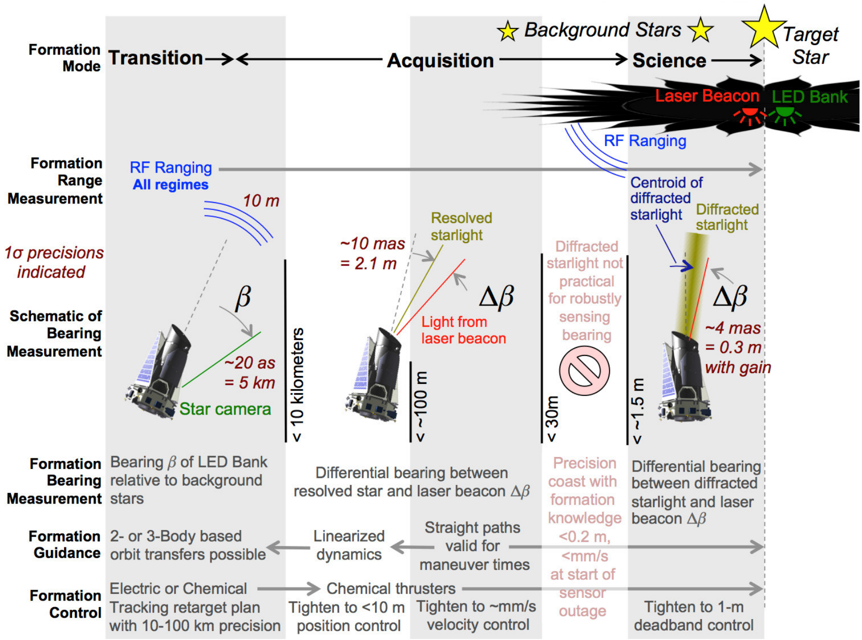

FIGURE 2 Formation sensing, guidance, and control by formation mode. β =bearing angle; RF = radio frequency. Image courtesy NASA/JPL-Caltech (NASA 2015a, Figure 6.3-2).

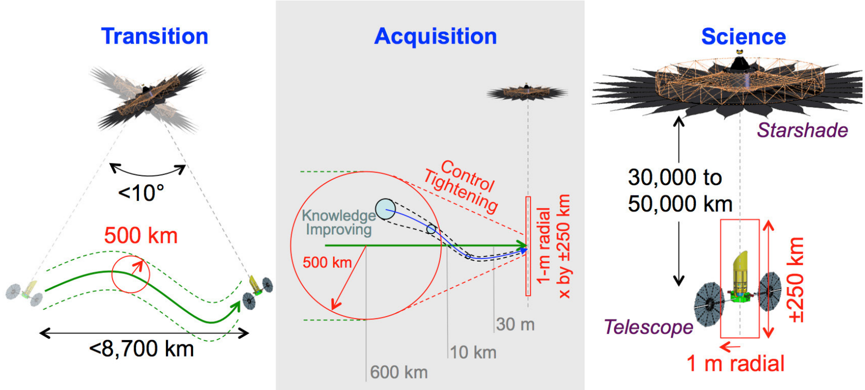

FIGURE 3 Formation flying modes and requirements. Image courtesy NASA/JPL-Caltech (NASA 2015a, Figure 6.3-1).

which will not be possible in the planned deep space orbits of the exoplanet direct imaging observatories.

As an alternative, computer vision–based sensing methods require only a camera and CPU and no a priori knowledge of the target. Star tracker–like subsystems on smaller spacecraft can also be used. Research so far in this field has developed a monocular simultaneous localization and mapping (SLAM) method for implementation on a proxops simulator.

The use of monocular SLAM produces a 3D map from relative motion while tracking the camera pose (location and orientation). Initial results of the space-specific monocular SLAM method show depth and pose estimates of targets, are robust to the dynamic backgrounds and highly reflective surfaces in the space environment, have an average error on the order of 0.31° (0.27° standard deviation, 0.46° RMS), and run in near real time of 11.8 frames per second. It is important to reiterate that no a priori target knowledge is required to generate 3D models. The results, samples of which are shown in Figures 4 and 5, pave the way for automated proxops on platforms as small as a CubeSat.

Autonomous state estimation and control, which mathematically represent physical quantities such as position and velocity, are also key technology enablers of the precision formation flying operations required for exoplanet direct imaging.

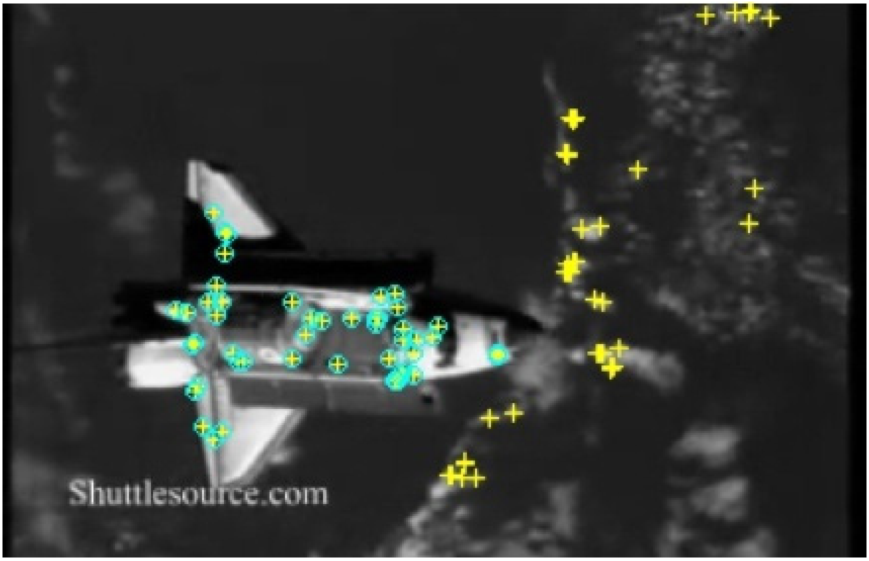

FIGURE 4 Frame from processed shuttle video showing identified feature points. Circles around crosses denote features detected on the shuttle. Crosses without circles denote features detected on the background (Kelly 2015).

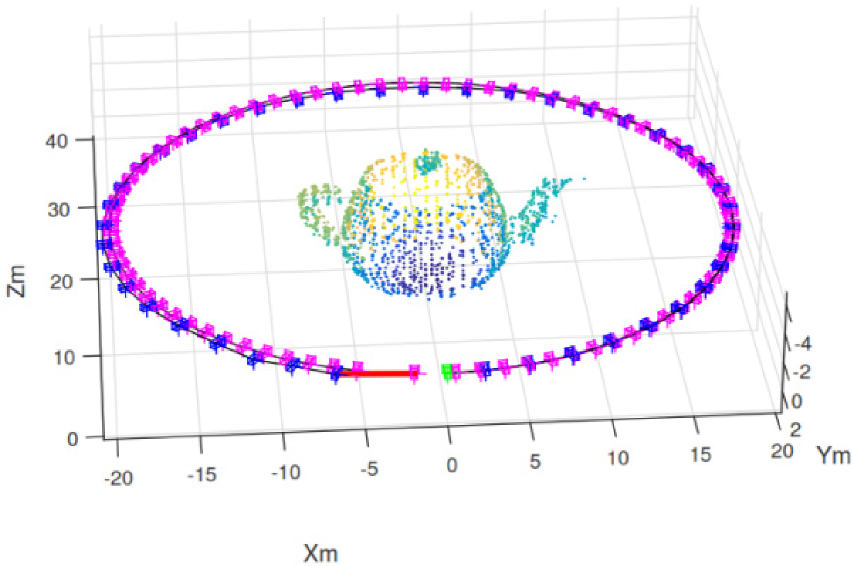

FIGURE 5 Sample monocular simultaneous localization and mapping (SLAM) results showing actual (blue boxes in ring) and estimated (pink boxes) camera position (3D location) and pose (direction the camera is pointing). Figure available in color at http://www.nap.edu/catalog/21825. SOURCE: Kelly (2015).

Work in this area provides a unique approach to autonomous fuel/time trajectory optimization for relative reconfiguration of two objects in perturbed elliptical orbits.

The relative motion model for the deputy satellite with respect to the chief[2] is well approximated using a fully nonlinear state transition matrix. Control is applied in the in-track and cross-track directions, making this system underactuated but reachable for any formation flying orbit. The system is discretized using a zero-order hold on the input and the control signals are computed using a linear program, resulting in a bang-off-bang[3] control profile. A balance between the time of flight and the required fuel is analyzed using a genetic algorithm. Combining the new estimation with the tracking control produces a novel, robust, and computationally light (autonomous) optimal terminal guidance capability, a sample of which is shown in Figure 6.

[2] Deputy refers to the maneuvering satellite over which there is control. Chief refers to the nonmaneuvering satellite being inspected.

[3] No throttle is available, and the thruster is either on or off.

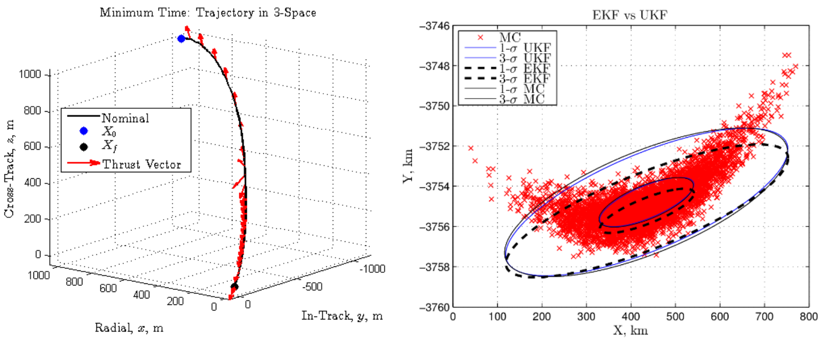

FIGURE 6 Control inputs for tracking minimum time trajectory (*left*) and high-order state estimation (*right*). EKF = extended Kalman filter; MC = Monte Carlo; UKF = unscented Kalman filter; X_0 = initial position; X_1 = final position.

FUTURE CHALLENGES AND DIRECTIONS

The sensing controls required to enable the precision formation flying of starshades and observatories tens of thousands of kilometers apart in deep space are an extreme engineering challenge. The current state-of-the-art technology development in formation flying and spacecraft autonomy for small satellites needs to be adapted for exoplanet direct imaging missions to ensure mission success. The sensing task will likely be coupled to the autonomous orbital and attitude control systems discussed here, and may also require advances in machine learning and cognition.

ACKNOWLEDGMENTS

The author thanks Andrew Rogers of Virginia Tech.

REFERENCES

Kelly S. 2015. A monocular SLAM method to estimate relative pose during satellite proximity operations. MS thesis (March). Available at http://www.dtic.mil/docs/citations/ADA616162.

NASA [National Aeronautics and Space Administration]. 2015a. Exo-S: Starshade Probe-Class Exoplanet Direct Imaging Mission Concept—Final Report (CL#15-1155). ExoPlanet Exploration Program, Astronomy, Physics, and Space Technology Directorate, Jet Propulsion Laboratory, March. Pasadena. Available at http://exep.jpl.nasa.gov/stdt/.

NASA. 2015b. Exo-C: Imaging Nearby Worlds—Final Report (CL#15-1197). Science and Technology Definition Team, ExoPlanet Exploration Program, Astronomy, Physics, and Space Technology Directorate, Jet Propulsion Laboratory, March. Pasadena. Available at http://exep.jpl.nasa.gov/stdt/.

Ogilvie A, Allport J, Hannah M, Lymer J. 2008. Autonomous robotic operations for on-orbit satellite servicing. Proceedings of SPIE 6958:09–12.

Optical and Mechanical Metamaterials

Optical and Mechanical Metamaterials

Jennifer Dionne
Stanford University

Luke Sweatlock
Northrop Grumman Aerospace Systems

The ability to engineer the properties of high-performance materials is critical for applications ranging from high-efficiency energy production and storage to advanced medical imaging and therapeutics. The principle of "metamaterials" refers to the design of composites whose properties derive as much from their structure as from their composition.

Metamaterials have energized many materials engineering disciplines, leading not only to the discovery of a powerful "toolbox" of new design methods but also to an expansion in fundamental understanding of the physics of materials. Metamaterials have been particularly impactful in the fields of mechanics and photonics, where they have prompted reevaluation of a number of conventionally accepted bounds on material performance and the discovery of an array of surprising, and often useful, properties.

Optical metamaterials, for example, have enabled control over both electric *and* magnetic fields of light, so that permittivities and permeabilities can be precisely tuned throughout positive, negative, and near-zero values. Through careful design of subwavelength "meta-atoms," optical metamaterials have enabled negative refraction, optical lensing below the diffraction limit of light, and invisibility cloaking. In addition, mechanical metamaterials, thanks to their micron-to-submicron structure, exhibit extraordinary responses to applied forces, including negative bulk moduli, negative Poisson's ratios, and negative mass densities. Such effects have been used to create solids that behave like liquids and ultralight, low-density materials with unprecedented strength.

This session highlighted recent scientific advances in metamaterials—fundamental breakthroughs, technological relevance, and impacts. Speakers discussed metallic and ceramic mechanical metamaterials, compliant mechanisms, new plasmonic and resonant dielectric optical metamaterials and metasurfaces, acoustic metamaterials, microelectromechanical devices, and advanced nano- and microscale manufacturing of large-area metamaterials.

The session began with a talk by Julia Greer (California Institute of Technology), who creates and studies advanced materials that derive extraordinary strength from 3D architecture and microstructure. She also studies recoverable mechanical deformation in compliant nanomaterials. By constructing nanolattices of a wide variety of constituents—from ceramics to metals, semiconductors, and glasses—her research enables new applications in thermomechanics and affects such disparate fields as ultralightweight batteries and biomedical devices. The second speaker, Chris Spadaccini (Lawrence Livermore National Laboratory), described the development of engineering materials with remarkably light weight and ultrahigh stiffness, as well as the relationship between nanostructure and designer properties such as negative thermal expansion and negative stiffness. Next, Andrea Alù (University of Texas, Austin) discussed metamaterial-based design engineering. His research highlights the connection between metamaterials' microscopic structural properties (e.g., symmetry and shape) and their macroscopic response, focusing on the creation of new useful devices that would not be possible with conventional materials. Examples of such devices are one-way antennas, "invisibility cloaks" that work over a wide spectral bandwidth, and acoustic circulators. The final speaker, Alexandra Boltasseva (Purdue University), talked about optical and infrared metamaterials and about metamaterial-enabled devices that could revolutionize optical technologies in communications, photovoltaics, and thermal radiation management. One of her research focus areas is the incorporation of high-temperature and functional materials as constituents of metamaterials. This is a critical frontier as optical metamaterials transition from the laboratory to specific real-world applications with challenging requirements.

Materials by Design: Using Architecture and Nanomaterial Size Effects to Attain Unexplored Properties

Julia Greer
California Institute of Technology

Engineers are actively trying to mimic hard biomaterials such as mollusk shells and beaks because of their resilience and damage tolerance, which are believed to stem from hierarchical arrangements in their design. Technological advances in fabrication methods make it possible to create architected structural metamaterials using materials whose design and dimensionality are similar to those found in nature and whose hierarchical ordering ranges from angstroms and nanometers at the level of material microstructure to microns and millimeters at the level of macroscale architecture.

The use of architectural features in defining multidimensional material design space will enable the independent manipulation of coupled physical attributes and the development of materials with unprecedented capabilities. The result is that the behavior and properties of architected structural metamaterials can no longer be defined solely by the properties of the constituent solid or by the structure or architecture; instead they benefit from the linked behavior of the material and the structure at very small dimensions. The material creation paradigm will shift from *structure→processing→property* to *property→architecture→fabrication*.

The feasibility of this "materials by design" approach is gated by the ability to understand and predict the mechanical responses of metamaterials, whose properties are controlled by engineered structure in addition to atomic composition, and whose feature size and geometry are critical design parameters. Current research is making significant headway testing, demonstrating, and understanding these properties.

INTRODUCTION

The concept of architecture in the design of materials is common in nature. For example, the properties of butterfly wings are a function of their multiscale, hierarchical construction. The nanoarchitecture of a butterfly wing interacts with light to create multicolored wings, and for shells such as nacre small ceramic platelets are key to their extreme resilience (Von Freymann et al. 2013). Many research groups have attempted to mimic and replicate the properties of these and other natural materials.

Hierarchically designed cellular materials have been used for many decades as the basis for mechanically robust engineered structures, such as the Eiffel Tower. The introduction of architectural elements enables the creation of structures that are both lightweight, because they use a fraction of monolithic material with the same dimensions, and strong, because the architecture provides a way to more efficiently distribute load-bearing capability.

Nanoarchitected structural metamaterials extend the concept of architecture to the micro- and nanometer length scale, often down to the atomistic level of material microstructure, enabling the use of material size effects to create metamaterials with amplified properties. Size effects that emerge in solids at the nanoscale—single crystalline metals become stronger, nanocrystalline metals become weaker, and metallic glasses and ceramics undergo brittle-to-ductile transition (Greer and De Hosson 2011)—can be exploited by using micron- to nanometer-scale building blocks to create larger structures.

STRUCTURAL METAMATERIALS: MANIPULATION OF CELLULAR PROPERTIES

The mechanical performance of architected solids on the macroscale is a function of their deformation mechanism, relative density, and constituent material properties (Deshpande and Fleck 2001; Fleck et al. 2010; Gibson and Ashby 1997; Meza et al. 2014; Ng et al. 2009; Schaedler et al. 2011; Valdevit et al. 2013; Wadley et al. 2003). Cellular solids offer a useful combination of light weight and mechanical integrity and have been studied at the macroscale experimentally, computationally, and theoretically. All of these studies assume that material properties of the constituent solid are constant and geometry is the single tunable parameter (Fleck and Qiu 2007; Fleck et al. 2010; Hutchinson and Fleck 2006; Symons and Fleck 2008).

Cellular solids can deform by either bending or stretching of the elements, both of which are dictated by lattice geometry and its nodal connectivity (Sun et al. 2013; Wang et al. 2003). A three-dimensional (3D) structure must have a connectivity of $Z = 6$ at the nodes to be rigid; a connectivity of $Z = 12$ to be stretching-dominated, which leads to the stiffest materials; and a connectivity of $6 \leq Z < 12$ to be bending-dominated, which results in more compliant materials (Wang et al. 2003).

The structural deformation mechanism, determined by the nodal connectivity, directly affects the stiffness and yield strength of the overall structure (Deshpande et al. 2001a,b). The yield strength and stiffness of 3D open-cell bending-dominated structures, such as honeycombs, scale as $\sigma_y = 0.3\bar{\rho}^{1.5}\sigma_{ys}$ and $E = \bar{\rho}^2 E_s$, where σ_{ys} and E_s are the yield strength and modulus of the constituent solid (Gibson and Ashby 1997). For 3D stretching-dominated structures, such as the octet truss, the yield strength and modulus scale as $\sigma_y = 0.3\bar{\rho}\sigma_{ys}$ and $E = 0.3\bar{\rho}E_s$, causing strength to decrease less rapidly (compared to bending-dominated structures) as relative density decreases.

Relative densities of cellular solids, along with stiffness and strength, can be modulated by using hollow tubes instead of solid rods in the same architecture or by creating hierarchical structures (Figure 1). When hollow tubes are used in low-density cellular solids, structural effects can be activated by changing the various ratios of geometric parameters that define the lattice tubes. For example, reducing the slenderness ratio from 1 to 20 causes a 2 order of magnitude reduction in relative density, and hollowing out the tubes reduces density by an additional order of magnitude. Replacing solid beams with self-similar, fractal-like elements gives a further 1.5 to 2 order of magnitude reduction in density.

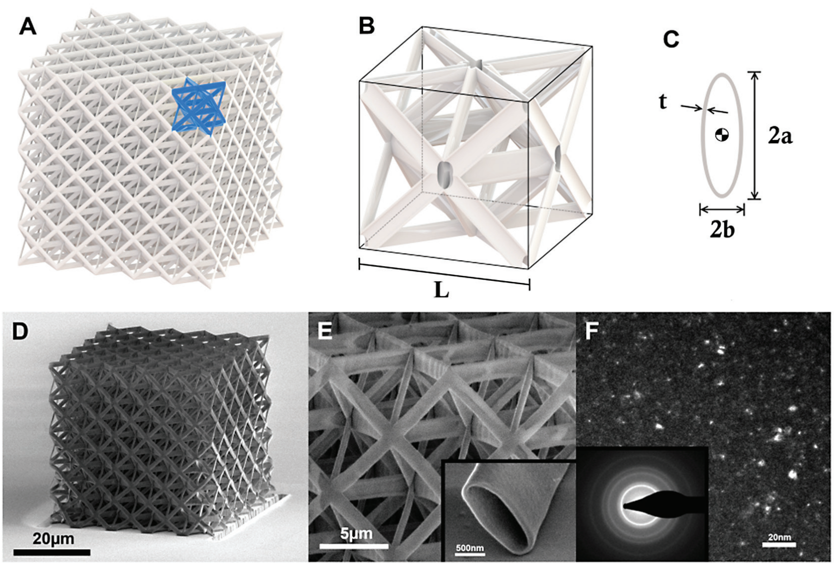

FIGURE 1 Octet truss design. (A) Single unit cell highlighted. (B) Cutaway of hollow octet truss unit cell. (C) Schematic of hollow elliptical nanolattice tube. (D) Scanning electron microscope image of alumina nanolattice. (E) Zoomed-in image of (D). (F) Dark field transmission electron microscope image with diffraction pattern: amorphous microstructure. Reprinted with permission from Meza et al. (2014).

DEMONSTRATED IMPROVEMENTS IN MECHANICAL BEHAVIOR

Recent work in our research group demonstrated the attainment of simultaneous light weight, high strength/stiffness, and recoverability by combining the architecture and material size effect of nanomaterials. We tested the mechanical behavior of octet alumina nanolattices, a class of new, lightweight durable materials that have optimized nano-sized induced material properties, high surface area, and 3D architectures and are desperately needed for a variety of applications (e.g., small-scale energy storage devices, biomedical devices, space travel vehicles, wind turbines, personnel protection, and lightweight thermal insulation).

One way to think about such nanolattices is a 2D nanoribbon wrapped around a 3D architecture (Figure 2). Ordered cellular solids such as the octet truss and kagome lattice have robust mechanical properties such as a combination of high strength and fracture toughness (Deshpande et al. 2001a,b). The geometry of these structures makes it possible to create materials with not only improved mechanical properties but also low densities and high surface area to volume ratios.

We found that they contain ~99 percent air recovered after compression of more than 50 percent (Jang et al. 2013; Meza and Greer 2014; Meza et al. 2014, 2015; Montemayor et al. 2014). Hollow rigid nanolattices have recently attracted much interest as they attained GPa-level stiffnesses at ~10 percent of the density of the parent solid (Jang et al. 2013; Meza and Greer 2014; Meza et al. 2014, 2015; Montemayor et al. 2014). At some critical wall thickness nearly all nanotrusses—ceramic, metallic, and metallic glass—fully recovered after <50 percent compression without sacrifice in strength. Our most recent work demonstrated that these nanolattices also exhibit insensitivity to flaws, that is, failure tolerance (Montemayor et al. 2015).

All of these developments position scientists favorably to address the challenge of developing lightweight and damage-tolerant materials that can have a suite of other valuable properties, such as thermal insulation (conductivity), as well as electronic and optical response.

