UNLEASHING OPPORTUNITY

a project of

NATIONAL AFFAIRS

Policy Reforms to Advance Innovation

Jim Manzi
Mark P. Mills
Peter Huber & Paul Howard
Christopher DeMuth
UNLEASHING OPPORTUNITY

A project of NATIONAL AFFAIRS

Policy Reforms to Advance Innovation

Jim Manzi
Mark P. Mills
Peter Huber & Paul Howard
Christopher DeMuth
Contents

Executive Summary 7

Innovation Policy for American Leadership 11
   in the 21st Century | Jim Manzi

Energy Policy for American Leadership 26
   in the 21st Century | Mark P. Mills

Biopharmaceutical Policy for American Leadership 52
   in the 21st Century | Peter Huber & Paul Howard

Wireless Telecommunications Policy for American Leadership in the 21st Century 78
   | Christopher DeMuth
Executive Summary

Anything done a second time is no longer an innovation, and therefore innovation resists planning and rational management. Consider three of the most successful American technological innovations of the past several decades: the IT revolution, the defeat of HIV/AIDS, and the shale energy revolution. Information technology has become central to contemporary economic growth; HIV/AIDS is the only instance in decades of our having defeated a major new disease on the scale of earlier triumphs over infectious diseases; and the United States has suddenly and unexpectedly moved to a position of global energy superpower. In each case federal investments, often over decades, were crucial to ultimate success. But in each case, the methods that created the breakthroughs themselves were either independent of, or actually contrary to, the dominant federal policy focus in each area.

This was in part because of the crucial role played by marginalized groups ranging from nerdy computer hobbyists to gay-rights activists. And it was in part because the solutions themselves were typically very threatening to established technology organizations, both public and private. In 1975, almost nobody in power foresaw that Bell Labs and IBM were about to be outcompeted by start-ups that didn’t yet exist. In 1985, almost nobody in power foresaw the rapid, iterative method for drug development that invented the algorithmic cocktails necessary to conquer AIDS. In 2005, almost nobody in power foresaw the shale revolution, which would transform the geopolitical situation of the U.S. within a decade. And in 2016, we can be confident that almost nobody in power can currently foresee, never mind manage, whatever future technology innovation will ultimately produce the greatest benefits for America and the world.

What then should those in power do? In short, they should try to create the conditions for innovation, rather than attempting to guide the
direction of innovation. This would be in keeping with the historical
American approach to achieving innovation, which at its most effective
has demonstrated an almost ruthless pragmatism in implementing the
core principles of free markets and strong property rights, overlaid with
decisive government investments in infrastructure, human capital, and
fundamental knowledge. The contemporary implementation of these
principles combines four specific elements: (1) a foundation of free mar-
kets and strong property rights; (2) a swarm of entrepreneurial start-ups
with independent venture financing; (3) competitive-cooperative re-
lationships between these start-ups and large incumbent companies;
and (4) support by massive government technology investments. This
paradigm first emerged in the information technology industry, but has
now become the foundation of more recent breakout successes in the
energy and life-sciences sectors.

Policymakers should attempt to reinforce each of these four ele-
ments of American innovation leadership. Focusing on the specific and
practical, several of the most attractive policy changes are the following:

1. Reallocate 5% of federal social welfare spending to double fed-
eral spending on science and technology.
2. Focus U.S. immigration policy on the recruitment of high-skill
workers, rather than on family-unification and humanitarian
goals.
3. Reorient energy policy around accelerating and exploiting the
shale energy revolution in the medium-term, and investing
in breakthrough fundamental research for alternative energy
sources for the long-term.
4. Update pharmaceutical regulation by modernizing patent law
and FDA approval of new therapies to reflect the new, increas-
ingly personalized nature of therapeutic innovation.
5. Liberalize regulation of the electromagnetic-frequency
spectrum to allow holders of spectrum licenses to use their
assigned frequencies for any purpose from radio broadcasting
to smartphone services.

The first chapter of this section is a general overview of the contempo-
rary American innovation process set in historical context, with several
resulting economy-wide recommendations. The three subsequent
chapters sequentially analyze three industries central to American innovation: energy, biopharmaceuticals, and wireless telecommunications. Each of these three chapters then makes practical, sector-specific policy recommendations. In total, this section provides an agenda for federal government policies that would likely enable accelerated innovation in the United States.
A mericans have become uncharacteristically pessimistic about the future. There is broad acceptance of the need for innovation, but a lack of belief that we are up to the job.

This feels like an unfamiliar challenge. But in fact, American innovation has always proceeded in cycles in which pessimism and near panic act as the emotional impetus for creative advances that yield a new economic order that thrives for a time and then loses its edge. That cycle has been essential to America’s unprecedented economic achievement over the past two-and-a-half centuries, and it is playing out again today. By seeing how we’ve managed, time and again, to remake America into an engine of innovation and prosperity, we can better understand the nature of the challenge we now face, the character of the opportunities we may have for addressing it, and the kinds of responses that are most likely to work.

The American System

Gross domestic product per capita is the fundamental measure of a society’s material standard of living over time. By this measure, the American colonies started poor but began getting richer very quickly. By 1820, the United States had achieved a level of GDP per capita roughly equal to that of Western Europe. By 1900, it had achieved a decisive global advantage compared to all potential strategic rivals. It has maintained a great advantage on that front ever since. Though Europeans certainly live very well, and though both China and India are rapidly improving their citizens’ standards of living, the United States remains indisputably ahead.

How has America achieved and sustained these great gains in living
standards? By definition, GDP per capita is equal to GDP per work-hour multiplied by the number of hours worked per capita. Because there are only 24 hours in a day, the only way to increase living standards indefinitely is by increasing the GDP per work-hour part of the equation—that is, by increasing labor productivity.

There are really only two ways to increase labor productivity. The first is to increase our use of other inputs like land or equipment. The second is to invent and implement new ideas for getting more output from a given set of inputs—that is, to innovate. In 1957, economist Robert Solow of MIT published the first modern attempt to measure the relative contributions of additional inputs versus innovation to increasing labor productivity. He looked in particular at the United States between 1909 and 1949 and estimated that only about 12.5% of all growth in output per work-hour over that period could be accounted for by increased use of capital; the remaining 87.5% was “attributable to technical change,” or innovation.

Intense debate has followed in the long wake of Solow’s conclusions, for which, along with related work, he was awarded a Nobel Prize. The participants in that debate have often offered arguments in which deeply rooted ideological beliefs have been passed off as technical assumptions, but in spite of these disagreements, there is widespread scholarly consensus that (as common sense would indicate) innovation, broadly defined, has been central to increasing American living standards. In summary, the root of American economic success hasn’t been luck, or land, or conquest; it has been innovation.

The nation’s approach to achieving innovation has varied with the times, but it has generally demonstrated an almost ruthless pragmatism in implementing the core principles of free markets and strong property rights, overlaid with decisive government investments in infrastructure, human capital, and new technologies.

Free markets and aggressive public investments in infrastructure exist, of course, in some tension. Generally speaking, the underlying system of economic organization in the United States has been not only a free-market system but one that is among the freest in the world. Independent economic agents own private property and engage in only loosely regulated contracting. Highly distributed trial-and-error learning motivated by enlightened self-interest has always been the key driver of innovation. The government’s role is mostly that of an armed
referee rather than a participant in the economy. In this sense, America’s
underlying innovation policy has been “no policy.” But significant government overlays have always existed to reinforce our free economy. Indeed, the federal government has been active in shaping specific kinds of innovation since the first months of the republic, when Alexander Hamilton published his epochal 1791 Report on Manufactures. The Report proposed subsidies and protections for developing manufacturing industries — the high-tech sector of its day — to be paid for by tariffs.

The debate about these recommendations was strikingly modern. Hamilton shared in the reigning American consensus in favor of free markets, but advocated for an exception in the case of manufacturing. His case was rooted in sophisticated externality arguments. Manufacturing, he argued, would allow for a far more efficient division of labor and better matching of talents and capacities to occupations, would create an additional market for agricultural products, and would encourage immigration to further extend each of these benefits. All of this would be immensely useful to the new nation, and it was only sensible for the government to actively encourage it. The opponents of Hamilton’s plan emphasized what we would call today the public-choice problems with subsidizing specific sectors and businesses, especially the potential for corruption and sectional favoritism. Broadly speaking, the tariffs were implemented, but not the subsidies. In fact, the tariffs quickly became much higher than those Hamilton had proposed, as political constituencies grew up around them.

Despite his suggestions being only partially implemented, Hamilton’s basic insight — that the enormous economic value that innovative industries could offer the nation merited public efforts to enable their success — has always had strong adherents in national politics. In the decades leading up to the Civil War, for instance, the federal government intervened strategically in markets to spur innovation, immediately and frequently exercised its constitutionally enumerated power to grant patents, and even encouraged and protected Americans who stole industrial secrets from Britain — at the time the world leader in manufacturing technology.

Much of the motivation for such policies was grounded in military priorities. The West Point military academy was founded in 1802 in large part to develop a domestic engineering capability, and armory
expenditures stimulated the growth of an indigenous manufacturing capability that by 1850 had in some sectors become the most advanced in the world. Other significant investments in infrastructure included financing, rights of way, and other support to build first canals and then railroads, which were essential to driving productivity improvements. The internet of the era was the telegraph, and it too benefited from public support early on. In 1843, Congress allocated the money to build a revolutionary telegraph line from Washington, D.C., to Baltimore that pioneered many of the important innovations—such as suspended wires—that would come to be used to build out the national telegraph network and later the telephone network.

Henry Clay called this program of tariffs, physical infrastructure, and national banking “The American System.” Its goal was to transform the United States from a group of sectionally divided agricultural states tightly linked to the British manufacturing colossus into a unified, dynamic industrial economy. Abraham Lincoln, who identified himself as “an Old-line Henry Clay Whig,” accelerated this process dramatically, both because of the exigencies of war and because the Southern-based opposition to the program was no longer in Congress during the Civil War years. The federal government moved aggressively. It expanded the infrastructure of railways and telegraphs, increased tariffs, established a system of national banks, founded the National Academy of Sciences, and established the Department of Agriculture and a system of land-grant colleges that ultimately created agricultural experiment stations to promote innovation on farms.

Federal investments in biology and health innovation began to accelerate rapidly in the late 19th century. A set of Navy hospitals with origins in the 1790s was organized into the Marine Hospital Service in 1870, and Congress allocated funds for the study of epidemics, with particularly significant innovation occurring in the study of malaria. This organization ultimately became the 20th-century Public Health Service and spawned the Centers for Disease Control and Prevention. Congress also established the Laboratory of Hygiene in 1887, which eventually became the National Institutes of Health.

A direct line ran from Hamilton to Clay to Lincoln, and this approach—a free-market base overlaid with specific interventions to provide infrastructure and to promote incremental, innovation-led growth—was the pattern for roughly the century that followed
Lincoln’s assassination. But as a political matter, it became increasingly populist over that time. Hamilton was often attacked as a pretentious would-be aristocrat. Though Clay lived as a gentleman, he made hay of his humble origins and is often credited with inventing the term “self-made man.” Lincoln was born in a log cabin and had a public image that was the opposite of aristocratic. The political energy behind this program remained nationalistic, but became increasingly focused on upward mobility, social striving, and maintaining the long-term legitimacy of the economic regime with the promise of opportunity for all.

This trend continued into the last century. Consequently, from roughly 1870 to 1970, the goals of conquering disease, educating the unschooled, and winning wars provided the strongest impetus for government investments in innovation. This remains central to the American ethos, and in certain respects the Progressive Era and the New Deal can be seen as extensions of this program, in which egalitarian ideas played an ever-increasing role.

This approach interacted with the rise of the mass institutions of mid-20th-century American life to create an innovation system that, in retrospect, was unusually centrally directed by American standards. World War II, of course, saw a level of defense-led government investment in technology that was unprecedented. At the conclusion of the war, Vannevar Bush—who personified this big institutional approach as dean of engineering at MIT, founder of the defense contractor Raytheon, and founding director of the federal Office of Scientific Research and Development through which almost all military R&D was carried out during the war—authored the pivotal July 1945 report “Science: The Endless Frontier.” The document laid the groundwork for the Cold War-era system of government sponsorship of science and engineering.

What is most immediately striking about this report to a contemporary reader is that Bush found it essential to define and defend federal investments in basic research. Previously, the United States had relied predominantly on exploiting fundamental scientific discoveries made in Europe, but in the post-war world, the nation would have to forge ahead on its own. In the same era, the G.I. Bill at the federal level, combined with increased spending at the state level, democratized access to higher education. Just as much of the rest of the world was catching up to America in secondary education, the United States began to pursue
mass higher education. This created the pool of people who would become known as the “knowledge workers” of the new economy.

Another striking thing about the Bush report was its emphasis on large institutions — both public and private. He lived in and described a world of mammoth organizations, centralized coordination, and scale economics. These were the institutions that had won the war and would win the peace. America in the wake of the Depression and the war exhibited a faith in such large institutions that was unmatched in our country before and has not been seen since.

The big-institution approach that Bush described and then helped to orchestrate was an incredibly successful program for innovation that created the conditions for the growth of the information economy, as well as much of the aerospace and biomedical industries and related fields. But it also planted the seeds of its own obsolescence (or at least a drastically reduced relevance). Ironically, the new economy that this post-war system helped to create has little room for the post-war system itself. Innovation in the new economy looks quite different, a fact that America is still struggling to understand.

This transformation came first to information technology.

THE INFORMATION TECHNOLOGY REVOLUTION

The U.S. IT industry has roots that are at best semi-capitalist. For decades, the Department of Defense was the primary customer for innovative information technology. The technology itself was mostly developed by government labs, universities, defense contractors, Bell Labs, and similar institutions. From World War II through about 1975, this public-private complex was at the frontier of innovation, producing (among many other things) the fundamental components of the software industry, as well as the hardware on which it depended. Government agencies collaborated with university scientists to develop the electronic computer and the internet. Bell Labs invented the transistor, the C programming language, and the UNIX operating system.

The roots of the sector’s transformation can be traced to the 1980s. The PC revolution was transforming all of computing in a consciously democratized and decentralized way. This began the ascendance of Silicon Valley over all other technology centers, with its more open, freewheeling start-up culture. The technology sector was also directly affected by changes elsewhere in the economy. Financial innovators both invented so-called
“rocket-science finance” and began the process of using debt and equity markets to break up and transform huge American manufacturing companies. Strategy consulting was a new industry, mostly created by people with science and engineering degrees who wanted to stimulate and directly profit from rapid change in large organizations to an extent that was typically infeasible as employees of the companies themselves. These new companies began to grow explosively. Increasingly, the best science and engineering graduates were drawn to start-ups, finance, and strategy consulting. The center of gravity of innovation moved decisively from the behemoths of the post-war era to newer, more nimble competitors.

A Defense Science Board report published in January 1987 quantified the resulting transformation of the industry, noting that commercial electronics such as computers, radios, and displays were one to three times more advanced, two to ten times cheaper, five times faster to acquire, and altogether more reliable than their equivalents from the Department of Defense. Extrapolating from the report, professor Steven Vogel correctly anticipated that “commercial-to-military ‘spin-ons’ are likely to boom while military-to-commercial ‘spin-offs’ decline.” In a complementary report published a month later, the Board acknowledged that “the Department of Defense is a relatively insignificant factor to the semiconductor industry” that it had originally midwifed.

Why did all of this happen at that time?

First, the cultural revolution of the 1960s and ’70s elevated independence and iconoclasm at the expense of the organization man. That ethic was essential to the early culture of the information-technology boom, and has in many respects remained quite important.

Second, information technology radically lowered many kinds of economic transaction costs, so efficient firm size became smaller for many of the most important industries most essential to economic growth. While industrial technology rewards controlled scale, information technology rewards decentralized networks. As a result, the smaller, insurgent information-technology firms had different internal systems and cultures that were less overtly hierarchical and more decentralized.

Third, the science and technology that undergirded these innovative sectors worked to the advantage of this new model. It took huge, integrated organizations to do things like build out the national telephony network, but small groups could exploit those achievements to facilitate other technologies, like the consumer modem.
Fourth, the same industry trends also allowed those who were creating improvements to secede from huge firms and form independent companies, using equity and equity-like vehicles to extract greater economic value. This caused a much bigger spread in compensation across companies than is practically possible within most large, traditional organizations. Thus, for those who believed they could create identifiable shareholder value, there were enormous financial incentives to migrate to start-ups, consulting firms, and new types of financial firms like private-equity shops. This created a virtuous cycle: As these firms grew more culturally attractive and remunerative, they increasingly attracted the strongest graduates, which reinforced both their economic and cultural appeal.

Fifth and finally, financing became far more available with the rapid growth of the venture-capital industry, which itself was very much part of this new world. An entire ecosystem of investors, lawyers, accountants, and even landlords (who would take equity instead of cash for initial rent) arose to encourage a venture-backed sector with the basic job of disrupting and displacing big, established companies.

These forces have created a new paradigm for innovation that almost always boils down to figuring out how to invent and use information technology creatively to re-engineer an ever-expanding range of activities. This isn’t surprising, as Moore’s Law (which observes that computer-chip performance doubles approximately every 18 months) points to by far the most sustained and significant increase in fundamental technical capacity over the past half-century. The institutional arrangements that have served to enable this wave of innovation consistently exhibit a four-part structure: (1) innovative entrepreneurial companies, (2) financed by independent investment firms, (3) competing and cooperating with established industry leaders, and (4) all supported by long-term government investments in infrastructure and R&D.

This combination characterizes large swaths of the key information-technology and biotechnology industries. Together, these two industries represent about 80% of cumulative U.S. venture-capital commitments over the past 30 years. And their success has not only benefitted those directly involved but has also yielded enormous advantages for Americans in general. According to the National Venture Capital Association, as of 2010 about 11% of all U.S. private-sector jobs were with venture-backed companies, including 95% of all software jobs, more than 70% of all
semiconductor and biotechnology jobs, and about half of all computer and telecommunications jobs.

This new approach to innovation does not exclusively apply to “high tech” sectors. In fact, perhaps the most important recent example of unexpected innovation following this approach has involved the extraordinarily quick and unexpected transformation of our energy economy—a transformation that has run directly contrary to what had long been the government-led strategic approach to energy innovation.

**The Energy Revolution**

America’s dependence on imported fossil fuels is widely acknowledged to be a source of many serious problems—from the enormous military expenditures required to keep supply lines open in dangerous parts of the world to the dangers of pollution and the threat of climate change.

As recently as six or seven years ago, these difficulties seemed intractable, as the rapid development of alternative energy sources seemed the only way out of our dependence on foreign oil, and such development seemed nowhere in sight. In 2008, renewables provided about 7% of all American energy, up less than one percentage point from their contribution a decade earlier. Nuclear-power use was also flat over that decade at about 8%, leaving the lion’s share—about 85%—of all American energy to be provided by fossil fuels. That same year, the International Energy Agency reflected widespread conventional wisdom when it projected that U.S. oil and natural-gas production would remain flat or decline somewhat through about 2030, therefore necessitating ever-growing imports. Despite all of the talk and plans, no progress at scale seemed to be feasible.

In the last decade, however, a technological revolution in the extraction of so-called unconventional fossil fuels has transformed this situation with breathtaking speed. The most important technology has been hydraulic fracturing, often called “fracking,” but other important developments have included tight-oil extraction, horizontal drilling, and other new applications of information technology. Since 2006, America’s output of crude oil, natural-gas liquids, and biofuels has increased by about the same amount as the total output of Iraq or Kuwait, and more than that of Venezuela. America has now become the world’s largest oil producer.

For Americans, the benefits of this change are enormous. First are the geopolitical benefits: North American energy independence would not mean that we would no longer have important interests in the Middle
East, Africa, and elsewhere, but it would mean that we would no longer be negotiating under duress. Second, because it has caused a net shift from coal to natural gas, the fracking revolution has produced tremendous environmental benefits. Since 2006, carbon-dioxide emissions have fallen more in the United States than in any other country. And third are economic benefits. Citigroup estimated in 2013 that within seven years the new energy revolution should add about 3% to GDP, create about 3 million more jobs, and reduce the trade deficit by about 2% of GDP.

America has led this technological revolution and as of today stands alone among the major world powers in this regard. A more detailed view of this energy revolution will be presented in a later chapter, but at a summary level the U.S. has succeeded because of the same combination of structural factors that work together to encourage innovation in information technology. First has been a foundation of free markets and strong property rights. Among the world’s key petroleum-producing countries, only the United States allows private entities to control large-scale oil and gas reserves and to set prices mostly freely. This combination has been particularly advantageous for fracking because shale formations tend to have great local variability. Second has been the network of independent petroleum producers, oil-field service companies, and specialized financing expertise that this regulatory structure has allowed to thrive in America over the past century. Many of the recent technological advances have been made through trial-and-error and incremental improvements, which map well to a Darwinian competition among a network of independent companies, as opposed to huge one-time projects by industry giants or quasi-governmental organizations. Finally, government technology investments have mattered. Direct subsidies for speculative energy technologies and research over at least 35 years have played some role in the development of 3-D seismology, diamond drill bits, horizontal drilling, and others. But government-led efforts that are less obviously related have actually been much more important, crucially the defense-related expenditures described earlier that enabled the U.S.-centered information-technology revolution, which has in turn created the capacity to more rapidly develop “smart drilling” technology.

