ABIGAIL BERGER, HASSAN KHAN, ANDREW SCHRANK, AND ERICA R. H. FUCHS

A New Policy Toolbox for Semiconductor Supply Chains

Instead of relying on production subsidies to shore up fragile semiconductor supply chains, policymakers should explore alternative approaches, including common design platforms and flexible architectures.

Until recently, the semiconductor industry was the poster child for the idea that globalized production networks are a win-win scenario for everyone. A single chip may be made using equipment, software, and other technologies sourced from countries across multiple continents. It may cross national borders dozens of times before being assembled into a final product. But during the COVID-19 pandemic, the industry's complicated production ecosystem revealed itself to be highly vulnerable to disruption. The economic and security impacts of semiconductor shortages awakened policymakers to the risks and fragility of global production networks.

In response, governments around the world have rushed to approve well over $100 billion in new incentives to boost production and strengthen domestic capabilities. In the United States, total support is estimated at $79 billion in manufacturing subsidies, research and development investments, and tax incentives. Globally, these public investments are heavily skewed toward manufacturing subsidies. Across advanced economies in the West and Japan, these investments are seen as an opportunity to rekindle advanced manufacturing capabilities that have increasingly moved offshore. For emerging markets like India, they represent a chance to ascend into a higher stage of economic complexity. In Taiwan and South Korea, they're a necessary investment to defend state conglomerates that have taken decades to build.

It will take years to see results from these investments, but even if successful, semiconductor industry supply networks will still harbor underlying weaknesses that need to be tackled directly. Instead of relying on domestic subsidies alone, policymakers should broaden their toolbox. We believe that rethinking product design could significantly change the terms of global competition and reduce supply chain fragility. We focus here on lessons from safety-critical robust semiconductors.

The history of semiconductor policies

The conventional accounts of the semiconductor industry center on great men and innovative firms: William Shockley and Bell Labs, Jack Kilby and Texas Instruments, Robert Noyce and Gordon Moore at Fairchild and Intel. Popular history also tends to reduce the role of government to simply financing R&D and being the first to buy expensive breakthroughs. Exhibit A is usually the Department of Defense, which purchased both silicon transistors and the integrated circuit when these were emerging and risky technologies. This version of the story overlooks the varied roles and contributions of public agencies as the semiconductor technology developed and matured.

One example of this underexplored history is the way that public agencies codified an open knowledge network both through antitrust requirements on Bell Labs as well as contracting requirements in defense programs, enabling knowledge about transistor technology to flow around the globe. Famously, open licensing of the published proceedings of a transistor symposium, a two-volume work that came to be known as "Ma Bell's Cookbook," enabled an upstart firm in Japan—Sony—to license transistor technology and revolutionize music consumption.

It will take years to see results from these investments, but even if successful, semiconductor industry supply networks will still harbor underlying weaknesses.

In fact, the government employed multiple policy levers to shape the industry. For example, interservice rivalry and the distinct needs of the different branches of the armed forces spurred numerous approaches to solving the so-called tyranny of numbers, which required soldering together more and more transistors to increase performance. Before silicon integrated circuits emerged as the solution, the Army prioritized the ruggedness and ease of repair of "micromodules," the Navy invested in the reliability of thin films, and the Air Force funded a radical approach called "molecular electronics." Public agencies additionally helped lay the technical foundation for the chip foundry model to support very large-scale integration design and a multiproject wafer service that lowered costs and eased piloting designs in commercial facilities. Ultimately, public officials directly shaped the industry's organizational structure and technical directions through procurement, while also indirectly influencing it by shepherding research and knowledge flows.

In contrast to the early days of the industry, government procurement and R&D now represent a mere fraction of the total electronics market. Consequently, government spending takes a back seat to commercial users in pushing the boundaries of chip fabrication. Absent significant market power, the government must seize other policy levers to strengthen myriad structural weaknesses in supply chains.