FABRICATION

Significant efforts are under way to develop architected materials with precisely designed (i.e., nonstochastic) geometries that can support the prototyping of exceptionally lightweight, strong, and tough metamaterials. Control of such nanolattice geometries has been made possible by advances in 3D micro- and nanofabrication techniques, ranging from polymer waveguides to microstereolithography to direct laser writing two-photon lithography (Zheng et al. 2014).

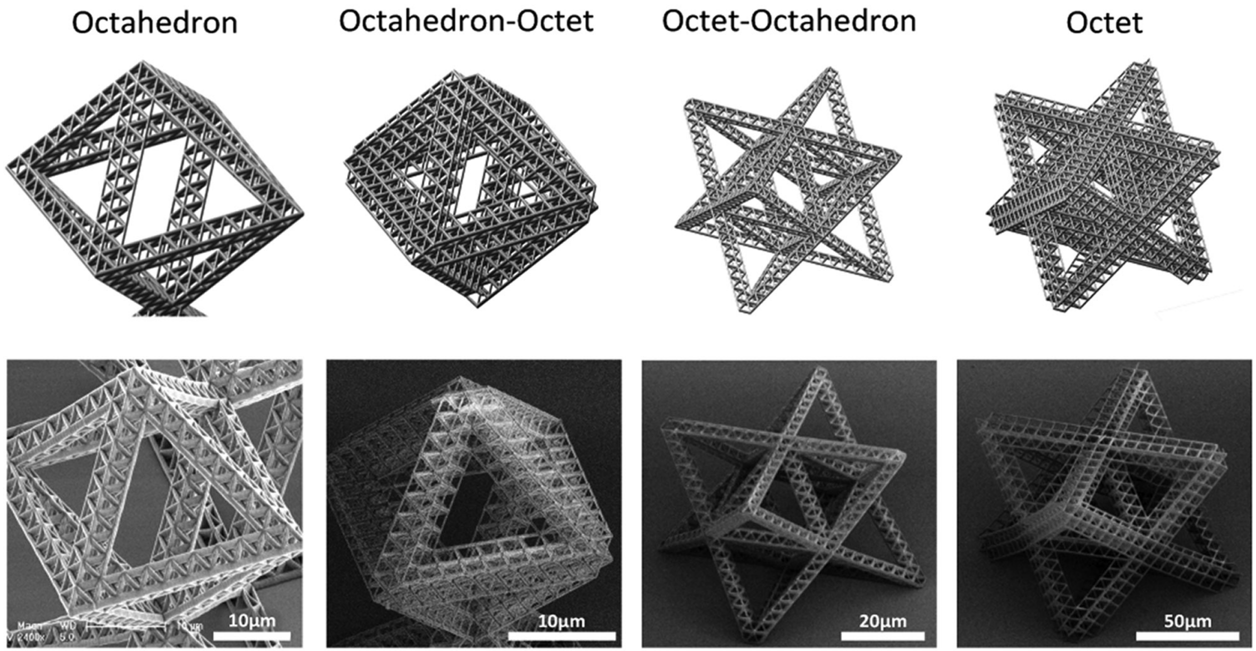

FIGURE 2 Computer-aided designs and scanning electron microscope images of fractal nanotrusses with various geometries specified above each set. Reprinted with permission from Meza et al. (2015).

Direct Laser Writing Two-Photon Lithography (DLW TPL)

Advantages of the TPL method over other 3D fabrication techniques are precise nanometer and submicron feature resolution, smooth beam surfaces, and extreme versatility in geometry (Fischer and Wegener 2013; Sun and Kawata 2004; Xiong et al. 2012). Nanolattices fabricated by TPL enable a reduction in size by up to 3 orders of magnitude compared to microlattices fabricated by 3D printing or by self-propagating polymer waveguides, and they are amenable to subsequent coating and scaffold removal. Their appearance is monolithic and may resemble an iridescent aerogel whose strength is comparable to that of metals and ceramics.

The individual building blocks that constitute nanolattices have dimensions that are below the resolution of the human eye, with strut lengths of 3–20 μm, diameters of 100 nm–1 μm, and wall thicknesses of 5–600 nm (Jacobsen et al. 2007; Jang et al. 2013; Meza and Greer 2014; Meza et al. 2014, 2015; Montemayor et al. 2014, 2015). Their useful properties arise from a combination of size-dependent nanomaterial properties and structural response of the discrete architecture; they cannot be predicted solely by scale-free continuum theories (Gibson and Ashby 1997; Torrents et al. 2012; Valdevit et al. 2013).

Figure 3 shows a schematic representation of the fabrication steps during the nanolattice creation procedure. Nanolattices were first fabricated from a negative photoresist (IP-Dip 780) using DLW TPL, which uses a 780 nm femtosecond

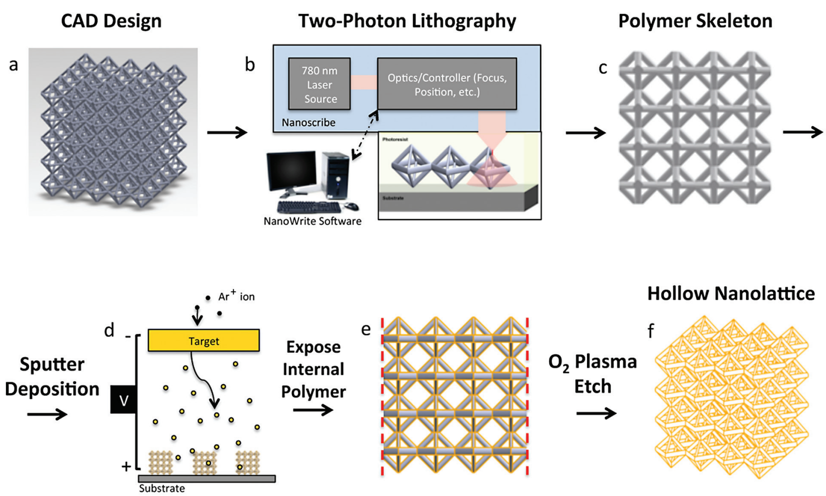

FIGURE 3 A schematic of the two-photon lithography technique followed by the deposition of material of interest and etching out of the internal polymer scaffold. Reprinted with permission from Montemayor et al. (2014).

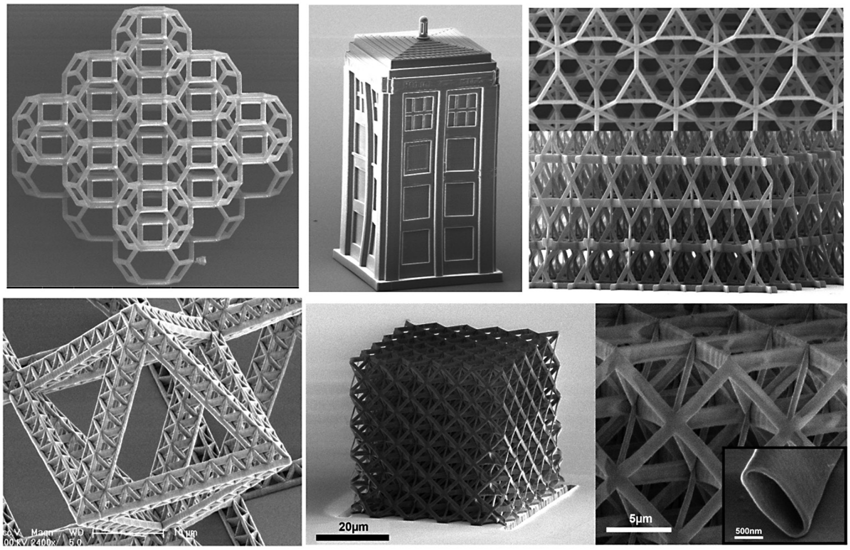

FIGURE 4 Examples of different architectures written by the two-photon lithography technique showing the versatility of this approach. Reprinted with permission from Meza et al. (2015) (*bottom row left*) and Meza et al. (2014) (*bottom row center and right*).

pulsed laser focused into a small volume element, a voxel, in the polymer. The energy is sufficient to cross-link the monomer and to harden the material within the voxel before the two photons are absorbed. The elliptical voxel is rastered in three dimensions in a droplet of photoresist to effectively sculpt the prescribed structure, which can have virtually any geometry; some examples are shown in Figures 1, 2, and 4.

Other Methods

To make nanolattices out of different materials, the polymer scaffolds are coated (as conformally as possible) with the specific material of interest (e.g., metals, semiconductors, oxides, metallic glasses, piezoelectric materials, other polymers). We have demonstrated the feasibility of sputtering different metals onto nanolattices up to ~300 nm (Montemayor and Greer 2015; Montemayor et al. 2014); and Meza and colleagues and several other research groups successfully utilized atomic layer deposition to apply ~5–60-nm-thick ceramic alumina (Al_2O_3) and titanium nitride (TiN) coatings onto polymer scaffolds (Jang et al. 2013). After deposition, the original internal polymer scaffold is exposed by slicing two of the six sides off the sample using a focused ion beam and then removed using an oxygen plasma, revealing a hollow-tube nanolattice that is made entirely of the coating material (Meza and Greer 2014; Meza et al. 2014, 2015; Montemayor and Greer 2015; Montemayor et al. 2014, 2015).

DEMONSTRATION OF ENHANCED PROPERTIES

Hollow Nanolattices

Our group has demonstrated that hollow Al_2O_3 nanolattices with octet truss geometry and relative densities of $\rho = 10^{-4} - 10^{-1}$ recovered up to 98 percent of their initial height after uniaxial compressions in excess of 50 percent (Jang et al. 2013; Meza and Greer 2014; Meza et al. 2014, 2015). Their ductile-like deformation and recoverability were attributed to the critical ratio between the wall thickness (t) and the semimajor axis (a) of the hollow elliptical strut cross-section. When (t/a) is less than the critical value (~0.03 for alumina), the nanolattices deformed via shell buckling and recovered after deformation; at greater (t/a) values, they failed catastrophically with little to no recovery (Meza and Greer 2014; Meza et al. 2014).

The constituent Al_2O_3 is a brittle ceramic; reducing the wall thickness to the nanometer level and "wrapping" the film around a 3D architecture enables the metamaterial to bypass the properties of the constituent solid and to exhibit a diametrically opposite property, recoverability. The ~10 nm thickness reduces the number and size of flaws in the Al_2O_3—the largest is only 10 nm! The combination of Weibull statistics flaw distribution-based fracture, a material size effect, and structural deformation mechanism as a function of (t/a) makes it possible for ceramic nanolattices to have such unique properties (Jang et al. 2013; Meza and Greer 2014; Meza et al. 2014, 2015).

Solid Nanolattices

The enhanced properties induced by the interplay of structural and material size effects have also been observed in TPL-created lattices with solid metal struts (Blanco et al. 2000). Copper (Cu) mesolattices were created using an inverse DLW TPL method: a negative of the scaffold was patterned in a positive-tone resist, and then Cu was electroplated into the exposed pores (Blanco et al. 2000). Removal of the polymer mold revealed a free-standing mesolattice with solid Cu beams whose thickness and grain size were ~2 μm. These mesolattices had relative densities between 40 percent and 80 percent and a unit cell of 6 μm and 8 μm, and at the highest density their compressive yield strength was ~330 MPa, which is a factor of ~2.5 higher than that of bulk copper, 133 MPa.

This amplification in strength is a result of the size effect in single crystalline metals ("smaller is stronger") (Blanco et al. 2000). These results imply that the architected Cu nanolattices outperformed bulk copper by a factor of ~3 for relative densities of $\rho > 0.6$—that is, when more than 40 percent of the material was removed from a monolithic cube of the same volume.

Other Structures and Materials

Current pursuits in the field of architected structural metamaterials include efforts to develop auxetic 3D structures (i.e., materials with a negative Poisson's ratio), which are fabricated using DLW TPL (Babaee et al. 2013; Lee et al. 2012). Typical monolithic materials expand in the direction orthogonal to the loading axis when uniaxially compressed, whereas auxetic materials contract in this direction.

Pentamode materials, or 3D solids that ideally behave as fluids, have also been designed and fabricated using TPL and provide a pathway to decouple bulk and shear moduli, enabling independent wave propagation through media (Rinne et al. 2007).

CHALLENGES OF SCALABILITY

The most significant obstacle to incorporating structural nanoarchitected metamaterials in useful technological applications is scalability, that is, the ability to manufacture either a large number of small-scale components or materials with large dimensions in a reasonable amount of time. Stochastic foams (e.g., nanoporous gold), in which the porosity distribution is random, have been explored as a possible vehicle for propagating material size effects to larger scales (Hodge et al. 2005). But these foams exhibit poor scaling of strength with relative density and are ultimately limited in the types of architecture they can create (Fleck et al. 2010).

Some technologies that could lead to scalability are roll-to-roll fabrication with nanoimprintable patterns, holographic lithography, and phase-shifting masks.

OUTLOOK

This selective overview highlights the state of the art of some structural metamaterials with dimensions that span microns to nanometers. Experiments on nanolattices have demonstrated that unique material size effects that emerge only at the nanoscale can be effectively propagated to macroscopic dimensions and have the potential to create new classes of bulk engineered materials with unprecedented properties.

Nanoarchitected metamaterials represent a new approach to "materials by design," making it possible to create materials with previously unattainable combinations of properties—light weight and enhanced mechanical performance—as well as unique thermal, optical, acoustic, and electronic attributes. Some realistic technological advances enabled by these metamaterials are untearable and unwettable paper, tunable filters and laser sources, tissue implants generated on biodegradable scaffolds that can be implanted and absorbed by the human body, battery-powered implantable chemical sensors, and extremely insulating and super-thin thermal lining for jackets and sleeping bags.

REFERENCES

Babaee S, Shim J, Weaver JC, Chen ER, Patel N, Bertoldi K. 2013. 3D soft metamaterials with negative Poisson's ratio. Advanced Materials 25(36):5044–5049.

Blanco A, Chomski E, Grabtchak S, Ibisate M, John S, Leonard SW, Lopez C, Meseguer F, Miguez H, Mondia JP, and 3 others. 2000. Large-scale synthesis of a silicon photonic crystal with a complete three-dimensional bandgap near 1.5 micrometres. Nature 405:437–440.

Deshpande VS, Fleck NA. 2001. Collapse of truss core sandwich beams in 3-point bending. International Journal of Solids and Structures 38:6275–6305.

Deshpande VS, Ashby MF, Fleck NA. 2001a. Foam topology: Bending versus stretching dominated architectures. Acta Materialia 49:1035–1040.

Deshpande VS, Fleck NA, Ashby MF. 2001b. Effective properties of the octet-truss lattice material. Journal of the Mechanics and Physics of Solids 49(8):1747–1769.

Fischer J, Wegener M. 2013. Three-dimensional optical laser lithography beyond the diffraction limit. Laser and Photonics Reviews 7(1):22–44.

Fleck NA, Qiu X. 2007. The damage tolerance of elastic-brittle, two-dimensional isotropic lattices. Journal of the Mechanics and Physics of Solids 55:562–588.

Fleck NA, Deshpande VS, Ashby MF. 2010. Micro-architectured materials: Past, present and future. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science 466:2495–2516.

Gibson LJ, Ashby MF. 1997. Cellular solids: Structure and properties. Cambridge, UK: Cambridge University Press.

Greer JR, De Hosson JTM. 2011. Plasticity in small-sized metallic systems: Intrinsic versus extrinsic size effect. Progress in Materials Science 56:654–724.

Hodge AM, Biener J, Hsiung LL, Wang YM, Hamza AV, Satcher JH. 2005. Monolithic nanocrystalline Au fabricated by the compaction of nanoscale foam. Journal of Materials Research 20:554–557.

Hutchinson RG, Fleck NA. 2006. The structural performance of the periodic truss. Journal of the Mechanics and Physics of Solids 54:756–782.

Jacobsen AJ, Barvosa-Carter W, Nutt S. 2007. Micro-scale truss structures formed from self-propagating photopolymer waveguides. Advanced Materials 19(22):3892–3896.

Jang D, Meza LR, Greer F, Greer JR. 2013. Fabrication and deformation of three-dimensional hollow ceramic nanostructures. Nature Materials 12:893–898.

Lee J-H, Singer JP, Thomas EL. 2012. Micro-/nanostructured mechanical metamaterials. Advanced Materials 24(36):4782–4810.

Meza LR, Greer JR. 2014. Mechanical characterization of hollow ceramic nanolattices. Journal of Materials Science 49(6):2496–2508.

Meza LR, Das S, Greer JR. 2014. Strong, lightweight, and recoverable three-dimensional ceramic nanolattices. Science 345:1322–1326.

Meza L, Zelhofer AJ, Clarke N, Mateos AJ, Kochmann DM, Greer JR. 2015. Resilient 3D hierarchical architected metamaterials. Proceedings of the National Academy of Sciences 112(37): 11502–11507.

Montemayor LC, Meza LR, Greer JR. 2014. Design and fabrication of hollow rigid nanolattices via two-photon lithography. Advanced Engineering Materials 16(2):184–189.

Montemayor LC, Greer JR. 2015. Mechanical response of hollow metallic nanolattices: Combining structural and material size effects. Journal of Applied Mechanics 82(7):071012.

Montemayor LC, Wong WH, Zhang Y-W, Greer JR. 2015. Insensitivity to flaws leads to damage tolerance in brittle architected meta-materials. Nature Scientific Reports (in revision).

Ng KY, Lin Y, Ngan AHW. 2009. Deformation of anodic aluminum oxide nano-honeycombs during nanoindentation. Acta Materialia 57:2710–2720.

Rinne SA, García-Santamaría F, Braun PV. 2007. Embedded cavities and waveguides in three-dimensional silicon photonic crystals. Nature Photonics 2:52–56.

Schaedler TA, Jacobsen AJ, Torrents A, Sorensen AE, Lian J, Greer JR, Valdevit L, Carter WB. 2011. Ultralight metallic microlattices. Science 334:962–965.

Sun H-B, Kawata S. 2004. Two-photon photopolymerization and 3D lithographic microfabrication. Advances in Polymer Science, vol. 170: NMR, 3D Analysis, Photopolymerization. Berlin: Springer-Verlag. pp. 169–273.

Sun J, Bhushan B, Tong J. 2013. Structural coloration in nature. RSC Advances 3(35):14862–14889.

Symons DD, Fleck NA. 2008. The imperfection sensitivity of isotropic two-dimensional elastic lattices. Journal of Applied Mechanics 75(5):051011.

Torrents A, Schaedler TA, Jacobsen AJ, Carter WB, Valdevit L. 2012. Characterization of nickel-based microlattice materials with structural hierarchy from the nanometer to the millimeter scale. Acta Materialia 60:3511–3523.

Valdevit L, Godfrey SW, Schaedler TA, Jacobsen AJ, Carter WB. 2013. Compressive strength of hollow microlattices: Experimental characterization, modeling, and optimal design. Journal of Materials Research 28:2461–2473.

Von Freymann G, Kitaev V, Lotsch BV, Ozin GA. 2013. Bottom-up assembly of photonic crystals. Chemical Society Reviews 42(7):2528–2554.

Wadley HNG, Fleck NA, Evans AG. 2003. Fabrication and structural performance of periodic cellular metal sandwich structures. Composites Science and Technology 63:2331–2343.

Wang J, Evans AG, Dharmasena K, Wadley HNG. 2003. On the performance of truss panels with kagome cores. International Journal of Solids and Structures 40:6981–6988.

Xiong W, Zhou YS, He XN, Gao Y, Mahjouri-Samani M, Jiang L, Baldacchini T, Lu YF. 2012. Simultaneous additive and subtractive three-dimensional nanofabrication using integrated two-photon polymerization and multiphoton ablation. Light: Science and Applications 1:e6.

Zheng X, Lee H, Weisgraber TH, Shusteff M, Deotte JR, Duoss E, Kuntz JD, Biener MM, Kucheyev SO, Ge Q. 2014. Ultra-light, ultra-stiff mechanical metamaterials. Science 344:1373–1377.

Mechanical Metamaterials: Design, Fabrication, and Performance

Christopher Spadaccini
Lawrence Livermore National Laboratory

Material properties are governed by the chemical composition and spatial arrangement of the constituent elements at multiple length scales. This characteristic fundamentally limits the properties with respect to each other and requires tradeoffs when selecting materials for specific applications. For example, strength and density are inherently linked so that, in general, the more dense the material, the stronger it is.

In my laboratory we are combining advanced microstructural design, using flexure and screw theory as well as topology optimization, with advanced additive micro- and nanomanufacturing techniques to create new material systems with previously unachievable property combinations. The performance of these *mechanical metamaterials* is controlled by geometry at multiple length scales rather than by chemical composition alone.

We have demonstrated designer properties of these mechanical metamaterials in polymers, metals, ceramics, and combinations thereof, yielding properties such as ultrastiff lightweight materials, negative stiffness, and negative thermal expansion. Our manufacturing techniques include projection microstereolithography (PμSL), direct ink writing (DIW), and electrophoretic deposition (EPD). With these methods, we can generate three-dimensional micro- and nanoscale architectures with multiple constituent materials in the same structure.

INTRODUCTION

Material properties can be controlled via intricate assemblies and structural organization at multiple length scales, as evidenced by naturally occurring cellular materials such as honeycombs (Ando and Onda 1999), trabecular bone (Ryan and Shaw 2013), plant parenchyma (Van Liedekerke et al. 2010), and sponges (Shen et al. 2013). It is the architecture of the material's structure at the micro- and nanoscale, as much as the chemical composition, that yields its mechanical properties. By designing highly ordered architectures in cellular solids, it is possible to engineer the mechanical response of these materials to create mechanical metamaterials (Deshpande et al. 2001; Gibson and Ashby 2001).

The ability to decouple properties via micro- and nanoarchitectural control can allow for unique material performance such as ultralightweight, high-stiffness, and high-strength materials (Bauer et al. 2014; Schaedler et al. 2011), negative Poisson's ratio (Lakes 1987), negative stiffness (Lakes et al. 2001), and negative thermal expansion coefficient (Sigmund and Torquato 1997). Paramount to achieving these often unnatural properties is an ability to design, fabricate, and characterize structures for the properties of interest. In fact, this method of choosing a unique property and engineering a material's performance through its architecture could be described as an *inverse design* problem. Normally, material properties are taken to be absolute and functional structures are then created from these materials. Mechanical metamaterials result from exactly the opposite approach.

A classic example of architectural control and the resulting unique material performance is the octet truss stretch-dominated lattice (Deshpande et al. 2001) shown in Figure 1. This structure, which contains b struts and j frictionless joints, satisfies Maxwell's criterion, where $b - 3j + 6 > 0$, which defines a stretch-dominated structure. Because the struts in the unit cell are designed to be

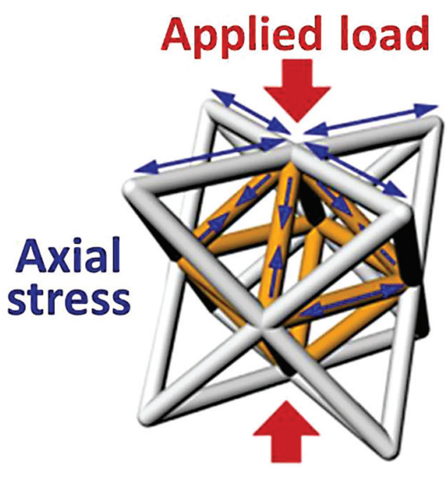

FIGURE 1 Octet truss stretch-dominated unit cell that can be tessellated in space to form a lattice. SOURCE: Zheng et al. (2014).

in either tension or compression under applied load, as opposed to bending, the lattice is mechanically efficient with a high stiffness-to-weight ratio (E/ρ). In fact, it is designed to have a linear scaling relationship between stiffness and density, $E/E_s \propto (\rho/\rho_s)$, where the subscript *s* denotes bulk properties.