RecommEnDAtions

Many of the economy-wide reforms most needed to stimulate further American innovation would be familiar in concept to Abraham
Lincoln, because they are merely an updated version of the American System. As in Lincoln’s day, they are motivated by an enlightened nationalism that seeks growing incomes and widely shared prosperity and opportunity, with direct investments focused on infrastructure, human capital, and new technologies. To make the most of this new American system in our time, policymakers should pursue four basic goals.

First, they should build infrastructure. While many claims about America’s deteriorating basic infrastructure are overblown, further investment in roads, bridges, dams, and railways are warranted. According to the World Economic Forum, American infrastructure quality now ranks 23rd best in the world after having been in the top ten less than a decade ago. This requires investment. As described in more detail in a later chapter, special attention should be paid to modernizing and hardening the electrical grid.

There is an adage among infrastructure engineers: “organization before electronics before concrete.” That is, we should always do the low-cost but unglamorous work of optimizing current fixed investments before making new ones. But there is also the need for classic big projects. The key remediable barriers to this type of improvement have to do with legal complexity: federal, state, and local regulations; private lawsuits; union work rules; and other legal hurdles. Congress might help to cut the red tape by creating a class of federal “special national infrastructure projects” that would be exempted from numerous regulatory and related barriers, granted presumptive immunity from specific classes of lawsuits, and given expedited eminent-domain rights. Congress should link funding for such projects to special exemptions from analogous state-level and local regulations in those areas that want to benefit from the projects.

The physical capacity for movement of digital data is the modern version of the telegraph and telephony networks. Development of infrastructure for technologies like these, which are themselves rapidly evolving during the build-out, presents special challenges. We have historically had an approach to this class of infrastructure that combines government investment in visionary projects, financing, right-of-way provision, and standard-setting with a heavy reliance on private-sector competition for the actual build-out. This always seems very messy in the short term, but when technologies are in a state of flux, the combination of public and private tends to create infrastructure better suited to economic success over time.

The crucial unknown today is the relative importance of high-speed
fixed-line broadband versus better mobile capacity. Unfortunately, we are leaders in neither technology. America has lower broadband subscriptions per capita than the major Western European economies and ranks 111th in the world in mobile subscriptions per capita. We should err on the side of overinvestment and seek leadership in both. And as described in more detail in a later chapter, we should modernize regulation of the wireless spectrum to allow far more flexible use of this resource.

Second, policymakers should invest in building visionary technologies. America’s technology strategy through most of its history was to commercialize the discoveries of European science. We began investing massively in basic research in the post-World War II era only because there was nobody else left to do it. Today, America is the global leader in basic science. Almost half of all the most cited scientific papers are produced in the U.S. But the world is changing, and in 2013 the U.S. represented only about 20% of world GDP and 28% of world R&D spending. Over roughly the past 20 years, the fraction of American scientific papers with a non-American co-author has grown from 12% to 32%. Science is becoming more international again. We should give ground grudgingly but recognize that over time more science will be done outside the United States. We should participate aggressively in research collaborations such as international space-exploration efforts and the European CERN particle-physics facility, and we should fund exchanges and other vehicles to ensure that we gain maximum benefit in return for our basic-research investments.

And we should think differently in this new world about what basic science we conduct here. We should bias basic-research funds not toward those areas that inherently hold the greatest promise, but toward those in which the long-run economic benefits are likely to remain in the United States, because they require the build-up of hard-to-transfer expertise or infrastructure that are likely to generate commercial spin-offs. University and research laboratory rules and the patent system should recognize the long-run desirability of researchers creating private wealth in part through the exploitation of knowledge created by these publicly supported institutions.

We have a long track record of doing this well and an existing civilian infrastructure that can be repurposed, including most prominently the Department of Energy’s national laboratories, the National Institutes of Health, and NASA. Each of these entities is to some extent adrift and
should be given bold, audacious goals. They should be focused on solving technical problems that offer enormous social benefit, but are too long-term, too speculative, or have benefits too diffuse to be funded by private companies.

Careful deliberation will be required to establish specific targets, but great technical organizations have a characteristic spirit that starts with goals that are singular, finite, and inspiring. That is, each organization should have one goal. The goal should be sufficiently concrete that we can all know if it has been achieved or not. And it should be sufficiently impressive that people are proud to work toward it, without being so obviously outlandish that it just inspires cynicism. Beyond goals, political leadership has the responsibility for selecting extremely able senior leadership, providing adequate resources, granting operational autonomy, and measuring progress.

Third, we should build human capital. As American statesmen have known for centuries, this is an essential building block of innovation and prosperity. And we are losing our edge.

Human capital is built by a combination of attracting and admitting immigrants who have it, and then helping everyone in the country to further develop their skills and abilities. Our immigration system, for instance, should be reconceived as a program of recruitment, rather than law enforcement or charity. We should select among applicants for immigration above all through assessments of skills and capabilities. Canada and Australia do this now, with excellent results. And, exactly as Hamilton argued 200 years ago, immigration and investments in visionary technology projects will tend to be mutually reinforcing. The projects will tend to attract and retain immigrants that have much to offer the country, and key immigrants will help drive these visionary projects forward.

Better schools and universities are also essential, of course, and they would again reinforce gains made through high-skill immigration, investments in technical projects, and investments in infrastructure. The challenges confronting meaningful education reform, however, seem almost intractable. Addressing them will require a serious commitment to encouraging competition and innovation within education itself.

Fourth and finally, therefore, we must significantly deregulate and encourage competition in the three sectors that have been most resistant to the new American system of innovation: government services, education, and medicine. Today, we treat these three sectors as hopeless victims of
Baumol’s Cost Disease: the idea that, because they involve services delivered by human beings, they will inevitably grow in relative cost versus those sectors that deliver value through technology. But there is nothing inevitable about this, and all three are too important to be disregarded as lost causes. We cannot just write off the 30% of the workforce that works in government services, education, and medicine from productivity gains, especially given how essential these sectors are to enabling innovation and productivity growth in other parts of the economy.

Driving productivity improvements in these areas, in theory, would simply require allowing the same IT-driven gale of disruptive innovation that has transformed other sectors. The practical reality, though, is that it is very difficult to create in these industries the sorts of incentives that have driven innovation in areas like software or energy. Resources are mostly politically controlled, and various participants use this to protect themselves from disruptive change. This is not entirely cynical. Our moral intuitions about them are very different, for one thing. It is not coincidental that, in the West, schools and medicine were provided for centuries through church-linked institutions by “professionals”—a religious term in its origin, implying that teachers, nurses, and doctors were expected to place service to others ahead of self-interest. And the provision of public services, too, is taken to be different in kind from that of other services in our economy.

But reform in these areas that recognizes these realities is both essential and possible. We should focus on a number of reforms: unbundling our various integrated welfare programs to allow more targeted piecemeal improvements in government services; permitting greater consumer choice in both education and health care; providing useful, standardized outcome measurements to enable more informed consumer decision-making in all these areas; encouraging new market entrants by loosening regulatory constraints and permitting profits; and, finally, funding demonstration projects for innovative application of technologies and methods that produce measurable gains in stated outcomes-per-unit cost. Realistically, however, the best we can hope for is to make the government, education, and health sectors more market-like. They will never be as efficient as other parts of the economy.

Simply making government, education, and medicine function more like markets does not seem like all that tall an order, but at the moment we are surely moving in the other direction—perhaps in
medicine most of all. It is important to change course. Transforming government, education, and medicine is essential not only because these sectors play such an important part in our economy but also because they are vital to innovation in every other arena. Today, they too often stand in the way of such innovation.

They do so not only by resisting innovation but also by denying it much-needed resources. Ultimately, the money for visionary infrastructure and technology investments will have to come from somewhere. Spending on the core welfare-state programs of health care, education, pensions, and unemployment insurance dwarfs spending on innovation programs. Even small improvements resulting from a modernized welfare system and a more competitive health-care system could free up the resources necessary to invest in our future without increasing taxes or exacerbating our budget problems. Our main limitation in pursuing an aggressive innovation program is a deficit of vision and confidence, not a lack of money.
Energy Policy for American Leadership in the 21st Century

Mark P. Mills
Senior Fellow, Manhattan Institute

There are three broad and immutable trends that should anchor American energy policies in the realities of our 21st-century circumstances. First, the world’s populations and economies are getting bigger and more interdependent. Increasing world trade increases petroleum’s geopolitical importance because oil supplies 95% of the energy used to move goods and people. Second, urbanization and ever-deepening societal dependence on digital systems are accelerating. These two trends increase the importance of electric-grid reliability and security in the face of natural disruptions and rising threats of both physical and especially cyber-attacks. Third, the astronomical scale of prospective global demand for all kinds of goods and services will create unprecedented stresses on land use and environmental conditions, calling for radical advances in basic sciences and not just incremental improvements in existing technologies.

Even the energy forecasts with the most aggressive expectations for rapid alternative energy growth see hydrocarbon use rising substantially and continuing to dominate overall supply. Thus, over-arching policy frameworks to meaningfully address these core future realities should do three things.

First, they should re-orient oil and natural-gas policies to capitalize on the economic and geopolitical opportunities from facilitating an American shale 2.0 revolution. The U.S. should develop policies—and mount dedicated trade missions—to support and accelerate the export infrastructures and abilities for our thousands of small and mid-sized oil and gas companies to compete in a low-price world. Niall Ferguson, Harvard professor and historian, has observed: “There are deleterious
Policy Reforms to Advance Innovation Policy

consequences if the leading power in the world abdicates its leadership role.”¹ America now has a generational opportunity to take a leadership rather than subservient role in a key aspect of global geopolitical stability.

Second, they ought to re-focus electricity policy around the primacy of security and reliability to reduce exposure to physical and cyber threats. Cyber-attacks across all sectors have been growing at a 60%-per-year rate; there has been a similarly rapid rise in attacks targeting America’s electrical and physical infrastructures. A physical cyber-security framework is needed, which moves at the speed of innovators and not of bureaucrats. Vinton Cerf, Google’s VP and Chief Internet Evangelist, recently and correctly noted: “[A]s dependent as we are on communication technology, we’re even more dependent on electricity.”² Information technology can be used to help secure the grid; it should not be (even if advertently) used to make the grid more vulnerable.

Third, they should restructure federal research funding towards a focus on basic science — new “miraculous” technologies won’t emerge from subsidies or corporate welfare where the Department of Energy spends money on industrial-class projects best left to industry. Federal support for basic science is waning not growing; that should be reversed. Bill Gates recently called for a tripling of basic research at DOE to pursue the long-term breakthroughs needed: “[W]e need innovation that gives us energy that’s cheaper than today’s hydrocarbon energy, that has zero CO₂ emissions, and that’s as reliable as today’s overall energy system. And when you put all those requirements together, we need an energy miracle.”³ Radical transformations in technology are hard to predict and impossible to order-up, but they inexorably emerge from a healthy curiosity-driven basic science enterprise.

THE MORE THINGS CHANGE . . .

With an energy-costly war behind us, a reevaluation of the world’s and the nation’s energy resources was timely…with the probability of long-continued international strain — or worse — and fiercer international competition for energy.

—From the April 12, 1948 meeting of the American Petroleum Institute⁴

Energy is fundamental to the operation of everything in society from propelling vehicles to powering the internet, from growing grain to growing silicon ingots, to illuminating TV screens and rooms. While energy policies over the stretch of modern history have been driven by geopolitical,
economic, and social interests, all possibilities are ultimately bounded by the physics and, derivatively, engineering realities of energy.

The fundamental energy sources available to power society have remained essentially unchanged for 75 years. The idea, often articulated now, that there are “a multitude” of new energy options to satisfy society’s needs is rhetorical hyperbole. There is no new physics in energy. And there are no new energy sources, just better ways to use those that exist.

The newest addition to the phenomenology of energy production dates to the 1954 invention of the solar-electric cell at Bell Labs, followed by the first megawatt-scale PV station built in 1982 in Hisperia, California. Nuclear fission was demonstrated in 1939 and the first power plant completed in 1957 at Shippingport, Pennsylvania. Oil and natural gas reach back a century and a half, coal’s history is storied, and water and windmills as sources of energy date back to the Middle Ages (indeed in some respects to pre-history), with the first megawatt-scale wind turbine built in 1941 in Vermont.

The most remarkable new, and unpredicted, change in the energy landscape has been the rise of shale technology. Oil and gas production from shale fields has added 400% more to the U.S. energy supply in the past decade than solar and wind combined. And that rapid and unsubsidized growth in shale hydrocarbon production contributed more than $1 trillion to the U.S. GDP since the end of the Great Recession and thus played a disproportionate role in keeping America from sliding back into negative GDP territory during the long recovery.

In economists’ terms, shale technology has been a beacon of success by achieving astonishing productivity growth in an economy where that key growth metric has been otherwise lagging for years. Federal policy played only a minor role in the shale revolution by providing some valuable (though relatively modest) R&D funding to shale pioneer George Mitchell’s company in the late 1990s. But it was a role that nonetheless offers lessons relevant to framing future energy policy.

Presidents are remembered for many things, and energy-related policies, while frequently important, rarely dominant presidential histories, except at pivots in history that are usually driven by geopolitical events. In that regard, the two iconic energy-related historical pivots thus far were President Eisenhower’s “beating swords into plowshares” with his “atoms for peace” following World War II and the discovery of nuclear fission, and President Nixon’s emergency measures (amongst them the ignominious
national imposition of the wildly unpopular 55 mph speed limit) in reaction to the epoch-setting 1973-74 Arab oil embargo that created gas lines and shot petroleum prices up 300% nearly overnight. No other events or policies of similar moment have occurred in modern history.

It is possible that another energy-related pivot in history could take place during the term of the next president if, for example, policies were enacted to facilitate the return of America as a geopolitical petroleum power. Or, on the other side of the equation, we could see an historical pivot if polices are enacted that increase the prospect for and result in a bad actor’s successful cyber-attack on an American city’s power grid. Either of these potential events entail significant geopolitical fallout.

But if the next president wants to shape events more than be shaped by them, a few key facets of energy innovation policy must be kept in mind.

THE WORLD AS IT IS

It is nearly impossible today to discuss energy policy without confronting the global-warming issue. Consequently, essentially all energy-policy proposals and debates can now be divided into a basic philosophical difference between two camps. On the one hand there are the ideas that seek to deal with the energy world as it is. On the other, there are those aspirational ideas and proposals that seek to reshape the energy world into what it should be to conform to a certain vision.

The energy world, as it is today, can be distilled into these essential facts: 85% of global energy needs are met with hydrocarbons.⁷ Technology has steadily, even radically expanded access to hydrocarbons at ever-lower costs. Over the past two decades 85% of the net additions to global energy supply came from hydrocarbons.⁸ The emergence of American shale oil and gas has been the single biggest change on the global energy landscape in decades.

Then there is the aspirational energy world — a world as it “should be” rather than as it is today — one that needs to become utterly free of hydrocarbon use. The “aspirational” worldview is animated by an expectation that adding more carbon dioxide to the atmosphere from burning hydrocarbons will cause not just somewhat undesirable but potentially catastrophic changes to the planet’s climate.⁹ An energy future devoid of hydrocarbons must overcome the following: All renewable energy sources collectively comprised just over 10% of net additions to global energy supply in the past two decades.¹⁰ And this comes after
two decades and roughly $1 trillion dollars of global subsidies so far.¹¹ Serious analysts, as opposed to aspirational advocates, have universally concluded that there are as yet no viable means to completely replace hydrocarbons at the cost and scale society needs.¹²

The two worldviews are not incompatible in theory. The first is the reality of today, where the second is aspirational. The key issue is the timeline. Energy policies potentially compatible with both worldviews are thus at odds regarding priorities. Conflicts and costs arise when policies seek to radically expand or accelerate our pursuit of the hydrocarbon-free aspirational worldview and ignore economic and derivatively social and geopolitical realities. In essence, the disputes distill into to whether policies establish as a priority the economic, employment, and social benefits from cheap energy, or force on society far higher known energy costs today in order to minimize putative theoretical costs that may arise in the future from carbon-dioxide-induced climate changes.

Reconciling the two worldviews is made difficult, if not impossible, by hyperbolic rhetoric and claims of an imminent or inevitable global apocalypse. The framing of energy policy is thus reduced to trading the actual welfare of people today (eliminating cheap energy) for the theoretical welfare of people in the far future. The only way this conflict could be mooted would be if non-hydrocarbon energy were in fact cheap, which it most certainly is not.

In a framing typical of the energy aspirational worldview, leading environmentalist Bill McKibben writes that companies and policies that support hydrocarbons are “helping push the planet over the edge and into the biggest crisis in the entire span of the human story.”¹³ Accepting such a proposition doesn’t leave much room for debate never mind compromise with regard to taking actions that could avoid such a calamity.

McKibben is far from alone in using such apocalyptic rhetoric, though he is one of the more articulate and effective campaigners for the proposition of completely abandoning hydrocarbons. News stories, studies, and proposals commonly use language invoking the “health of the planet and the survival of its natural systems.”¹⁴ One campaign pursuing oil companies as “climate culprits” claims policies and businesses that support hydrocarbons have “pushed humanity (and all creation) toward climate chaos and grave harm.”¹⁵

The apocalyptic thesis has been gaining far more visibility in recent years, though it has been around for decades. Twenty-five years ago, the
then-secretary of state of Brazil said: “The specter of global warming unites humanity in a common task.”¹⁶ In the same vein, U.N. Secretary General Ban Ki-moon called the 2015 Paris climate agreement “a monumental success for the planet and its people.”¹⁷ A poll of global business and political leaders attending the 2016 Davos confab found climate change ranked as the number one concern for the first time, ahead of regional wars, weapons of mass destruction, pandemics, and water shortages.¹⁸ It’s notable that this ranking stands in stark contrast to general public opinion, which ranks climate at the bottom of a long list of concerns.¹⁹

For a significant proportion of both U.S. and foreign policymakers and policy influencers, global warming does in fact take precedence over nearly any other consideration. Consequently, we are told that climate solutions will require “new supra- and transnational institutions,” that there is an urgent need to “transform world economies,”²⁰ for mandates “for global governance of energy,” and for policies “limiting final energy demand.”²¹ Indeed, for some, a climate-centric energy policy requires challenging the very nature of the American government: “We need a new conversation about the appropriate role of government” and the “weakness of the original Articles of Confederation, in the structure of the U.S. Constitution.” Advocates of the apocalypse thesis believe there is an “unreasonable reliance on free markets.”²² Recognizing the implications of such an assertion, they are quick to assert: “We have to reject the canard that addressing climate change threatens our liberty.”

In a climate-at-all-costs worldview, energy policy will leave liberties in place, but they will just be constrained by very different and far more limited energy choices and (much) higher energy costs. “For the climate accord to work, governments must resist the lure of cheap fossil fuels in favor of policies that encourage and, in many cases, require the use of zero-carbon energy sources. But those policies can be expensive.”²³

A new formulation for energy policy has thus emerged and is gaining traction. The proposition, in a nutshell, is that theoretical calculations of putative future costs relating to theoretical future climate consequences trumps all other cost considerations in meeting today’s energy needs. Period. The argument is that the key to the very survival of civilization requires that society completely and rapidly replace anything that resembles traditional energy policy, and even traditional structures of governance and national sovereignty.

The climate forecasts, of course, are all based on computer projections
associated with the indisputable fact that carbon dioxide is being added to the atmosphere from burning the hydrocarbons that supply 85% of the world’s, and America’s, energy.

It bears noting that carbon dioxide occupies a unique place in the pantheon of “pollutants” since it is an essential nutrient for all flora on earth without which there would be no life. The climate-science debate is not about this irrefutable fact. A higher carbon-dioxide concentration, for example, enhances plant growth and is used as a technique in commercial greenhouses. (A reality many cannabis growers have eagerly embraced.)²⁴ And there is no dispute over the fact that mankind’s emissions of carbon dioxide must be evaluated, keeping in mind that the earth’s natural annual flux of carbon dioxide is 20-fold greater than civilization’s contribution.²⁵ The debate is about whether or not humans are creating some critical and catastrophic concentration that constitutes a tipping point in an ostensibly delicate and perfectly balanced planetary system.²⁶

The climate apocalyptics are nothing if not breathtakingly ambitious. But we find that actor Leonardo DiCaprio may have most succinctly and accurately summarized the prospects for the kind of transformations proposed. In a recent interview about climate change, DiCaprio concluded: “Are we going to come together as a world community? Are we going to evolve as a species and actually combat this issue? The human race has never done anything like that in the history of civilization.”²⁷

One needs no knowledge about or position on the veracity of extreme claims about the future climate to know that there are two things the human race has never done, ever, in history: Come together as a world community, whether voluntarily or by coercion, for any reason, much less to form the kind of world governance imagined by the climate apocalyptics; or allow a sustained trajectory toward deliberately more expensive energy.

Thus, insofar as U.S. energy policy is concerned, the issues that matter can be reduced to one central question: Given that world governance over energy, or anything else, will not happen, and that there is no magic wand to make non-hydrocarbons radically cheaper any time soon, what policies make sense in the energy world as it is?