The underlying causes of supply chain failures

The public became aware of trouble in semiconductor supply chains in mid-2021, when a perfect storm of the COVID-19 pandemic, bad weather, and industrial accidents created a shortage. Auto manufacturers, who had canceled orders at the onset of the pandemic, were surprised by strong consumer demand for new cars in late 2020. Just as they were recalibrating their supply chains, a severe February freeze in Texas caused power failures that knocked four fabrication facilities in the state offline. A month later, a fire at Japanese manufacturer Renesas further reduced global capacity.

In addition to these obvious causes of the chips crisis, there were important underlying weaknesses: the limited number and geographic concentration of influential firms, plus a market that is vertically disintegrated, overspecialized, and unable to accommodate substitutions. Among these liabilities, geographic concentration has received the most attention from policy makers. However, other structural weaknesses ultimately turned a simple supply crunch into price shocks, product shortages, and fodder for political backlash. In a world where the missing chips could have been swapped with others, or manufactured in other factories, the perfect storm might have been just a rainy day.

Understanding distal causes of this fragility reveals new options for policymakers. The structure of semiconductor markets has been optimized to value performance above all other considerations. To meet tight performance requirements, chips are often designed to be produced by a single, specialized fabrication facility, or fab—sometimes just a single line in a single fab. While understandable, this implicitly forces a trade-off in supply resiliency; latent bottlenecks quickly become constrictions, which become shortages.

Constrictions are most acute in robust, safety-critical semiconductors, a class of chip often used in automobiles, industrial machinery, and defense applications. Designs for these applications require chips with high reliability and tight performance windows to ensure they can operate long-term under demanding environmental conditions. These performance requirements incentivize chip designers and manufacturers to finely tune their product offerings for specific use cases, manufacturing a wide variety of chips that can rarely be substituted for one another. For small production runs of a bespoke design—common in defense programs requiring small numbers of unique chips with high performance standards—designers and foundries may not have enough cycles of learning to optimize designs and processes, which can lead to higher numbers of defective chips. In this environment, manufacturers become locked into specific chips and the costs of switching to another are very high.

But this lock-in creates weaknesses in supply chains that are often invisible until they break. Original manufacturers and defense contractors rarely have deep or systemic knowledge of their own supply chains. When a chip cannot be delivered, firms are left with few good options. An entire project can be held up because a subcontractor chose a unique chip design that can only be sourced from a single factory and selecting an alternate may require lengthy revalidation. Assuming that manufacturing capacity is available, end-to-end production timelines for chips may be three months or more. And if production schedules are booked out—as was the case during the most acute shortages in 2021—delivery schedules are even further delayed. Without intensive preplanning, the inability to substitute either chips or factories is a problem that is difficult to solve, particularly given the long qualification timelines.

These complex design and production dynamics are unlikely to be solved by today's policy initiatives to increase overall manufacturing capacity and reduce geographic concentration of production. Likewise, other proposals, including stress tests, resiliency requirements, and stockpiling, are too limited in scope to repair supply chains with limited transparency, high costs, and little flexibility. Instead, what is needed are policies and approaches that embed resiliency into the design process, giving manufacturers and end users more alternatives and room to maneuver during crises.

Designing for resiliency

Because supply chain fragility for safety-critical chips is rooted in the need for highly specific high-performance requirements, building a truly resilient semiconductor supply chain will require a change upstream of manufacturing—in product design. We believe that adopting common design platforms could boost supply chain resiliency as well as the value of manufacturing subsidies. But successful approaches to create more resilient supply chains will require policy engagement to overcome existing incentive structures.

One approach would be for multiple original equipment manufacturers (OEMs) to adopt a standardized chip architecture and so reduce the heterogeneity of chips used in similar products with similar safety requirements. This common platform could be coordinated by the government, an independent third party, or a public-private partnership. A potential starting point could be for defense agencies to mandate the use of common designs across their various contractors, especially in sub-modules that share key technologies. Similar initiatives, backed by a trusted third party, could be adopted for commercial applications. For example, automakers today are hesitant to share data about their microelectronics bill-of-materials with competitors, but a third party could enable data sharing, allowing firms to evaluate where they share existing common designs. Importantly, one benefit of such transparency is that it would allow semiconductor customers to gain market power with semiconductor suppliers.