Most naturally occurring materials with stochastic porosity have a quadratic or even cubic relationship: for every order of magnitude decrease in density there is a corresponding 2 to 3 order of magnitude decrease in stiffness. The architected design fundamentally changes the scaling relationship of the lattice material through geometry rather than composition. This concept can be further advanced by taking advantage of nanoscale size effects. Strength to density relationships can be effectively manipulated with control at size-scales below the critical flaw and crack dimensions (Jang et al. 2013).

DESIGN

Numerous methods can be used to solve the inverse design problem for mechanical metamaterials. In our laboratory and with key collaborators, we have primarily been developing and using two techniques, one analytical and the other computational. The analytical method is known as *freedom, actuation, and constraint topologies* (FACT) and it relies on design of flexure and screw elements to create unit cells and lattices with prescribed properties (Hopkins and Culpepper 2010a,b). The computational method we have been using is *topology optimization* (TO), which involves optimizing a unit cell's layout subject to an objective function and boundary conditions.

The FACT Method

The FACT method relies on use of a previously developed, comprehensive library of geometric shapes that define fundamental flexure and screw motion. These shapes enable the designer of a unit cell to visualize all the regions wherein various microstructural elements may be placed to achieve desired bulk material properties.

As an example, consider a two-dimensional (2D) unit cell design with negative thermal expansion that was derived using FACT and is shown in Figure 2 (Hopkins et al. 2013). In this unit cell, two materials (shown in red and grey) plus void space are required to achieve the negative property. As the unit cell heats up, the red material, which has a larger thermal expansion than the grey material, volumetrically expands more relative to its grey counterpart. Consequently, the red angled component pulls inward the center of the flexure element, which makes up the sidewall of the unit cell, while simultaneously pushing the corners of the unit cell outward. When arranged in a lattice connected at the midpoint of each sidewall, the corners grow into the void space while the sidewalls are pulled inward, resulting in an overall contraction of the lattice and hence negative thermal expansion.

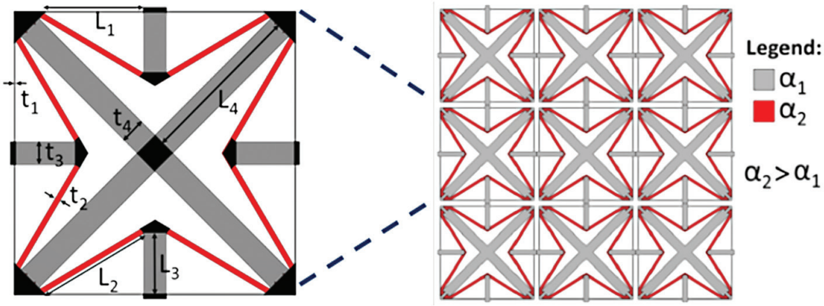

FIGURE 2 Unit cell and lattice with negative thermal expansion designed using freedom, actuation, and constraint topologies (FACT). α = coefficient of thermal expansion; L = length; t = thickness. Adapted from Hopkins et al. (2013).

The FACT technique can be used to design other mechanical metamaterials with properties such as negative Poisson's ratio and nonlinear responses.

Topology Optimization

Topology optimization is a computationally driven inverse design method. In our implementation, we utilize a finite element solver as the core physics engine and an optimization algorithm subject to an objective function and constraints to evolve the design.

A negative thermal expansion metamaterial again serves as an example. In a typical implementation, we begin with a unit cell with three phases randomly distributed throughout the space: a high thermal expansion constituent material, a relatively lower thermal expansion constituent material, and void space. An objective function such as a specific target thermal expansion for the unit cell is defined along with quantitative constraints such as stiffness and volume fraction bounds.

Initially, the finite element solver calculates the material properties for the random distribution of phases. These properties will likely be far from the target and may also violate the constraints. At this point, the optimization algorithm, which in this case is a gradient-based method, will redistribute the three phases by some small amount and the finite element solver again calculates properties. The new properties are then evaluated against the target and the previously calculated values and the optimization algorithm again redistributes material based on this information in an attempt to approach the target.

This iterative process is repeated until it converges to a design that minimizes the objective function and satisfies all constraints. An example of a topology-optimized negative thermal expansion unit cell design is shown in

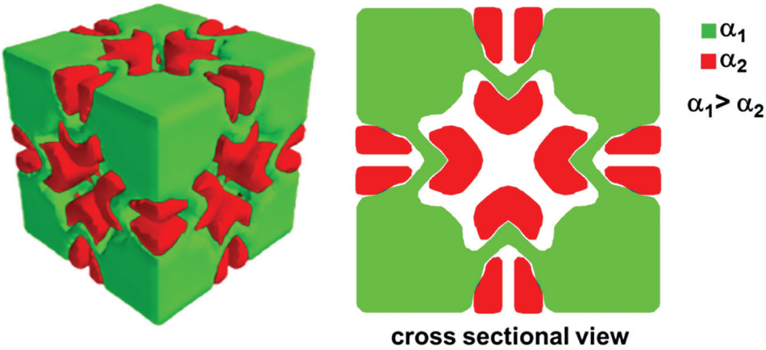

FIGURE 3 Negative thermal expansion metamaterial unit cell designed using topology optimization. The red and green represent two different materials with different thermal expansion coefficients (α). Figure in color at http://www.nap.edu/catalog/21825.

Figure 3. In some cases convergence may not be achieved because the problem is overconstrained and/or has poor initial conditions.

There are significant limitations to TO methods, including a lack of knowledge about practical manufacturing constraints in the algorithm, a propensity to converge to a local minimum solution rather than a global one, and, for more sophisticated design problems, the need for expensive high-performance computing resources.

FABRICATION

Physical realization of mechanical metamaterials requires a suite of fabrication processes with unique capabilities. Additive manufacturing (AM) methods are particularly well suited to the geometric complexity of these structures and lattices. But some features and geometries in these structures are not attainable with commercially available AM tools, and so we have developed our own custom processes and materials, described and illustrated in the following sections.

Projection Microstereolithography

Projection microstereolithography (PμSL) uses a spatial light modulator—a liquid crystal on silicon (LCoS) or digital micromirror device (DMD)—as a dynamically reconfigurable digital photomask to fabricate three-dimensional materials in a layer-by-layer fashion.

A 3D computer-aided design (CAD) model is first sliced into a series of closely spaced horizontal planes. These 2D image slices are sequentially transmitted to the reflective LCoS chip, which is illuminated with ultraviolet (UV) light from a light-emitting diode (LED) array. Each image is projected through a reduction lens onto the surface of a photosensitive resin. The exposed liquid cures, forming a layer in the shape of the 2D image, and the substrate on which it rests is lowered, reflowing a thin film of liquid over the cured layer. The image projection is then repeated with the next image slice until the desired number of layers has been fabricated to complete the 3D structure. A schematic of the basic system is shown in Figure 4 along with a photo of an example structure (Zheng et al. 2014).

We have also recently developed a scanning version of this concept, enabling us to rapidly fabricate structures approaching 10 cm in size while maintaining features as small as 10 microns.

Direct Ink Writing

Another method that we have been utilizing to fabricate these metamaterials is direct ink writing. DIW is a layer-by-layer printing approach in which concentrated inks are deposited in planar and 3D layouts with lateral dimensions (minimum ~400 nm) that are at least an order of magnitude lower than those achieved by conventional extrusion-based printing methods. Paramount to this approach is the creation of concentrated inks that can be extruded through fine deposition nozzles as filaments and then undergo rapid solidification to maintain their shape, as shown in Figure 5 (Lewis 2002; Smay et al. 2002). In many cases they can even span significant gaps across unsupported regions (Ahn et al. 2009).

Techniques such as DIW offer an attractive alternative to conventional manufacturing technologies because of the low cost of the printing equipment, ease of manufacture, and flexibility of material systems and dimensions.

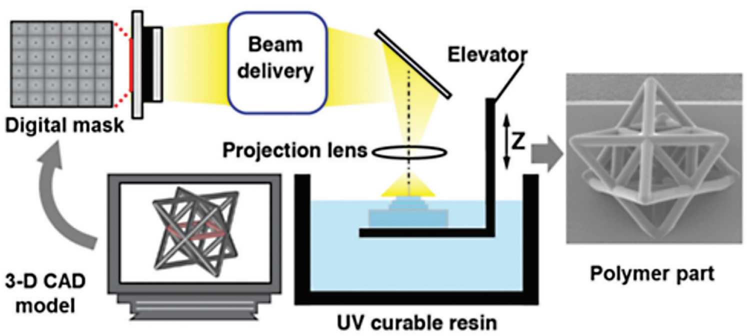

FIGURE 4 Schematic (*left*) of projection microstereolithography (PμSL) and photo (*right*) of a fabricated component. CAD = computer-aided design; UV = ultraviolet. SOURCE: Zheng et al. (2014).

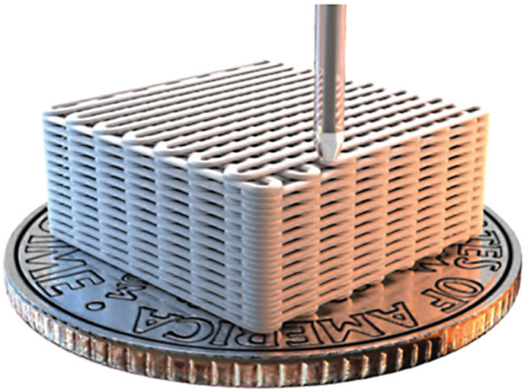

FIGURE 5 Direct ink writing is an extrusion-based process.

Electrophoretic Deposition

A third fabrication method that allows for the deposition of a range of materials is electrophoretic deposition (EPD), a bottom-up fabrication process that utilizes electric fields to deposit charged nanoparticles from a solution onto a substrate (Besra and Liu 2007; Pascall et al. 2014).

EPD can be used with a wide range of nanoparticles, including oxides, metals, polymers, and semiconductors. Once the particles are deposited the green body can be dried and/or sintered to adhere the particles together into a fixed structure. A schematic and fabricated nanostructure are shown in Figure 6.

EPD has traditionally been used for coating applications, such as depositing ceramic materials onto metal tooling. We have expanded the technology to enable patterning of mesoscale multimaterial structures with micron-scale tailoring. Our modifications include automated sample injection during deposition to tailor the material composition, the use of dynamic electrodes to controllably vary the electric field profile on the deposition plane and precisely pattern geometries, and in-depth process modeling to predict the deposition parameters required to achieve a specific packing structure.

Postprocessing Techniques

Postprocessing techniques can help expand the palette of usable materials and/or improve final component properties. Thermal treatments such as sintering and hot isostatic pressing are frequently used. The use of polymer structures as templates for other materials is another common method for achieving an expanded material set.

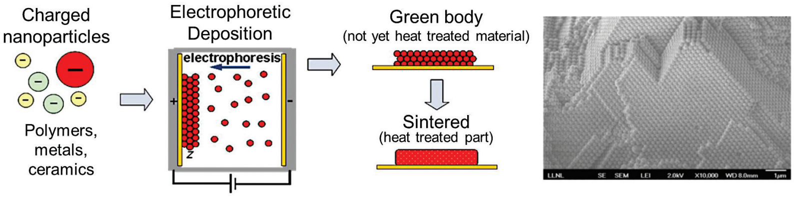

FIGURE 6 Schematic (*left*) of the electrophoretic deposition process and (*right*) photo of a fabricated structure.

For example, a polymer structure fabricated with any one of the methods described above could be coated via electroless plating, atomic layer deposition (ALD), or chemical vapor deposition (CVD) processes. The polymer in these hybrid structures could then be thermally or chemically removed, converting the structure to pure metallic or ceramic hollow structures.

Finally, nanoparticles can be suspended in the base feedstocks such as liquid monomers, resulting in a hybrid structure with particles distributed throughout the polymer. Thermal processing can then be used to remove the polymer and densify the particles, leaving a relatively pure, nonpolymeric structure.

PERFORMANCE

Mechanical metamaterials are becoming a reality thanks to advanced fabrication and design methods. Two examples of simple lattice-based materials with unique properties are ultralight, ultrastiff microlattices and elastomeric cellular architectures with negative stiffness.

Figure 7 shows four types of octet truss stretch-dominated lattices fabricated at the microscale using PμSL. The images in the first column show a basic polymeric lattice made from hexanediol diacrylate (HDDA) with solid microscale struts and 11 percent relative density. In the next column, the same polymer structure was electrolessly plated with nickel phosphorus (Ni-P), and the polymer core was then removed via thermal processing, resulting in a lattice with 0.5 percent relative density. A hollow tube ceramic lattice, shown in the third column, was formed via ALD and similar polymer removal. This structure represents the lightest fabricated material in this test series, with a relative density of 0.025 percent and wall thickness less than 50 nm. Finally, we fabricated a lattice with alumina nanoparticles suspended in the polymer and conducted sintering procedures to remove the polymer and densify the ceramic, yielding a solid ceramic lattice with 8 percent relative density (far right column; Zheng et al. 2014).

The performance of these mechanical metamaterials is illustrated in Figure 8, where nondimensional stiffness is plotted versus relative density. The fact that the

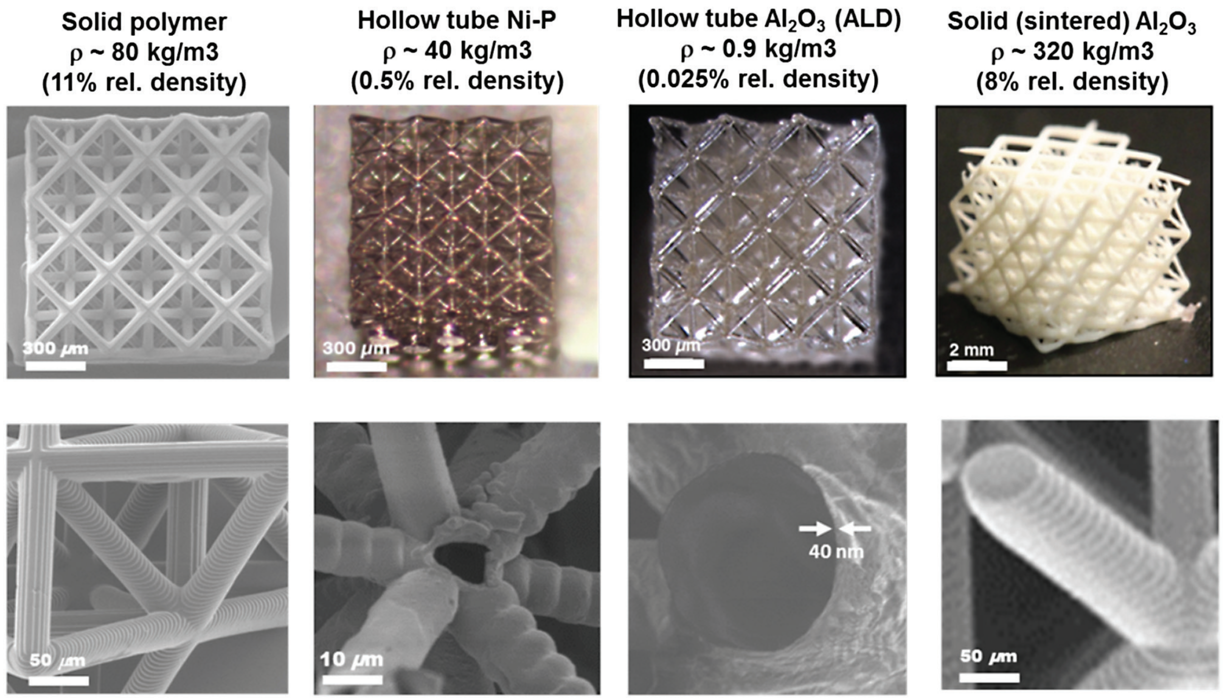

FIGURE 7 Octet truss–based mechanical metamaterials of varying materials and configurations. See text for description of techniques used to produce these structures. Al_2O_3 = aluminum oxide; ALD = atomic layer deposition; Ni-P = nickel phosphorus. SOURCE: Zheng et al. (2014).

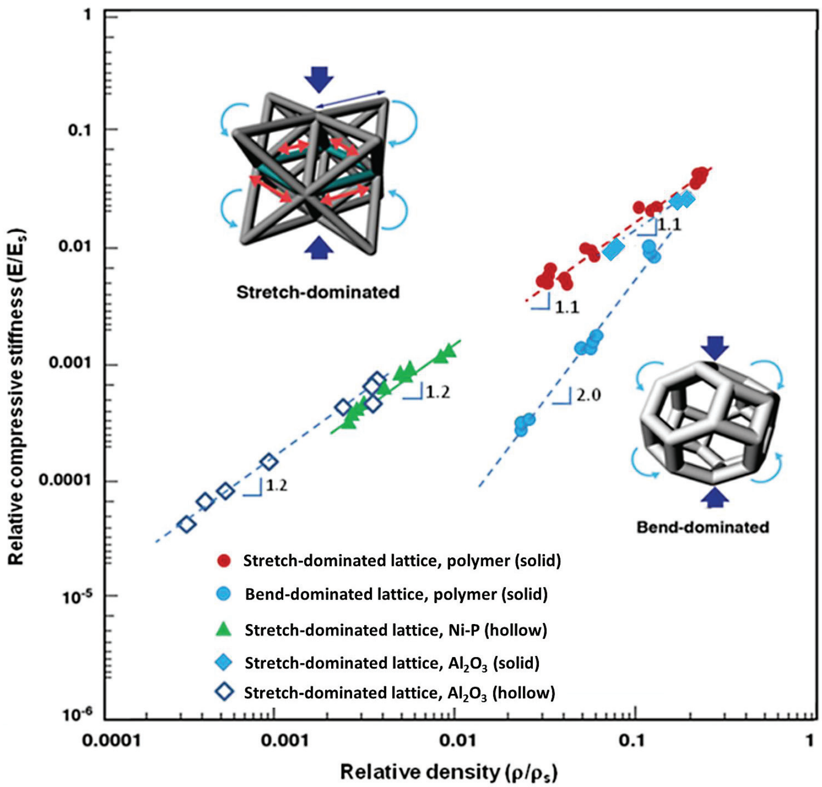

FIGURE 8 Nondimensional performance of ultralight octet truss stretch-dominated lattices. AI_2O_3 = aluminum oxide; HDDA = hexanediol diacrylate; Ni-P = nickel phosphorus. Adapted from Zheng et al. (2014).

measured scaling relationship between these two parameters is approximately linear across all constituent material types, all relative density regimes, and regardless of hollow tube or solid strut configurations clearly demonstrates the impact of the stretch-dominated architecture. For comparison, a bend-dominated Kelvin foam architecture displayed the classic quadratic relationship when tested. Furthermore, in absolute terms, the hollow tube alumina lattices have densities approaching aerogels (known as some of the lightest materials in the world), but with 4 to 5 more orders of magnitude in stiffness due to the architected structure (Zheng et al. 2014).

Another interesting mechanical metamaterial is shown in Figure 9. These "woodpile" structures were fabricated out of silicone with the DIW process. Varying the architecture between a simple cubic (SC) type structure and a face-centered tetragonal (FCT) configuration results in different bulk-scale mechani-

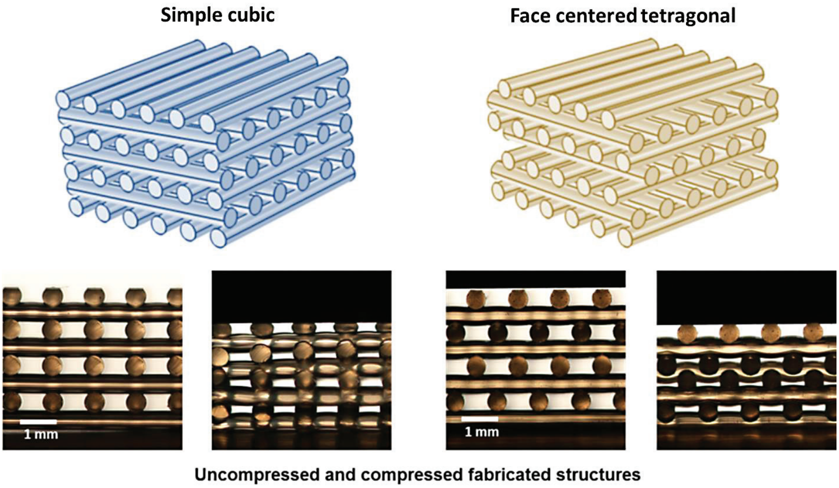

FIGURE 9 Schematics (*top*) and images (*bottom*) of silicone structures fabricated using direct ink writing. Adapted from Duoss et al. (2014).

cal properties. The cross-sections in the bottom row of the figure clearly show that the two structures will have different compressive behavior: the SC layout will be stiffer than the FCT-structured material because of the aligned versus staggered arrangement of the nodes (Duoss et al. 2014).

However, what is not as obvious is the difference in shear response of the two materials. Figure 10 elucidates this difference, which manifests as "negative stiffness" and can be seen in the negative slope of the stress-strain response in the SC material. The negative stiffness is a unique "snap-through" property that can be engineered into the material. It is now possible not only to control the compressive response but also to independently design and control the shear response, perhaps even with negative stiffness (Duoss et al. 2014).

We have also begun to explore *multifunctional metamaterials*, combining normally disparate physics into lattice-type architectures. Figure 11 shows an early example, a printed graphene aerogel structure (Zhu et al. 2015). This simple woodpile lattice has enhanced mechanical properties and, because of the nature of graphene, significant electrical and/or chemical functionality as well. Supercapacitors with high compressibility and durability, for example, may become possible.

FUTURE DIRECTIONS

By combining mechanical metamaterials with inverse design methods and custom micro- and nanoadditive manufacturing techniques, we have been able

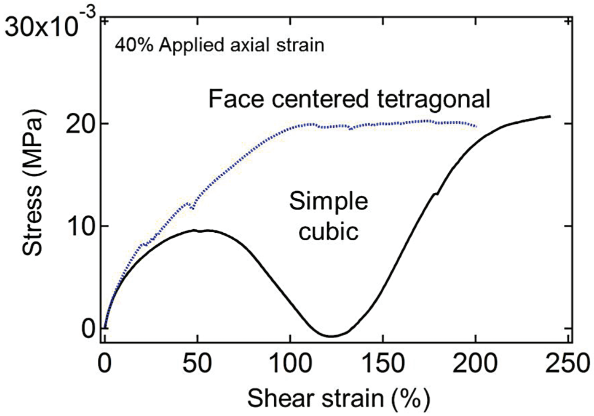

FIGURE 10 Shear response of elastomeric cellular architectures showing negative stiffness. MPa = megapascals. Based on Duoss et al. (2014).

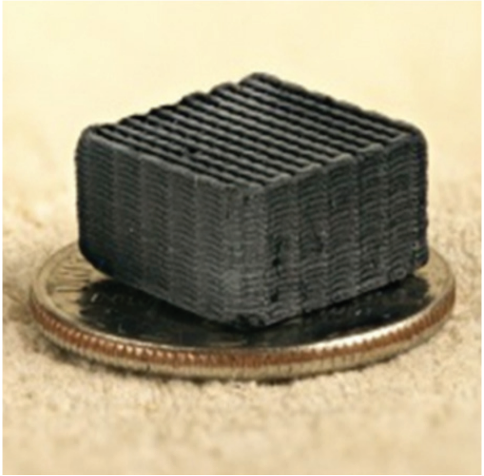

FIGURE 11 Image of printed graphene aerogel, positioned atop a quarter to show size. SOURCE: Zhu et al. (2015).

to develop unique properties not previously attainable in known materials. This is just the beginning of a powerful new methodology for approaching material design and realization.