**Scale and Demand**

Religious zeal and moral certainty have frequently plagued American politics. To be sure, such fervor helped sustain the civil-rights movement itself in its darkest hours. But since the 1960s, these tendencies, and the
rigidity and inflexibility associated with them, have become pervasive and institutionalized. And in our decades-long debate over immigration, these political dynamics have encouraged immigration advocates to not take their opponents seriously, indeed to cavalierly dismiss them. Physics dictates that energy is essential to everything that makes life and society possible. This means, a priori, that energy policies reach into and affect every aspect of society. And, unlike climate models and forecasts, the physics of energy and the engineering economics of energy systems are both clear and dispositive.

The energy equivalent of ten gallons of oil are needed, for example, to fly or drive one person about 300 miles, or power one lecture hall for one hour, or produce the beef for 15 hamburgers, or deliver 100 GB to a smartphone. The annual consumption of these kinds of activities is measured in the billions and trillions of air and road miles, lecture-hall hours, pounds of food, and gigabytes.

In general, three activities account for nearly all energy consumption: transporting people or goods, consuming or using goods (food, housing), and consuming data. The first is utterly oil-dominated (95%), and the second is largely electric-dependent (65%). The third category—information-communications technology (ICT)—used little to almost no energy until recent history and is now nearly entirely electric-dependent.

The ICT ecosystem is now not only a major driver of economic growth, but it has also become a significant energy-consuming sector in its own right. Global ICT activities today use more energy than global aviation.²⁸ In different terms, since ICT energy is almost exclusively consumed as kilowatt-hours, global ICT now uses more electricity than that produced by the entire grids of Japan and Germany combined. This calculus excludes, by the way, the energy cost related to manufacturing info-tech products.²⁹ Digital-centric products require roughly 1,000 times more energy per kilogram to manufacture than the materials that dominated the 19th- and 20th-century economies, and the world produces tech products by the megatons per year.³⁰

Because of energy’s profound importance, energy policy has traditionally focused first on issues of security, reliability, and cost. For nearly all of human history, obtaining adequate fuel and food consumed most of a nation’s or a family’s income, a condition still true for much of humanity today. In the U.S., however, the cost and availability of energy is no longer a primary economic worry. This has been a monumental
technological achievement. Even at $100 per barrel oil (an episodic price that has never lasted for long), spending on all forms of energy combined accounts for less than 10% of America’s GDP.

But while energy costs have receded into the economic background, the absolute consumption of energy has not. The relationship between economic growth and rising energy use is long-standing with no evidence of a de-linking at the global level.31 As economies and populations continue to grow, now accelerated by the efficiencies from the still-widening ICT revolution, global demand for energy will rise. Most of the people in the world today live in energy poverty; billions have no car, little or no electric illumination, no internet connection, and no air conditioning. The potential world electric demand to run residential air conditioning is 45 times greater than that used for the same purpose in the U.S.32 And both global air-miles and road-miles are forecast to more than double in the next two decades.

Thus it’s unsurprising that every respected forecast reaches roughly the same conclusion: World energy demand over the coming two decades will rise by the equivalent of adding another United States worth of consumption. Even though energy demand is expected to grow far more slowly in the mature economies of the U.S. and Europe, such growth is on top of a base of already enormous consumption. Thus there will inevitably be increasing competition for, or opportunities to supply, more energy to meet rising global needs.

From the “world as it is” perspective, the policy challenges necessarily involve how to ensure a reliable and low-cost supply of energy to keep lights and computers on, cars and planes moving, and factories and homes humming. For the aspirational “world as it should be” posture, the focus is first on how to use fewer hydrocarbons everywhere.

The latter worldview was for years reinforced by the Malthusian conviction that oil and gas were severely limited resources and that coal simply was too old and too dirty to use, regardless of its abundance. The fact that modern clean-coal technology can make the extraction and combustion of coal far more environmentally acceptable for all of the relevant and regulated effects considered in policies of the past half-century is in the aspirational worldview now obviated by concerns over carbon-dioxide emissions. In that worldview, even low-cost clean coal is unacceptable.

Low-cost oil and natural gas are also unacceptable in the aspirational worldview. Now that oil and gas are obviously abundant, the aspirational
energy worldview has adopted a keep-it-in-the-ground posture. And, since the aspirational worldview has had to surrender its idea that alternative energy forms can become quickly cheaper than hydrocarbons, in order to convince (or force) society to avoid using abundant, low-cost hydrocarbons, policy positions now focus on a combination of continued or expanding subsidies together with the imposition of new taxes on hydrocarbons.

The carbon-tax argument has been successful at gaining some bipartisan support, in part as a way to thread the needle to avoid the global-warming debate, and in some cases as a way to engage a political trade for reducing middle-class income taxes. The challenges with a carbon tax reside in economic reality: Because 85% of all U.S. and global energy supply comes from hydrocarbons, by definition a carbon tax will be highly regressive, affecting lower-income citizens the most (in relative terms). The tax code would have to be used to offset “unfairness”—the effect of which will mean little if any reduction in hydrocarbon energy demand, since unwinding the regressivity will just keep energy cheap for the segment of the population where the potential to influence demand is most significant. But if it were implemented in a “progressive” way, the carbon-tax burden would fall mainly on the wealthy where energy-use behavior is far less sensitive to price because energy is such a small share of spending.

Subsidies in one form or another are the preferred virtual taxation method for overcoming the economic disadvantages of non-hydrocarbon energy. America’s wind and solar tax credits and subsidies were extended at the end of 2015 (in a political exchange for the elimination of the ban on exporting American crude). Subsidies of course have significant costs, and at sufficient scale lead to overall higher costs for taxpayers. And, like carbon taxes, they are highly regressive. A recent analysis at U.C. Berkeley found that the aggregate $18 billion in U.S. subsidies provided between 2006 and 2014 for various alternative energy choices (electric cars, rooftop solar, and the like) have accrued to the benefit of the wealthy over the working class; the top income quintile received 60% of overall subsidies and 90% of electric-vehicle subsidies.³³ In Britain, Germany, and Spain, where subsidies and credits for wind and solar have been more extensive and aggressive than in America, electricity costs have increased over 100% in recent years.³⁴ Energy subsidies are now under attack or in retreat in many European nations.³⁵

The keep-it-in-the-ground movement has recently sought to promote
the idea that today’s hydrocarbon energy costs must include an “externality” fee or tax that reflects a future estimated cost arising from the theoretical consequences of more carbon dioxide in the atmosphere. This “externality” argument is not a new concept. But in this case it is just another tax with a new name.

The policy debates around subsidies and taxes will continue. However, from the broad perspective of forging national energy policies, the core issue is whether it is possible, at any cost, for alternative energy technologies to radically reduce, never mind eliminate, global hydrocarbon fuel use. It is here that physics and engineering establish clear boundaries.

THE CHALLENGES OF ALTERNATIVES TO HYDROCARBONS

Those advocating the aspirational worldview point to what they consider to be promising alternatives to the use of hydrocarbons in transportation and in the generation of electricity. None of them, however, can meaningfully reduce the use of accessible, low-cost hydrocarbons. With transportation, the problem is primarily scale; with electricity, the challenges are far more complicated.

The two main options for displacing petroleum used in transportation are biofuels and batteries. For biofuels, even setting aside the subsidies, the inherently high economic costs, and the environmental impacts (such as prodigious water use), the basic fact remains that with 40% of America’s corn harvest now distilled into ethanol, farmers still supply under 5% of domestic transportation energy (or about 1.5% in global terms).³⁶ Even if every single kernel of American corn along with all Brazilian sugarcane were used to make ethanol for cars, the need for petroleum would not be materially altered.

Meanwhile, although global biodiesel production rose 400% in the past four years, it remains far more expensive than ethanol, never mind petroleum, and displaces less than 0.1% of world oil use.³⁷ Even an unaffordable 100-fold increase in biodiesel use, which no serious analysis forecasts or anticipates, would be strategically irrelevant.

Then there is the promise of electric cars, for which the technology has improved radically in the past decade. There is no doubt electric vehicle (EV) sales are on track to grow substantially as batteries continually get better. Indeed it is likely the U.S. will, in due course, exceed by ten-fold President Obama’s 1 million electric-car goal, (though it won’t
happen until long after the original target date). But even that will displace less than 10% of U.S. petroleum use.

In the carbon-free vision for EVs, the electricity would need to come mainly from windmills, the least expensive non-hydrocarbon alternative (other than conventional hydroelectricity of course). A barrel-equivalent of energy is produced by a wind turbine once per hour, while a shale well (which costs roughly the same to create as a wind turbine) produces an actual barrel of oil every two minutes. And, to be useful for vehicles, one needs storage, and it takes about $400,000 worth of Tesla-type batteries to store a single barrel’s worth of wind electricity.³⁸ Even if batteries become twice as good as they are today — which is not on any production roadmap — that still won’t overcome such enormous economic and physics disadvantages. Such huge disparities can’t be hidden by subsidies for long.

The inherent characteristics of the molecules used to store energy determine what engineers can accomplish for transportation. Pound for pound, the chemicals comprising gasoline store at least 40 times more energy than the best chemicals in batteries. Pounds matter in all transportation, and they are utterly determinative for aviation. Liquid hydrocarbons are not just denser but also are remarkably safe, and easy to store and move. In biofuels and synthetic chemistry, the essential goal is to invent a synthetic, organic, oil-like molecule at the same or lower cost as a “natural” one from hydrocarbons.

The disparities in theory are revealed in practice. Batteries are also consumed (i.e., they have a finite useful life), albeit more slowly than oil. If the battery cost is amortized over its useful life and calculated in terms of the cost per fill-up, then driving 200 miles in an electric car uses about $5 of electricity and about $50 worth of the battery. It’s the inverse for a gasoline car, in which 200 miles uses $15 of gasoline but $0.25 of the amortized cost of the steel fuel tank. This does not obviate the potential for significant applications for EVs in dense urban markets; it just means that even batteries twice as good as those of today are not going to displace a major share of transportation fuel for the foreseeable future.

It is somewhat more difficult to summarize the challenge with displacing hydrocarbons in the production of electricity because of the idiosyncratic physics of electricity. Unlike the transportation sector where there are precious few options at scale, there are many ways to make electricity. The key issues around the grid, while ultimately about
costs, originate with a critical singular fact about electricity: nearly all kilowatt-hours consumed need to be generated at the instant they’re used. This is not the case for any other energy form (nor for practically any other product in our economy).

On average, there are months’ worth of annual national demand in storage at any given moment for every key commodity from oil and natural gas to grains and metals. Electricity is the exception. The total amount of electricity stored at any given moment in all the batteries in the market for all purposes is countable in minutes — there is not even a day’s worth of national demand in storage.

The technical and economic issues around the electric grid are and will be vigorously debated and now litigated (shortly, at the Supreme Court) because of the Obama administration’s 1,560-page Clean Power Plan (CPP).

The CPP lays out a roadmap, consistent with the aspirational worldview, which will require national electric supply to radically reduce the use of hydrocarbons and increase the use of wind and solar power. The costs arising from the CPP will of course be relevant to state decisions; national average electric rates have been rising since 2005, reversing an earlier 25-year trend of declining rates.³⁹ But cost aside, the central practical challenge for wind and solar is the self-evident fact that neither can produce energy continuously.

There are two critical technical aspects to the episodic nature of wind and solar: capacity factor and availability. To use automobiles as a stand-in for electricity, “capacity factor” measures how often a car engine operates. “Availability” is the probability that the car will actually start when needed. Both features have practical and economic relevance for electric grids.

Capacity factors determine the inherent economics of a power plant. Wind turbines have low capacity factors compared to conventional power plants: a megawatt wind turbine delivers one-third as much energy as a megawatt gas turbine.⁴⁰ Solar electric facilities are similar. Simplistically, three wind or solar megawatts of capacity are needed to equal the energy produced by one megawatt of gas turbine capacity. (The exact ratio depends on the wind or sun conditions at a specific locale.) This means that it is entirely inaccurate to claim a solar or wind plant with a capital cost per “nameplate” megawatt equal to a conventional power plant has achieved what is termed “grid parity.” Grid parity is achieved only when both capacity and availability are taken into account. And even if you build extra wind and solar capacity, that extra capacity is worthless if it’s not available when it’s needed.
In order for the grid to deliver power continuously and nearly instantaneously following normal daily and hourly demand cycles, and in the face of inevitable challenges (plant failures, weather, and the like), operators must have access to unused capacity that is available to be called upon—or “dispatched,” in utility jargon—any time. Wind and solar have not only low average availability compared to conventional power plants but, more important, zero availability for many hours at a time every day—periods when, for reasons of weather or time of day, wind and solar can’t be dispatched. (There are other important technical issues to consider as well, including those relating to maintaining grid stability.)

Today, 90% of America’s power comes from highly available sources: A total of 65%, roughly equally split, comes from coal and natural gas, 20% from ageing nukes, and 5% from big, old hydroelectric dams.

It is availability that matters when it comes to the engineering challenges, and derivatively the economic challenges, of keeping a grid continuously operating and stable. To return to the automobile analogy: If an individual car is not available, one is forced to borrow, rent, or catch a ride in another that is available. That’s how solar and wind successfully operate on the grid today. For example, in Iowa, the nation’s second biggest wind-producing state, coal still produces 50% of electricity.

The standard answer to managing the wind and solar availability challenge is to propose greater use of batteries, transmission, and the internet (the latter to control demand).

Lithium batteries in particular have become far better and cheaper and are widely touted as the solution to storing grid-scale amounts of energy. But to illustrate the challenge of storing electricity at the grid-level, consider the enormous $5 billion Tesla battery factory under construction in Nevada—the so-called “gigafactory.” Once completed (if completed), it is slated to produce more than all of the world’s existing lithium-battery factories combined—a quantity of batteries each year that can store 30 billion watt-hours of electricity. The U.S. economy uses about 10,000 billion watt-hours every day. Thus, it would take more than 100 years of production from the gigafactory to produce enough batteries to build storage capable of holding one-third of a day’s electric demand for when the wind or sun are not available. This says nothing about the high cost or short lifespan of batteries, which can be counted in years rather than the decades needed for grid-scale power systems.
More transmission is another solution for the low-availability problem for wind and solar power. Since it is (almost) always windy or sunny somewhere on the continent, perhaps we could simply build a big enough grid and enough extra capacity everywhere. But the economics of this option are inescapable: Every extra $2 million-per-mile of long-haul transmission built to offset a low-availability source is a de facto additional cost of that source. This says nothing about the need to spend capital on excess wind and solar capacity so that it is available for long-range sharing, and the collateral reliability challenges for a far longer grid.

The other solution for the episodic nature of supply from wind and solar is to use modern information technology to encourage (or force) matching episodic reductions in demand—so-called Demand Side Management (DSM). Utility DSM programs, dating back over a half-century, have long been used to encourage big electric users to shut down when grid demand peaks or supply is lost for any reason. DSM programs offer discounted “interruptible” rates for large consumers. That the internet now makes DSM easier is useful, but much of the easily harvested industrial savings have long ago been captured. What’s left are consumers with minimal significant uses of interruptible demand, or many business operations that cannot be turned on and off. Data centers are one good example.

The bottom line is that, despite billions already invested in efficiency to stifle demand growth, and despite a devastating recession that did stifle growth, U.S. electric demand today is about 10% greater than in 2001. That seemingly modest rise at the scale of America’s grid represents a demand increase equal to Italy’s entire electric grid. And despite depressed electric growth in recent years (almost certainly a hangover from the Great Recession and continued anemic GDP growth), the EIA forecasts at least another 10% rise in overall U.S. demand by 2030. That increased demand will require adding capacity equal to Germany’s entire current grid.

The challenge for states and state utility regulators will be that the CPP pushes planning toward power sources that will make the grid less reliable at precisely the time when modern society needs greater reliability.

The demand for “always on” power to keep the digital and information-centric economy lit has never been greater. The share of the U.S. GDP associated with information—which is entirely dependent on
electricity—is now three times bigger than the share associated with the oil-dependent transportation sector. Meanwhile, the average incidence of grid outages has been rising at an 8% to 10% annual rate since 1990. And the duration of outages has also been rising by about 14% per year. There is also increasing concern about grid cyberattacks, an entirely new class of risk. Aside from the social and human costs and inconveniences arising from electricity outages, the overall economic costs to the U.S. from outages are estimated at $150 billion a year.

Finally, a brief observation with respect to another proposition offered within the aspirational worldview: Advocates of rooftop solar-battery generation suggest that the “old” utility model is due for “disruption” by distributed generation and “smarter” grids. It has become popular to assert in various ways that America has a 19th-century grid for a 21st-century world.

Yet just over ten years ago a seminal National Academy of Engineering report ranked the invention of the electric grid at the top of a list of the 20 greatest inventions of the 20th century—not just one of the great engineering achievements, but first among them all. The Academy ranked the internet 13th. There have been no changes in technology that would suggest that today’s electric utility model is “obsolete,” despite popular media claims otherwise.

In fact, the “utility” model, with its enormous, well-managed, low-cost central plants supporting distributed users, represents precisely the evolutionary direction for the ICT industry. Distributed, small data centers in businesses, industries, research, and academic institutions are rapidly giving way to far less expensive, more efficient, more powerful, and more massive central computing, connecting to users on a grid of glass and radio frequency “wires.” The ICT community refers to this as the “cloud” architecture—functionally a synonym for the utility model.

The ascendant challenge for both electric and information utilities in the future will be ensuring reliability and security. A future electric grid that is both more expensive and less reliable will be economically destructive and politically toxic.

Moore’s Law and ‘Moonshots’

A common response to all of the aforementioned observations about the limits to disrupting the status quo is to propose that the government launch the energy equivalent of the Manhattan Project or the Apollo
Program—a “moonshot” energy program. But fueling society is not like putting a man on the moon (or building a military weapon). It’s more like putting everybody on earth permanently on the moon. The former was a one-time engineering feat; the latter would take miraculous technology. With time, advances in science do create the equivalent of miracles, but they are not common or predictable, and can’t be summoned on demand.

There is no doubt there is much to be discovered, many new technologies yet to be invented and companies to be created. But this reality is frequently conflated in energy domains with a kind of “irrational exuberance” associated with the engineering prowess of Silicon Valley. To note just one iconic example: Vinod Khosla, one of the prominent and successful pioneers of Silicon Valley venture capital, made billion-dollar venture bets on biofuels, asserting in 2007 that “I have no doubt that 100 percent of our gasoline use can be displaced in the next 25 years.”⁵³ He has since recanted. The world is nowhere close to seeing that goal achieved.

Khosla was not alone then, or now, in making such assertions. Many tech entrepreneurs still believe that “disruptive” energy innovation is imminently achievable. The analogy commonly offered is the “miracle” disruption of the legacy landline phone business with the advent of cell phones—or the disruption of taxis by Uber, or hotels by AirBnB. Oil “disruptors” believe, in effect, that the engine in a Pontiac can follow the same tantalizing technology trajectory as the Pentium in a laptop. The problem is that the physics of information doesn’t translate into the world of energy.

Moore’s Law—which describes the relentless and astonishing gains in computing power—has yielded technologies that do seem miraculous. Today’s smartphones are more powerful than a room-sized IBM mainframe from 30 years ago.

The essence of digital-silicon technology is that more and more information can be stored and transported in ever-smaller ways that, individually, use profoundly less energy. On top of this, software engineers, enabled by increasingly powerful microprocessors, can use clever mathematical codes to parse, slice, and shrink information itself, compressing it without loss of essence. The combination is powerful. Compared to the dawn of computing, today’s information-moving hardware consumes one hundred million times less energy for a logic operation and can store data in a physical space one million times smaller.

But in the world of atoms and aircraft, as opposed to algorithms
and Amazon, the hardware tends to expand not shrink when more is needed, whether it’s more speed or greater carrying capacity. The energy needed to move a ton of people, or heat a ton of steel, emerges from properties of nature with immutable boundaries dictated by laws of gravity, inertia, friction, mass, heat transfer, and the like.

A Moore’s Law type of energy disruption isn’t just unlikely, it can’t happen with the physics we have today. If energy technology could follow a Moore’s Law trajectory, today’s Pontiac engine would produce a thousand-fold more horsepower and collapse to the size of an ant. Engineers can build ant-sized engines, but they produce 100 billion times less power than a Pontiac.

No amount of innovating will cause an aircraft or car engine to disappear into your wallet. Nor will the quantity of fuel needed to power it. And in the physical world there is no analog to compression software, the mathematical trickery that puts more information more efficiently into smaller spaces. Only in science fiction can you digitize, transmit, and then re-assemble physical objects or humans.

Such physical realities do not mean that Silicon Valley and information technology have no potential to make dramatic impacts on the energy landscape. But the changes and improvements will come from new materials, some that have never existed in nature (designed by supercomputers) and have radically superior control systems (the marriage of sensors, networks, and algorithms). There is enormous potential to wring far more efficiency out of physical and energy resources. But all these gains will accrue to all energy sources—including and especially to those that have inherent physical advantages.

For example, shale technology has provided over 100 times more energy supply to America in the past decade than has solar.⁵⁴ Measured in terms of energy produced per unit of capital spent, solar technology is about 300% cheaper now than it was 15 years ago. By the same measure, shale gas and oil rigs improved 300% in five years. Both will continue to improve. But the former (solar) has no known path to totally displace the latter (hydrocarbons).