There are many potential future directions for advancing the state of the art, including through continued exploration of size-scale effects and efforts that push the boundaries of multimaterial design and fabrication. This could lead to unique new materials with groundbreaking mechanical properties and electrical or photonic functionality all in one.

ACKNOWLEDGMENT

This work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344 (LLNL-CONF-677195).

REFERENCES

Ahn BY, Duoss EB, Motala MJ, Guo X, Park S, Xiong Y, Yoon J, Nuzzo RG, Rogers JA, Lewis JA. 2009. Omnidirectional printing of flexible, stretchable, and spanning silver microelectrodes. Science 323(5921):1590–1593.

Ando K, Onda H. 1999. Mechanism for deformation of wood as a honeycomb structure I: Effect of anatomy on the initial deformation process during radial compression. Journal of Wood Science 45:120.

Bauer J, Hengsbach S, Tesari I, Schwiger R, Kraft O. 2014. High-strength cellular ceramic composites with 3D microarchitecture. Proceedings of the National Academy of Sciences 111(7):2453–2458.

Besra L, Liu M. 2007. A review on fundamentals and applications of electrophoretic deposition (EPD). Progress in Materials Science 52(1):1–61.

Deshpande VS, Fleck NA, Ashby MF. 2001. Effective properties of the octet-truss lattice material. Journal of the Mechanics and Physics of Solids 49(8):1747–1769.

Duoss EB, Weisgraber TH, Hearon MK, Zhu C, Small W IV, Metz TR, Vericella JJ, Barth HD, Kuntz JD, Maxwell RS, Spadaccini CM, Wilson TS. 2014. Three-dimensional printing of elastomeric, cellular architectures with negative stiffness. Advanced Functional Materials 24(31):4905–4913.

Gibson JL, Ashby FM. 2001. Cellular Solids: Structure and Properties. Cambridge, UK: Cambridge University Press.

Hopkins JB, Culpepper ML. 2010a. Synthesis of multi-degree of freedom, parallel flexure system concepts via freedom and constraint topology (FACT)—Part I: Principles. Precision Engineering 34(2):259–270.

Hopkins JB, Culpepper ML. 2010b. Synthesis of multi-degree of freedom, parallel flexure system concepts via freedom and constraint topology (FACT)—Part II: Practice. Precision Engineering 34(2):271–278.

Hopkins JB, Lange KJ, Spadaccini CM. 2013. Designing the microstructure of thermally actuated materials using freedom, actuation, and constraint topologies. Journal of Mechanical Design 135(6):061004.

Jang D, Meza LR, Greer F, Greer JR. 2013. Fabrication and deformation of three-dimensional hollow ceramic nanostructures. Nature Materials 12:893–898.

Lakes R. 1987. Foam structures with a negative Poisson's ratio. Science 235(4792):1038–1040.

Lakes R, Lee T, Bersie A, Wang YC. 2001. Extreme damping in composite materials with negative-stiffness inclusions. Nature 410:565–567.

Lewis JA. 2002. Direct-write assembly of ceramics from colloidal inks. Current Opinion in Solid State and Materials Science 6(3):245–250.

Pascall AJ, Qian F, Wang G, Worsley MA, Li Y, Kuntz JD. 2014. Light-directed electrophoretic deposition: A new additive manufacturing technique for arbitrarily patterned 3D composites. Advanced Materials 26:2252–2256.

Ryan TM, Shaw CN. 2013. Trabecular bone microstructure scales allometrically in the primate humerus and femur. Proceedings of the Royal Society B: Biological Sciences 280:20130172.

Schaedler TA, Jacobsen AJ, Torrents A, Sorensen AE, Lian J, Greer JR, Valdevit L, Carter WB. 2011. Ultralight metallic microlattices. Science 334(6058):962–965.

Shen JH, Xie YM, Huang XD, Zhou SW, Ruan D. 2013. Compressive behavior of luffa sponge material at high strain rate. Advances in Engineering Plasticity XI 535-536:465–468.

Sigmund O, Torquato S. 1997. Design of materials with extreme thermal expansion using a three-phase topology optimization method. Journal of the Mechanics and Physics of Solids 45(60):1037–1067.

Smay JE, Cesarano J, Lewis JA. 2002. Colloidal inks for directed assembly of 3-D periodic structures. Langmuir 18(14):5429–5437.

Van Liedekerke P, Ghysels P, Tijskens E, Samaey G, Smeedts B, Roose D, Ramon H. 2010. A particle-based model to simulate the micromechanics of single-plant parenchyma cells and aggregates. Physical Biology 7(2):026006.

Zheng X, Lee H, Weisgraber TH, Shusteff M, DeOtte J, Duoss E, Kuntz J, Biener MM, Ge Q, Jackson J, Kucheyev SO, Fang NX, Spadaccini CM. 2014. Ultralight, ultrastiff mechanical metamaterials. Science 344(6190):1373–1377.

Zhu C, Han TY, Duoss EB, Golobic AM, Kuntz JD, Spadaccini CM, Worsley MA. 2015. Highly compressible 3D periodic graphene aerogel microlattices. Nature Communications 6:6962.

Metamaterial-based Device Engineering

Andrea Alù
The University of Texas at Austin

Metamaterials are artificial materials with unusual bulk properties (Engheta and Ziolkowski 2006) based on suitably designed arrays of complex resonant subwavelength elements. By definition, metamaterials have properties not available in any of their constituents (or in any natural material). These properties emerge from strong wave-matter interactions and carefully engineered mesoscopic structures.

In the past decade, metamaterials have opened several exciting directions in basic science, such as the realization of artificial plasmas at microwaves, artificial magnetism in optics, negative refraction, cloaking and extreme scattering manipulation, and large wave control over surfaces significantly thinner than the wavelength. These features not only are shedding light on fascinating new areas for basic research in optics, electromagnetics, acoustics, and beyond, but also are yielding important, direct applications in more applied engineering contexts.

I briefly review recent impacts of metamaterials in a few engineering fields where they have helped overcome some long-standing challenges for technology.

CLOAKING AND RADIO-TRANSPARENT ANTENNAS

Invisibility has been a tantalizing concept in human culture for several centuries. In engineering, the ability to cloak objects is very important for camouflaging and noninvasive biomedical sensing, to name just two applications. With recent developments in metamaterial science and technology, the possibility of cloaking objects against electromagnetic radiation has escaped from the realm of science fiction to become a technological reality (Alù and Engheta 2005; Schurig et al. 2006).

Given inherent challenges in substantially suppressing the scattering from objects that are many wavelengths large with a passive coating (Monticone and Alù

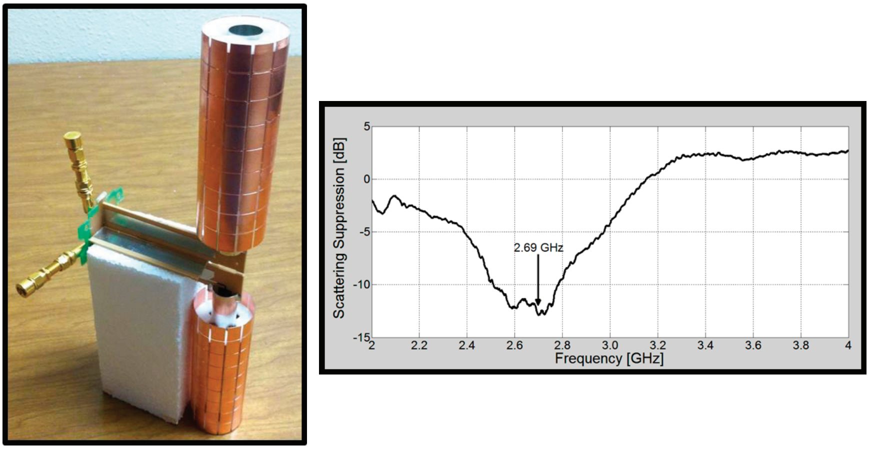

FIGURE 1 *Left*: A cloaked dipole antenna, which can transmit and receive signals while producing significantly less scattering than a bare antenna. *Right*: Measured scattering gain, defined as the ratio (in decibels, dB) between a bare dipole and a cloaked one. Adapted from Soric et al. (2015).

2013), a viable application of cloaking technology is in the area of radio frequency–cloaked antennas or radiators, whose size is comparable to the wavelength and therefore allows considerable scattering suppression over large bandwidths.

In our group, we have worked for several years on radiowave cloaking applications (Alù 2009; Monti et al. 2012) and have recently shown that conventional antennas can be cloaked with suitably designed metasurfaces to greatly reduce scattering from radio waves while preserving the possibility of transmitting and receiving signals over large bandwidths of operation (Soric et al., 2015). Figure 1 shows a cloaked dipole antenna for cellular communications; it largely suppresses radar cross-section at all angles, compared to a bare dipole, while transmitting and receiving radio frequency signals with good matching and isolation performance over the entire cellular band.

AN INVISIBLE ACOUSTIC SENSOR

Cloaking a sensor or receiving antenna is challenging because the action of sensing requires extracting a portion of the impinging signal and thus creating a shadow (Fleury et al. 2014a). Schemes that involve active (instead of passive) circuit elements to realize more advanced metamaterials provide a way around this limitation.

Inspired by recent advances in quantum mechanics in the area of parity-time (PT) symmetry, we have shown that it is possible to realize an invisible acoustic

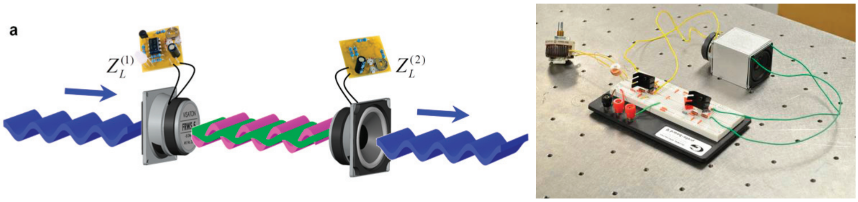

FIGURE 2 An invisible acoustic sensor based on parity time (PT) symmetric meta-atoms in the form of loudspeakers loaded by circuit components with conjugate impedances. The setup enables a shadow-free efficient acoustic sensor. Z_L represents the loading impedances on the two loudspeakers: the one on the left is passive, and the one on the right is its time-reversed image and therefore active. Adapted from Fleury et al. (2015a).

sensor with strong absorption properties by pairing a resonant sensor with its time-reversed image (Figure 2; Fleury et al. 2015a). This device can absorb the entire incoming signal without creating any shadow or reflection. The system was built by pairing in free space two identical loudspeakers loaded by circuits with conjugate impedances: one passive, for converting the impinging sound into a voltage across a resistor, the second active, for emitting a signal, in sync with the impinging one, that suppresses shadows and reflections.

This class of PT-symmetric sensors can also be used as the basis of PT-symmetric metamaterials, to realize loss-free negative index aberration-free planar lenses (Fleury et al. 2014b) and advanced cloaks (Sounas et al. 2015).

NONRECIPROCAL MAGNET-FREE DEVICES

Sound, light, and many other waves tend to travel symmetrically in space: if a signal can be sent from point A to B, then it can typically be sent from B to A. This symmetry, known as reciprocity, is due to the fact that wave propagation in conventional media, including light and sound, is time reversible.

Reciprocity is not necessarily desirable, however, especially when one wants to isolate a source from its echo or separate signal flows travelling in opposite directions. To enable full-duplex communication—the transmission and receipt of signals from the same transducer on the same frequency channel—which may lead to more efficient radiowave communications or better ultrasound imaging devices, it may be necessary to break reciprocity.

The most common way to break reciprocity is based on magnetic bias, but this method has several challenges, including the use of scarce materials and difficulty integrating them on-chip. We recently proved that reciprocity can be broken by applying an angular momentum bias to a metamaterial cavity, obtained by loading a hollow cavity with small fans, enabling the first-of-its-kind circulator for acoustic waves (Figure 3; Fleury et al. 2014c).

FIGURE 3 A table-top circulator that breaks the symmetry of acoustic wave travel in air. It is formed by a circularly symmetric cavity loaded with three CPU fans (*upper right*) that move air at a moderate velocity that is enough to break reciprocity and enable large isolation. The cavity is connected to three acoustic waveguides (*bottom left*) carrying sound signals. Adapted from Fleury et al. (2014c).

The system is a basic three-port device that allows one-way rotation of the input signals—from port 1 to 2, 2 to 3, and 3 to 1—while preventing transmission in the opposite direction. Very large isolation (over 40 decibels) was realized for airborne acoustic waves using a suitably designed subwavelength acoustic ring cavity in which air was rotated simply by using fans, and the cavity was symmetrically coupled to three acoustic waveguides, which formed the input and output channels of the device.

While it is interesting to see how such a basic active component can modify the way sound propagates, a mechanical motion of the material in the cavity may not always be convenient or practical, especially when translating these effects to other types of waves, such as radio signals, which travel much faster than sound. Estep and colleagues (2014) and Fleury and colleagues (2015b) extended these concepts to an equivalent meta-atom in which fluid motion was replaced by spatio-temporal modulation of three strongly coupled resonators, realizing magnet-free circulators for radio and ultrasound waves. Magnet-free circulators may enable cell-phone and handset devices with non-reciprocal, circuit-based components, a crucial development to realize full-duplex communications, that is, to be able to transmit and receive through the same antenna, on the same frequency channel, and at the same time.

CONCLUSIONS

The field of metamaterials has opened exciting directions in basic research in the past 15 years, but only recently has it become a platform for important opportunities in applied technology, with implications for many engineering fields. In this work, we have reviewed the application of metamaterials to scattering suppression and non-reciprocal wave propagation, with specific attention to its impact on practical applications of interest to the engineering world.

ACKNOWLEDGMENT

This work was supported by the Department of Defense and the National Science Foundation.

REFERENCES

Alù A. 2009. Mantle cloak: Invisibility induced by a surface. Physical Review B 80(24):245115.
Alù A, Engheta N. 2005. Achieving transparency with plasmonic and metamaterial coatings. Physical Review E 72(1):016623.
Engheta N, Ziolkowski RW. 2006. Metamaterials: Physics and Engineering Explorations. Hoboken, NJ: John Wiley & Sons.
Estep N, Sounas D, Soric J, Alù A. 2014. Magnetic-free non-reciprocity based on parametrically modulated coupled-resonator loops. Nature Physics 10(12):923–927.
Fleury R, Soric J, Alù A. 2014a. Physical bounds on absorption and scattering for cloaked sensors. Physical Review B 89(4):045122.
Fleury R, Sounas D, Alù A. 2014b. Negative refraction and planar focusing based on parity-time symmetric metasurfaces. Physical Review Letters 113(2):023903.
Fleury R, Sounas DL, Sieck CF, Haberman MR, Alù A. 2014c. Sound isolation and giant linear nonreciprocity in a compact acoustic circulator. Science 343(6170):516–519.
Fleury R, Sounas DL, Alù A. 2015a. An invisible acoustic sensor based on parity-time symmetry. Nature Communications 6:5905.
Fleury R, Sounas DL, Alù A. 2015b. Subwavelength ultrasonic circulator based on spatio-temporal modulation. Physical Review B 91(17):174306.
Monti A, Soric J, Alù A, Toscano A, Vegni L, Bilotti F. 2012. Overcoming mutual blockage between neighboring dipole antennas using a low-profile patterned metasurface. IEEE Antennas and Wireless Propagation Letters 11:1414–1417.
Monticone F, Alù A. 2013. Do cloaked objects really scatter less? Physical Review X 3(4):041005.
Schurig D, Mock JJ, Justice BJ, Cummer SA, Pendry JB, Starr AF, Smith DR. 2006. Metamaterial electromagnetic cloak at microwave frequencies. Science 314(5801):977–980.
Soric J, Monti A, Toscano A, Bilotti F, Alù A. 2015. Dual-polarized reduction of dipole antenna blockage using metasurface cloaks. IEEE Transactions on Antennas and Propagation 63(11):4027–4034.
Sounas DL, Fleury R, Alù A. 2015. Unidirectional cloaking based on metasurfaces with balanced loss and gain. Physical Review Applied 4(1):014005.

Catching Light Rays: Refractory Plasmonics for Energy Conversion, Data Storage, and Medical Applications

Alexandra Boltasseva
Purdue University and Nano-Meta Technologies, Inc.

Urcan Guler
Nano-Meta Technologies, Inc.

The influence of optical technologies on the development of modern society cannot be overestimated. From conventional mirrors, lenses, microscopes, telescopes, optical sensors, and high-precision measurement systems to lasers, fiber-optic communications, and optical data storage systems, optical instruments have enabled revolutionary advances and novel concepts in many disciplines, including astronomy, manufacturing, chemistry, biology (particularly bio- and chemical sensors), medicine (particularly ophthalmology and optometry), various engineering fields, and information technology.

INTRODUCTION

The rise of a new generation of optical technologies is fueled by the emergence of nanophotonics, the study of the behavior of light at the nanometer scale and the interaction of nanometer-scale objects with light. A major focus of the field is development of a new class of plasmonic structures and "metamaterials" as potential building blocks for advanced optical technologies.

Plasmonics involves metallic components that can focus and manipulate light at the nanometer scale via excitation of surface plasmons, collective oscillations of free electron clouds (found in metallic materials) coupled to light (Maier 2007). Plasmonics can "squeeze" light into tiny volumes much smaller than the wavelength of light, thus offering light confinement beyond the usual diffraction limit as well as extreme light enhancement in nanometer-scale areas known as plasmonic "hot spots" (Lindquist et al. 2013; Schuller et al. 2010).

Plasmonic devices are expected to transform optoelectronics, microelectronics, on-chip optical communication, and data transmission by enabling low-power, nanometer-scale photodetectors; fast light modulators; and nanoscale, power-efficient lasers and light sources. Plasmonics is paving the way for optical microscopy and photolithography with nanometer-scale resolution, novel concepts for data recording and storage, improved energy harvesting through optimized light-capturing techniques, single-molecule sensing, and advanced spectroscopy.

PLASMONIC CONSTITUENT MATERIALS: CHALLENGES

It is now possible to design plasmonic structures and metamaterials—artificial composite surfaces or materials that use plasmonic building blocks as their functional unit cells—with versatile properties that can be tailored to fit almost any practical need. But new plasmonic technologies will require the resolution of significant limitations associated with the use of metals as constituent materials. There are difficulties in the fabrication and integration of metal nanostructures with existing semiconductor technology, and the materials need to be more precisely tuned to have the optical properties needed for the required functionality.

Current Materials

In the devices demonstrated so far, too much light is absorbed in the metals (e.g., silver and gold) commonly used in plasmonic structures and metamaterials. Moreover, such metals are soft materials with relatively low melting points. Thus plasmonic devices cannot meet the challenges that real industry applications face, particularly those characterized by high operational temperatures (e.g., energy harvesting or data recording applications, as explained below), high pressure, and harsh chemical environments. For these reasons, plasmonics is a "what happens in the lab stays in the lab" area of research.

The development of practical devices utilizing plasmonic concepts will depend on new constituent materials that both exhibit the plasmonic properties required to capture and manipulate light at the nanoscale and provide durable, chemically, mechanically, and thermally stable solutions for the realization of rugged optical instruments.

New Materials: Metal Nitrides

Plasmonic ceramic materials have recently been proposed as the basis for practical, low-loss, complementary metal-oxide semiconductor (CMOS)–compatible plasmonic devices, an important advance for the field (Naik et al. 2013).

Transition metal nitrides such as titanium nitride (TiN) and zirconium nitride (ZrN) have been suggested as refractory plasmonic materials that are capable of sustaining high temperatures (the TiN melting point is 2930°C) and exhibit good optical properties along with bio- and CMOS compatibility, robustness, chemical stability, corrosion resistance, mechanical strength, and durability (Guler et al. 2014a, 2015b). The attractiveness of TiN for practical device applications is illustrated by its extensive use in semiconductor manufacturing (as both a gate layer and a diffusion barrier), large-scale integrated microelectronics, microelectromechanical systems (MEMS), and biotechnology.

APPLICATIONS OF PLASMONICS

Energy Conversion

Improved efficiency of light harvesting is among the major engineering challenges for the upcoming decade. Photovoltaics (PV) is considered an important potential energy source, but its development is hampered by problems connected to efficiency and stability degradation when the device heats during the absorption of solar radiation.

Various techniques based on plasmonic effects have been proposed to improve solar cell efficiencies via field concentration and hot electron generation (Polman and Atwater 2012). Plasmonic metamaterials have been investigated as broadband absorbers and spectrally engineered emitters for solar thermophotovoltaic (STPV) systems (Li et al. 2014; Molesky et al. 2013). The STPV concept involves a perfect absorber designed for broad absorption of solar radiation while a selective emitter (designed to emit light in a narrow energy band just above the semiconductor bandgap in the PV cell) can be heated by the absorber through an intermediate layer or via chemical, nuclear, or waste heat sources (Bauer 2011; Fan 2014).

The beauty of the STPV approach is that the system can be used in a hybrid mode (hence the name), but high-temperature operation again introduces the problem of material degradation (Guler et al. 2015a). High operational temperatures (well above 800°C) have hindered STPV progress because of low melting points for noble metals, poor optical performance, and lattice imperfections for refractory metals.

Refractory plasmonic ceramics such as TiN represent a unique platform for realizing STPV as an energy conversion concept that promises efficiencies of up to 85 percent (Guler et al. 2015a). TiN absorbers have already been shown to provide high optical absorption (about 95 percent) over a broad range while being extremely durable under exposure to heat and strong illumination (Li et al. 2014).

Figure 1(a) gives a schematic representation of an STPV system and exemplary absorber and emitter metamaterial designs. Figure 1(b) shows the absorption spectra measurements from identical plasmonic metamaterials made of TiN and gold (Au). TiN metamaterial provides broader absorption and retains its optical properties after 8 hours of annealing at 800°C, whereas Au metamaterial has narrower resonance peaks and its absorption properties degrade after just 15 minutes annealing at the same temperature (Li et al. 2014).

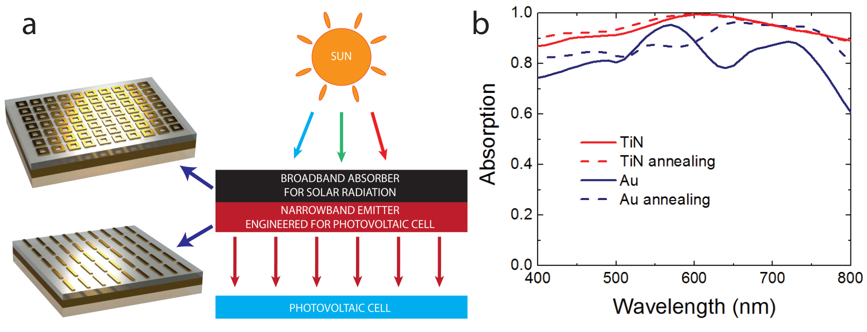

FIGURE 1 (a) A solar thermophotovoltaic system consists of a broadband solar absorber (*top left*) and a spectrally selective emitter (*bottom left*) engineered to match the bandgap of a photovoltaic cell. Adapted with permission from Guler et al. (2015a). © Elsevier (2015). (b) Titanium nitride (TiN) metamaterial provides better absorption compared to an identical gold (Au) absorber and retains its properties after exposure to high temperatures (800°C). Adapted with permission from Li et al. (2014). © John Wiley & Sons, Inc. (2014).