Google engineers reached similar conclusions. Six years after launching a project to develop renewable energy that would be cheaper than coal (titled “R<C”), Google closed it down. The lead engineers made it clear their task was not physically possible: “Incremental improvements to existing [energy] technologies aren’t enough; we need something truly
disruptive.... We don’t have the answers. Those technologies haven’t been invented yet.”⁵⁵ Even the most climate-policy-centric forecasts for accelerating and subsidizing renewables see far more hydrocarbons consumed in all future scenarios.⁵⁶ All of this is consistent with the position that Bill Gates has recently articulated in a high-profile set of interviews, lectures, and meetings. Gates concluded: “[W]e need innovation that gives us energy that’s cheaper than today’s hydrocarbon energy, that has zero CO₂ emissions, and that’s as reliable as today’s overall energy system. And when you put all those requirements together, we need an energy miracle.”⁵⁷ Gates went on to clarify that he didn’t view energy “miracles” as impossible, but that the options don’t yet exist, and thus the single most important policy action is for a radical increase in support for basic scientific research.

**Geopolitical Realities**

The realities of what is possible in domestic energy policy have foreign policy and geopolitical implications.

Oil’s centrality to global commerce is the reason that the 1973 Arab oil embargo—the geopolitical event that framed 40 years of American energy policy—shocked both the United States and the world. Following that embargo, the U.S. became increasingly import-dependent, and America lost the geopolitical petroleum power that it had enjoyed for the previous half-century. It is understandable that for the past four decades U.S. policy has been fixated on achieving “energy independence” through conservation and the pursuit of petroleum alternatives. But the world has changed. America is now far less import-dependent, while the world is now far more oil-dependent.

Consider the key oil-consuming changes since 1973: Global automobile use has increased by 300%; maritime shipping has risen over 300%; and global air travel has grown 700%. And oil fuels about 95% of all the transport of all goods and people. While one-third of the world’s GDP was involved in trade in 1973, that share is now over 60%.⁵⁸ World trade and commerce are thus more oil-dependent than ever before in history.

Until recently, the OPEC nations and Russia were the dominant sources expected to meet rising global oil demand. The entirely unexpected emergence of America’s shale industry not only doubled U.S. oil production, returning it to levels last seen 50 years ago and cutting imports by 60%, but it also accounted for three-fourths of the new global
oil supply over the past decade. Even before the U.S. begins exporting crude and natural gas in quantity, American production already rocked markets, triggering price collapses for both fuels because a rapid decrease in imports glutted the markets—and the prospects for imminent exports from the U.S. triggered renegotiations of long-term contracts from traditional exporters. Next begins the re-emergence—after a half-century of absence—of the U.S. as a significant exporter of those hydrocarbons.

Shale technology has reversed America’s geopolitical posture as a supplicant state to one with the potential to influence the global hydrocarbon trade. U.S. policymakers and strategists now have the ability to think in terms of restoring “soft” power as a vital option in America’s arsenal, and as an alternative to the costs and risks of over-dependence on “hard” power in domains where energy geopolitics are in play.

This is happening at a critical time. Petroleum and geopolitics are intertwined from the Middle East and Russia to Central and South America. Geopolitical tensions pivot around oil precisely because petroleum, and increasingly natural gas, are so critical. Wishful thinking about the world using fewer hydrocarbons and nonsense phrases such as “addiction to oil” don’t erase the realities, or the opportunities.

**Energy Priorities**

The world will use more energy in the future and will burn more, not less, hydrocarbons regardless of subsidies or policies that aim to persuade countries to do otherwise. This reality is the consequence of laws of physics, economics, and human behavior, claims of an impending climate apocalypse notwithstanding.

Even so, nothing about this reality obviates a growing future role for non-hydrocarbon energy sources. It is possible, though it would be remarkable, for solar and wind technologies to grow from supplying about 2% of America’s energy today to, say, 20% or 30% in the coming decades. Such growth would represent a staggering increase in the scale of wind and solar industries and, assuming it was achieved at close to cost parity with hydrocarbons, would also constitute astonishing profits for investors. But such an unprecedented rise in wind and solar would not obviate the need for low-cost hydrocarbons to supply the other 70% to 80% of energy needs.

Over the decades, the U.S. has developed a vast labyrinth of federal energy policies and programs. Some policies have been important and
effective in achieving strategic, economic, and social goals. But too many policies have emerged that are now duplicative and too often counterproductive. And invariably many policies, even if initially well-structured and well-intentioned, suffer from mission creep or outmoded rationale.

As a first order of business, the next administration should form a task force to undertake a thorough inter-agency and inter-policy review looking for opportunities for consolidation and elimination of energy policies that are counterproductive or have outlived their original purpose. Only then can the administration create bold new policies that seize the opportunities created by technologies that exist while dealing with the realities of the world as it is. In doing so, it will be critical to sort through the inevitable proliferation of issues and objectives that will continue to clutter the inherently broad domain of energy policy.

As the next administration thinks toward such policies, it should frame its priorities in a way that addresses the three central macro trends of the 21st century.

First, the growth in world populations and economies will increase the importance of global trade. And because oil supplies 95% of the energy used to move goods and people in trade, petroleum’s importance will increase in coming years—not just because oil is the largest single traded commodity, but because it is inherently central to commerce and geopolitical stability. Trade in oil and natural gas and derivative chemicals comprises 25% of all global trade in all goods of all kinds. Oil-consuming services—transportation and travel—comprise 50% of all global trade in all services of all kinds. And both of those domains have been the fastest growing aspects of world trade in the past decade.⁵⁹

Second, continued urbanization and ever-deepening societal dependence on ICT and digital systems will not only increase the demand for electricity, but also increase the criticality of grid reliability and security in the face of every-present natural disruptions and rising threats of both physical and especially cyberattacks on grids. Over the coming two decades, EIA forecasts 40% of additions to global electric supply will come from heavily subsidized renewables and 45% from gas and coal (twice as much from the former as the latter), but by then 60% of global kilowatt-hours will still come from burning hydrocarbons.⁶⁰

Third, the astronomical scale of prospective global demand for all kinds of goods and services will create unprecedented increases in stresses on land use and environmental goals. This will make it more
important to find radical, not just incremental, improvements in the energy ecosystem and, ideally, even “miraculous” new energy technologies that can address two issues relating to social “justice”: cheap energy that enables and allows access to goods and services, and a radically smaller environmental footprint for society’s energy requirements.

**Key Policy Actions**

To address these three realities, the next administration should focus its overarching policy frameworks on three key policy actions. First, it should re-orient oil and natural gas policies to capture the economic and geopolitical benefits from stimulating a Shale 2.0 revolution. Second, it should re-focus electricity policy around the primacy of security and reliability to decrease exposure to rising physical and cyber threats. And third, it should restructure federal support for research by increasing the focus on basic science — new “miraculous” technologies are ultimately inevitable but certainly won’t emerge from subsidies or corporate welfare.

None of these recommendations obviates the need for policies that support today’s non-hydrocarbon energy sources, or improvements in efficiency and conservation. Rather, all such “alternative” domains have for all practical purposes become the “conventional” approach to energy policy and enjoy either more than adequate funding or are over-funded in terms of meeting stated goals in meaningful timeframes. Indeed, over the past decade, 80% of all federal energy support has been directed at renewables and efficiency, and the total spending has been more than twice as great in that decade as the cumulative total spending directed at hydrocarbons in the prior two decades.⁶¹ The proposals offered herein are thus an alternative to “business as usual,” and are based on the realities of what the world will look like in the near term, and a realism about what it takes, and how long it takes, to effect transformational changes in the energy landscape.

The three framework directives above each suggest some specific policy actions. First, to harness the benefits for the shale revolution, the next administration should consider how to enhance shale hydrocarbon technology and infrastructure.

Over the past half-dozen years, the United States became the world’s fastest growing oil and natural-gas producer, without incentives, special subsidies, grants, or stimulus. Policies should now focus on taking advantage of this unprecedented, unplanned, and largely unsupported
revolution in shale hydrocarbons. The potential to have an impact on America’s economy and geopolitical posture is unparalleled in modern times. Policymakers no longer need to think in terms of minimizing economic and strategic import dependencies, but instead can focus on maximizing future domestic and geopolitical opportunities from petroleum and natural-gas abundance. The U.S. has a substantial lead over all other nations in unlocking its underlying Saudi-level hydrocarbon resources in domestic shale fields. This advantage has created unprecedented opportunities for trade arrangements with our allies and others to reduce geopolitical dependence on high-risk sources or high-threat transit routes for oil (and natural gas). But there are as yet no organized geopolitical policies or principles designed to take advantage of, rather than simply ride (or tolerate), the shale revolution. At the same time, the technologies that underlie the shale revolution are new and have only just begun to unfold, and the global price war now in play is putting substantial financial stress on an industry that is dominated by small and mid-sized firms.

There are a number of specific federal oil and gas policies a new administration could implement to help the U.S. take advantage of its position as the leader of the shale movement. For one, it should implement a time-out on imposing more regulatory constraints on the tens of thousands of small and mid-sized business that are responsible for the shale revolution and that now collectively produce 75% of America’s oil and natural gas.⁶²

The next administration should also create an interagency review of the state of U.S. seaports and related infrastructure relevant for crude and natural-gas exports in order to identify impediments to and opportunities for expedited expansion. For example, the Louisiana Offshore Oil Port is well positioned for rapid conversion into a major export terminal; built in 1982 to import crude, it is the only U.S. port capable of berthing low-cost supertankers.⁶³ ⁶⁴ (Note: Exports constitute an important but stymied opportunity for U.S. coal producers as well.)

The federal government should work with industry to develop a near-term and long-term plan for trade missions designed to provide our allies and other nations with new, stable, long-term sources of critical oil and natural gas, in order to offset the geopolitical risks associated with many nations’ rising dependence on the Middle East and Russia.

Furthermore, the next administration should facilitate the demonstration and validation of emerging shale technologies — sensors,
advanced materials, analytics, robotics, and control systems—that are key to enabling an expansion of domestic industries at the “new normal” of low-priced oil and natural gas.⁶⁵

In addition to the cost-neutral resetting of priorities within existing budgets, additional funding to support the above proposals can be achieved by freeing up capital inherent in the excess quantity of petroleum now in the Strategic Petroleum Reserve (SPR). The SPR was established in 1985 to ensure that sufficient oil was on hand in the event of “significant disruptions” to U.S. supply (both the domestic supply and specifically imports). Thanks to the productivity of the shale industry, the SPR now holds nearly double the 90 days of imports considered necessary for disruption protection, and holds four times more than needed if imports from Canada are not included in the dependency calculus.⁶⁶ The 2015 Bipartisan Budget Act directed the sale of 100 million barrels, about 12% of the SPR, to free up funds for deficit reduction and SPR maintenance.⁶⁷ At least another 20% to 30% of the excess petroleum in the SPR could be sold (in a measured, strategic fashion) without compromising the strategic utility of the reserve, freeing up billions of dollars to meet the above goals at no cost to taxpayers.

The second major point of action on energy policy for the next administration should be securing the electric grid. Electric power is for modern society the fundamental infrastructure on top of which the rest operates. It enables more than lights and heat; electricity pumps gasoline, water, and sewage, keeps food cool and elevators moving, and powers citizen and emergency communications and the entire internet ecosystem. The electric dependency of every aspect of modern society is hypertrophied in cities.

Physical and cyber threats to the grid are increasing at the same time that reliability and resilience are more critical for a more electric-dependent economy. While there are those who claim that one can’t do much planning in the face of so-called “black swan” events, Stanford University professor and risk expert Elisabeth Paté-Cornell says that “perfect storms” are “lame excuses for bad risk management.”⁶⁸ Threats from cyber terrorists to Mother Nature are within the scope of our imagination. The U.S. Department of Energy has spent less than $150 million over the past decade on cybersecurity, compared to $25 billion on smart grid⁶⁹ programs and over $100 billion funding cleantech.⁷⁰

Thus, the next administration needs a specific plan to guard against threats to the electric grid, and there are a number of electricity policies
the federal government should pursue. For one, it should formulate, in collaboration with industry, a program to create a certification protocol for “leadership” focused on cyber and physical resilience. The program can be modeled on the principles that underpin energy efficiency goals, such as LEED (Leadership in Energy & Environmental Design) certifications.

It should formulate public-private partnerships with Silicon Valley software and cybersecurity firms to determine how to develop next generation Information of Things and grid cybersecurity. Physical cybersecurity has to advance at the speed of entrepreneurs and not a bureaucratic crawl.

The next administration should form an interagency working group to apply cybersecurity lessons learned from the Department of Defense Cyber Command. It is not reasonable to expect private companies to defend themselves from nation-state, or nation-state-sponsored, cyberattacks any more than from physical invasions from the same.

It should also re-examine the Critical Infrastructure Protection (CIP) requirements for the long-haul grid to ensure they fully address the threat of physical threats to the grid (natural threats and terrorism), and the operational reality of grids as they exist. In addition, mechanisms are needed to fund “insurance” for warehousing long-lead-time grid hardware, and, for the longer-term, to fund R&D to develop power-electronic solutions for inter-operability of critical grid equipment.

Finally, the third major action the new administration should take to advance energy policy is to radically increase basic R&D. In the modern era, basic scientific research has been foundational to innovation broadly and thus to economic growth and social progress. But it is exactly this open-ended, basic research that can yield the kinds of fundamental or “miraculous” breakthroughs sought to revolutionize everything from health care and security to energy and the environment.⁷¹

The federal government has long been and continues to be the primary supporter of basic research. While the private sector spends far more on R&D in general, at best about 5% of that spending goes to basic research.⁷² The vast majority of support for basic science comes from federal funding, most of that directed to universities. Over 80% of federal civilian R&D spending is concentrated in four agencies: NIH, DOE, NASA and the NSF.⁷³ But lately, federal agencies are increasingly focusing on applied research—emphasizing near-term problems and projects—competing, in effect, with the private sector which already
spends 400% more on applied R&D. This alarming trend represents a de facto conversion of federal R&D policy into industrial policy, and it drains money away from the opportunity to fund undirected, transformational, basic science.

Thus, if the next administration wants to pursue the aspirational, and by definition long-term, goal of finding radical new energy technologies through R&D, it should focus on a few specific policies. For one, it should reform and give priority to basic research and collaterally reduce spending on all types of energy-related industrial-class projects within Department of Energy R&D budgets. Cutting the latter in half would, on average, double the spending on basic science, with no increase in the overall budget.

It should also increase the spending allocated to basic sciences at DOE, rather than specific technologies, devices, or products. These basic sciences are the domains where the equivalent of the discovery of the photovoltaic cell may emerge, or perhaps a radical new catalysts that could convert gases to liquids. This kind of discovery would have obvious applications for methane (natural gas) or carbon dioxide.

Specifically, the next administration should increase the spending allocated to basic sciences associated with shale hydrocarbons, including geophysics, geology, chemistry, and related analytics. DOE takes credit for having played an early supporting role in the basic research that helped pave the way for America’s shale revolution. But there are many features in the underlying science that remain poorly understood; better science can lead to better technologies. (Less than 8% of the DOE’s energy R&D budget relates to hydrocarbons, the fuel sources that supply 85% of U.S. energy.)

The administration should also adopt the Hughes Medical Research model, wherein support for (most) basic research is directed at talented scientists in basic disciplines, rather than at projects with specific directed outputs.

These actions would create a policy environment that, by taking a realistic, world-as-it-is approach, could bring us closer to a revolutionary energy innovation.
Biopharmaceutical Policy for American Leadership in the 21st Century

*Peter Huber & Paul Howard*

**Senior Fellows, Manhattan Institute**

Since the 1980s, the United States has led the world in medical innovation, and our unmatched skill at developing lifesaving medicines has delivered enormous economic value and health improvements to the American economy and patients. In a 2013 report prepared for the Pharmaceutical Research and Manufacturers of America, Battelle, a research organization, estimates that the biopharmaceutical sector adds more than $789 billion of value into the American economy annually and employs more than 813,000 workers, whose average annual wages exceed $110,000, more than double the U.S. private-sector average. Those salaries generate billions in state and federal tax revenues. Further, biopharmaceuticals are one of the country’s leading exports, grossing $50 billion in 2014.

The industry is also — by far — one of the nation’s most R&D intensive, with global R&D spending equal to about 18% of sales. That investment ($51 billion as of 2014) helps the U.S. maintain its status as home to the world’s most prolific life-sciences industry, claiming 53% of global patents granted to pharmaceutical technology — twice that of our nearest competitor, the European Union (26%), and five times that of third-place Japan (10%).

The U.S.’s position at the forefront of global biomedical innovation is, however, far from permanently assured. Europe claimed that position as recently as the 1980s, but was overtaken by the U.S. in the following three decades. Experts attribute the shift to the greater attractiveness of the U.S. market in a variety of areas, including stronger protections for intellectual property and high levels of funding for basic biomedical research through a competitive grant process operated by the National Institutes of Health (NIH).
But what has undoubtedly made the industry one of the crown jewels of America’s high-tech industries has been a relatively free pricing environment for patented medicines. Premium returns from the sale of new and innovative medicines encourages high levels of R&D investment and a vibrant private venture-capital market that funds numerous small start-up biotech companies—an increasing source of groundbreaking new treatments.⁸⁰

Nonetheless, growing financial pressures associated with spending on health-care entitlements at the state and federal level, and a siloed health-care system that focuses on short-term drug prices, rather than the long-term role that medicines play in improving productivity and reducing other health-care costs, are creating a U.S. political environment that is more skeptical of the benefits the industry provides, and more open to European-style price controls, than at any time in recent memory.

Without public-policy reforms designed to sustain and advance biomedical innovation in the U.S.—especially by reducing the cost, time, and risk associated with bringing new medicines to market while also aligning drug prices with real-world outcomes—America could easily cede its leadership of this vital industry to more nimble competitors in Asia or Europe.⁸¹

**PRICE CONTROLS ARE NOT THE CURE**

The current U.S. drug-pricing regime is certainly not without its flaws. Third-party payment systems can desensitize patients from considering the marginal benefits and costs of new medicines (a criticism equally applicable to medical devices, as well as hospital and physician services). Paying by the pill, rather than the outcome, provides perverse incentives to both manufacturers and insurers to focus on bulk discounting—or, conversely, across-the-board utilization restrictions, such as narrow formularies—that ignore the wide variation in patient responses to medicines. Some patients are thus exposed to potential side effects without the prospect of offsetting benefits, while other patients are denied access to products that produce greater gains than average. The tools are increasingly available for clinicians and patients to customize treatment protocols, but these protocols are discouraged by one-size-fits-all reimbursement schemes.

Of equal concern is the rise of high-deductible health plans, often without offsetting Health Savings Accounts⁸², and narrow or tiered formularies with unified deductibles for hospital, physician, and pharmacy
care that are exposing more patients to higher out-of-pocket costs from medicines used to treat serious chronic ailments.

While most Americans are faced with only modest co-pays, a small number of patients (2%) with serious chronic diseases—cancer, multiple sclerosis, or HIV—account for 30% of all out-of-pocket payments, meaning significant financial pressures are put on the patients who may be least able to bear them. Patients with higher out-of-pocket drug costs (through insurance co-pays, deductibles, and co-insurance) are also much more likely to discontinue drug therapy, leading to higher costs and worse health in the long run.⁸³

As discussed below, aligning insurance payments with patient outcomes—delivering the right treatment, to the right patient, at the right time—in a competitive, consumer-driven market would undoubtedly help lower net costs and lead to better outcomes for patients and payers. It is also likely to promote the path-breaking innovations that the biopharmaceutical industry is uniquely positioned to deliver through precision medicines and diagnostics.

Policymakers routinely decry the fact that drug companies concede larger discounts to public payers outside the U.S., especially in the U.K., Canada, and Europe. However, government intervention in other wealthy nations’ pharmaceutical markets effectively creates monopsony pricing power. Without that intervention, it is likely that other wealthy nations would pay prices closer to those of the U.S. (on a GDP per capita basis). The resulting increase in global industry revenues would incentivize even greater investments in innovation. As the U.S. Department of Commerce noted in a 2004 report:

...[I]n the Organization for Economic Cooperation and Development (OECD) countries studied in this report, governments have relied heavily on government fiat rather than competition to set prices, lowering drug spending through price controls applied to new and old drugs alike. Such controls, when applied to new drugs, reduce company compensation to levels closer to direct production costs, leaving less revenue for R&D. As OECD countries individually seek to reduce spending on drugs through price controls, their collective actions reduce R&D that would provide substantial health benefits to all.⁸⁴ [emphasis added]
Another study, from the RAND Corporation, found that if the U.S. adopted European-style price controls, the result would be significantly less future drug innovation in return for only marginally lower prices today.\textsuperscript{85} So while the E.U. is undoubtedly free riding on U.S. funding of global R\&D, the U.S. could not adopt similar tactics without significantly dampening R\&D efforts—and significantly harming future U.S. patients.

Indeed, pharmaceutical innovation—and its benefits—have become so commonplace that we risk taking them for granted. HIV/AIDS—once a death sentence—is now a manageable chronic illness. Beginning in 1996, Highly Active Antiretroviral Therapy led to an eventual 85\% decline in HIV mortality rates in the U.S., with an estimated 862,000 premature deaths avoided. The newest combination therapy has been shown to prevent infection in high-risk individuals. One study from Truven, a health-care analytics firm, found that better HIV/AIDS treatments available from 1996-2010 produced more than $600 billion in economic value, net of costs.\textsuperscript{86}

While cancer remains the nation’s second leading cause of death, there has been a 23\% decline in death rates since the 1990s, with over two-thirds of cancer patients now surviving at least five years. The pace of decline in cancer deaths has also accelerated in recent years, declining by 15.1\% from 2000 to 2011 compared to 75\% between 1990 and 2000, driven by improved treatment and detection efforts.

With some cancers, we can even begin to speak of effective cures. Ninety percent of women diagnosed with breast cancer can now expect to live at least five years—up from just 75\% in 1980.\textsuperscript{87} Until the approval of Gleevec (imatinib) in 2001 patients with chronic myelogenous leukemia had a five year survival rate of 31\%—after imatinib was approved, it rose to 90\%.\textsuperscript{88} CML patients who respond to imatinib can have similar life expectancy to that of the general population.\textsuperscript{89} Second and third line therapies are also available for patients who don’t respond, or whose cancers become drug resistant.

Mortality rates for cardiovascular disease have fallen by over 50\% since 1980, with much of the gain attributable to better drug treatments for risk factors such as high LDL cholesterol, and blood pressure, and clot-busting drugs designed to reduce future risks for patients who experience a first heart attack.

Hepatitis C, a chronic liver infection that afflicts millions of
Americans and which can eventually cause severe liver scarring and liver cancer or liver failure, has seen its cure rate more than double in just five years—from 40% in 2010, to 95-96% today, thanks to safer and far more powerful antiviral treatments developed by industry.

Ironically, calls for price controls are mounting even as new medicines are having a greater impact on patient prospects for long-term survival and healthier lives. Rather than slowing innovation to a crawl in the hopes of curtailing short-term costs, American policymakers should find ways to lower the costs and risks of drug development, thus accelerating the pace of innovation, while also spreading the costs of new innovations across more lives and longer periods of time.

As we discuss later, the development of biomarker science and its use in developing precision medicine—targeted drugs and protocols for their prescription to precisely selected cohorts of patients—are increasingly allowing companies and researchers to identify and attack the molecular roots of serious and life-threatening ailments, pointing to a future in which we will be able to prevent, delay, or mitigate the impact of life-threatening diseases such as cancer, Alzheimer’s, diabetes, and Parkinson’s—and thus lower the health-care costs associated with prolonged disability and reduced productivity.

Developments like these will have a large positive effect on overall health-care spending because, as the economist Michael Mandel has written, “the single biggest driving force for increased health-care spending in the U.S. is the rising cost of labor, not drugs.” He goes on to note that “the cost of labor amounts to more than 40% of the increase in the total cost of personal health-care spending since 2007, while the cost of prescription-drugs amounts to only 10% of the increase.”

Accelerating the development and adoption of precision medicines and diagnostics that compress serious disability to an ever shorter portion of the human lifespan is the best and most far reaching cost-control strategy Washington could adopt in the health-care sector.

**Drug Development and Excessive Caution**

High-profile safety scandals from Thalidomide to Vioxx and Avandia, have left the U.S. Food and Drug Administration institutionally inclined toward risk aversion. That, in turn, has led the FDA to require longer and larger clinical trials designed to identify rare side effects before a new drug is approved, particularly drugs that are used for primary-care
indications, such as heart disease and diabetes, which often must be taken indefinitely by large patient populations.

But longer and more demanding clinical trials come with real costs to industry, patients, and payers. According to the Tufts Center for the Study of Drug Development, including the cost of capital, it now takes $2.6 billion and approximately 10 years to bring a single new FDA-approved medicine to market. Tufts researchers have also found that, from 2003 to 2011, total procedures per FDA clinical-trial protocol increased by 57%, the investigator site work burden by 64%, eligibility criteria by 48%, and length of trial treatment by 25%.⁹¹

All of these increases make it more difficult, complex, and costly to bring new therapies to patients. Fewer than 12% of medicines that enter Phase I clinical trials (the first phase of human testing for safety required by the FDA) end up being approved. This means that the industry must recoup its costs and profits from a relatively small number of marketed products, often for diseases, such as some cancers or cystic fibrosis, that treat smaller patient populations (at least as compared to previous blockbuster treatments for the primary prevention of heart attacks, i.e., lowering high LDL cholesterol).

Demanding more information and longer trials pre-launch from a relatively small number of approved therapies over small patient populations increases the pricing pressures that payers often decry.

Regulatory costs and barriers sharply limit new entry and market competition, because only a few large pharmaceutical firms have the capital, and regulatory acumen, to navigate ever-expanding FDA evidentiary requirements. In 2011, Michael Rawlins, at the time head of the U.K.’s National Institute for Health and Care Excellence and a frequent critic of industry pricing, noted that the regulatory requirements in both the U.S. and U.K. “[had] increased hugely.”⁹² He pointed out that in the 1990s the median number of patients exposed to a new drug in clinical trials was about 1,500; by 2011, that number had grown to 12,000. “It is a huge increase with not much gain, not much benefit from these increased numbers,” Rawlins noted. “And of course, it puts up the cost of drug development hugely.” He went on to estimate that clinical trials accounted for well over 50% of the cost of new drugs.

International regulators are beginning to recognize that the high costs and obstacles to competition attributable to the regulatory system’s trial protocols can and should be sharply scaled back. The executive director
the European Medicines Agency (EMA), Europe’s counterpart to the 
FDA, noted⁹³ that the new “adaptive pathways initiative” that the EMA is 
developing could reduce “by years” the time it takes to win approval, 
and EMA’s “expectation is that companies will reflect this by reducing 
the price of medicines for the benefit of patients and for the sustainability of 
our healthcare systems.”⁹⁴ In short, if new medicines are allowed to reach 
market faster at lower cost, more firms can compete in the field, leading 
to more pricing competition without reducing incentives to innovate. 

The FDA reforms we propose below are particularly important be- 
cause the regulatory status quo isn’t just less than optimal. Failure to 
develop the science during FDA-mandated drug trials has undesirable 
consequences for patients. Adaptive trials are considerably more effi-
cient—they can achieve statistically robust results when they involve 
fewer patients, ensuring that fewer patients are treated with a drug 
that cannot in fact help them while its side effects may harm them. 
Recognizing that a drug is ineffective earlier also allows researchers and 
patients to shift scarce time and resources towards other, potentially 
more productive treatment strategies.

Smaller adaptive trials can also be shorter than conventional tri- 
als—many years shorter according to at least one estimate. The implied 
lower cost of capital per FDA-approved medicine should allow innova-
tors to embrace more flexible pricing contracts with payers, without 
reducing net profit margins. Of course, faster trials also mean earlier 
patient access to successful new life-saving drugs.

The synergies of molecular-biological science and high-power comput-
ing discussed below are beginning to deliver rapid-cycle innovation in 
the biopharmaceutical industry. Ongoing advances in our understanding 
of human genomics and related disciplines (epigenetics, proteomics, and 
systems biology) are allowing researchers to test promising new drugs in 
patient cohorts identified by molecular profiles (biomarkers) that make 
those patients most likely to respond well to the drug and least likely to 
experience serious side effects. Integrated into clinical trials, these tools 
can accelerate and lower the costs of the drug-approval process and place 
it on a much more solid scientific foundation than is provided by the one-
dimensional statistical correlations traditionally relied on by the FDA.

This approach is already being implemented by oncologists on a 
learn-as-you-go, patient-by-patient basis as oncologists practice truly per-
sonalized precision medicine. Tumor biology is carefully analyzed and
Policy Reforms to Advance Innovation Policy

drugs designed to home in on specific molecular targets are prescribed only to patients who present them. Data gathered from patients is stored in large databases, and sophisticated analytical algorithms then analyze the data and recommend optimum treatments for future patients. This process allows physicians to prescribe off-label treatment regimens when biologically appropriate regardless of how the drug was tested during the FDA approval process.

The targeted drugs involved in this “rapid learning” pharmacology, however, will remain expensive—or not get approved at all—if the FDA’s drug-approval process isn’t changed to accept the full implications of the advent of biomarker-guided drugs that make precision medicine possible. Ensuring that biomarker-guided drug development is a sustainable path for innovators, payers, and patients will also require rethinking other elements of the U.S. health-care system to better align value with reimbursement.

The Precision Medicine Revolution

In our generation, biochemists have acquired the tools to gather reams of molecular data about the rogue human cells and microbes that propel the diseases that kill us. They have also developed a remarkable array of new tools for designing precisely targeted drugs. Advances in structure-based drug design, monoclonal antibodies, and, most recently, gene-editing technologies have given biochemists the tools to design drugs that can modulate specific molecular targets or reprogram immune system T cells and stem cells that protect, repair, spawn, and maintain tissues throughout our bodies.

Using these tools to cure diseases, however, hinges on working out the causal connections between what we can see and control in the lab and the clinically defined disorders that we wish to control in patients.

Recently acquired diagnostic tools have revealed the roots of the safety and efficacy conundrums that often lead regulators—in correctly—to binary, one-size-fits-all regulatory decisions when reviewing medicines. At the molecular level, many seemingly common disorders—such as diabetes or depression, conventionally defined by their clinical symptoms—are in fact clusters of biochemically distinct disorders.

Understanding how to mine this information is the next challenge—one we are already overcoming.

The National Institute of Health’s 1000 Genomes Project reported in
2012 that its study of 14 population groups in Europe, Africa, East Asia, and the Americas had identified 38 million “single nucleotide polymorphisms” (“SNPs”)—single letter variations—in their DNA. Another study, completed a few months earlier, analyzed SNPs in the potential “drug target genes” of 14 thousand individuals thought to be particularly susceptible to heart attacks, strokes, obesity, and other health problems. On average, each subject was found to carry about 14 thousand SNPs, about 12 thousand of which were exceedingly rare. Each subject carried an estimated 300 genes with variants found in less than 0.5% of the population that would probably disrupt a protein’s structure in ways likely to undermine health and affect how the protein would respond to targeted drugs.

To further complicate the picture, some of our diseases—cancers most notably—involve cells that mutate rapidly and thus quickly learn to evade drugs prescribed to treat them. Late-stage cancers mutate so fast that they are rarely beaten by a single drug—“cocktail cures” are required instead. A drug’s performance can also depend on how it is metabolized in the patient’s liver or interacts with molecular bystanders in other organ systems to cause unwanted side effects. As noted above, the molecular chemistry involved in all of these processes can vary significantly across patients.

Precision medicine depends on systematically working out how a complex array of molecular factors can propel a disease and affect its response to targeted drugs. This strategy hinges on developing and analyzing large databases that include molecular and clinical information collected from large and diverse arrays of patients—not one-off drug trials for regulatory approvals.

The development of those databases is already well underway. The director of the Genetic Variation Program in the National Institutes of Health’s National Human Genome Institute recently estimated that there were “about 2,000 separate databases” addressing genetic links to various diseases. The NIH itself has compiled a Cancer Genome Atlas. The NIH is also funding many other studies of genetic variations that affect health, among them a project that pools data supplied by a consortium of genetic researchers from around the world. It is also working directly with ten big drug companies and eight non-profit organizations that focus on specific diseases, to unravel the molecular pathways that lead to Alzheimer’s, Type 2 diabetes, rheumatoid arthritis, and lupus—and to investigate new methods to track a disease’s
progress that could provide early reads on how a drug is affecting it.⁹⁷

The private sector is also deeply involved. Independent researchers and doctors have set up databases of their own in which they pool and analyze molecular and clinical data collected during the treatment of patients with approved drugs. Increasingly, these databases are being analyzed using software designed to recommend drug prescriptions — on label or off — that match the molecular pathway that is propelling the patient’s disorder with the pathway that a drug was designed to modulate. The managers of these systems and services often receive in return information on how things worked out, and the constant feedback steadily improves the quality of future treatment recommendations.

Google and Illumina, the leading supplier of gene-sequencing machines, among others, recognized the converging, synergistic power of the biochemical and digital revolutions some time ago.⁹⁸ And they already have broad access to customers and the tools to collect the data quickly and efficiently — hence their rapidly rising interest in developing huge databases of molecular and clinical information and analytical engines that can unravel the complex causal chains and identify the signaling systems that propel cancers and other diseases.

Given enough data and computing power, modern statistical tools can map out complex causal networks, and assess the importance of key nodes and links. In analyzing genomic databases, they have already demonstrated their ability to deal successfully with “hierarchical” pathways, identifying the relatively small number of genomic variations that play dominant roles — as hubs linked to other, less important, variations — and excluding the many variations that play no role at all. An analysis of this kind, for example, provided what has, until recently, been the standard categorization of breast cancers into four subtypes. A more recent analysis of more data revealed at least ten subtypes.⁹⁹, ¹⁰⁰

But the FDA has made clear that it will almost never approve a new drug on the basis of a pathophysiological demonstration that the drug can shut down or repair a disease-propelling pathway. The FDA asserts — correctly — that a drug’s demonstrated effect on a single, disease-specific molecular pathway often fails to predict its ultimate clinical effect on patient health. But much of the time we already know why, or can find out if we wish to.

And the analysis of disease-causing molecular pathways will never be complete because it cannot preclude the possibility that we have
not yet identified all possible variations in that pathway nor the development of further variations in that pathway. Bruce Johnson—a researcher at Boston’s Dana-Farber Cancer Institute and one of the doctors involved in the original trials of Iressa, a drug developed to target the epidermal growth factor receptor (EGFR) on non-small-cell lung cancer—remarked in 2005, “For us as investigators, at this point, there are at least 20 different mutations in the EGF receptors in human lung cancers, and we don’t know if the same drug works as well for every mutation…which is why we want as many EGFR inhibitor drugs available as possible for testing.”¹⁰¹ And however precisely targeted it may be, a drug’s overall impact will often also depend on how it interacts with other parts of the patient’s body.

In sum, advances in biological science have revealed that the generally accepted symptom-based taxonomy of diseases—still relied on by the FDA in the drug-approval process—is obsolete, and antithetical to the advance of the precision medicine of targeted drugs in the real world of complex patients. As the National Research Council (NRC) put it, we need a “new taxonomy of disease.”¹⁰² We would add that we need a new FDA capable of viewing itself as the curator of that taxonomy, rather than a gatekeeper for drug approvals based on clinical signs and symptoms.

Complex diseases like cancers are among those poorly served by the FDA’s reliance on traditional clinical-trial designs. The National Center for Biotechnology Information has said “cancer research is…poorly served because of the many existing clinical trials from which we currently learn almost nothing.” Instead we should “consider the possibility of linking the efforts of physicians, researchers, and patients in advancing cancer research…. Increasingly, randomized trials will be forced to share the stage with innovative trials that deeply investigate cancer within individuals.”¹⁰³

What we see emerging here is the inevitable and essential response to the advent of the science and technologies of precision medicine. The only way to develop the science that tells us when a drug will perform well or badly when prescribed is to study patients and their responses to drugs in the real world of clinical practice.

Critics of such an approach would counter that it is unethical to use patients as guinea pigs, but conventional clinical-trial protocols already do so—slowly and at enormous cost. Doctors do so again, when they
Policy Reforms to Advance Innovation Policy

don’t prescribe medicines according to a drug’s FDA-approved label. The difference is between entering into the process blindfolded or with our eyes wide open and determined to learn as much as possible at every step of the process. The unethical option is to cling to outdated drug-trial protocols that, when there are no other good treatments available, rob patients of the possibility of truly informed consent.

Learn-as-you-go medicine

Ongoing analysis of how the patient-side chemistry can affect a drugs’ performance should populate a large database that continues to grow in predictive power and relevance as more patients are treated with the drug and treating doctors continue to gather molecular and clinical data from every patient treated.

Clinical-trial protocols that facilitate the development of such databases and complementary analytical tools can, as noted earlier, not only place the drug-approval process on a much more solid scientific foundation, they can continue to be updated using post-market data-collection tools, creating a seamless interface between the clinic and the research laboratories searching for the next molecular scalpel to attack previously unknown disease variants.

Most of this information can’t be obtained until the drug starts getting prescribed to significant numbers of patients — but the first opportunity to frame the right questions to ask begins with the drug-approval clinical trials required by the FDA. Unfortunately, the development of biomarker science does not in fact happen under the agency’s existing, narrowly defined clinical-trial protocols.

This creates a Catch-22. Doctors can’t take the lead in working out how to prescribe a drug to the right patients until the drug has been approved; but the drug won’t perform at its best and get approved until someone works out how to prescribe it to the right patients. For doctors, the one notable exception is their authority to prescribe an already-approved medicine off-label. The high cost of running the FDA’s current trial designs ensures that if the medicine fails the first time around, it is often simply discarded.

In well formulated adaptive trials, by contrast, on-the-fly study of patient-side molecular biomarkers that account for different responses can allow progressively better selection of patients who will respond well. This allows the trial to converge much more quickly and effectively
on the patients who do respond and develops scientific criteria for identifying such patients in the clinic. Non-responders can spur additional opportunities for drug development.

Allowing early access to experimental medicines to expert physicians who specialize in the disease of interest (through a strategy called “conditional approval” or “adaptive licensing”) can allow the development of precision prescription protocols through an iterative process that improves at every step—simultaneously lowering the costs and risks facing developers, while expanding access to promising treatments for patients who have no other good options.

Current clinical trials account for over 50% of estimated drug-development costs. Adaptive trials are considerably more efficient—they can achieve statistically robust results when they involve fewer patients and thus cost less. And by homing in progressively on the information needed to prescribe the drug to the patients who are most likely to respond well, they are also more likely to culminate in the approval of the drug.

An additional advantage of these smaller adaptive trials is that fewer patients are treated with a drug that cannot in fact help them while its side effects may harm them. Yet another, often overlooked advantage of trials that focus from the outset on the molecular etiology of the disease being treated is that they can lead to the enormous economies of drug “repurposing”—using a drug that has been approved to treat one particular disease to treat another, quite different disease. This is quite common in oncology, because the same molecular targets and pathways are often involved in driving two or more types of cancers that develop in different tissues or organs. Oncologists know this and quite often investigate repurposing possibilities by prescribing approved drugs off-label.

If adaptive trials are integrated into clinical treatment conducted in centers—such as the major cancer centers and cancer cooperative groups designated by the National Cancer Institute—that specialize in treating particular diseases, the trial protocols can also be flexible enough to exploit the unmatched expertise of these doctors to investigate such things as dosages and combination, multi-drug therapies, and other aspects of how the new drug is used in environments that will better approximate real-world conditions.

Consciously and strategically blurring the line between experimental and FDA-approved medicines will address the tremendous unmet medical need of the millions of patients who do not respond to currently
available therapies for life-threatening diseases such as cancer, Duchenne muscular dystrophy, and Alzheimer’s.

Adaptive trials integrated into clinical treatment can have a further salutary effect on costs and prices. It is reasonable to charge patients who are receiving treatment even when the patients are also paying in the other currency of providing data that will help deliver higher-caliber precision medicine. As the precision-medicines databases grow and the analytical tools improve, a novel process for setting drug prices based on outcomes can be systematically explored and eventually become the norm for pricing drugs once they reach the market. The precision-medicine databases will steadily improve their ability to predict how much a patient is likely to benefit from the treatment, and that knowledge can be starting point for outcomes-based pricing.

Another, as-yet-ungrasped opportunity is moving new preventive medicines through adaptive trials in post-market settings. Developing preventive therapies through conventional trial protocols is often prohibitively expensive because the trials must continue for as long as it takes the disease to materialize and progress. Trials focused on the molecular etiology of diseases and ongoing analysis of a drug’s ability to disrupt a disease pathway can establish efficacy much more quickly.

Finally, the rise of precision-medicine databases and analytical tools that can tell doctors and patients how best to match a specific drug to a specific disease may well help solve the new drug problem of sticker shock.

Most of the cost of developing a new drug is incurred before the drug comes to market, and must be recovered before applicable patents expire. This means loading the huge front-end costs on early adopters, the first cohort of patients who are treated with the drug. Prices routinely plummet when patents expire and cheap, generic substitutes flood the market. But for doctors to prescribe the generics well, they will need access the precision-medicine database. By imposing a modest fee for access, drug companies or others who have taken charge of assembling and managing the database could spread the up-front costs of drug development over a broader group of patients and thus sharply lower up-front prices.

NEW REGULATORY PARADIGMS ARE A COMPETITIVE ADVANTAGE

America’s international competitors recognize that integrating clinical research with patient care can be done while still maintaining scientific
and statistical rigor. Creating seamless lab-bench-to-bedside protocols can both accelerate patient access to effective therapies and create a more attractive environment for international biotechnology investment.

The United Kingdom has announced plans to dispense with traditional clinical trials by mining genomic information from patients’ electronic health records to identify novel targets for drug development and match patients with tailored therapeutics. It is also devoting £300 million to sequencing the genomes of 100,000 patients with cancer and rare diseases by 2017, and has set up a biobank with samples and clinical histories from 500,000 patients as a resource for academic and medical researchers to identify previously unknown disease pathways and potential biomarkers.

At the same time, Innovate UK, a government agency whose mission is to promote economic growth “by working with companies to de-risk, enable and support innovation,” has created innovation centers called Catapults, each designed to “accelerate and simplify the path from research to commercial products,” including one focused on precision medicine. The precision-medicine Catapult operates with the explicit goal of “making the UK the leading place worldwide to develop and launch new solutions” for precision medicine through the use of cutting-edge diagnostics and Big Data algorithms.