Durable, refractory TiN also holds great promise to enable efficient, TPV-based waste heat recovery (Bauer 2011). Efficient heat energy harvesting could have a transformative effect on a number of industries (e.g., metal casting, aerospace, and gas and oil) and lead to the development of fossil fuel–based power generation (including diesel and gas engines), radioisotope-based cells, fuel-fired cells, and portable power generators for civil and military needs. TiN properties are also well suited for solar thermoelectric generators (Kraemer et al. 2011), plasmon-mediated photocatalysis (Clavero 2014), and plasmon-assisted chemical vapor deposition (Boyd et al. 2006).

MEDICAL TREATMENT

Other heat-generating applications of plasmonic nanoparticles are in health care. Photothermal therapy utilizes a unique property of metallic nanoparticles to concentrate light and efficiently heat a confined nanoscale volume around the plasmonic structure (Loo et al. 2004). Nanoparticles thus delivered to a tumor can be heated via laser illumination at near-infrared wavelength in the biological transparency window. Hyperthermia is known to induce cell death in diseased and other tissues and has been shown to increase both local control of treatment and overall survival in combination with radiotherapy and chemotherapy in randomized clinical trials.

Gold nanoparticles are emerging as promising agents for cancer therapy and are being investigated as drug carriers, photothermal agents, contrast agents, and radiosensitizers. But gold nanoparticles resonate at light wavelengths that lie outside the biological transparency window and therefore require larger dimensions and complex geometries such as nanoshells and nanorods (Huang et al. 2008). The larger sizes in turn affect the nanoparticles' pharmacokinetics, biodistribution, and in vivo toxicity.

TiN-nanofabricated particles have been shown to exhibit both plasmonic resonance in the biological transparency window and higher heating efficiencies than gold (Guler et al. 2013). More importantly, TiN obviates the need for complex geometries and provides simple, small-size particles that optimize cellular uptake and clearance from the body after treatment (Guler et al. 2014b).

Medical Imaging

Figure 2(a) shows a high-resolution transmission electron microscope image and optical transmittance data from a colloidal single crystalline TiN sample. Lattice parameters of the nanoparticle closely match the tabulated single crystalline bulk values of TiN samples, and the optical transmittance data show the plasmonic extinction dip at the biological transparency window.

Figure 2(b) shows a comparison between the calculated absorption efficiencies of TiN (left-hand graph) and Au (right-hand graph) nanodisks. The dipolar resonance peak of Au is located around 520 nm, where the excitation light is strongly attenuated in biological samples. The TiN dipolar resonance peak, at about 800 nm, allows the use of small particles.

TiN is a very contamination-safe material (and therefore widely used in surgical tools, food-contact applications, and medical implants), so TiN colloidal particles could become a next-generation solution for tumor-selective photothermal therapy, medical imaging, and other biomedical applications.

Data Recording

Refractory plasmonic materials are considered the best candidates for applications that require nanometer-scale field enhancement and local heating. An example of such an application is heat-assisted magnetic recording (HAMR; Challener et al. 2009), a nanophotonic next-generation data recording technology that will significantly increase the amount of data on a magnetic disk by using a laser light tightly focused on a magnetic material. The tight focusing to a subwavelength spot is achieved via a plasmonic nanoantenna.

In contrast to noble metals that are prone to deformations such as melting and creep because of material softness and melting point depression in nanostructures, any degradation of refractory plasmonic materials can be avoided with the proper material integration (Guler et al. 2015a; Li et al. 2014). TiN antennae have recently been shown to satisfy the stringent requirements for an optically efficient, durable HAMR near-field transducer, paving the way for next-generation data recording systems (Guler et al. 2015a).

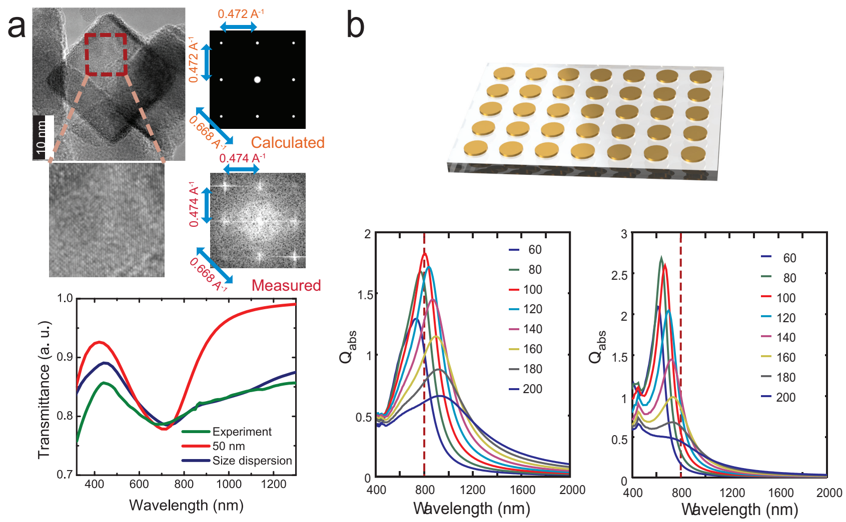

FIGURE 2 (a) Actual measures of lattice parameters of a single crystalline titanium nitride (TiN) nanoparticle closely match calculated values from bulk TiN samples. The graph below shows the transmittance data from a colloidal TiN sample with a plasmonic extinction dip at the biological transmittance window. a. u. = arbitrary unit. Adapted with permission from Guler et al. (2014b). (b) Absorption efficiencies calculated for TiN (*left*) and gold (Au; *right*) nanodisks according to the design at the top of the figure. Small TiN nanodisks provide enhanced absorption at 800 nm (dashed vertical line) while large nanodisks of Au are required at the same wavelength because of spectral mismatch. Q_{abs} = absorption efficiency (optical cross-section normalized by the geometric cross-section). Adapted with permission from Guler et al. (2013). ©American Chemical Society (2013).

More generally, the use of refractory plasmonic ceramics can greatly expand the realm of tip-based applications, including near-field scanning optical microscopes and other local field–enhanced signal measures, opening up measurement capacities in previously unavailable frequency ranges and operational regimes (Boltasseva and Shalaev 2015; Guler et al. 2014a).

Other Applications

The durability and refractory properties of TiN and ZrN could also make them the only material building block for high-temperature, harsh environment optical sensors, flat photonic components such as ultrathin lenses, and spatial light modulators using the concepts of the emerging field of metasurfaces. Refractory flat optical components last longer in harsh environments, provide more reliable data, and offer ultracompactness combined with a planar fabrication process that is large-scale, robust, and low-cost. In oil and gas industries, ultracompact, extremely durable plasmonic sensors could replace electrical sensors and enable novel measures for pressure, flow, drill bit temperature, and breakage detection.

The thermal, mechanical, and chemical stability of TiN, together with its high conductivity and corrosion resistance, make it an ideal material for nanofabrication. TiN can be used for making ultradurable imprint stamps with unparalleled hardness and resistance to wet chemistry processes. When combined with emerging plasmonic nanolithography schemes, TiN films can be used to create durable multiple-use master molds and novel fabrication concepts for large-scale sub-10-nanometer resolution patterning.

Finally, CMOS-compatible refractory plasmonic materials are considered a platform for next-generation on-chip hybrid photonic-electronic devices such as subwavelength photodetectors, optical interconnects, and modulators with unprecedented compactness, speed, and efficiency (Kinsey et al. 2015).

CONCLUSION

With an excellent combination of hardware performance properties, appealing optical properties, durability, and contamination safety, plasmonic ceramics in general and TiN in particular have the capacity to enable highly robust, ultracompact, CMOS-compatible optical devices capable of addressing numerous application-specific challenges. As such, they are promising building blocks for advanced optical technologies, including data processing, exchange, and storage; new concepts for energy conversion, including improved solar cells; nanoscale-resolution imaging techniques; a new generation of cheap, enhanced-sensitivity sensors; and novel types of light sources.

REFERENCES

Bauer T. 2011. Thermophotovoltaics: Basic Principles and Critical Aspects of System Design. Berlin: Springer-Verlag.

Boltasseva A, Shalaev VM. 2015. All that glitters need not be gold. Science 347(6228):1308–1310.

Boyd DA, Greengard L, Brongersma M, El-Naggar MY, Goodwin DG. 2006. Plasmon-assisted chemical vapor deposition. Nano Letters 6(11):2592–2597.

Challener WA, Peng C, Itagi AV, Karns D, Peng W, Peng Y, Yang XM, Zhu X, Gokemeijer NJ, Hsia Y-T, Ju G, Rottmayer RE, Seigler MA, Gage EC. 2009. Heat-assisted magnetic recording by a near-field transducer with efficient optical energy transfer. Nature Photonics 3(4):220–224.

Clavero C. 2014. Plasmon-induced hot-electron generation at nanoparticle/metal-oxide interfaces for photovoltaic and photocatalytic devices. Nature Photonics 8(2):95–103.

Fan S. 2014. Photovoltaics: An alternative "Sun" for solar cells. Nature Nanotechnology 9(2):92–93.

Guler U, Ndukaife JC, Naik GV, Nnanna AGA, Kildishev AV, Shalaev VM, Boltasseva A. 2013. Local heating with lithographically fabricated plasmonic titanium nitride nanoparticles. Nano Letters 13(12):6078–6083.

Guler U, Boltasseva A, Shalaev VM. 2014a. Refractory plasmonics. Science 344(6181):263–264.

Guler U, Suslov S, Kildishev AV, Boltasseva A, Shalaev VM. 2014b. Colloidal plasmonic titanium nitride nanoparticles: Properties and applications. Nanophotonics 4(1):269-276.

Guler U, Shalaev VM, Boltasseva A. 2015a. Nanoparticle plasmonics: Going practical with transition metal nitrides. Materials Today 18(4):227–237.

Guler U, Kildishev AV, Boltasseva A, Shalaev VM. 2015b. Plasmonics on the slope of enlightenment: The role of transition metal nitrides. Faraday Discussions 178:71–86.

Huang X, Jain P, El-Sayed I, El-Sayed M. 2008. Plasmonic photothermal therapy (PPTT) using gold nanoparticles. Lasers in Medical Science 23(3):217–228.

Kinsey N, Ferrera M, Shalaev VM, Boltasseva A. 2015. Examining nanophotonics for integrated hybrid systems: A review of plasmonic interconnects and modulators using traditional and alternative materials. Journal of the Optical Society of America B 32(1):121–142.

Kraemer D, Poudel B, Feng H-P, Caylor JC, Yu B, Yan X, Ma Y, Wang X, Wang D, Muto M, McEnaney K, Chiesa M, Ren Z, Chen G. 2011. High-performance flat-panel solar thermoelectric generators with high thermal concentration. Nature Materials 10(7):532–538.

Li W, Guler U, Kinsey N, Naik GV, Boltasseva A, Guan J, Kildishev AV. 2014. Refractory plasmonics with titanium nitride: Broadband metamaterial absorber. Advanced Materials 26(47):7959–7965.

Lindquist NC, Jose J, Cherukulappurath S, Chen X, Johnson TW, Oh S-H. 2013. Tip-based plasmonics: Squeezing light with metallic nanoprobes. Laser and Photonics Reviews 7(4):453–477.

Loo C, Lin A, Hirsch L, Lee MH, Barton J, Halas N, West J, Drezek R. 2004. Nanoshell-enabled photonics-based imaging and therapy of cancer. Technology in Cancer Research and Treatment 3(1):33–40.

Maier SA. 2007. Plasmonics: Fundamentals and Applications. New York: Springer.

Molesky S, Dewalt CJ, Jacob Z. 2013. High temperature epsilon-near-zero and epsilon-near-pole metamaterial emitters for thermophotovoltaics. Optics Express 21(S1):A96–A110.

Naik GV, Shalaev VM, Boltasseva A. 2013. Alternative plasmonic materials: Beyond gold and silver. Advanced Materials 25(24):3264–3294.

Polman A, Atwater HA. 2012. Photonic design principles for ultrahigh-efficiency photovoltaics. Nature Materials 11(3):174–177.

Schuller JA, Barnard ES, Cai W, Jun YC, White JS, Brongersma ML. 2010. Plasmonics for extreme light concentration and manipulation. Nature Materials 9(3):193–204.

Forecasting Natural Disasters

Forecasting Natural Disasters

James Done
National Center for Atmospheric Research

Amir AghaKouchak
University of California, Irvine

Decades of research investment have produced significant advances in the predictive skill of natural hazard forecasts such as tropical cyclone track and intensity and flash flooding. Such advances have arisen from enhanced understanding of the Earth system, advances in Earth observing systems, and the growth of computational power.

Yet catastrophic infrastructure failures are becoming more frequent. The world is entering a new era of natural disasters that are causing more damage than in the past. Economic losses attributed to natural disasters have ballooned from $75.5 billion in the 1960s to $659.9 billion in the 1990s, a compound annual growth rate of 8 percent. Total worldwide insured losses are now dominated by natural catastrophes.

The principal driver of these increasing losses is increasing exposure, meaning that natural disasters are anything but "natural." The population of Florida, for example, rose 690 percent between 1950 and 2010, putting many more people at risk given the state's extensive coastal exposure and subtropical location. Moreover, the potential for the hazards themselves to become more damaging in the future with climate change will compound the effects of increasing exposure.

The speakers in this session laid out frontiers in forecasting natural disasters and looked into a future of useful forecasts, effective messaging, and avoidance of catastrophic failure. The first speaker, Ning Lin (Princeton University) outlined the future of probabilistic, quantitative natural hazard risk assessment in support of risk management. Next, Jeffrey Czajowski (University of Pennsylvania) outlined the critical importance of accounting for human behavioral biases to move from risk assessment to risk reduction. Rebecca Moore (Google) concluded the session with a description of the Google Earth Engine platform and how it provides a deeper understanding of changes to the environment.[1]

[1] Paper not included in this volume.

An Integrated Approach to Assess and Manage Hurricane Risk in a Changing Climate

Ning Lin
Princeton University

Hurricanes, with their strong winds, heavy rainfall, and storm surges, cause much damage and loss of life worldwide. Recent disasters, such as Hurricanes Katrina in 2005 and Sandy in 2012, Cyclone Nargis in 2008, and Typhoon Haiyan in 2013, underscore the significant vulnerability of the United States and the world to landfalling hurricanes. Furthermore the impacts of these storms may worsen in the coming decades because of rapid coastal development coupled with sea-level rise and possibly increasing hurricane activity due to climate change.

Major advances in hurricane risk management are urgently needed. Given the inherent uncertainties in hurricane activity, such management should be strongly informed by probabilistic risk assessment. Furthermore, hurricane risk assessment cannot rely solely on historical records: to account for projected future changes, it should integrate physical knowledge and models with observational data.

INTRODUCTION

A physically based probabilistic hurricane risk assessment framework should integrate analysis of storm activity, hazards, and risk. Because of the limitations of historical records and the complexity of the problem, Monte Carlo (MC) methods, based on numerous synthetic simulations, are often used.

In an MC approach, large numbers of synthetic but physically possible storms, characterized by their track, intensity, and size, are simulated (with their annual frequencies estimated) under observed or climate model–simulated climate conditions. Hazard models are then used to estimate the wind, surge, and rainfall-induced flooding associated with the simulated storms.

Given the estimated hazards and coastal exposure, vulnerability models can be applied to estimate storm-induced consequences (e.g., damage and/or economic losses) and thus risk. The risk assessment can in turn inform risk management.

The following sections review the main components—hurricane activity, hazards, and risk—of a physically based hurricane risk assessment framework and its application to evaluating risk mitigation strategies.

HURRICANE ACTIVITY

Various MC methods have been developed to simulate storms that depict hurricane activity and climatology. Most of these methods (e.g., Hall and Sobel 2013; Toro et al. 2010; Vickery et al. 2000) create simulations based on the statistics of the historical storm records.

In my laboratory we apply the statistical-deterministic model developed by Emanuel and colleagues (2006, 2008). It simulates storm environments statistically but generates synthetic storms deterministically (with physical models). The large samples of synthetic storms generated by the model are in statistical agreement with the (albeit limited) observations. Moreover, as the synthetic hurricane environments can be generated for any given climate state, the model can simulate storms not only in current and past climates but also in projected future climates.

This model has been used to simulate storms in various ocean basins under projected climates over the 21st century to investigate how storm intensity and frequency may change with the changing climate (Emanuel 2013). It has also been used to simulate storms at city scales—for New York City (NYC; Lin et al. 2010a, 2012; Reed et al. 2015); Miami (Klima et al. 2011), Apalachee Bay (Lin et al. 2014), and Tampa (Lin and Emanuel 2015) in Florida; Galveston, Texas (Lickley et al. 2014); Cairns, Australia (Lin and Emanuel 2015); and Dubai in the Persian Gulf (Lin and Emanuel 2015). As an illustration, Figure 1 shows a sample of 5,000 storms simulated for NYC. These city-scale simulations can be used to analyze local hazards and risk.

HURRICANE HAZARDS

Given a storm's characteristics, hazard models can be applied to estimate the wind, surge, and rainfall-induced flooding during the storm's landfall. Because large numbers of simulations are required for the MC-based risk analysis, the hazard models should be (computationally) "simple" (often there is a balance between accuracy and efficiency).

Wind

Various simple parametric methods have been developed to model the wind. In such an approach, one estimates the storm wind field using a parametric wind

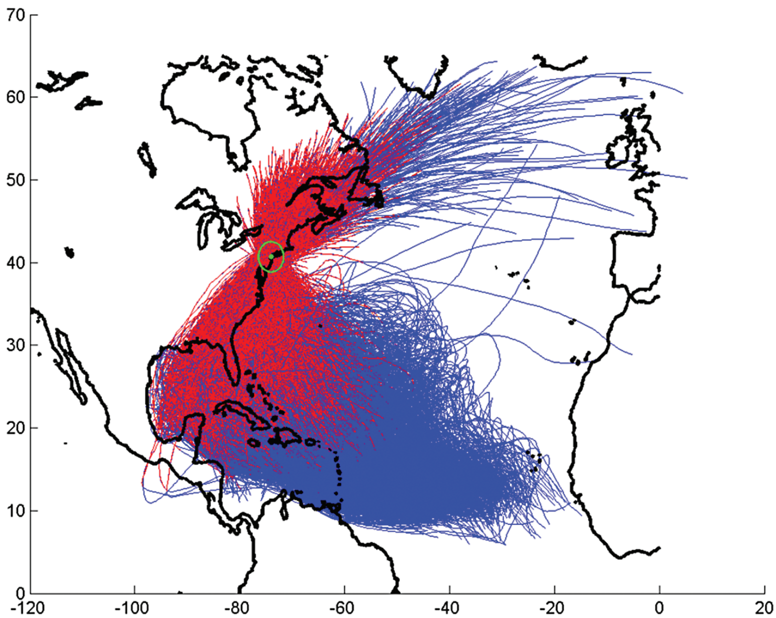

FIGURE 1 Simulation, using the Emanuel et al. (2006) model, of 5,000 synthetic storms that pass within 200 km of the Battery in New York City with a maximum wind speed greater than 21 m/s, under the observed climate of 1981–2000 and with an estimated annual frequency of 0.34. The green circle shows the 200 km radius around the Battery. Each blue curve shows a storm track, and the red portion of each track shows the 100-hour period before and during landfall (the main time period considered for hazard modeling). Longitude and latitude are shown on the axes. Figure in color at http://www.nap.edu/catalog/21825.

profile (e.g., Holland 1980; Jelesnianski et al. 1992) and adds an estimated background wind (Lin and Chavas 2012) to obtain the total wind field. My lab has recently developed a new wind profile (Chavas et al. 2015), motivated by the physical understanding that the canonical wind fields of mature hurricanes, although approximately circularly symmetric, cannot be described by a single mechanism.

Emanuel (2004) developed a physical model of the outer nonconvecting region of the storm, and Emanuel and Rotunno (2011) established an analytical profile that is physically valid only for the inner convecting region. We mathematically merged these two theoretical solutions to develop a complete wind profile for the entire domain of the storm (Chavas et al. 2015). This new physical model, evaluated and calibrated with various observational datasets, will have broad applications in hurricane hazard analysis.

Storm Surges

Storm surges, driven mainly by the storm surface wind and pressure, are also sensitive to coastal bathymetry and topography. Hydrodynamic surge models, basically solving coastal shallow water equations, include the SLOSH (Sea, Lake, and Overland Surges from Hurricanes) model (Jelesnianski et al. 1992), used by the National Hurricane Center for real-time forecasting, and the Advanced Circulation (ADCIRC) model (Westerink et al. 2008). The SLOSH model is computationally more efficient, but the ADCIRC model can better resolve the physical processes and produce results with higher resolutions.

Both the SLOSH and ADCIRC models are applied in my lab, depending on applications. Figure 2 shows that, as an example, the simulated storm surges in the NYC area from Hurricanes Irene (2011) and Sandy, using the ADCIRC model in this case, compare very well with the tidal gauge observations. In these simulations, we used high-resolution bathymetry and topography data, observed storm characteristics, as well as our new complete wind profile (Chavas et al. 2015) and a simple parametric pressure model (Holland 1980).

Rainfall

Hurricane rainfall is comparatively difficult to model because of its large spatial and temporal variation. Thus, most hurricane rainfall modeling applies full numerical weather prediction models (e.g., Lin et al. 2010b; Tuleya and DeMaria 2007). However, this approach requires large quantities of input data and has a high computational cost, so it is not effective for risk analysis.

Recently, simpler parametric models have been developed based on historical rainfall statistics (e.g., Lonfat et al. 2007; Tuleya et al. 2007) and physical principles (e.g., Langousis and Veneziano 2009). The basic physics of hurricane rainfall is that it is determined mainly by the combination of environmental moisture and the speed of the storm updraft. The latter depends on low-level convergence due to surface friction, storm intensification, interactions with topography, and the background baroclinic state. A model that describes these processes has been shown to generate rainfall statistics comparable to the observations (Zhu et al. 2013).

Research in my lab is ongoing to evaluate and further develop this physical rainfall model, which can then be coupled with a hydrologic model (e.g., Cunha et al. 2012) to simulate inland flooding.

HURRICANE RISK

Hazard models can be applied to simulated synthetic storms to generate large samples of hazards from which hazard probabilities can be estimated. For example, the ASCE building code has used such an approach to establish design wind maps (showing wind speeds for various return periods) for the entire US

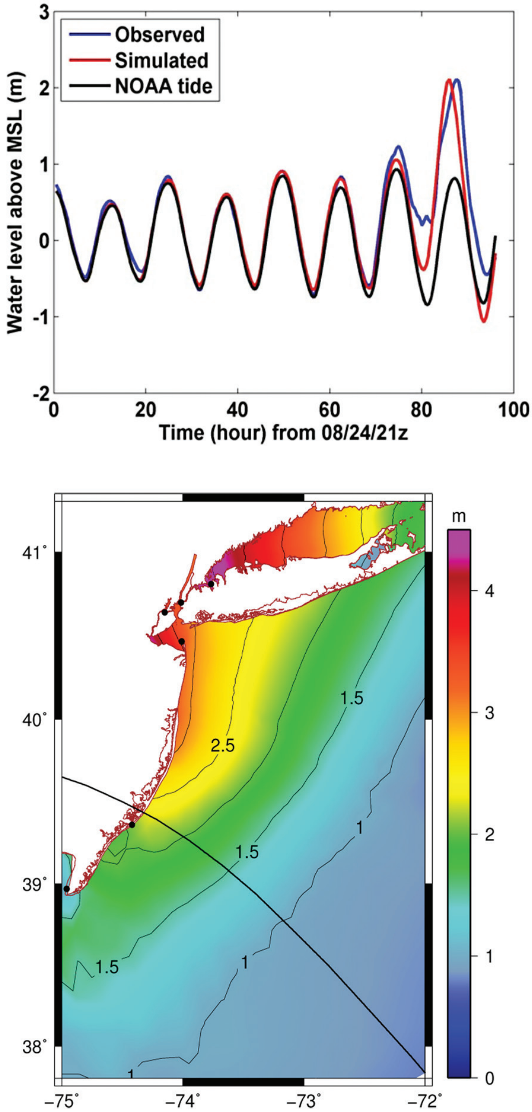

FIGURE 2 Storm surge estimation for New York City for Hurricanes Irene (*above*; time series of water level at the Battery) and Sandy (*below*; spatial distribution of the simulated peak surge; the estimated peak surge at the Battery, about 2.9 m, is close to the observed value of about 2.8 m above high tide; black curve shows the track of Sandy). Figure in color at http://www.nap.edu/catalog/21825. MSL = mean sea level; NOAA = National Oceanic and Atmospheric Administration.

coast. Similarly, the Federal Emergency Management Agency (FEMA) developed flood maps depicting 100- and 500-year floodplains as a basis for the federal flood insurance policy. (Different storm and hazard models were used in these different applications.)