The U.K. government recognizes that close cooperation between stakeholders — patients, academic researchers, innovative drug and diagnostic companies, regulators, and payers — will be necessary to create a rapid adoption of precision-medicine technologies by lowering barriers to product commercialization. The precision-medicine Catapult functions as “trusted neutral party . . . by offering a critical mass of multidisciplinary expertise, infrastructure and services” to companies operating at the cutting edge of science.¹⁰⁴

At the heart the U.K.’s embrace of rapid cycle, patient-focused innovation is database-driven drug development, including the ability to rapidly share knowledge across various health-care providers. In a March 2015 report, the Association for the British Pharmaceutical Industry explained how the approach could work to accelerate innovation while also enhancing pricing flexibility by reducing regulatory costs and risks:

Novel, matched case controlled studies which include real world data of patient relevance can utilise health databases to more quickly identify and recruit subjects, and allow data capture and
analysis in real time. Adaptive designs, with prospective and in-stream stratification, can increase targeting and further personalise medicines development. Time and cost savings are achieved through expedited recruitment, reduced study complexity and use of fewer investigator sites.

As evidence accrues through post-approval continuation of studies, the value proposition will change based on the evidence generated. The price paid for a medicine should thus adapt to account for the value it brings. Ultimately, greater cost-effectiveness and affordability should result. With lowered development costs, a reduced price can maintain profitability, increase development portfolio cost efficiency, and allow the progression of a greater number of promising projects at reduced cost.

The ABPI report calls for the U.K. to seize the opportunity to “set a new [global] regulatory standard and take a lead in enhancing patient care through medicines evaluation and uptake.” Advancing global regulatory standards through database-driven drug development that matches promising medicines to patients in the clinic would enable “wider applicability, including utilization of data generated substantially in the UK at improved speed and cost,” producing a “major incentive for UK life science investment.”¹⁰⁵

Not to be outdone, in 2014 the European Union launched the second phase of its Innovative Medicines Initiative (IMI2), which recognizes that the “availability of the complete sequence of the human genome, the growing body of ‘-omic’ data sets and epigenetic markers in health and disease, the availability of patients’ electronic medical records, next generation genetics for target identification, and sophisticated bioinformatics techniques offer the opportunity to revolutionise the current medicines development process.”¹⁰⁶

Biomarker development is one of the four key priorities identified by IMI2, which will include an effort to “identify and validate biological markers, tools and assays (biochemical, functional and imaging) to support disease reclassification and patient stratification approaches, monitor disease progression, provide proof of pharmacological response, predict and monitor the efficacy and safety of drugs and vaccines, as well as biomarkers that may serve as surrogate markers in clinical trials.” IMI2 is intended to run for 10 years with a budget of €3.276 billion, focusing on a broad range of diseases and drugs,
including antimicrobials, cardiovascular disease, oncology, psychiatric diseases, and autoimmune diseases.

For patients, the advantage of rapidly incorporating biomarkers and surrogate endpoints into clinical treatment are clear: It allows greater access to targeted treatments by patients who are most likely to benefit, moving beyond the artificial confines of randomized controlled trials in which some patients, even when they fit the molecular profile of the intervention, are randomized to receive the standard of care and are denied access to targeted medicines—which can amount to a death sentence. In one 2010 trial of a targeted cancer medicine for metastatic melanoma, two cousins with the disease were randomized, with one receiving the treatment and the other receiving a “notoriously ineffective” chemotherapy. Even after his disease progressed, the patient was not allowed to switch over to the treatment arm of the trial. The patient who received the drug survived, while his cousin died. Some oncologists have called randomization of patients in such circumstances unethical.¹⁰⁷

America’s competitors recognize that embracing the full potential of molecular medicine to transform both drug development and accelerate the adoption of precision-medicine technologies gives them the best opportunity to overtake the U.S. biotech industry by simply modernizing their drug-approval process faster than we do.

A ROADMAP FOR AMERICAN LEADERSHIP

To encourage the development of precision-medical treatments and biomarker-based diagnostics that can revolutionize the health-care system and lower costs for both private and public payers, Congress and the next administration should focus on four key reforms.

The first and most complicated of these reforms is perhaps the most important: They should advance the FDA’s toolkit for approving new medicines based on biomarkers, surrogate endpoints, adaptive clinical trials, and real-world data. As discussed above, the FDA should give greater deference to the external scientific community in developing evidentiary standards for incorporating biomarkers into the drug-development process and embracing adaptive clinical-trial designs. This approach can significantly lower the cost and time needed to bring new products to patients, expanding the number of therapeutic options available while also increasing competition based on price and outcomes. Congress should also direct the FDA to develop a rapid-learning
drug-approval process that would allow a drug that modulates a known disease-promoting pathway to be used, after initial safety testing, in clinical treatment by expert physicians and medical centers that specialize in treating the condition of interest. These researchers would then take responsibility for gathering data and developing treatment protocols for the use of these products and for identifying patients most likely to benefit, with predefined endpoints that would indicate when enough evidence had been collected to allow for full FDA approval.

Frustration with the current clinical-trials system is also reflected in the “Right to Try” movement, which advocates state-based legislation that would allow terminally ill patients to obtain experimental medicines after Phase I trials, if the manufacturer agrees to grant access. Right to Try legislation has passed in 27 states to date and reflects the fact that patients are determined to take more control of their own choices when faced with a terminal illness.

One approach that could both expand access for patients without effective treatment options and retain experimental rigor is a conditional-approval pathway or pilot for oncology medicines. This paradigm would rely on expert oncologists learning to use candidate compounds in targeted cohorts of patients using precision-diagnostic and bioinformatics platforms that help them rapidly match patients to treatments they are likely to respond to, based on the patients’ molecular profiles.

Under this approach, compounds would be given conditional approval after demonstrating significant activity in early-stage trials — after demonstrating (through a variety of pre-clinical and clinical tests) their ability to modulate molecular pathways (biomarkers) or surrogate endpoints that are implicated in tumor growth or proliferation in specific cohorts of patients or disease indications.

The compounds would then be made available through the NCI’s network of comprehensive cancer centers or networks of cancer clinical trials like the cancer cooperative groups (such as the Southwest Oncology Group) or any participant with the bioinformatics platforms (EMRs, decision support tools, standardized high-quality assays) and experience in running sophisticated clinical trials. This would rapidly put promising compounds in the hands of oncologists with the requisite expertise and the most experience in treating patients with these characteristics. This infrastructure would allow them to collect real-world outcomes data in a variety of settings and treatment combinations that can be analyzed
to validate the clinical effects predicted by the biomarkers or surrogates. They could then develop precision prescription protocols.

If the candidate medicines fail to meet pre-specified endpoints (in either combination treatment regimens or as single-arm therapy) FDA will have the authority to expeditiously withdraw them from market, but sponsors will be allowed to continue development through the traditional approval pathway. If medicines meet pre-specified endpoints (based on trial designs accepted jointly by the sponsor, NCI, and FDA), they will be given full approval and permission to be marketed outside the cooperative groups.

While oncology has made the most progress towards embracing a precision-medicine paradigm — and can rapidly provide a “proof of concept” that this strategy is viable — conditional approvals should not be confined to it. Heterogeneity is a biological phenomenon that is seen in most, if not all, complex human diseases.

Conditional approvals would not only slash the time and cost needed to bring new treatment options to patients who have run out of options — they would also generate vital data on how new medicines perform in real-world patients, data often lacking today. Ideally, participation in clinical trials should also become the standard of care for off-label treatments, to spur the development of large oncology-patient registries and seamless integration of patients into Phase I studies.

One paradigm for this type of approach is the recently announced PrECISE international consortium for prostate cancer, called the Project to Construct Computational Models to Improve Prostate Cancer Treatment, Care. The members of this consortium include IBM Research, Technikon, Technical University of Darmstadt, Aachen University Hospital, ETH Zurich, University of Zurich, Baylor College of Medicine, Curie Institute, and AstridBio Technologies. The aim of the consortium “is to develop different algorithms that allow us to understand tumor heterogeneity, understand better why drugs work and don’t work, and come up with more effective therapies [and] in particular combination therapies.”¹⁰⁹

Consortium members will also “develop computational approaches that integrate genomic, epigenetic, transcriptomic, proteomic, and clinical information,” including data from publicly available datasets and published in scientific journals. Members will use the resulting models to “investigate prostate cancer’s molecular mechanisms and to try to
predict new targets for therapy.” By homing in on aggressive prostate-cancer subtypes, the consortium will allow clinicians “to classify patients according to risk, minimizing patients exposure to unnecessary surgery or other treatments, reducing spending as a result.”¹¹⁰

 Consortia like these allow oncologists to rapidly test and validate new treatment approaches across a variety of disease settings and patient cohorts—learning much more about a drug’s performance than is possible under traditional clinical-trial designs that offer binary succeed-or-fail outcomes. This could also “avoid unnecessary replication of either positive or negative experiments…[and] maximize the amount of information obtained from every encounter”¹¹¹ and thus allow every treatment to become “a probe that simultaneously treats the patient and provides an opportunity to validate and refine the models on which the treatment decisions are based.”¹¹²

 The only thing missing from this platform—which Congress could supply—is a conditional-approval pathway matching promising cancer-drug candidates with the patients who are likely to respond in a data-rich environment.

 Experts have been advocating the adoption of this type of approach for nearly a decade. In 2007, a group of health-care experts convened by the Institute of Medicine coined a phrase for it: “Rapid learning health care.” In brief, the workshop participants proposed a process for continuously improving drug science using data collected by doctors in the course of treating patients, with a particular focus on groups of patients not usually included in drug-approval clinical trials.

 Patient access in this environment blurs the line between experimental treatment and FDA approval, but we should also recognize that the high incidence of off-label treatment of cancer has already blurred it substantially. The time has come to make a virtue of necessity and formalize a conditional-approval approach that would grant access to larger cohorts of patients in a structured environment.

 Researchers at MIT, who have done pioneering work on conditional approvals for drugs more generally, write that a conditional-approval pathway linked to post-marketing surveillance could have a “profound effect” on drug development by “allowing smaller development programs to achieve greater success.”

 They estimate that development costs could be reduced by 90% and development time by 50%, “if the threshold for initial approval
were defined in terms of efficacy and fundamental safety.” “Requiring high-quality and transparent patient registries for independent safety monitoring, would be a more informative and cost-effective approach,” compared to traditional strategies.¹¹³ Christopher McKenna, general manager of discovery science at Thomson Reuters, believes that “identifying targets for drug discovery and identifying patients for clinical studies early in the process will reduce drug development cost and cycle times sufficiently” to enable “biopharma portfolios . . . filled with hundreds of drugs that each generate $40 million to $50 million” as opposed to a dozen or so blockbusters that generate $1 billion or more annually.¹¹⁴

Over the long term, the FDA’s approval system should shift from clinical-symptom-based approvals and labeling to molecular-indication-based labeling, with additional data collected in the post-market environment that would progressively improve clinicians’ ability to prescribe drugs with high precision in the safest and most effective manner concomitant with each patient’s molecular profile.

A slow, smooth transition to integrating drug-approval trials with clinical-patient treatment could begin with the recognition that the high incidence of off-label prescription in treating cancer has already substantially blurred the line between experimental treatment and FDA approval. The doctors and medical centers that have already developed and begun to use rapid-learning databases and analytical systems should review their protocols and analytical tools with the FDA. The medical centers and FDA should then cooperate in the development of uniform standards. Then FDA could formalize a conditional-approval approach that would grant access to new drugs to larger cohorts of patients in a structured environment and that would allow drug companies and the FDA to rely on the work of doctors at medical centers to approve off-label uses and amend labels accordingly.

After the development of new drug-approval pathways, there remain three other key reforms that Congress and the next administration should pursue to support the development of precision-medical treatments and biomarker-based diagnostics. The next should be to encourage a new market-based pricing system for innovation that rewards companies for developing new precision treatments and diagnostics.

Government regulations — such as Medicaid’s “best price” provision and FDA restrictions on the communication of off-label prescription information — often prevent innovator companies from entering into “pay
for performance” contracts with insurers and pharmacy benefit managers (PBMs) that would link reimbursement to real-world outcomes based on molecular biomarkers or other diagnostic criteria. The Centers for Medicare & Medicaid Services should create a safe-harbor for such contracts, and the FDA should promulgate guidance that would allow companies to inform physicians and payers of any relevant molecular information or pharmaco-economic data that would allow them develop more personalized prescription protocols. The free-flow of scientifically reliable information among sophisticated payers and purchasers, along with the freedom to experiment with novel value-based reimbursement contracts, would do much to align drug prices with their value given their overall impact on the total cost of care for a given disease state, the patient’s quality of life and risk preferences, or any other factor that innovators, expert physicians, payers, and patients recognize as valuable.

Under the next administration, Washington’s third step should be to reform the U.S. corporate tax system to make it more attractive to investors and innovative firms. An economic barrier to sustaining and expanding U.S.-based innovation is the country’s corporate tax rate, which ranks among the world’s highest. The United States is unique among developed countries, moreover, in taxing the worldwide earnings of its global firms; other countries tax only the earnings from sales within their borders. The U.S. tax on foreign earnings is deferred until the money is repatriated, but that gives life-sciences firms a perverse incentive to keep their profits offshore, rather than use them to fund further investment in the United States. This means that low-tax nations will continue to attract the infrastructure for innovation (labs, manufacturing facilities, and the like) in preference to the U.S., and foreign-based firms will also have greater access to offshore capital in the competition to acquire the most promising U.S.-based companies and their associated technologies and drug pipelines.

Tax reform will be become increasingly important as our competitors in Asia and other emerging economies develop the expertise necessary to compete in innovative R&D projects. Congress should ensure that tax policy attracts investors and companies to our shores — instead of driving them away.

Fourth, the next administration should expand the FDA’s platform for crowd-sourcing new regulatory standards. One of the most persistent problems facing innovators in the 21st century is a regulatory structure
and mindset at the FDA that hearkens back to the mid-20th century, an era defined by mass manufacturing and hierarchical command-and-control structures—the seeming hallmarks of successful corporate and military organizations that dominated war fighting and international economic competition throughout much of the 20th century.

That regulatory model was defined by two realities: first, the mass delivery of drugs and medical devices to homogenous populations defined by clinical symptoms; and second, the extremely high cost and long timelines associated with conducting “gold standard” medical research, meaning the randomized controlled trial.

The advent of distributed high-performance computing, the rapidly falling cost of whole-genome sequencing and novel gene-editing technologies, and access to high-quality public data sets allow researchers to conduct much more nimble and targeted experiments on the fly, answering far more nuanced (and clinically relevant) questions at far less cost.

But the FDA’s system for developing regulations and regulatory guidance for new technologies remains overly centralized and slow moving—and is sometimes outdated by the time it is completed. It can take between 425 days and 797 days to finalize draft FDA guidance, leaving them “languishing in unfinished form for years, even as new scientific developments or broader shifts in policy render them irrelevant.”¹¹⁵

One approach to closing the gap between regulation and innovation would be to crowd-source regulations through a Wiki-like commons where academic researchers (including the NIH and other federal research agencies), industry, regulators, and patient groups could come together to establish performance standards for novel technology platforms, innovative clinical-trial designs, and even advanced manufacturing technologies.

The prototype for this is the FDA’s existing precisionFDA platform, a public-private venture operated by the FDA and DNA Nexus for developing standards for next-generation sequencing platforms. The FDA’s chief informatics officer explains:

PrecisionFDA is an online, cloud-based, portal that will allow scientists from industry, academia, government and other partners to come together to foster innovation and develop the science behind a method of “reading” DNA known as next-generation sequencing (or NGS)….
precisionFDA users will have access to a number of important tools, including reference genomes, such as “Genome in the Bottle,” a reference sample of DNA for validating human genome sequences developed by the National Institute of Standards and Technology. Users will also be able to compare their results to previously validated reference results as well as share their results with other users, track changes and obtain feedback.

Over the coming months we will engage users in improving the usability, openness and transparency of precisionFDA. One way we’ll achieve that is by placing the code for the precisionFDA portal on the world’s largest open source software repository, GitHub, so the community can further enhance precisionFDA’s features.

Platforms for rapid-cycle regulatory innovation are increasingly important as we transition from an era of “one test-one disease paradigm” towards simultaneous scans of a patient’s microbiome, genome, transcriptome, and exome that produce massive amounts of data that could “potentially detect multiple conditions in a single test.” DNA Nexus Chief Medical Officer David Shaywitz writes that precisionFDA represents a “novel and forward thinking approach to regulation” in this data-rich environment:

Rather than envisioning governmental regulators as the folks who will define and then impose a specific set of performance standards, precisionFDA instead sees the government as providing the platform that will enable the NGS community to evolve the standards on their own — organically and transparently.

...the ability to design, refine, and deploy this platform in such a rapid and agile fashion reflects in part the value of well-conceptualized public-private partnerships, in this case between the FDA and DNAnexus. By intentionally leveraging the skills and capabilities of a company like ours, the FDA was able to implement and realize their exciting and ambitious vision.¹¹⁷ [emphasis in the original]

Importantly, FDA staff will also be able to interact much less formally and more flexibly with the members of the precisionFDA community, which include 23andMe, the Baylor College of Medicine, Intel, the
Human Longevity Institute, and the National Institutes for Standards and Technology (NIST)/Genome in a Bottle consortium, to name just a few platform collaborators.

While the precisionFDA effort is ongoing, Congress should scale up these types of virtual platforms for generating timely regulatory standards for innovative technologies by creating a public-private consortium for regulatory innovation with a remit for developing such standards, particularly for regenerative medicine, biomarkers, nanotechnology, and Bayesian trial designs.

This consortium should also have the authority to pilot promising approaches in a rapid-cycle approach in collaboration with industry, NIH, NCI, NIST, and DARPA for developing breakthrough innovations for unmet medical needs including neurological injuries, Alzheimer’s, rare and ultra-rare diseases, and drug-resistant cancers.

A pilot approach would address the FDA’s reluctance to promulgate new standards because of its inability to access the needed expertise internally regarding novel technologies, as well as generate funding needed to pilot these approaches in a rigorous way.

**THE FUTURE OF AMERICAN BIOPHARMACEUTICAL INNOVATION**

U.S. policy should encourage the development of more paradigm-shifting precision medicines and protocols for approving drugs for off-label use faster and more efficiently, and with more detailed guidance on which patients benefit most from their use. A modernized FDA drug-development and approval framework would improve industry productivity by reducing the risks and costs associated with bringing new medicines to market and allowing more precise prescription of targeted drugs. Washington should also reform or eliminate regulations that currently prevent drug companies and payers from aligning drug prices with the value they deliver to patients and, by extension, the entire health-care system through value-based contracts linked to real-world outcomes. Competition between targeted therapies based on their real-world value would also help to address concerns regarding drug pricing, without reducing incentives to innovate. Reforming FDA trial protocols to accelerate the drug-approval process would allow patients with serious and life-threatening diseases to avoid having to wait in excess of a decade for access to better therapies.
The platform for 21st century innovation we’ve outlined looks beyond current drug-pricing controversies and focuses on reforms that would pay dividends for the U.S economy and patients for decades to come. The opportunity remains to be seized, by the U.S.—or by our competitors abroad.
Wireless Telecommunications Policy for American Leadership in the 21st Century

Christopher DeMuth
Distinguished Fellow, Hudson Institute

Within a single generation, human connectedness has become radically more intense, purposive, and far-ranging than ever before. Early forms of desktop computers and mobile cellular telephones were successfully introduced in the United States in the 1980s; smartphones, tablets, mobile broadband internet, and Wi-Fi appeared in the 2000s. Today 90% of adult Americans own a mobile phone (mostly smartphones), 70% own a desktop or mobile personal computer, 45% own a mobile tablet, and 75% of households have high-speed internet (65% with Wi-Fi); and all of those technologies are ubiquitous at work and in public places. With them, we exchange more than 7 billion text and email messages every day, routinely correspond at social-media sites, shop and bank and navigate, conduct business, share documents and photos and videos, report emergencies, access news, sports, entertainment, and personal health information, and through internet search have the entirety of recorded human knowledge at our fingertips and voice command. On average, we are online six hours a day.¹¹⁸

This transformation has profoundly affected personal, family, and social life, business and finance, science and engineering, and politics and government. It has powered innovation in critical fields such as the biological sciences and energy exploration and development. And it is at the heart of many innovations now under development, such as autonomous cars and other vehicles, long-distance medicine and learning, and the “Internet of Things” — the deployment of remote sensors throughout transportation networks, water systems, farms, factories, buildings, hospitals, households, the electric grid, and the atmosphere for purposes of continuous monitoring, coordination, and adjustment.
The communications revolution is, however, being hobbled by outdated government policies. All wireless communications, whether between persons or things, employ channels of the electromagnetic-frequency spectrum, which have long been allocated and regulated by the Federal Communications Commission. FCC policies crafted in the age of radio broadcasting have proven wholly inadequate to the far more intense spectrum demands of universal high-capacity wireless networks. The commission has responded with important policy innovations—since the early 1990s, it has liberalized the terms of certain of its spectrum licenses, which has made way for the construction of our current cellular telephone and broadband infrastructure; conducted spectrum auctions, which have moved unused and underused spectrum into the new wireless economy; and reserved ample “unlicensed spectrum” for short-range uses such as Wi-Fi. Yet regulatory innovation has not kept pace with wireless innovation and is falling progressively further behind.

The time is ripe, and urgent, for the fundamental next step in spectrum liberalization—all-purpose spectrum licenses. Holders of licenses would be permitted to use their frequencies for any purpose. Current license restrictions, which limit each holder to a narrow purpose, would simply be removed. The reform would dramatically extend and accelerate the FCC’s recent reforms. Its social and economic benefits would be immediate and palpable. It is politically feasible and could be accomplished in a stroke.

**BACKGROUND**

“Radio waves” is the conventional term for the portion of the electromagnetic spectrum suitable for transmitting information (with much lower frequencies and longer wavelengths than those of visible light).¹¹⁹ Radio waves are the medium for all wireless communication—radar, broadcast television, garage-door openers, sending photo images from Pluto back to Earth. Almost all uses require a spectrum license from the FCC (not including garage-door openers, an example of “unlicensed spectrum”). License details vary from case to case, but they typically specify the spectrum frequency band and, within it, the bandwidth the licensee may use; the forms of signal modulation and other methods for encoding and transmitting information on the designated spectrum; transmission power (“electromagnetic energy radiated”); the type of transmitter and antenna equipment (sometimes down to a particular brand and model);
the location (by geographic coordinates), ground level, height, radiation pattern, and geographic range of transmission; and (sometimes) hours of operation.

In addition, licenses are limited to specific purposes, such as television and radio broadcasting, mobile telephone and smartphone service, various satellite links (satellite to ground, ground to satellite, satellite to satellite), and a host of narrower purposes such as police radio, maritime navigation, and meteorological satellites. Finally, licenses specify licensees’ organizational and business forms—such as amateur radio, non-profit educational, for-profit corporation, and advertising versus subscription supported.

Spectrum licenses are regularly bought and sold, but the restrictions on a seller’s spectrum continue to apply to the buyer—so that, for example, the license of an AM radio station may be sold only for AM broadcasting by someone else. If you want to transmit a certain kind of information from A to B, but the FCC has already allocated all of the spectrum it has “zoned” for that kind of information in that place, and the incumbent license holders are not interested in selling, you are out of luck.

Different radio frequencies are better suited to different applications, depending on such variables as distance, transmission capacity, power availability, and “propagation properties” (lower frequencies generally transmit information more slowly but travel further and are better at penetrating walls and other objects, but these tendencies are affected by transmission power and other factors). The FCC’s zoning scheme takes account of these technical considerations but is also based on estimates of market demand for various uses. Thus, the commission allocates spectrum between television broadcasting and mobile broadband, meteorological and geostationary positioning satellites, and a host of other competing uses of technically suitable spectrum according to its assessment of the need for each service. It maintains more than 100 “high level service categories.”

This “economic planning” feature of spectrum licensing has proven increasingly problematic with the emergence of many new forms of wireless communication and many new techniques for sharing and combining frequency channels. In recent years, the problems have become severe. The fantastic growth of smartphones, tablets, and laptop computers and the now-routine use of video-streaming, personal navigation, internet “cloud” storage, and other data-intensive applications have far outstripped the FCC’s spectrum allocation for wireless broadband.
At the same time, the growth of cable and satellite television, which now reach the vast majority of households, has left a great deal of spectrum for old-fashioned broadcast television underused or dormant. (The TV-broadcast zone, first established in the early 1950s, still maintains generous allocations for local UHF, or ultra-high frequency, television stations, which are now little used outside some rural areas.) The spectrum designated for broadcast TV is ideal for wireless broadband, yet much of it is lying fallow. In addition, federal agencies have exclusive use of one-third of the most suitable spectrum (administered by the National Telecommunications & Information Administration, part of the Department of Commerce), and much of it is lightly used.¹²³

Recent and current FCC spectrum auctions are designed to alleviate these imbalances by taking unallocated spectrum, underused federal spectrum, and spectrum purchased from television broadcasters and selling it to wireless-broadband suppliers through competitive bidding. The auctions have helped, and indeed have been a landmark improvement over previous schemes of allocation by administrative hearings or lotteries. But, as we shall see, they have been highly complex and slow moving, and are falling increasingly behind the explosive growth in demand for wireless broadband.

Centralized administrative allocation of rights to radio transmission was first conceived by Secretary of Commerce Herbert Hoover in the mid-1920s, in response to the first appearance of commercial radio broadcasting stations, and was then legislated in the Radio Act of 1927 and the Communications Act of 1934. Federal regulation displaced the development, then underway, of property rights in radio spectrum and legal rules to settle conflicts among different users and interference between users of adjacent spectrum channels. Many experts in communications technology and economics believe the regulatory approach was a mistake (the economist R.H. Coase received the Nobel Prize in 1991 in part for demonstrating that this was so).¹²⁴

It was, however, an understandable mistake. Radio was then a strange new phenomenon, useful mainly for public purposes such as broadcasting, maritime navigation, and military communications: It was an invisible frontier that, Hoover and many others believed, the government should develop for the national good, just as it had the western physical frontier. Moreover, the known uses for radio were few in number, so designating frequencies for particular uses was a simple matter.
But radio has long since become well developed and ubiquitous. Almost everyone uses it several times every day, mostly for purposes that are private and personal—searching, scheduling, reading, listening to music, watching movies and videos, networking and sharing with family, friends, and colleagues. Even radio and television programming is now distributed mainly through dedicated wireless channels and landline networks rather than traditional broadcasting. And 80 years of experience and discovery have generated innumerable new uses of more and more spectrum (at progressively higher frequencies), along with new methods for data compression, sharing frequency bands for multiple purposes, and combining different frequencies for the same purposes.¹²⁵ The effort to limit particular frequencies to unique uses and transmission methods is increasingly out of step with the dynamics of communications and information technology, and has become, as a practical matter, simply unmanageable.

The Wireless-Broadband Shortage

In recent years, the problem of spectrum misallocation has centered on the shortage of spectrum available for smartphones, tablets, and other devices that rely on wireless broadband. (Technically, “broadband” means a radio channel that encompasses a range of adjacent frequencies used to transmit multiple signals simultaneously, but it is now generally used to refer to internet-access standards of increasing speed and capacity—3G, 4G, 4G-LTE, and now 5G and even faster ones to come.) The broadband shortage has generated task-force reports, agency blueprints (in particular the FCC’s 2010 “National Broadband Plan”), industry white papers, congressional hearings, and presidential proclamations.¹²⁶ Everyone agrees that the shortage is seriously retarding innovation in a critical sector of the economy. But all of the proposed solutions are highly complex and many are highly partial—addressing only part of the problem, and doing so in ways that serve its proponent’s interests.

The shortage of spectrum for wireless-broadband applications is indeed serious. It is needlessly raising the costs and retarding the speed and quality of personal communications. Wireless providers such as Verizon and AT&T have been obliged to raise prices and reduce speeds selectively for heavy users of video and data applications, leading to charges of “discrimination” that the FCC has taken seriously in its net-neutrality and other initiatives. It is also fostering wasteful commercial strategies, such
as AT&T’s ill-fated 2013 attempt to acquire T-Mobile, which was really a desperate attempt to acquire spectrum. At the same time, the shortage is slowing the introduction of long-distance learning and medicine, improvements in air- and highway-traffic control, and innumerable business applications, all with immense potential for social betterment.

The current evolution from M2M (Machine-To-Machine, meaning point-to-point connections among integrated machines and between sensors and controllers) to the Internet of Things (wide sharing and analysis of data transmitted from machines and sensors, usually through the internet cloud) is making the problem much more severe, for devices can be proliferated far beyond the size of the human population. Many on-the-horizon applications, such as continuous remote monitoring of medical patients, self-driving cars, and greatly strengthened cybersecurity for personal and commercial data, simply will not get beyond the pilot stage without large additions of spectrum.¹²⁷

New techniques for sharing and combining spectrum and speeding transmission are sometimes touted as cures for the spectrum shortage.¹²⁸ Their effects, however, are actually ambiguous, because they lead to new uses for spectrum and increased competition in the supply of spectrum-dependent services, which increase spectrum demand. To date, improvements in transmission technology have been accompanied by huge increases in wireless-broadband demand, with causation surely running in both directions.¹²⁹ But the steady increase in prices paid for broadband spectrum in recent (post-2007) FCC spectrum auctions, properly controlled for other factors affecting price, suggests that the demand for spectrum is outpacing improvements in transmission efficiency.¹³⁰

The wireless-broadband shortage points to a problem that is larger still. A central administrative agency such as the FCC cannot possibly know the relative values, among multifarious and ever-changing uses, of a resource as pervasive and versatile as radio waves. The commission has erred many times in the past. In the 1940s and 1950s, it delayed the introduction of FM radio (with vastly superior quality to AM) by more than a decade;¹³¹ in the 1970s and 1980s, it delayed the introduction of mobile cellular telephones even longer.¹³² Even when its judgments are approximately correct for the time being, it lacks the flexibility to take account of varying local circumstances — its usage zones are nationwide, so a given frequency generally cannot be employed, for example, for financial exchanges in Manhattan and mountain-rescue in Colorado.
And the FCC’s errors are not random: It is naturally attentive to incumbent firms that know the agency ropes and support its budget, and less so to newbies with unfamiliar ideas that could disrupt the settled plans of its licensees and its staff. Many of the past delays in disruptive innovations, and many of the distortions in current spectrum allocations, are the result of lobbying by incumbent licensees and FCC favoritism. Some of the commission’s own efforts to counteract the inefficiencies of its zoning scheme have been defeated by political machinations. An example is its perennial proposal to charge license holders a substantial annual fee in order to discourage hoarding of unused or underused spectrum—a problem created by the narrow use restrictions in its spectrum licenses—which license holders have consistently quashed in Congress.¹³³ The FCC’s initial moves toward spectrum auctions were obstructed by the television broadcasters, who feared that auctions would be a device for raiding their treasure troves of spectrum for reassignment to the new wireless applications.¹³⁴

**Spectrum Liberalization to Date**

The FCC has nevertheless made significant progress in mitigating the harms and inefficiencies of its spectrum zoning system and rigid technical license specifications. It has done so through spectrum auctions, license liberalization, and unlicensed spectrum.

Beginning in 1994, the FCC has allocated most newly available spectrum licenses by competitive auctions. During the 21-year period ending in September 2015, it had completed 101 auctions of a total of more than 85,000 spectrum licenses, collecting $52.2 billion for the U.S. Treasury ($53.6 billion in auction revenues offset by $1.4 billion in auction expenses) with substantial additional receipts expected.¹³⁵ Allocating licenses by the price system, in place of the former approach of allocation by administrative hearings or lotteries followed by regulated secondary-market transactions, has undoubtedly speeded the movement of a considerable amount of spectrum to more productive, highly valued uses. In particular, the auctions have allocated approximately 600 MHz of highly valued frequencies to various cellular telephone and wireless-broadband uses—much of it in auctions after 2004 aimed at alleviating the broadband shortage, and much of it subject to the commission’s new license-liberalization policies discussed below.¹³⁶ The commission’s “broadcast incentive auction,” begun in March 2016—consisting of a reverse auction to purchase spectrum
Policy Reforms to Advance Innovation Policy

from TV broadcasters followed by a forward auction to sell that spectrum to wireless-broadband suppliers\textsuperscript{137} — could transfer another 126 MHz of spectrum to wireless broadband, all of it under liberal licenses.

The FCC auctions have, however, been beset by a host of difficulties. They are highly bureaucratic and at least mildly politicized — involving hundreds of pages of rules, arcane restrictions on who may bid, special credits and preferences for certain bidders, special obligations for purchasers of some spectrum, and other features that have permitted or encouraged collusion and strategic behavior, suppressed bids, and led to lengthy delays.\textsuperscript{138} The 2016 broadcast incentive auction, which is particularly complex, will take most of the year to execute, followed by at least another three years for the commission to relocate many of the selling broadcasters to other broadcast spectrum.\textsuperscript{139} But there are deeper problems that cannot be remedied with improved procedures or larger FCC budgets and staffs. The auctions are conducted within the inherited structure of assigned usage zones and extreme spectrum fragmentation into scores of thousands of individual licenses, which complicates auction procedures, suppresses participation and bids, and severely limits the potential for moving spectrum to better uses.\textsuperscript{140} And the auctions are absurdly over-centralized and episodic — as if sales and purchases of thousands of parcels of valuable real estate throughout the United States were restricted to occasional blunderbuss now-or-never dramas in Washington.

The greatest improvements in spectrum efficiency during the auctions era have come not from the auctions themselves but rather from the FCC’s concurrent re-zoning of spectrum in response to new cellular and broadband technologies. Indeed, the auctions, with their elaborate and time-consuming procedures, have slowed the migration of spectrum to more valuable uses by years or decades compared to the alternative of all-purpose licenses proposed in this paper; at their current pace, the auctions may never catch up with the still-expanding demand for wireless-broadband spectrum. In the light of history, the auctions era will probably be viewed as a transition — an initial step from the commission’s command-and-control traditions to full liberalization of spectrum usage, and one that eased the way forward by revealing the immense value of spectrum to the modern economy.

FCC license liberalization, which began in the late 1980s and continued through the 1990s, was a response to the development of cellular networks as the most efficient means of mass wireless communication, and to the
rapid evolution of technologies of spectrum transmission and management within and around the cellular architecture. The commission’s licensing regime had been designed for three primary types of communication: radio and television broadcasting, characterized by one-way transmission from a fixed point to innumerable passive receivers; point-to-point transmission in terrestrial microwave networks and among satellites and ground-stations; and a few precursors to today’s personal mobile communications, such as radio dispatch and walkie-talkies, which operated in local environments rather than as parts of integrated networks. Its traditional license specifications of transmitter location, power, equipment, antenna direction, and other particulars were ill-suited to the construction of thousands and then tens of thousands (now well over 200,000) of communications cells of widely varying size and geography, each one populated by hundreds and then thousands of moving transmitter-receiver devices in constant use, and each cellular network requiring use and coordination of many different frequencies within and among cells. Traditional FCC specifications of bandwidths and methods of signal encoding could not keep pace with innovations in spectrum sharing, spread-spectrum jumping, and technologies for transmitting data at progressively faster rates. And the appearance of new uses and applications (from voice to data and video; social media; M2M) and a variety of commercial arrangements (fee-based, advertising-based, free), all of them coexisting on the same communications networks, made a hash of the commission’s traditional zoning of spectrum “service categories.”

The FCC’s response took the form of (essentially) four kinds of license liberalization.¹⁴¹ First, beginning with the appearance of digital-transmission technologies in the late 1980s, the commission permitted licenses designated for analog transmission to be upgraded to digital, and thereafter pursued a permissive approach to accommodating progressive improvements in transmission technologies. Second, in the early 1990s the commission established a broad new service category, “Commercial Mobile Radio Service” (CMRS), which subsumed several narrower categories such as Specialized Mobile Radio, Personal Communications Services, Business Radio, and Common Carrier Paging, and thereafter incorporated additional service categories into CMRS.¹⁴² Third, the commission relaxed or abandoned many of its specifications of transmission power and equipment and antenna location, giving suppliers flexibility to adjust them to the varying demands of individual cells. Fourth, through
auctions and other means, the commission permitted mobile-service suppliers to “overlay” existing licenses for point-to-point and broadcast services — to use portions of the allocated spectrum (so-called “white spaces”) in ways that did not interfere unduly with the incumbents’ uses. These spectrum-sharing policies set the stage for private transactions among incumbent and overlay licensees to adjust their business models, relocate to other frequencies, manage radio interference, and otherwise economize on the use of spectrum for competing uses.

Taken together, these policies have produced the closest approximation to date of a private property regime for spectrum allocation, where wireless service suppliers have been freed from narrow license restrictions to respond to evolving technology and market demand. Regulated only by straightforward recordkeeping and reporting requirements, suppliers have chosen their own transmission methods and equipment, service offerings, and business models; optimized signal power from location to location; shared and exchanged spectrum frequencies; deployed several generations of progressively faster and more proficient network technologies; and collaborated with (and subsidized) manufacturers of commensurately more proficient mobile phones, tablets, and computers that now continuously monitor network conditions and adjust frequencies and power levels.

It is difficult to derive a precise estimate of the economic value of license liberalization from trends in prices of CMRS spectrum at successive FCC auctions, given the importance of many independent variables such as geography, spectrum quantity, and improvements in transmission capacity and physical infrastructure.¹⁴³ It is clear, however, that, taking account of the independent factors, liberal licenses are substantially more valuable than traditional licenses with narrow specifications of use and technology.¹⁴⁴ This is strong evidence of the further benefits of extending the FCC’s reforms to all-purpose spectrum licenses. In the meantime, liberal licenses for wireless broadband have been the sine qua non of the construction of a $1.4 trillion¹⁴⁵ communications network that has yielded continuously falling consumer prices for mobile services and devices¹⁴⁶ and is now the backbone of an economic sector that adds $1-2 trillion in value (5% to 10% of GDP) to the American economy each year.¹⁴⁷

Not all of the spectrum is licensed; the FCC has long reserved portions of the electromagnetic spectrum for short-range, low-power uses such as microwave ovens and remote control of television sets and
garage doors. The reserved portions are called “unlicensed spectrum” because manufacturers are not required to obtain licenses to transmit over the designated frequencies but must simply observe the commission’s limits on range, power, and transmission methods. Users must accept any radio interference they receive and, on complaint, correct any interference they cause to others.

In 1985, the commission established expedited procedures and standards for a wider array of “Part 15” (unlicensed spectrum) devices, and in the 1990s a host of new applications were introduced, such as local-area wireless networks of phones and computers, cordless landline telephones, and wireless microphones. Then, beginning in 1999, computer and mobile-phone manufacturers began to introduce Wi-Fi and Bluetooth technology into their products—the former primarily for voice and data connection to the cellular network through local “hotspots” the size of a residence, business, or public facility; the latter primarily for shorter-range connections such as wireless computer keyboards and music speakers.¹⁴⁸

Bluetooth and Wi-Fi proved to be immensely useful and popular, especially after the introduction of smartphones and tablets in the mid-2000s. With improvements in the speed and capacity of “last mile” landline connections from homes and businesses to cellular networks (such as through replacing copper wires with optical fiber), Wi-Fi became a good substitute for wireless-network subscriptions for many people—it was limited to places with Wi-Fi equipment but less expensive than more “nomadic” wireless service available throughout network cells. As Wi-Fi technology itself improved, and more and more businesses, commercial establishments (famously beginning with Starbucks), and transportation terminals and carriers installed the equipment, the locational limitations of Wi-Fi decreased. Soon tablets and smartphones were enabled to switch automatically between Wi-Fi and direct cellular connections depending on the availability of good Wi-Fi connections. And Wi-Fi, once installed, could be used for a growing number of additional purposes within homes and offices, such as connecting fixed desktop computers, music and video systems, and thermostat and security systems. Today, two-thirds of American households have their own Wi-Fi, as do more than 10 million shops, hotels, and other public facilities. Most wireless-broadband traffic connects to users through Wi-Fi (especially in urban areas), and wireless-service providers are beginning to employ unlicensed
spectrum within their cellular networks of mostly licensed spectrum.¹⁴⁹ At the same time, Bluetooth and similar very-short-range technologies have increasingly replaced wires at desks and in cars and kitchens. In response to these developments, the FCC has allocated increasing spectrum for use by unlicensed devices—from 235 MHz of spectrum in 1985 to 955 MHz at the end of 2008, which was more than twice the spectrum it had allocated to liberal licenses for wireless broadband.¹⁵⁰

These and other uses of unlicensed spectrum have been tremendous successes, and the emerging Internet of Things will make more intense use of it. The advantages of unlicensed spectrum have, however, been exaggerated by proponents of a “spectrum commons” — in which increasingly intelligent devices will put an end to spectrum scarcity and unlicensed spectrum will progressively displace and eventually replace licensed spectrum.¹⁵¹ While unlicensed spectrum has employed many advanced, spectrum-economizing transmission technologies, such as spread-spectrum techniques for moving continuously among different frequencies, recently liberalized licensed spectrum has employed these technologies as well. For every example of creative use of spectrum sharing in unlicensed space, such as spectrum overlays in broadcasting “white spaces,” there is an example of equally creative use in liberally licensed space (the latter include wireless delivery of Kindle e-books and iTunes music, and GM’s OnStar navigation system, all of which piggyback on licensed broadband by private agreement). While unlicensed spectrum advocates emphasize that Wi-Fi has increased demand for wireless-broadband services and reduced the costs of cellular networks by offloading some of their traffic, the existence of the cellular networks has itself been a predicate for the demand for Wi-Fi to connect to them — in practice, the two are strong economic complements.¹⁵²

We need not adjudicate the merits of the licensed-unlicensed debates in any detail. For purposes of evaluating the proposal to move to all-purpose spectrum licenses, three general, relatively uncontroversial propositions will suffice.

First, the proper reference point for evaluating unlicensed spectrum is not traditional, highly restricted licensed spectrum such as that for broadcast television — characterized by low spectrum usage, slow innovation, and inflexibility in the face of changing technology and consumer demand. Instead it is liberally licensed spectrum, which was introduced during the same time period as unlicensed spectrum, and
in response to the same technological developments. Although liberal licenses for wireless broadband have been granted much less spectrum than unlicensed uses, they too have provided the framework for rapid innovation, much more intense spectrum usage, massive investments in physical infrastructure and devices, falling consumer prices, and very large consumer surpluses (the value consumers receive from goods and services above what they pay for them).¹⁵³

Second, unlicensed spectrum is not an unregulated commons free of the constraints of property rights, but rather is regulated differently than licensed spectrum. The FCC regulates the power levels, transmission methods, and ranges of unlicensed spectrum devices to localize their use and control radio interference. Users own and manage the devices and, for almost all applications to date, the real property in which they are used. Residential and coffee-shop Wi-Fi is regulated by inherent limits on numbers of users and by precautions on use by neighbors. More expansive systems, such as those deployed throughout airports, hotels, buildings, and college campuses, are actively regulated by local administrators through such means as router placements; passwords; user fees or indirect charges; separate user categories of guests, employees, club or department memberships, and user location. Rural townships that have introduced unlicensed local systems have employed similar methods within their jurisdictions.