If the hurricane model used to generate synthetic storms draws on climate model–projected climate environments (Emanuel et al. 2008), then one can estimate probabilistic hazards under future climates. We have performed such analysis for various coastal cities; for example, Figure 3 shows our estimations of the storm surge level for NYC as a function of return period, under the observed current climate as well as climate model–estimated current climates and climate model–projected future climates. The results indicate a potentially significant increase of surge floods in the future due to climate change.

The hazard probabilities can also be combined with the estimated consequences of the hazards to quantify the risk. (The consequent damage/losses can be estimated with vulnerability models such as the Hazus model developed by FEMA.) The risk is often expressed by the expectation (mean) of the loss in a year (e.g., Aerts et al. 2013), but the full probabilistic distribution of the loss, if available, is more informative. In the context of climate change and coastal development, this risk is likely increasing.

To obtain a temporally integrated measure, the overall loss is also typically quantified by its present value (PV), the sum of all discounted losses occurring over a given time horizon (e.g., the next 100 years). Then the risk can be considered as the mean or, better, the probability distribution of the PV of future losses.

BENEFITS AND COSTS OF RISK MITIGATION STRATEGIES

The PV also provides a convenient metric for comparing the benefit and cost of risk mitigation strategies. The benefit can be considered as the PV of the future losses prevented by mitigation, and the cost is the PV of the total cost of the mitigation (including construction and maintenance). While the cost is largely deterministic, the benefit is random.

Most studies have focused on comparing the cost and the mean of the benefit (e.g., Aerts et al. 2014). We present a more informative probabilistic cost-benefit analysis, applied to various coastal flood mitigation strategies proposed for NYC. As shown in Figure 4, for each strategy we estimate the full probability distribution of the benefit and plot its exceedance probability function to compare with the cost. The crossing of the curve of the benefit exceedance probability and the line showing the cost indicates the probability that the benefit is greater than the cost. The probabilities of getting any higher or lower benefits can also be easily read from the curve.

The probabilistic benefit-cost analysis thus provides adequate information for making management decisions for any specific risk tolerance (decisions made based on the mean implicitly assume "risk neutral"). Also, this risk management

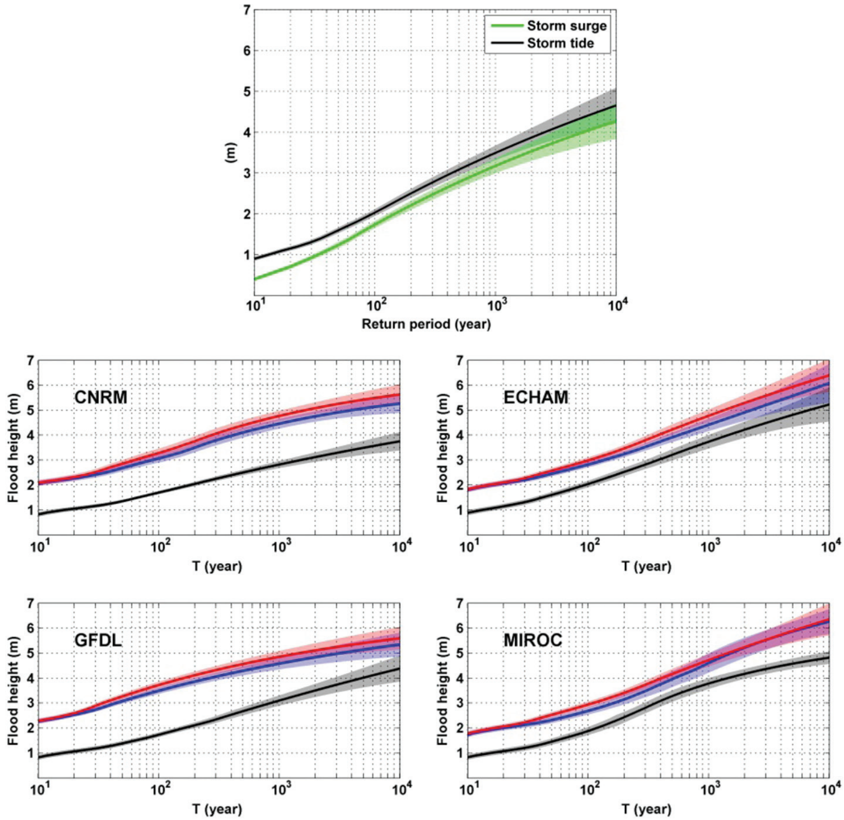

FIGURE 3 Estimated storm surge return level curves for New York City, under (*top*) the observed climate of 1980–2000 and (*below*) four climate model–estimated climates for 1980–2000 (black) and 2080–2100 (blue and, when accounting also for potential changes in storm size, red; with projected 1 meter sea level rise accounted for). Each curve is based on 5,000 synthetic storms (e.g., the curves in the top panel are based on the 5,000 storms shown in Figure 1). Four climate models are applied: CNRM (CNRM-CM3; Centre National de Recherches Météorologiques, Météo-France), ECHAM (ECHAM5; Max Planck Institute), GFDL (GFDL-CM2.0; NOAA/Geophysical Fluid Dynamics Laboratory), and MIROC (MIROC3.2, Model for Interdisciplinary Research on Climate; University of Tokyo Center for Climate System Research (CCSR)/National Institute for Environmental Studies (NIES; Japan)/Frontier Research Center for Global Chance (FRCGC), Japan Agency for Marine-Earth Science and Technology (JAMSTEC)). Figure in color at http://www.nap.edu/catalog/21825. SOURCE: Lin et al. (2012).

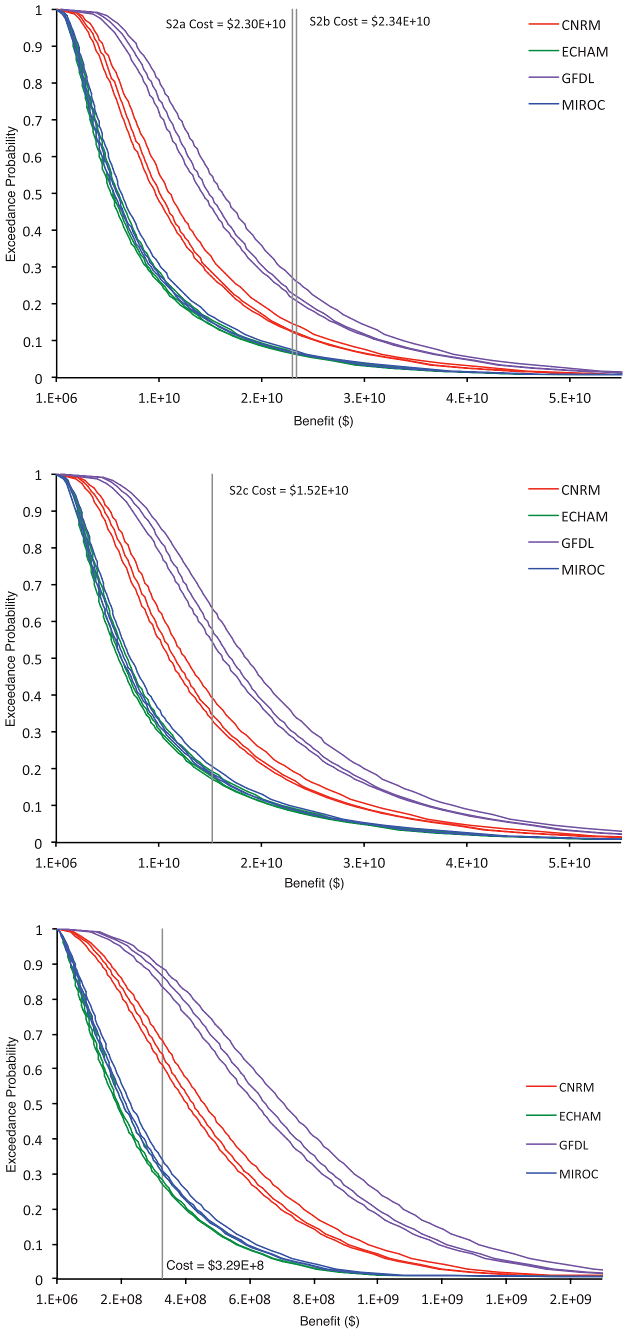
S2a Cost = $2.30E+10
S2b Cost = $2.34E+10
S2c Cost = $1.52E+10
Cost = $3.29E+8
CNRM
ECHAM
GFDL
MIROC
Exceedance Probability
Benefit ($)
1.E+06
1.E+10
2.E+10
3.E+10
4.E+10
5.E+10
2.E+08
4.E+08
6.E+08
8.E+08
1.E+09
2.E+09

FIGURE 4 Estimated exceedance probability of the benefit (curve) compared to the cost (vertical line) for strategies S2a and S2b (*top*), strategy S2c (*middle*), and a strategy of elevating new houses on the floodplain by 6 feet (*bottom*) for New York City (NYC). S2a consists of three barriers to close off parts of NYC and New Jersey but preserves wetland dynamics of Jamaica Bay. S2b expands on S2a by adding a fourth barrier that closes off Jamaica Bay. S2c replaces three barriers from S2b with one large barrier in the outer NY harbor to protect a larger area. The details of these mitigation strategies are discussed in Aerts et al. (2014). The analyses account for projected coastal development and changes in storm activity and sea level over the 21st century. Current and future building stock data are obtained from the NYC Office of Emergency Management. Synthetic storm surge events are obtained from Lin et al. (2012) for the four climate models (CNRM, ECHAM, GFDL, and MIROC; see Figure 3 for definitions), as shown by the return level curves in Figure 3. The probabilistic sea level rise projection is based on Kopp et al. (2014); the three curves of the same color show results with sea level rise projected for the three IPCC AR5 (Intergovernmental Panel on Climate Change Fifth Assessment Report) emission scenarios (RCP 2.6, RCP 4.5, and RCP 8.5). Figure in color at http://www.nap.edu/catalog/21825.

analysis has relied on the physically based risk assessment to account for the dynamic evolution of urban development, storm climatology change, and sea level rise.

FUTURE RESEARCH

Although much remains unknown about how hurricanes, especially their frequency and size, will vary with the climate, risk assessment should continue to incorporate the state-of-the-art science to support risk management. Effective risk analysis will require more physical or physical-statistical methods for simulating synthetic storms. Hurricane rainfall models, especially those based on physics, need to be developed to estimate inland flood risk in a changing climate. Hurricane hazards are correlated (e.g., hurricane wind affects both storm surge and rainfall; coastal and inland flooding may interact), and multihazard approaches are needed to estimate how hazards will jointly evolve and how to deal with the joint risk.

In addition to engineering measures, urban planning and federal and private insurance play increasingly important roles in coastal risk mitigation. Establishing systematic and more integrated strategies may be the future direction for hurricane risk management.

ACKNOWLEDGMENT

The author acknowledges the support of National Science Foundation grant EAR-1520683.

REFERENCES

Aerts JC, Lin N, Botzen W, Emanuel K, de Moel H. 2013. Low-probability flood risk modeling for New York City. Risk Analysis 33(5):772–788.

Aerts JCJH, Botzen WJW, Emanuel K, Lin N, de Moel H, Michel-Kerjan EO. 2014. Evaluating flood resilience strategies for coastal megacities. Science 344(6183):473–475.

Chavas DR, Lin N, Emanuel K. 2015. A model for the complete radial structure of the tropical cyclone wind field. Part I: Comparison with observed structure. Journal of the Atmospheric Sciences 72(9):3647–3662.

Cunha LK, Mandapaka PV, Krajewski WF, Mantilla R, Bradley A. 2012. Impact of radar-rainfall error structure on estimated flood magnitude across scales: An investigation based on a parsimonious distributed hydrological model. Water Resources Research 48(10).

Emanuel K. 2004. Tropical cyclone energetics and structure. In: Atmospheric Turbulence and Mesoscale Meteorology, ed. Fedorovich E, Rotunno R, Stevens B. Cambridge, UK: Cambridge University Press. pp. 165–192.

Emanuel KA. 2013. Downscaling CMIP5 climate models shows increased tropical cyclone activity over the 21st century. Proceedings of the National Academy of Sciences 110(30):12,219–12,224.

Emanuel K, Rotunno R. 2011. Self-stratification of tropical cyclone outflow. Part I: Implications for storm structure. Journal of the Atmospheric Sciences 68:2236–2249.

Emanuel K, Ravela S, Vivant E, Risi C. 2006. A statistical deterministic approach to hurricane risk assessment. Bulletin of the American Meteorological Society 87:299–314.

Emanuel K, Sundararajan R, Williams J. 2008. Hurricanes and global warming: Results from downscaling IPCC AR4 simulations. Bulletin of the American Meteorological Society 89:347–367.

Hall TM, Sobel AH. 2013. On the impact angle of Hurricane Sandy's New Jersey landfall. Geophysical Research Letters 40(10):2312–2315.

Holland GJ. 1980. An analytic model of the wind and pressure profiles in hurricanes. Monthly Weather Review 108(8):1212–1218.

Jelesnianski CP, Chen J, Shaffer WA. 1992. SLOSH: Sea, lake, and overland surges from hurricanes. NOAA Technical Report NWS 48. Silver Spring, MD: National Weather Service, National Oceanic and Atmospheric Administration.

Klima K, Lin N, Emanuel K, Morgan MG, Grossmann I. 2011. Hurricane modification and adaptation in Miami-Dade County, Florida. Environmental Science and Technology 46:636–642.

Kopp RE, Horton RM, Little CM, Mitrovica JX, Oppenheimer M, Rasmussen DJ, Strauss BH, Tebaldi C. 2014. Probabilistic 21st and 22nd century sea-level projections at a global network of tide-gauge sites. Earth's Future 2:383–406.

Langousis A, Veneziano D. 2009. Theoretical model of rainfall in tropical cyclones for the assessment of long-term risk. Journal of Geophysical Research 114:D02106.

Lickley MJ, Lin N, Jacoby HD. 2014. Analysis of coastal protection under rising flood risk. Climate Risk Management 6:18–26.

Lin N, Chavas D. 2012. On hurricane parametric wind and applications in storm surge modeling. Journal of Geophysical Research 117:D09120.

Lin N, Emanuel K. 2015. Grey swan tropical cyclones. Nature Climate Change. doi:10.1038/nclimate2777.

Lin N, Emanuel KA, Smith JA, Vanmarcke E. 2010a. Risk assessment of hurricane storm surge for New York City. Journal of Geophysical Research 115:D18121.

Lin N, Smith JA, Villarini G, Marchok TP, Baeck ML. 2010b. Modeling extreme rainfall, winds, and surge from Hurricane Isabel (2003). Weather and Forecasting 25:1342–1361.

Lin N, Emanuel K, Oppenheimer M, Vanmarcke E. 2012. Physically based assessment of hurricane surge threat under climate change. Nature Climate Change 2:1–6.

Lin N, Lane P, Emanuel KA, Sullivan RM, Donnelly JP. 2014. Heightened hurricane surge risk in northwest Florida revealed from climatological-hydrodynamic modeling and paleorecord reconstruction. Journal of Geophysical Research: Atmospheres 119(14):8606−8623.

Lonfat M, Rogers R, Marchok T, Marks FD Jr. 2007. A parametric model for predicting hurricane rainfall. Monthly Weather Review 135:3086–3097.

Reed AJ, Mann ME, Emanuel KA, Lin N, Horton BP, Kemp AD, Donnelly JP. 2015. Increased threat of tropical cyclones and coastal flooding to New York City during the anthropogenic era. Proceedings of the National Academy of Sciences 112(41):12610–12615.

Toro GR, Resio DT, Divoky D, Niedoroda AW, Reed C. 2010. Efficient joint-probability methods for hurricane surge frequency analysis. Ocean Engineering 37(1):125–134.

Tuleya RE, DeMaria M. 2007. Evaluation of GFDL and simple statistical model rainfall forecasts for US landfalling tropical storms. Weather and Forecasting 22:56–70.

Vickery P, Skerlj P, Twisdale L. 2000. Simulation of hurricane risk in the US using empirical track model. Journal of Structural Engineering 126(10):1222–1237.

Westerink JJ, Luettich RA, Feyen JC, Atkinson JH, Dawson C, Roberts HJ, Powell MD, Dunion JP, Kubatko EJ, Pourtaheri H. 2008. A basin- to channel-scale unstructured grid hurricane storm surge model applied to southern Louisiana. Monthly Weather Review 136:833–864.

Zhu L, Quiring SM, Emanuel KA. 2013. Estimating tropical cyclone precipitation risk in Texas. Geophysical Research Letters 40:6225–6230.

Moving from Risk Assessment to Risk Reduction: An Economic Perspective on Decision Making in Natural Disasters

Jeffrey Czajkowski
University of Pennsylvania

A significant aim of natural disaster research is to improve the science, or the hazard assessment, of risk associated with such disasters. This goal could be achieved by, for example, enhancing the accuracy of short-term extreme weather or long-term climate forecasts or by increasing the validity of the hazard component of natural disaster catastrophe models.

It is assumed either that users of the information will fully understand the scientific data and incorporate that understanding in rational decisions based on a systematic analysis of tradeoffs between benefits and costs, or that losses will be better predicted and managed based on the enhanced scientific aspects of a catastrophe model. Yet, although hazard assessments have improved, many forms of losses from natural disasters have increased over time, associated with innumerable instances of inadequate investments in loss reduction measures and poor decision making before and after events.

As reported by the United Nations International Strategy for Disaster Reduction, "Experience has shown that a purely technical assessment of risk, however sophisticated and cutting-edge, is by itself unlikely to trigger actions that reduce risk. Successful risk assessments produce information that is targeted, authoritative, understandable, and usable" (UNISDR 2015, p. 148). Research provides empirical evidence of individuals exhibiting systematic behavioral biases and using simplified decision rules when making choices with respect to low-probability/high-impact events such as natural disasters. The findings show that such choices and the resulting behavior are significantly influenced by individual interpretation, which is dependent on how the scientific information is framed and presented, so it is essential to incorporate understanding of decision biases in the assessment and subsequent communication of natural hazard risks.

Unfortunately, little of this behavior-based knowledge has been incorporated into natural disaster risk assessment and catastrophe modeling. The use of appropriate risk management strategies based on such knowledge could reduce natural disaster losses.

INTRODUCTION

Progress in Forecasting and Modeling

Recent decades have seen significant progress in the ability not only to observe and understand the weather but also to provide more accurate forecasts (Hirschberg et al. 2011; NRC 2010). This is congruently true for extreme weather events such as hurricanes, as evidenced by the National Hurricane Center's reduced annual average track forecast errors from 1970 to 2014 (Figure 1). For example, the 72-hour track forecast error improved from nearly 450 nautical miles on average in 1970 down to less than 100 nautical miles in 2014, as shown

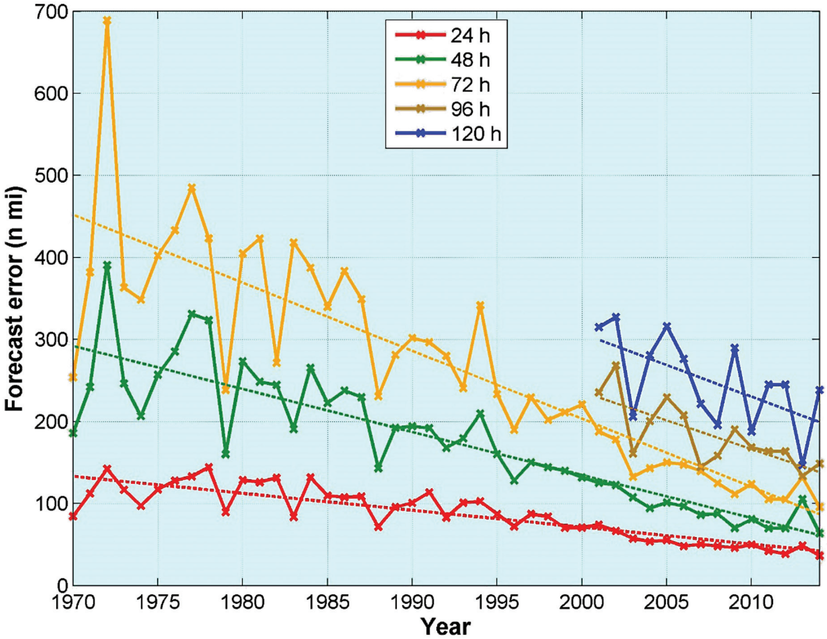

FIGURE 1 National Hurricane Center annual average official track errors as a function of forecast lead time for Atlantic Basin tropical storms and hurricanes, 1970–2014, with least squares trend lines superimposed. Figure in color at http://www.nap.edu/catalog/21825. n mi = nautical miles. SOURCE: http://www.nhc.noaa.gov/verification/verify5.shtml.

by the yellow least squares trend line. These forecast improvements have been credited with a number of associated benefits, such as a substantial reduction in the number of direct fatalities thanks to more timely evacuation (Gladwin et al. 2007; Rappaport 2000, 2014).

Major advances in the science and modeling of extreme weather hazards (see Lin et al. 2012 for a discussion of storm surge modeling) have led to the widespread use of catastrophe models for natural hazard risk assessment since the early 1990s by the insurance industry. This usage has in turn led to the further implementation of natural hazard risk transfer mechanisms such as reinsurance and capital markets (Grossi and Kunreuther 2005), allowing for the relatively uneventful absorption of natural hazard economic losses by the insurance industry in recent years.

Persistent Challenges

There remain serious concerns, however. First, the evidence suggests an upward trend in economic losses from natural disasters worldwide (Figure 2), to an estimated annual average of about $250 billion (UNISDR 2015). This rise is correlated with population and exposure growth in high hazard areas (UNISDR 2013, 2015), leading to more people affected by natural disasters, interdependencies in economic and social systems that increase vulnerability to

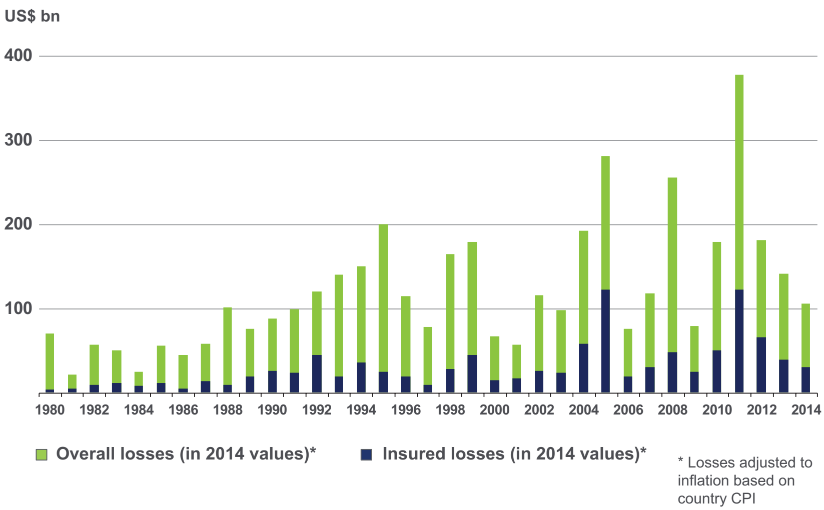

FIGURE 2 Overall and insured losses associated with worldwide natural catastrophes, in billions of dollars, 1980–2014. CPI = consumer price index. © 2015 Münchener Rückversicherungs-Gesellschaft, Geo Risks Research, NatCatSERVICE. Reprinted with permission.

disruptions, and potentially exacerbated hazard risks from climate change impacts (UNISDR 2015).

Second, reduced mortality benefits have been limited to select developed countries, largely because of a lack of capacity to forecast disasters and provide early warnings in developing countries (UNISDR 2015). Moreover, in developing countries noninsured and nondirect property and other losses, and the costs associated with recovery, are difficult to quantify and hence thought to be substantially underestimated (UNISDR 2015).

Finally, even in a relatively sophisticated natural disaster risk management landscape such as the United States there have been both inadequate investments in hazard-related loss reduction measures (e.g., with Hurricane Katrina in 2005 and Hurricane Sandy in 2012) and poor decision making. As an example of the latter, during the 2013 Oklahoma City tornado residents should have sheltered in place but were advised by a local meteorologist to evacuate south in their cars.

Illustrative examples show how a traditional natural hazard forecast risk context (i.e., risk "space") may be placed in a broader overview of event risk in time, importantly including behavioral implications of intertemporal decision making. The paper further describes an economic (i.e., benefit-cost) model of decision making in this risk space, highlighting potential sources of bias demonstrated in recent research.

DEFINING THE NATURAL HAZARD FORECAST RISK SPACE

Natural hazard risk is defined as the probability of a natural hazard event occurrence and its expected impact (Kunreuther and Useem 2010). Thus, the concept of natural hazard risk has two key components, hazard probability and impact, each of which has an element of uncertainty associated with it. Geoff Love and Michel Jarraud of the World Meteorological Organization provide a schematic of this natural hazard risk space (Figure 3), with the probability of the hazard on the y-axis and the impact on the x-axis (Love and Jarraud 2010); uncertainty is represented by the shaded shapes surrounding the three illustrative risks shown (e.g., C would have more uncertainty in impact vs. likelihood given the size, shape, and position of the circle).

Physical scientists working in the area of natural hazard forecast risk often concentrate on the likelihood of occurrence (y-axis), such as the return period for a flood event. Thus, as an illustrative example of this predisposition, only 0.6 percent of the National Oceanic and Atmospheric Administration's 2008 budget of $4 billion was directed to social science activities, which are more likely to focus on the impact side of a risk (NRC 2010).

Likewise, in a catastrophe modelling framework of combined hazard, exposure, and vulnerability components leading to loss (Grossi and Kunreuther 2005), emphasis is typically on hazard, although the other components may significantly affect losses and hence overall risk. For example, a Risk Management Solutions

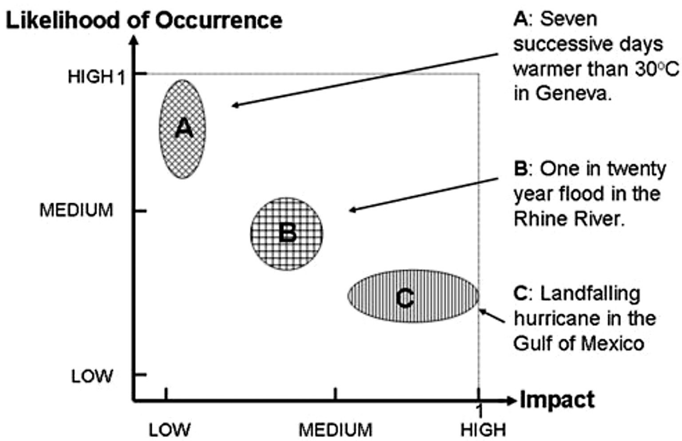

FIGURE 3 Natural hazard forecast risk space. Uncertainty is represented by the size, shape, and position of the shaded region surrounding the three risks illustrated. SOURCE: Kunreuther and Useem (2010), printed and electronically reproduced by permission of Pearson Education, Inc., New York.

study found that loss estimates could change by a factor of 4 when property exposure data gaps were filled or inaccurate information was corrected (RMS 2008).

There is a clear need to better understand the impact side of the natural hazard risk equation—including perceptions of expected impact—if overall risk reduction is the goal (Kunreuther and Useem 2010). For example, Botzen and colleagues (2015) find that in New York City flood risk perception is influenced by underestimation of hazard impact. Significantly, since the US tornado tragedies in 2011, impact-based warnings for tornadoes (Figure 4) have been implemented by the National Weather Service (NWS).[1]

Integrated loss modeling between the physical sciences and other disciplines such as engineering and the social sciences is critical to enhance understanding of the numerous factors behind natural hazard risk (Kunreuther and Useem 2010; Morss et al. 2011; Tye et al. 2014). Integrated disciplines (e.g., physical scientists and economists) are evident in natural hazard impact assessment research on hurricane risks in coastal locations (Czajkowski and Done 2014) and on inland flooding from tropical cyclones (Czajkowski et al. 2013).[2] In addition, a 2010

[1] Information about NWS impact-based warnings is available at http://www.weather.gov/impacts/. For an assessment of the impact-based tool see Harrison et al. (2014).

[2] A number of recent impact-focused assessments for extreme events are available: for example, Chavas et al. (2012), Malmstadt et al. (2009), Mendelsohn et al. (2012), Murnane and Elsner (2012), Murphy and Strobl (2010), Nordhaus (2006, 2010), Schmidt et al. (2009, 2010), Strobl (2011), and Zhai and Jiang (2014).

```
903
WFUS53 KFSD 050022
TORFSD
IAC035-050100-
/O.NEW.KFSD.TO.W.0020.131005T0022Z-131005T0100Z/

BULLETIN - EAS ACTIVATION REQUESTED
TORNADO WARNING
NATIONAL WEATHER SERVICE SIOUX FALLS SD
722 PM CDT FRI OCT 4 2013

...TORNADO EMERGENCY FOR WASHTA...

THE NATIONAL WEATHER SERVICE IN SIOUX FALLS HAS ISSUED A

* TORNADO WARNING FOR...
  CHEROKEE COUNTY IN NORTHWEST IOWA...

* UNTIL 800 PM CDT

* AT 720 PM CDT...A LARGE AND EXTREMELY DANGEROUS TORNADO WAS
  LOCATED NEAR WASHTA...AND MOVING NORTHEAST AT 30 MPH.

  THIS IS A TORNADO EMERGENCY FOR WASHTA. TAKE COVER NOW. THIS
  IS A PARTICULARLY DANGEROUS SITUATION.

  THIS IS A PARTICULARLY DANGEROUS SITUATION.

  HAZARD...DAMAGING TORNADO.

  SOURCE...EMERGENCY MANAGEMENT CONFIRMED TORNADO.

  IMPACT...YOU ARE IN A LIFE THREATENING SITUATION. FLYING
           DEBRIS MAY BE DEADLY TO THOSE CAUGHT WITHOUT SHELTER.
           MOBILE HOMES WILL BE DESTROYED. CONSIDERABLE DAMAGE
           TO HOMES...BUSINESSES AND VEHICLES IS LIKELY AND
           COMPLETE DESTRUCTION IS POSSIBLE.

* THE TORNADO WILL BE NEAR...
  QUIMBY AROUND 730 PM CDT.
  CHEROKEE AROUND 745 PM CDT.
  AURELIA AROUND 750 PM CDT.

PRECAUTIONARY/PREPAREDNESS ACTIONS...

HEAVY RAINFALL MAY HIDE THIS TORNADO. DO NOT WAIT TO SEE OR HEAR
THE TORNADO. TAKE COVER NOW.

&&
LAT...LON 4259 9585 4291 9565 4291 9550 4283 9538
      4269 9539 4256 9569 4256 9577
TIME...MOT...LOC 0023Z 225DEG 27KT 4260 9567

TORNADO...OBSERVED
TORNADO DAMAGE THREAT...CATASTROPHIC
HAIL...1.50IN

$$
```

FIGURE 4 National Weather Service tornado impact–based warning example for a catastrophic tornado warning. SOURCE: NWS Impact Based Warnings, http://www.weather.gov/impacts/#.VjyjjrerS72.

overview of the literature on integration of socioeconomic considerations in weather research cites six successful programs (NRC 2010, pp. 34–35).

NATURAL HAZARD FORECAST RISK IN TIME

Pre- and Postevent Planning

Forecasts of natural hazard risks are directly tied to an event, whereas the extent of overall impacts is a function of risk over time in the affected areas.

While most activity surrounding a natural hazard event is focused on the crisis management stages of preparation and response (during and immediately afterward), the socioeconomic impacts are tightly linked with the pre-event prevention, mitigation, and recovery planning activities and the postevent long-term recovery process. Narrowly focusing risk reduction efforts on the event itself will likely not support optimal total risk reduction efforts. Herman Leonard and Arnold Howitt of the Kennedy School of Government at Harvard provide a time-oriented view (Figure 5) that extends to include the oft-underappreciated stages of pre-event preparation and postevent recovery (Leonard and Howitt 2010).

The ability to expand the timescale of the natural hazard risk event space to earlier and later stages is critical. For example, how would warning messages of a potential natural hazard event risk in the relatively distant future affect current pre-event preparation activities (NOAA 2015)?

Accounting for Behavioral Biases

Although expanding the timescale of the risk space is essential, interjecting the notion of time is potentially problematic given temporal behavioral biases such as underweighting the future through hyperbolic discounting[3] (Kunreuther et al. 2012). In this way of thinking, although the costs of pre-event preparation and mitigation are immediate and certain, the benefits associated with action are in the distant future and therefore uncertain in both time and return. Even if properly discounted benefits that accrued over time (i.e., at a constant and appropriate discount rate) outweighed the upfront costs, individuals would tend to disproportionately discount the future given their aversion to delayed gratification (Kunreuther et al. 2012).

Other intertemporal behavioral biases (Kunreuther et al. 2012) that could hamper optimal pre-event mitigation are myopic planning (a limited time horizon of only the next few years), underestimation of the risk (the probability or impact of a hazard), and affective forecasting errors (poor predictions of emotional states

[3] Hyperbolic discounting rapidly discounts valuations for small time periods and slowly discounts valuations for longer periods. Exponential discounting, on the other hand, discounts by a constant factor per unit delay, regardless of the length of the delay.

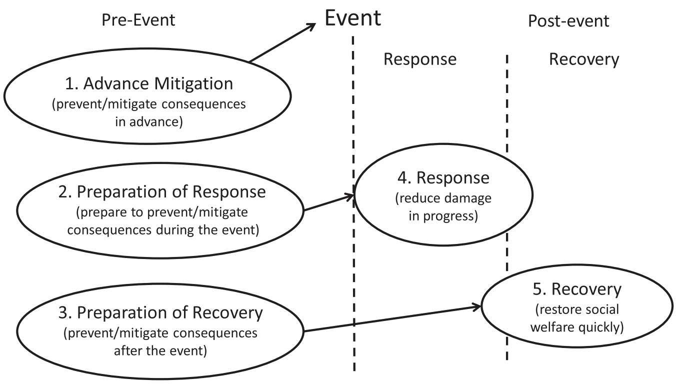

FIGURE 5 Overall timeline of natural disaster risk. SOURCE: Kunreuther and Useem (2010), printed and electronically reproduced by permission of Pearson Education, Inc., New York.

based on feelings today).[4] These biases make clear the importance of behavioral tendencies in decision making in the natural hazard risk context.

DECISION MAKING IN NATURAL HAZARD RISK CONTEXT

Intuitive vs. Deliberative Decision Making

From a rational economic perspective in the natural hazard risk space, individual decisions at a point in time are based on expected utility theory.[5] According to this theory, an individual confronted with the need for a decision with uncertain outcomes will decide based on the outcome with the greatest expected utility.

Table 1 shows the application of expected utility theory in the context of a natural disaster. When one is deciding to evacuate from a forecasted hurricane,

[4] Intertemporal bias of duration neglect (Kunreuther et al. 2012) may also exist in the postevent recovery phase, when there is a tendency to overestimate the time to recover and hence future protection would be overvalued.

[5] Other social science theories of decision making in a natural hazard context include the psychometric paradigm of psychology (perception of hazards taking into account qualitative information [e.g., dread] rather than just statistical [i.e., probability]); the cultural theory of risk in anthropology (social and cultural influences on risk perception); the mental models approach of psychology and risk (individuals have a “mental model” of reality, influenced by social interactions and experiences, that they use as a lens to view risky situations); the protection motivation theory of psychology (people protect themselves based on their perception of severity, probability, effectiveness of protective action, and self-efficacy); and the social amplification of risk framework of geography (risks are amplified or attenuated due to individual, social, and cultural factors) (NOAA 2015).

TABLE 1 Evacuation Payoff Matrix

Outcome Action	Landfall strike ($p = 0.3$)	Landfall miss ($p = 0.7$)	Expected utility
Stay	−2000	0	(0.3 × −2000) + (0.7 × 0) = −600
Evacuate	1500	−500	(0.3 × 1500) + (0.7 × −500) = 100

The utility of evacuating for the landfall strike of a disaster would be the avoided injuries net of the cost of evacuation; the disutility of staying would be the expected injuries or mortality; and the disutility of evacuating when no strike occurs would be the costs of evacuation. Expected utility is the summation by row of the probability of each outcome multiplied by its illustrative utility/disutility. p = probability.

utility (or disutility) is assigned to each possible future state (landfall hit or miss) given the possible action (stay or evacuate), and each future state is assigned a probability (p) with all probabilities summing to one. The choice of staying or evacuating is determined by selecting the action with the highest expected outcome across all possible states; in Table 1, the choice is to evacuate.

In reality, however, the decision-making process is often quite complex (especially over multiple forecast periods)[6] and rarely do the people under the warning act rationally. Rather, the combination of systematic behavioral biases coupled with simplified decision rules leads to choices that differ from those predicted by expected utility theory (Kunreuther and Useem 2010; Kunreuther et al. 2012).

Kahneman (2011) highlights the difference between intuitive and deliberative thinking, documenting extensive research on intuitive biases that operate in lieu of ideal deliberative decision making and result in suboptimal choices for low-probability/high-consequence events such as natural disasters. For example, the availability bias estimates the likelihood of disaster occurrence based on the saliency of the event as opposed to objective hazard probabilities. Or protective action is not taken because the subjective probability of expected impact is below some threshold level of concern. Kunreuther and Useem (2010) discuss a host of other behavioral biases revealed by the research, many of them associated with group behavior, risk culture, fear and other emotions, and trust.[7]

[6] For an illustration of this decision over time from a dynamic perspective, where for each forecast period the individual may choose to evacuate or wait for an additional forecast, see Czajkowski (2011).

[7] Chapter 4, "Cognitive Constraints and Behavioral Biases," discusses these in more detail as does Chapter 5, "The Five Neglects: Risks Gone Amiss," from an expected utility perspective.

Factors That Drive Positive Behavior

A considerable amount of research has sought to identify what factors drive positive behavior in this context, controlling for behavioral biases. Meyer and colleagues (2013) used a realistic simulated storm environment to better understand risk perception and decision making, and in a subsequent study they interviewed more than 2,000 respondents in real time under the threat of hurricane strikes during the 2010–2012 hurricane seasons (Meyer et al. 2014). Beatty and colleagues (2015) used a big data approach to analyze water bottle sales before and after a hurricane.

In a recent review and assessment of risk communication and behavior, NOAA (2015) points to work by Mileti and colleagues (2006) that identified a number of factors and categories consistently found to matter in the context of warning response: sociodemographic (female, white, more education, and children present); personal (experience, knowledge of hazard and actions, self-efficacy, fear, risk and vulnerability perception, more resources available, large and strong social network); source/channel (environmental or social cues present, official source, in person, familiar source, multiple sources); information (specific, credible, certain, frequent, consistent, and with guidance on actions); and threat (less lead time available, greater severity, close, confirmed).

Importantly, however, little of this behavior-based knowledge has been incorporated in natural disaster risk assessment or mitigation planning. Possible approaches are briefly articulated in the next section.

MOVING FORWARD

Despite significant advances in recent decades in observing, understanding, and forecasting extreme weather, the impacts and threats from natural disasters remain extensive. This paper has provided a definition and context for decision making in the natural disaster risk space where behavioral biases play a significant role. Efforts to reduce natural disaster risk will have to incorporate appropriate risk management strategies based on this behavior-based knowledge. The following measures are recommended based on this overview:

- Develop warning and forecast products that assess and communicate risk from the perspective of both probability and impact, including the notion of uncertainty.
- Extend the timescale of the risk forecast space to pre-event preparation/mitigation and postevent recovery planning.
- Account for the behavioral biases documented in the socioeconomic research literature when designing risk communication tools or incentivizing more proactive preparation, mitigation, and/or recovery activities.

- Extend catastrophe models to include risk perception and behavior components.

REFERENCES

Beatty TK, Shimshack JP, Volpe RJ. 2015. Disaster preparedness and disaster response: Evidence from bottled water sales before and after tropical cyclones. Working paper, University of Virginia, Charlottesville.

Botzen W, Kunreuther H, Michel-Kerjan E. 2015. Divergence between individual perceptions and objective indicators of tail risks: Evidence from floodplain residents in New York City. Working paper, Wharton Risk Management and Decision Processes Center, Philadelphia.

Chavas DR, Yonekura E, Karamperidou C, Cavanaugh N, Serafin K. 2012. US hurricanes and economic damage: An extreme value perspective. Natural Hazards Review 14(4):237–246.

Czajkowski J. 2011. Is it time to go yet? Understanding household hurricane evacuation decisions from a dynamic perspective. Natural Hazards Review 12(2):72–84.

Czajkowski J, Done J. 2014. As the wind blows? Understanding hurricane damages at the local level through a case study analysis. Weather, Climate, and Society 6(2):202–217.

Czajkowski J, Villarini G, Michel-Kerjan E, Smith JA. 2013. Determining tropical cyclone inland flooding loss on a large scale through a new flood peak ratio-based methodology. Environmental Research Letters 8(4):044056.

Gladwin H, Lazo JK, Morrow BH, Peacock WG, Willoughby HE. 2007. Social science research needs for the hurricane forecast and warning system. Natural Hazards Review 8(3):87–95.

Grossi P, Kunreuther H. 2005. Catastrophe Modeling: A New Approach to Managing Risk, Vol. 25. Boston: Springer Science and Business Media.

Harrison J, McCoy C, Bunting-Howarth K, Sorensen H, Williams K, Ellis C. 2014. Evaluation of the National Weather Service Impact-based Warning Tool. WISCU-T-14-001 Report. Available at http://www.seagrant.sunysb.edu/Images/Uploads/PDFs/Hurricanes-NWS-IBW_finalreport.pdf.

Hirschberg PA, Abrams E, Bleistein A, Bua W, Delle Monache L, Dulong TW, Gaynor JE, Glahn B, Hamill TM, Hansen JA, and six others. 2011. Weather and climate enterprise strategic implementation plan for generating and communicating forecast uncertainty information. Bulletin of the American Meteorological Society 92:1651–1666.

Kahneman D. 2011. Thinking, Fast and Slow. New York: Farrar, Straus and Giroux.

Kunreuther H, Useem M. 2010. Learning from Catastrophes: Strategies for Reaction and Response. Upper Saddle River, NJ: Prentice Hall.

Kunreuther H, Meyer R, Michel-Kerjan E. 2012. Overcoming decision biases to reduce losses from natural catastrophes. In: The Behavioral Foundations of Public Policy, ed. Shafir E. Princeton, NJ: Princeton University Press.

Leonard HB, Howitt AM. 2010. Acting in time against disasters: A comprehensive risk management framework. In: Learning from Catastrophes: Strategies for Reaction and Response, ed. Kunreuther H, Useem M. Upper Saddle River, NJ: Prentice Hall.

Lin N, Emanuel K, Oppenheimer M, Vanmarcke E. 2012. Physically based assessment of hurricane surge threat under climate change. Nature Climate Change 2(6):462–467.

Love G, Jarraud M. 2010. Forecasting and communicating the risk of extreme weather events. In: Learning from Catastrophes: Strategies for Reaction and Response, ed. Kunreuther H, Useem M. Upper Saddle River, NJ: Prentice Hall.

Malmstadt J, Scheitlin K, Elsner J. 2009. Florida hurricanes and damage costs. Southeastern Geographer 49:108–131. Available at http://fsu.academia.edu/JillMalmstadt/Papers.

Mendelsohn R, Emanuel K, Chonabayashi S, Bakkensen L. 2012. The impact of climate change on global tropical storm damages. Nature Climate Change 2:205–209.

Meyer R, Broad K, Orlove B, Petrovic N. 2013. Dynamic simulation as an approach to understanding hurricane risk response: Insights from the Stormview lab. Risk Analysis 33(8):1532–1552.

Meyer RJ, Baker EJ, Broad KF, Czajkowski J, Orlove B. 2014. The dynamics of hurricane risk perception: Real-time evidence from the 2012 Atlantic hurricane season. Bulletin of the American Meteorological Society 95:1389–1404.

Mileti DS, Bandy R, Bourque LB, Johnson A, Kano M, Peek L, Sutton J, Wood M. 2006. Annotated Bibliography for Public Risk Communication on Warnings for Public Protective Action Response and Public Education (rev. 4). Available at http://www.colorado.edu/hazards/publications/informer/infrmr2/pubhazbibann.pdf.

Morss RE, Wilhelmi O, Meehl G, Dilling L. 2011. Improving societal outcomes of extreme weather in a changing climate: An integrated perspective. Annual Review of Environment and Resources 36(1):1–25.

Murnane RJ, Elsner JB. 2012. Maximum wind speeds and US hurricane losses. Geophysical Research Letters 39(16):L16707.

Murphy A, Strobl E. 2010. The impact of hurricanes on housing prices: Evidence from US coastal cities. Federal Reserve Bank of Dallas Research Department Working Paper 1009.

NRC [National Research Council]. 2010. When Weather Matters: Science and Service to Meet Critical Societal Needs. Washington, DC: National Academies Press.

NOAA [National Oceanic and Atmospheric Administration]. 2015. Risk Communication and Behavior Assessment: Findings and Recommendations. Internal Report. Silver Spring, MD.

Nordhaus W. 2006. The economics of hurricanes in the United States. NBER Working Paper 12813. Cambridge, MA: National Bureau of Economic Research. Available at http://www.nber.org/papers/w12813.

Nordhaus W. 2010. The economics of hurricanes and implications of global warming. Climate Change Economics 1(1):1–20.

Rappaport EN. 2000. Loss of life in the United States associated with recent Atlantic tropical cyclones. Bulletin of the American Meteorological Society 81(9):2065–2073.

Rappaport EN. 2014. Fatalities in the United States from Atlantic tropical cyclones: New data and interpretation. Bulletin of the American Meteorological Society 95:341–346.

RMS [Risk Management Solutions]. 2008. A Guide to Catastrophe Modelling. The Review: Worldwide Insurance, London. Available at http://forms2.rms.com/rs/729-DJX-565/images/rms_guide_catastrophe_modeling_2008.pdf.