Third, unlicensed spectrum is to date almost entirely a phenomenon of small, short-range networks and “hotspot” cellular internet connections within privately (or municipally) owned premises. Most efforts to establish broader, public-use systems (such as “Muni Wi-Fi”) have foun- dered on range limitations and difficulties in controlling use and radio interference among larger and less well identified numbers of devices; examples of successful deployments are strikingly few and far between.¹⁵⁴ Given this experience, it is a tall order to translate ideas about a wide-ranging communications commons into practical reality. A city, state, or nation blanketed by interconnected Wi-Fi hotspots begins to look like the cellular network that we have, and raises the question of whom, in the absence of proprietary owners of spectrum rights, would make the massive investments to build and manage the system.¹⁵⁵ A world where millions of far-flung device users simultaneously exchange millions of data-intense communications across unlicensed spectrum is a world in possession of technologies utterly beyond current knowledge. No doubt
there will be many further, dazzling improvements in information and communications technology. But they will come in increments, each one with useful applications in both licensed and unlicensed spectrum architecture, and affecting the relative advantages of the two systems only at the margin. What we know so far is that unlicensed spectrum is advantageous for local connectivity but does not scale up easily, while licensed spectrum under liberal technical and usage rules exhibits large economies of scale and scope.

In sum, licensed and unlicensed spectrum have both come to play vital roles in modern communications and should be regarded as mutually beneficial rather than mutually exclusive systems. There is, of course, important competition between them—in use, when it comes to choosing one or another for a particular element of communications networks, and in policy, when it comes to allocating an additional increment of spectrum to one or the other. Allocating new spectrum to either system imposes costs on the other in the form of less available spectrum and greater radio interference from the favored system; the choice should be governed by judgments of the net benefits of new spectrum in licensed versus unlicensed uses. It is, however, impossible to make that comparison intelligently today, when only a small portion of licensed spectrum is subject to liberal, flexible-use rules. The relative benefits of licensed and unlicensed spectrum can be observed only when licensed users are as free as unlicensed users to deploy different technologies and different spectrum bands for different purposes.

**The Next Step: All-Purpose Spectrum Licenses**

All-purpose spectrum licenses is a simple idea for a policy field rife with complexities, many of them unnecessary. The FCC would allow license holders to use their spectrum for any valid purpose, liberalize technical specifications as it has done for wireless-broadband licenses, and permit spectrum to be bought and sold with only minimal restrictions. The commission could do this immediately. While spectrum auctions require authorization from Congress, spectrum zones and technical license restrictions are the commission’s own creations and can be revised as it sees fit (as it did in the case of wireless broadband). The commission would proposal a rule, which should take no more than a year to adopt through notice-and-comment rulemaking procedures, that removed the usage zones and most technical restrictions
in practically all existing spectrum licenses, and the NTIA would take similar steps for government spectrum. From that moment on there would be no spectrum shortage.¹⁵⁶

Broadcast-television licensees could sell spectrum to mobile-broadband providers, or not, or some spectrum but not all, depending on which application appeared more valuable to the parties involved. So could licensees in many other areas of misallocated spectrum that are not in the headlines and that the FCC doesn’t even know about. Frequencies could be used for different applications in different locales and at different times of day. Or they could be used for different applications in the same place from minute to minute, relying on technologies that deploy spectrum among different uses in real time according to usage patterns. Or different frequencies could be used in tandem for purposes now forbidden because some of the frequencies are in the wrong zone. The FCC would continue with its auctions of unallocated spectrum, but without restrictions on use.

Over time, spectrum use would become akin to private property, just as it has for wireless broadband. It would be subject to the same laws— contract, nuisance, antitrust — that govern the use of land, buildings, and other tangible assets. The military, the police, and other government agencies would own and employ radio spectrum for public purposes and buy and sell increments as necessary, just as they do other resources. The FCC would operate the national equivalent of a county land-title office, where buyers and sellers could assure themselves of good title and register rights and obligations affecting other owners.¹⁵⁷ The entire process could be online and searchable, as could spectrum transactions themselves, whether by direct sale, auction, brokerage, or organized exchange.

All-purpose spectrum is entirely permissive. In contrast to the many proposals for top-down spectrum reallocation, it doesn’t require anyone to do anything — it leaves it to license holders to bear the expenses and take the consequences, profit or loss, of whatever they decide to do. It simply opens up new opportunities. For this reason, and because of the explosive pace and unpredictability of innovation in communications and information technology, the social benefits of all-purpose spectrum cannot be estimated with any precision. From the prices paid in recent FCC auctions and private spectrum transactions, from the returns on recent investments in wireless services, and from empirical data on
currently unused and underused spectrum, we can get a glimpse of the benefits of relieving the current shortage in wireless broadband.

Academic and industry studies using these data find short-term economic benefits of many hundreds of billions of dollars.¹⁵⁸ There are large ranges of uncertainty in these estimates, but if one looks at expenditures and economic value (to both consumers and producers) in the initial stages of wireless growth, and at the many high-value, technically feasible applications now under development, the estimates are more than plausible. In any event, the economic benefits of all-purpose spectrum licenses would be much greater, because the reform would improve spectrum use throughout the spectrum and permit continuous improvements over time that the current system insensibly obstructs.

Most of all, the social, economic, and personal gains of all-purpose spectrum licenses would begin to be realized almost immediately, rather than years in the future as under the FCC’s desultory auction program. They would be large enough to show up in aggregate measures of national economic performance. And, because they would arise from the uncorking of new opportunities, they would not be subject to the zero-sum political wrangling that dooms so many beneficial regulatory reforms.

Objections to all-purpose spectrum licenses

All-purpose spectrum proposals build on recent FCC reforms that have acquired substantial constituencies and political momentum, but its potential for unleashing new rounds of disruptive innovation means that it would be sure to generate controversy. Fortunately, the recent experience of spectrum auctions, license liberalization, and unlicensed spectrum provides answers to the most important objections that might be made. What might have been considered a radical departure 25 years ago is today more evolutionary and grounded in practical experience.

Consider first the objection that all-purpose licenses would “privatize” an invaluable national resource—permitting license holders to profit from spectrum that belongs to the public, and to abscond with government revenues the FCC’s spectrum auctions could raise. The first part of this argument is a fallacy. Whether they paid for it or not, licensees have always profited from whatever economic value they could produce with spectrum (from their own use or sale to others). Their increased returns from the lifting of service and narrow technical restrictions from their licenses would depend on their using the flexibility
to find more complete and valuable uses for the covered spectrum, and consumers would realize a substantial share of that increased value.

The second part is misguided—good policy aims to increase social welfare, not just government revenues. In any event, simply permitting all-purpose spectrum licenses by rule would almost certainly generate higher revenues than the alternative of attempting to auction the removal of license restrictions. The $52 billion raised to date from 21 years of FCC spectrum auctions (in recent cases known by bidders to be accompanied by license liberalization) must be a small fraction of the taxes paid on the higher-valued commercial activities that they permitted. Tax revenues from capital gains on all-purpose licenses, sales taxes on goods and services that better-used spectrum made possible, and taxes on added personal and corporate income would probably far exceed auction revenues as well, and would continue with the flexibility to repurpose use over time across the radio spectrum.

And we know from the auction experience that those revenues would be collected years or decades sooner than if all-purpose licenses were enmeshed in auction procedures. The demands of government revenue collection often conflict with private-sector productivity, but there is no such conflict in the choice between spectrum auctions and all-purpose licenses, because the taxable private gains from the latter would come sooner and be orders of magnitude greater than auction revenues.

A second objection is that license holders, freed of the FCC’s usage zones, would employ spectrum in ways that created radio interference with other licensees in adjacent spectrum bands or geographic areas. But radio interference is technical phenomenon, not an artifact of one or another allocation scheme. It is legally actionable and can be resolved by direct negotiation among users of adjacent spectrum. This already happens under the current zoning scheme and has become routine in the build-out of cellular broadband networks under liberal licenses, where transactions to move or share frequency bands have become everyday business, often conducted by engineers.

In contrast, the FCC’s traditional approach to radio interference has been clumsy and terribly wasteful, requiring the preservation of large “white spaces”—buffers of unused spectrum between active bands. That approach led to the commission’s 2012 decision, based on radio-interference objections from GPS service providers, to revoke Lightsquared’s permission to establish a new wireless broadband network after the firm had already
invested $4 billion in the venture.¹⁵⁹ It was a policy debacle of the first order: The social value of the new broadband network would have been vastly greater than the costs of resolving any GPS interference problems.

With all-purpose licenses, spectrum that is now warehoused as buffer would be deployed much faster and more completely than under FCC auctions or administrative procedures, and border conflicts would be left to straightforward commercial and technical resolution rather than lobbyist-infested political resolution. Spectrum-sharing and other technologies plus the law of contract would move much more spectrum into productive use — and reveal, through experience, the most practical approaches to managing radio interference among competing uses.

While the FCC’s recent policy reforms have diminished the force of privatization and interference objections to all-purpose licenses, they have raised a potential new one. Exclusive spectrum rights, it might be objected, should not be liberalized but rather supplanted by unlicensed spectrum with no exclusive usage rights at all — an extension of the arguments for unlicensed spectrum discussed in the previous section. But these arguments have focused mainly on the waste and inefficiencies of traditionally restricted licenses, and the experience with liberal licenses has cast the licensed-versus-unlicensed question in a new light. Although many influential corporations and industry leaders have favored increased allocations of unlicensed spectrum, this has been in the context of the FCC’s established regime of narrow, inflexible licenses for most spectrum. When the comparison is instead to all-purpose licenses across the spectrum, the balance of advantages for different technical and commercial circumstances will change.

As a general matter, almost everyone in the communications industry wants a regime of flexible, decentralized decision-making where spectrum frequencies may be employed according to considerations of technology, economic combination with non-spectrum factors of production, and market demand. Unlicensed spectrum offers these advantages — but exclusively licensed all-purpose spectrum does as well. So current preferences for more unlicensed spectrum do not logically translate into opposition to all-purpose spectrum licenses, and may translate into support. In any event, as the practical possibilities of all-purpose licenses sink in, industry positions as well as those of government officials and academic analysts should shift toward the more
informed comparison of incremental licensed versus unlicensed spectrum suggested at the end of the previous section.

It appears that this may already be happening. In late 2014, the Aspen Institute convened a two-day roundtable on spectrum policy with 26 leading industry, government, and academic experts, and asked them to evaluate a “general-purpose spectrum regime” (their term for all-purpose spectrum). By design, the participants included vigorous, knowledgeable proponents of licensed spectrum, unlicensed spectrum, and “shared spectrum” (where, as mentioned earlier, frequencies are licensed but transmission rights are shared by more than one party).

In the words of the group’s rapporteur, Dorothy Robyn: “Roundtable participants were unanimous in embracing a general-purpose spectrum regime as a long-term, ‘aspirational’ goal, although they emphasized the importance of incremental gains, and individual participants viewed their preferred spectrum management model as providing the best transition path.”¹⁶⁰ It is unclear from the report why the goal was regarded as long-term and aspirational rather than, as proposed here, immediate and actual. There appeared to be general agreement that the current regime of fragmented, narrowly defined, technologically limited spectrum is fossilized and is impeding innovation and competition in the here-and-now; that “exclusive, flexible rights has worked extremely well for CMRS carriers”; and that the distinction between unlicensed Wi-Fi and licensed CMRS is being eroded by new transmission technologies.¹⁶¹

When a subgroup of technical experts was asked about the continuing need for usage zones, they concluded that, from an engineering standpoint, “no applications will require single-purpose spectrum in the future.”¹⁶² And when the group as a whole was asked to specify what restrictions should be placed on general-purpose spectrum, they concluded that there is no need to limit a particular band to a specific use, and that the only technical requirements should be “operating rights” and “admission control.”¹⁶³ By operating rights they meant transmission and interference standards based on system performance rather than technical inputs, and applicable to receivers as well as transmitters; by admissions control they meant procedures for determining who may access a spectrum band at any given time. Those, one might add, are exactly the issues for determination on a
spectrum-wide, usage-indifferent basis in an FCC rulemaking proceeding to inaugurate all-purpose spectrum licenses.

THE POLITICS OF ALL-PURPOSE SPECTRUM LICENSES

The beneficiaries of all-purpose spectrum licenses would include producers who need spectrum now but lack the necessary licenses or use flexibility; consumers who would soon receive new, better, and lower-cost services; and producers and consumers of unforeseeable innovations stimulated by dynamic spectrum markets—effectively all of us. But the most immediate beneficiaries would be those who currently hold spectrum licenses with traditional use and technology restrictions who could now devote them, by use or sale, to a far wider range of purposes.

And therein lies an important political advantage. It is incumbent licensees who have been most threatened by previous proposals to improve spectrum use through annual spectrum fees, claw-backs of underused spectrum, and reduced spectrum buffers. The licensees have been highly effective in delaying or defeating those proposals, and some of them have been successful in gaming spectrum auctions and using administrative proceedings to selfish advantage. But all-purpose spectrum, by greatly increasing the economic value of currently licensed spectrum, turns the incentives around—transforming incumbent licensees into an interest group for what is also in the public interest. Its permissiveness turns the attentions of those most directly involved from political rent-seeking to economic value seeking.

All-purpose spectrum licenses hold the promise of promptly correcting a serious waste of one of nature’s most valuable resources, generating profound economic benefits, and spurring new rounds of innovation in a field where recent innovations are widely understood and popular. That it is also highly feasible as a political matter makes it an opportunity not to be missed.
Notes

ENERGY POLICY FOR AMERICAN LEADERSHIP
IN THE 21ST CENTURY


8. Ibid.


11. Recent, reliable totals for global green subsidies are hard to come by, but see for example: United Nations Conference on Trade and Development (UNCTAD), “Building a development-led green economy,” June 2011, http://unctad.org/en/Docs/presspb20111_en.pdf: “…countries allocated more than $470 billion in their 2008-2009 economic recovery packages towards green investments.” OECD/IEA, World Energy Outlook 2012: “Renewable subsidies were $88 billion in 2011; over half the $4.8 trillion required to 2035 has been committed....”
Policy Reforms to Advance Innovation Policy


Unleashing Opportunity · Part I


26. With regard to climate science, there is a scientific consensus, perhaps even the oft-cited “97 percent”—though the papers that claim that statistic are suspect—about the idea that humans can impact and have impacted climate in some fashion. As has been recently publicized, Exxon scientists wrote (publicly, not secretly) as much in 1977: “[...]there is general scientific agreement that the most likely manner in which mankind is influencing the global climate is through carbon dioxide released from the burning of fossil fuels.” The key scientific question is not whether there is some impact, but how significant and how soon. About such questions there are great uncertainties and precious little that constitutes scientific evidence instead of speculation. But when scientists who accept the theory of human-induced climate change publish research that questions the validity of the apocalyptic position (notably for example, Judith Curry, Georgia Tech professor, and Steven Koonin, former Obama Administration DOE official and former CalTech Provost), they find themselves labeled as “climate deniers” or worse.


Note: Data center energy use has risen far faster than aviation energy use since the Gartner 2007 estimate: e.g., see “Scaling Up Energy Efficiency Across the Data Center Industry,” NRDC Data Center Efficiency Assessment, Issue Paper,
Policy Reforms to Advance Innovation Policy


Mills, *The Cloud Begins With Coal*.


Ibid.


Ross Koningstein and David Fork, “What It Would Really Take to Reverse


Unleashing Opportunity · Part I


70. Jenkins et al., Beyond Boom & Bust.


72. Ibid.

73. Note: This excludes Department of Defense R&D budget, of which 3 percent is allocated to basic science.


76. EIA, “Fossil fuels have made up at least 80% of U.S. fuel mix since 1900,” EIA, July 2, 2015, https://www.eia.gov/todayinenergy/detail.cfm?id=21912.

Biopharmaceutical Policy for American Leadership in the 21st Century


In a striking development considering the industry’s origins in Germany, France, and Switzerland, the past fifteen years have witnessed a significant shift in the center of power of the pharmaceutical industry: of the fifteen largest global firms in 2005, nine were headquartered in the United States, whereas one was in France, two were in Switzerland, and the sole German firm to make the group came in the fourteenth position.

http://fortune.com/2016/05/13/big-pharma-biotech-startups/. Alsever writes that: The majority of drugs approved in recent years originated at smaller outfits—64% of them last year, according to HBM Partners, a health care investing firm.”

81. Daemmrich, “Where Is the Pharmacy to the World?” Daemmrich writes that: Whereas safety and efficacy regulation were seen as causes for the industry’s decline in the 1970s, its subsequent turnaround has been attributed largely to price control policies in Europe and their absence in the United States. New product innovation, sales, and decisions on where to carry out clinical trials together paint a picture of a pharmaceutical industry in decline in Europe relative to the United States, though England, France, and Switzerland remain significant due to part to one or two very large firms in each country.


89. “CML Patients Taking Imatinib Have Similar Mortality Rates to People in General Population,” JNCI Journal of the National Cancer Institute 103, no. 7
Unleashing Opportunity · Part 1


94. Ibid.


Policy Reforms to Advance Innovation Policy

oxfordjournals.org/content/97/4/249.full.


110. Ibid.


112. Ibid.


WIRELESS TELECOMMUNICATIONS POLICY FOR AMERICAN LEADERSHIP IN THE 21ST CENTURY


119. The portion of the spectrum suitable for wireless radio communication is usually described as running from frequencies of 3KHz (wavelength 100 km) through 300 GHz (wavelength 1 mm), although there are some technical uses for lower and higher frequencies. The best known forms of communication — commercial radio and television; mobile telephones, smartphones, and computers; short-range links such as Wi-Fi and Bluetooth; and satellite communications for navigation and other purposes — use intermediate frequencies from 1 MHz through 3000 MHz (or 3 GHz), with the 300–3000 MHz range considered “beachfront property.” Frequency numbers are also used to designate bands or portions of spectrum — so, for example, when it is said that mobile
Policy Reforms to Advance Innovation Policy

broadband services “need 200 MHz more spectrum,” that does not mean that they need spectrum with a frequency of 200 MHz but rather that they need spectrum of any suitable frequency that totals 200 MHz in bandwidth. That might be spectrum running from 600 to 700 MHz plus that from 1750 to 1850 MHz or, more realistically, combinations of many more, narrower spectrum bands. In these cases, frequency numbers are used as a shorthand measure of capacity to transmit information.


121. The FCC posts two searchable databases of its approximately 2,111,600 spectrum licenses—License View (http://reboot.fcc.gov/license-view/) and Spectrum Dashboard (http://reboot.fcc.gov/reform/systems/spectrum-dashboard)—which present some but not all of the specifications in individual licenses. Both postings are designated “beta”; Spectrum Dashboard, launched with fanfare in 1910, has not been updated since July 2014, but License View is evidently being kept current. The Commission’s “License View Data Dictionary” (September 2010), downloadable at its License View site, details most but not all of its license terms and specifications.

122. See the FCC’s “License View Data Dictionary,” 13–16.


Unleashing Opportunity · Part I


131. Stanley M. Besen, “AM Versus FM: The Battle of the Bands,” Industrial and
Policy Reforms to Advance Innovation Policy


136. The 600 MHz estimate is derived from FCC Staff Paper, “Mobile Broadband: The Benefits of Additional Spectrum,” 15–16; Thomas W. Hazlett and Evan T. Leo, “The Case for Liberal Spectrum Licenses: A Technical and Economic Perspective,” Berkeley Technology Law Journal 26, no. 2 (Spring 2011): 1040–1041, 1048–1049, http://scholarship.law.berkeley.edu/cgi/viewcontent.cgi?article=1900&context=btlj; and, for more recent years, data on individual auctions listed on the FCC’s “Auctions Summary.” Bazelon and McHenry, “Mobile Broadband Spectrum,” 1, 7–8, estimate and list 645.5 MHz of spectrum allocated and licensed to wireless broadband as of 2015; some of it, however, had been allocated by lottery in the pre-auction period and subsequently purchased by wireless providers.


144. Wallston, “Quantifying the Factors Affecting Spectrum License Value.”


Policy Reforms to Advance Innovation Policy


155. For a review of the challenges facing the installation of 5G transmission technologies, see Doug Brake, “5G and Next Generation Wireless: Implications for Policy and Competition,” Information and Technology Innovation Foundation, June 2016, www2.itif.org/2016-5g-next-generation.pdf?mc_cid=93a4e5cba3&mc_eid=606730a1e2.

156. For antecedents and elaborations of the idea of all-purpose spectrum licenses,

157. Continued development of the FCC’s License View and Spectrum Dashboard beta projects, see note 4, would be a step in this direction.

158. An example is Bazelon and McHenry, “Mobile Broadband Spectrum,” which concludes that the total value of currently licensed mobile broadband spectrum is almost $500 billion.


161. Ibid., 11–12, 4, 6.

162. Ibid., 24.

163. Ibid., 25–27.