Schmidt S, Kemfert C, Hoppe P. 2009. Simulation of economic losses from tropical cyclones in the years 2015 and 2050: The effects of anthropogenic climate change and growing wealth. DIW Discussion Paper 914. Berlin: Deutsches Institut für Wirtschaftsforschung e.V.

Schmidt S, Kemfert C, Hoppe P. 2010. The impact of socio-economics and climate change on tropical cyclone losses in the USA. Regional Environmental Change 10:13–26.

Strobl E. 2011. The economic growth impact of hurricanes: Evidence from US coastal counties. Review of Economics and Statistics 93(2):575–589.

Tye MR, Holland GJ, Done JM. 2014. Rethinking failure: Time for closer engineer-scientist collaborations on design. Proceedings of the Institution of Civil Engineers 168(2):49–57.

UNISDR [United Nations International Strategy for Disaster Reduction]. 2013. From shared risk to shared value: The business case for disaster risk reduction. Global Assessment Report on Disaster Risk Reduction. Geneva.

UNISDR. 2015. Making development sustainable: The future of disaster risk management. Global Assessment Report on Disaster Risk Reduction. Geneva.

Zhai AR, Jiang JH. 2014. Dependence of US hurricane economic loss on maximum wind speed and storm size. Environmental Research Letters 9(6):064019.

Appendixes

Contributors

Amir AghaKouchak is an assistant professor of civil and environmental engineering at the University of California, Irvine (UCI) and a member of UCI's Center for Hydrometeorology and Remote Sensing. His research group bridges the disciplines of hydrology, climatology, and remote sensing to address critical global water resource issues, with particular interest in the effects of climate change and variability on the terrestrial water cycle and hydroclimate extremes. His long-term research objective is to utilize continuously growing satellite and climate datasets along with ground-based observations to develop integrated water resources modeling, prediction, and decision support systems.

Andrea Alù is an associate professor of electrical and computer engineering and the David and Doris Lybarger Endowed Faculty Fellow at The University of Texas at Austin. His research interests span a broad range of technical areas, including applied electromagnetics, nano-optics and nanophotonics, microwave, THz, infrared, optical and acoustic metamaterials and metasurfaces, plasmonics, nonlinearities and nonreciprocity, cloaking and scattering, acoustics, optical nanocircuits, and nanoantennas.

Jeremy Banik is a senior research engineer in the Space Vehicles Directorate at the Air Force Research Laboratory. His work involves advanced deployable space structures and platforms, including radio frequency reflectors, planar antennas, solar sails, booms, solar arrays, and the supporting material systems, analytical approaches, and testing methodologies.

Jonathan Black is an associate professor of aerospace and ocean engineering at Virginia Tech and an associate director of research for aerospace systems at the Hume Center for National Security and Technology. His research interests include smallsats, mission-oriented satellite constellation design, spacecraft bus engineering, orbital dynamics, aerospace structures and systems, cognitive mission management, joint optimization, engineering design for system payloads, and advanced sensing technology.

Alexandra Boltasseva is an associate professor of electrical and computer engineering in the Birck Nanotechnology Center at Purdue University and adjunct associate professor of photonics engineering at DTU Fotonik at the Technical University of Denmark. Her research is in nanophotonics, plasmonics, plasmonic materials, integrated optics, nanoantennas, optical metamaterials, sensing, and nanofabrication.

Robert Braun is the David and Andrew Lewis Professor of Space Technology at the Georgia Institute of Technology. His research focuses on advanced flight systems and technologies for planetary exploration, and he has contributed to the design, development, test, and operation of multiple space flight systems. He is an elected member of the National Academy of Engineering.

Corale L. Brierley is principal and founder of Brierley Consultancy LLC, which provides technical and business consultation to the mining and chemical industries and government agencies. She is an elected member and vice president of the National Academy of Engineering.

David Brumley is an associate professor of electrical and computer engineering and computer science at Carnegie Mellon University and the CEO of ForAllSecure. With an emphasis on software security, offensive computing, assurance, and defense, his goal is to develop automation using scaling verification and program analysis techniques that check software for exploitable weaknesses.

Jeffrey Czajkowski is the Willis Re Research Fellow at the Wharton Risk Management and Decision Processes Center of the University of Pennsylvania. His primary research areas are natural hazard event direct economic loss modeling, quantifying risk-based catastrophe insurance premiums, estimating the benefits of short- and long-term natural hazard preparation and mitigation, and other economic impact analysis of natural hazards and environmental factors.

Jennifer Dionne is an assistant professor of materials science and engineering at Stanford University. She develops new optical materials and tools to visualize nanoscale systems and phenomena, particularly those relevant to renewable energy and biology.

James Done is a project scientist II and Willis Research Fellow at the National Center for Atmospheric Research Earth System Laboratory. His research area is the physical processes leading to extreme weather and climate events and societal impacts. Recent research focused on predictability limits and variability of high-impact weather events across a range of temporal and spatial scales, and drivers of societal impacts incorporating weather, climate, exposure, vulnerability, and human behavioral factors.

Kevin Fu is an associate professor of computer science and engineering and a Sloan Research Fellow at the University of Michigan, where he directs the Archimedes Center for Medical Device Security and the Security and Privacy Research Group. His lab investigates technology and policy to increase the trustworthiness of embedded computer systems such as medical devices. He co-founded Virta Labs, Inc., which manufacturers a power outlet to detect anomalies and malware by using machine learning and power analysis.

Julia Greer is a professor of materials science and mechanics at the California Institute of Technology and the Kavli Nanoscience Institute, where she focuses on nanoscale mechanics, microstructure, in-situ mechanical deformation, dislocations, and novel nanoscale testing techniques.

Ning Lin is an assistant professor of civil and environmental engineering at Princeton University. Her research integrates science, engineering, and policy to study tropical cyclones and associated weather extremes, how they change with the climate, and how their impacts on society can be mitigated.

Amy Lo is a systems engineer at Northrop Grumman Aerospace Systems, where she focuses on mathematical simulations, mission architecture development, proposal development, and optical systems engineering.

Rebecca Moore is a computer scientist at Google, where she conceived and leads the development of Google Earth Engine. This platform makes trillions of scientific measurements from the world's satellite imagery data dating back over 40 years available online with analysis tools to detect changes, map trends, and quantify differences on Earth's surface. Applications include detecting deforestation, classifying land cover, estimating forest biomass and carbon, and mapping the world's roadless areas.

Daniela Oliveira is an associate professor of electrical and computer engineering at the University of Florida. Her research focus is interdisciplinary computer security, where she applies concepts from psychology, biology, and other disciplines to make computer systems more secure while studying the nature of software vulnerabilities.

Bryan Payne is the engineering manager of platform security at Netflix, with research interests that include host-based intrusion detection and prevention, operating system security, virtualization security, usable security, live and forensic memory analysis, and trusted platforms.

Franziska Roesner is an assistant professor of computer science and engineering at the University of Washington. Her research focuses on improving the computer security and privacy of existing and emerging technologies. She is particularly interested in designing and building computer systems that address security and privacy challenges faced by end users of technologies such as the web, smartphones, and emerging augmented reality platforms.

Dmitry Savransky is an assistant professor in the Sibley School of Mechanical and Aerospace at Cornell University. His research interests include the detection and characterization of extrasolar planets, optimal control of optical systems, statistical analysis of large astronomical surveys, and design and optimization of space missions.

Sara Seager is the Class of 1941 Professor of Physics and Planetary Science at the Massachusetts Institute of Technology. Her research focuses on theory, computation, and data analysis of exoplanets, including models of interiors and biosignatures that led to the first detection of an exoplanet atmosphere. She is an elected member of the National Academy of Sciences.

Christopher Spadaccini is a principal investigator and research engineer in the Center for Micro- and Nanotechnology at Lawrence Livermore National Laboratory. He develops 3D, multi-material fabrication processes spanning size scales that can generate materials with previously unobtainable properties and functionalities by controlling microstructure. High-impact applications include energy-efficient materials and lightweight structures using processes with potential manufacturing production volumes.

Luke Sweatlock is a research scientist in the nanophotonics and metamaterials laboratories at Northrop Grumman Aerospace Systems, where he leads interdisciplinary research in engineered optical materials. He focuses on functional metamaterials for active and nonlinear devices as well as plasmonic materials for nanophotonics and infrared metamaterials for applications in aerospace and communications technology.

Tomas Vagoun is cybersecurity R&D technical coordinator at the National Coordination Office for Networking and Information Technology R&D, providing expertise, technical leadership and management, and guidance, particularly relating to cybersecurity R&D coordination within the federal government.

Mitchell Walker is an associate professor of aerospace engineering at the Georgia Institute of Technology. His primary research interests lie in electric propulsion, plasma physics, and hypersonic aerodynamics/plasma interaction. His current research activities involve advanced spacecraft propulsion systems, diagnostics, plasma physics, helicon plasma sources, space debris mitigation, magnetoplasmadynamic thrusters, and pulsed inductive thrusters.

Program

NATIONAL ACADEMY OF ENGINEERING

2015 US Frontiers of Engineering Symposium
September 9–11, 2015

Chair:
Robert Braun, Georgia Institute of Technology

CYBERSECURITY AND PRIVACY

Organizers:
David Brumley, Carnegie Mellon University, and
Daniela Oliveira, University of Florida

Security at Different Layers of Abstractions:
Application, Operating Systems, and Hardware
Bryan Payne, Netflix

Computer Security and Privacy Where Human Factors Meet Engineering
Franziska Roesner, University of Washington

Interdisciplinary Security: Medical Devices
Kevin Fu, University of Michigan

Challenges of Engineering Cybersecurity: A Government Perspective
Tomas Vagoun, National Coordination Office for
Networking and Information Technology R&D

***

ENGINEERING THE SEARCH FOR EARTH-LIKE EXOPLANETS

Organizers:
Sara Seager, Massachusetts Institute of Technology, and
Mitchell Walker, Georgia Institute of Technology

Engineering the James Webb Space Telescope
Amy Lo, Northrop Grumman Aerospace Systems

Starlight Suppression: Technologies for Direct Imaging of Exoplanets
Dmitry Savransky, Cornell University

Realizing Large Structures in Space
Jeremy Banik, Air Force Research Laboratory

Sensing Controls for Space-based Planet Finding
Jonathan Black, Virginia Tech

***

OPTICAL AND MECHANICAL METAMATERIALS

Organizers:
Jennifer Dionne, Stanford University, and
Luke Sweatlock, Northrop Grumman Aerospace Systems

Materials by Design:
3-Dimensional Architected Nanostructured Meta-Materials
Julia Greer, California Institute of Technology

Mechanical Metamaterials: Design, Fabrication, and Performance
Christopher Spadaccini, Lawrence Livermore National Laboratory

Metamaterial-based Device Engineering
Andrea Alù, University of Texas at Austin

Catching Light Rays:
Refractory Plasmonics for Energy Conversion,
Data Storage, and Medical Applications
Alexandra Boltasseva, Purdue University

***

FORECASTING NATURAL DISASTERS

Organizers:
Amir AghaKouchak, University of California, Irvine, and
James Done, National Center for Atmospheric Research

Assessing and Managing Hurricane Risk in a Changing Climate
Ning Lin, Princeton University

The Economics of Natural Disasters:
Moving from Risk Assessment to Risk Reduction
Jeffrey Czajkowski, University of Pennsylvania

Google Earth Engine:
A New Platform for Global-Scale Disaster Risk Resilience
Rebecca Moore, Google

***

DINNER SPEECH

The Black Swan
Corale Brierley, Brierley Consultancy

Participants

NATIONAL ACADEMY OF ENGINEERING

2015 US Frontiers of Engineering Symposium
September 9–11, 2015

Amir AghaKouchak*
Assistant Professor
Department of Civil and Environmental Engineering
University of California, Irvine

Alina Alexeenko
Associate Professor
School of Aeronautics and Astronautics
Purdue University

Andrea Alù**
Associate Professor, David and Doris Lybarger Endowed Faculty Fellow in Engineering
Department of Electrical and Computer Engineering
University of Texas at Austin

Jorge Arinez
Lab Group Manager
Manufacturing Systems Research Lab
General Motors Research & Development

David Baek
Senior Staff Engineer
Qualcomm

Debasish Banerjee
Principal Scientist
Materials Research Department
Toyota Research Institute of North America

Jeremy Banik**
Senior Research Engineer
Space Vehicles Directorate
Air Force Research Laboratory

Behnam Bastani
Lead Staff Display Engineer
Display Division/ Google [X]
Google

*Organizing Committee
**Speaker

Sarah Bedair
Electronics Engineer and Team Leader
Sensors and Electron Devices Directorate
Power Components Branch
Army Research Laboratory

Peter Bermel
Assistant Professor
School of Electrical and Computer Engineering
Purdue University

Barron Bichon
Manager, Probabilistic Mechanics
Materials Engineering Department
Mechanical Engineering Division
Southwest Research Institute

Jonathan Black**
Associate Professor
Department of Aerospace and Ocean Engineering
Virginia Tech

Lars Blackmore
Principal Rocket Landing Engineer
Guidance Navigation and Control
SpaceX

Jose Blanchet
Associate Professor
Department of Industrial Engineering and Operations Research
Columbia University

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

Robert Braun*
David and Andrew Lewis Professor of Space Technology
Daniel Guggenheim School of Aerospace Engineering
Georgia Institute of Technology

David Brumley*
Associate Professor
Departments of Electrical and Computer Engineering and Computer Science
Carnegie Mellon University

Alberto Giovanni Busetto
Assistant Professor
Department of Electrical and Computer Engineering
University of California, Santa Barbara

Robert Candler
Associate Professor
Department of Electrical Engineering
University of California, Los Angeles

Krystel Castillo
GreenStar Endowed Assistant Professor in Energy
Department of Mechanical Engineering
University of Texas at San Antonio

Judy Cha
Assistant Professor
Department of Mechanical Engineering and Materials Science
Yale University

Julie Chambon
Engineer
Geosyntec Consultants

Ovijit Chaudhuri
Assistant Professor
Department of Mechanical Engineering
Stanford University

Ann Chiaramonti Debay
Materials Research Engineer
Material Measurement Lab
Applied Chemicals and Materials Division
National Institute of Standards and Technology

Duncan Coffey
Technical Manager
Industrial Biosciences
DuPont

Andrea Corrion
Research Staff Scientist
RF MMIC Technologies
HRL Laboratories

Jeffrey Czajkowski**
Senior Research Fellow
Wharton Risk Management and Decision Processes Center
University of Pennsylvania

Samantha Daly
Associate Professor
Departments of Mechanical Engineering and Materials Science and Engineering
University of Michigan

Anthony D'Amato
Research Engineer
Modern Control Methods
Ford Motor Company

Michael DePierro
Process Engineering Specialist
Silanes Process Engineering
Dow Corning Corporation

Shana Diez
Manager for Reusability Development
Vehicle Engineering
SpaceX

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

Jennifer Donahue
Associate
Geosyntec Consultants

James Done*
Project Scientist II and Willis Research Fellow
NCAR Earth System Laboratory

Ramakrishna Dontha
Director, Advanced Controls Development
Cummins Emission Solutions
Cummins Inc.

Adam Engler
Associate Professor
Department of Bioengineering
University of California, San Diego

Anna Erickson
Assistant Professor
George W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology

Larry Fahnestock
Associate Professor and CEE Excellence Faculty Fellow
Department of Civil and Environmental Engineering
University of Illinois at Urbana-Champaign

Waleed Farahat
Director of Mechatronics and Control
Rethink Robotics

Philip Freeman
Technical Fellow
Advanced Production Systems
Boeing Company

Kevin Fu**
Associate Professor
Department of Computer Science and Engineering
University of Michigan

Yun Fu
Associate Professor
Department of Electrical and Computer Engineering
Northeastern University

Jeannette Garcia
Research Staff Member
Chemistry and Advanced Materials Research
IBM Almaden Research Center

Kevin Geary
Senior Research Staff Engineer and Department Manager
Apertures Department, Applied Electromagnetics Laboratory
HRL Laboratories

Christopher Geiger
Market Segment Chief Engineer
Enterprise Test Solutions
Lockheed Martin

Brian German
Langley Associate Professor
Daniel Guggenheim School of Aerospace Engineering
Georgia Institute of Technology

Julia Greer**
Professor of Materials Science and Mechanics
Division of Engineering and Applied Sciences
California Institute of Technology

Eric Guyer
Principal Engineer and Office Director
Materials and Corrosion Engineering
Exponent, Inc.

Javier Guzman
Research Associate
Corporate Strategic Research
ExxonMobil Research and Engineering

Karl Haapala
Associate Professor
School of Mechanical, Industrial, and Manufacturing Engineering
Oregon State University

Xue Han
Assistant Professor
Department of Biomedical Engineering
Boston University

He (Helen) Huang
Associate Professor
UNC-NCSU Joint Department of Biomedical Engineering
North Carolina State University
University of North Carolina, Chapel Hill

DeShawn Jackson
Marketing Manager
Product Enhancement
Halliburton

Jennifer Jordan
Program Officer, Dynamic Materials and Interactions
Engineering and Information Sciences Branch
Air Force Office of Scientific Research

Christine Keenan
Program Manager
Graphics Communication Program Management Group
Xerox

Jeffrey Lee
Member of Technical Staff
Bell Laboratories/Alcatel-Lucent

Li Li
Associate Professor
Department of Energy and Mineral Engineering
Pennsylvania State University

Zhuo Li
Software Architect
Digital Signoff Group
Cadence Design Systems

Chien-Chi Lin
Assistant Professor
Department of Biomedical Engineering
Indiana University – Purdue University

Ning Lin**
Assistant Professor
Department of Civil and Environmental Engineering
Princeton University

Sarah Lipscy
Principle Engineer
Systems Engineering
Ball Aerospace and Technologies Corp.

Jennifer Listgarten
Researcher
MSR New England
Microsoft

Haowei Liu
Staff Research Engineer
Perceptual Computing
Intel

Amy Lo**
Systems Engineer
Northrop Grumman Aerospace Systems

Hussam Mahmoud
Assistant Professor
Department of Civil and Environmental Engineering
Colorado State University

Christienne Mancini
Systems Engineering Lead
Civil Systems Division, Information Systems Sector
Northrop Grumman

Megan McCain
Gabilan Assistant Professor
Department of Biomedical Engineering
University of Southern California

Bryan McCloskey
Assistant Professor
Department of Chemical and Biomolecular Engineering
University of California, Berkeley

Elia Merzari
Principal Nuclear Engineer
Departments of Mathematics and Computer Science and Nuclear Engineering
Argonne National Laboratory

Jeetain Mittal
Assistant Professor
Department of Chemical and Biomolecular Engineering
Lehigh University

Rebecca Moore**
Engineering Manager
Google Earth Engine
Google

Patrick Norris
Product and Technology Development Engineer
Cardiovascular Products
W. L. Gore & Associates

Grace O’Connell
Assistant Professor
Department of Mechanical Engineering
University of California, Berkeley

Daniela Oliveira*
Associate Professor
Department of Electrical and Computer Engineering
University of Florida

Timothy Ombrello
Research Aerospace Engineer
Aerospace Systems Directorate
Air Force Research Laboratory

David Parkes
George F. Colony Professor of Computer Science
John A. Paulson School of Engineering and Applied Sciences
Harvard University

Bryan Payne**
Engineering Manager
Platform Security
Netflix

Lisa Poyneer
Research Engineer
Lawrence Livermore National Laboratory

Jeff Putnam
Technical Director
Broadband Communications Group
Broadcom Corporation

Jun Qu
Senior R&D Staff Scientist
Materials Science and Technology Division
Oak Ridge National Laboratory

Varun Raghunathan
Research Engineer
Life Science and Pathology Instrumentation
Agilent Research Laboratories

Mabel Ramirez
Technical Staff
Air and Missile Defense Technology Division
Advanced Concepts and Technology Group
MIT Lincoln Laboratory

Franziska Roesner**
Assistant Professor
Department of Computer Science and Engineering
University of Washington

John Rose
Engineering Manager and Chief of Staff, Engineering
The Boeing Company

Scott Saponas
Researcher
Microsoft Research

Dmitry Savransky**
Assistant Professor
Sibley School of Mechanical and Aerospace Engineering
Cornell University

Sara Seager*
Class of 1941 Professor of Physics and Planetary Science
Departments of Earth, Atmospheric, and Planetary Sciences and Physics
Massachusetts Institute of Technology

Julie Shah
Associate Professor
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology

Maryam Shanechi
Assistant Professor and Viterbi Early Career Chair
Department of Electrical Engineering
University of Southern California

Sameer Sonkusale
Associate Professor
Department of Electrical and Computer Engineering
Tufts University

Christopher Spadaccini**
Principal Investigator and Research Engineer
Center for Micro- and Nanotechnology
Lawrence Livermore National Laboratory

Todd Stewart
Manufacturing Technology Senior Engineer/Scientist
DuPont Protection Technology
DuPont

Luke Sweatlock*
Research Scientist
Nanophotonics and Metamaterials Laboratory
Northrop Grumman Aerospace Systems

Stephanie TerMaath
Assistant Professor
Department of Mechanical, Aerospace, and Biomedical Engineering
University of Tennessee

Peter Tessier
Richard Baruch M.D. Career Development Associate Professor
Department of Chemical and Biological Engineering
Rensselaer Polytechnic Institute

Tomas Vagoun**
Cybersecurity R&D Technical Coordinator
National Coordination Office for Networking and Information Technology R&D

Mario Valenti
Lead Structural Engineer
Transportation and Infrastructure Division
STV Inc.

Jason Valentine
Assistant Professor
Department of Mechanical Engineering
Vanderbilt University

Anthony Vamivakas
Assistant Professor
Institute of Optics
University of Rochester

Peggy Van Eepoel
Associate Principal
Structures
Weidlinger Associates, Inc.

Mandana Veiseh
Senior Scientist
Electronic Materials and Devices Laboratory
Palo Alto Research Center (Xerox PARC)

Enrique Vivoni
Professor
School of Earth and Space Exploration
Arizona State University

Richard Vuduc
Associate Professor and Associate Chair
School of Computational Science and Engineering
Georgia Institute of Technology

Mitchell Walker*
Associate Professor
Daniel Guggenheim School of Aerospace Engineering
Georgia Institute of Technology

Chonggang Wang
Member Technical Staff/Senior Manager
Innovation Lab
InterDigital Inc.

Joshua White
Research Scientist
Atmospheric, Earth, and Energy Division
Lawrence Livermore National Laboratory

Laurent White
Research Associate
Corporate Strategic Research
ExxonMobil Research and Engineering

Kathryn Whitehead
Assistant Professor
Departments of Chemical Engineering and Biomedical Engineering
Carnegie Mellon University

Kristin Williams
Principal Engineer
Fabric and Home Care
Research and Development
Procter and Gamble Company

Guihua Yu
Assistant Professor
Department of Materials Science and Engineering
University of Texas at Austin

Roseanna Zia
Assistant Professor
School of Chemical and Biomolecular Engineering
Cornell University

Stephen Zingelewicz
Senior Engineer
Software Sciences and Analytics
GE Global Research

Dinner Speaker

Corale Brierley
Principal
Brierley Consultancy LLC

Guests

William (Bill) Hayden
Vice President
The Grainger Foundation

Sohi Rastegar
Senior Advisor
Emerging Frontiers and Multidisciplinary Activities
Directorate for Engineering
National Science Foundation

National Academy of Engineering

Alton D. Romig, Jr.
Executive Officer

Janet Hunziker
Senior Program Officer

Sherri Hunter
Program Coordinator