Critics may raise concerns over the potential inflexibility of these designs or worry about being locked in to a suboptimal technology. However, common designs need not recreate the lock-in dynamics endemic to the industry today. For commercial users, designs may be more akin to standards that ensure interoperability of multiple firms' chips for similar use cases as defined by a consortium of users (e.g., auto OEMs). In the case of defense users, long timelines of large projects already lock in specific design choices that fail to account for resiliency concerns—as we have recently experienced.

An alternate approach would be to design chips so that they could be built at multiple fabs. Currently, chips are designed from the outset with the capacities of a specific fab in mind. Relaxing designs to accommodate multiple manufacturers could reduce foundry lock-in. Furthermore, an open design process would enable stakeholders to better understand trade-offs in performance and supply resiliency. Firms could, for example, decide whether enabling production from multiple foundries is worth relaxing some performance standards. Many systems today are overdesigned with an eye to "future-proofing" or to accommodate degradation and process variability. Thus, incorporating resilience into the design process may have limited impact on real system performance. That said, fab flexibility will not enable drop-in replacements during supply crunches—customers requiring safety-critical and robust chips would still need to qualify each fab's output. But the cost and time to do so would fall with more open designs and fab flexibility.

> What is needed are policies and approaches that embed resiliency into the design process, giving manufacturers and end users more alternatives and room to maneuver during crises.

The advantage of standardizing chip architecture and broadening fab flexibility is that, unlike increasing overall production capacity, both approaches help stabilize structural weaknesses in the semiconductor supply chain. They also have the potential to reduce vendor lock-in and temporary monopoly power while increasing chip reliability. They might even raise the low yields endemic to designs for defense programs and increase the market

The advantage of standardizing chip architecture and broadening fab flexibility is that, unlike increasing overall production capacity, both approaches help stabilize structural weaknesses in the semiconductor supply chain.

power of semiconductor users. Reducing heterogeneity and finding design commonalities may also help to improve knowledge of specific components because they will have more things in common. Finally, this approach would complement existing subsidy programs aimed at reducing geographic concentration to improve overall resiliency.

We recognize that these approaches may feel alien to industry participants. There are good reasons the industry evolved to prioritize performance. Consider, however, an alternate history of the semiconductor industry where computing—rather than defense—and an emphasis on speed were the primary motivations for funding and early designs. It's possible that alternate materials may have beaten out silicon, which prevailed in part because of its reliability. Similarly, prioritizing resiliency in the design process may open up new, heretofore unknown technical pathways.

Diverse supply chains need holistic policy

Even as global policymakers continue to allocate funding to stimulate manufacturing capacity, the underlying causes of the pandemic's supply chain shocks have been left largely unresolved. Policymakers should take this opportunity to explore a broader set of policy and technical options to enhance resiliency and ensure that increased government spending has the desired effect.

Unlocking opportunities for policy innovation for semiconductors requires an understanding of both market dynamics and technical capabilities to consider immediate and underlying causes of weakness in supply chains. Hence, our proposal for common design platforms and expanding fab options, specifically in the case of robust safety-critical semiconductors, takes a broad view of trade-offs in design, performance, and flexibility.

In other sectors, alternative policy tools may prove more effective than subsidies for mitigating shortages. For example, in a synthetic material industry, where supplies of a critical mineral come from a single source in a geopolitically sensitive country, incentivizing innovations in production and extraction techniques might prove most effective. Alternatively, in an industry with limited manufacturing slack and extended lead times, incentivizing innovations in manufacturing techniques to more quickly ramp up or switch production may be better options.

Going forward, expanding the policy toolbox to alleviate supply chain shortages requires a deeper and more complete analysis of why they occur. By focusing on identifying both proximate and—importantly—distal causes of shortages, researchers can build a taxonomy for the types of failures that occur. Policymakers can use this analysis to design ways to get at the root causes of supply chain failures in multiple industries.

The disruptions from the last several years offer an opportunity to reconsider which metrics are prioritized in production design processes, and why. For decades, firms have considered performance, cost, and time to delivery as overriding priorities. Over time, production networks and design processes have evolved to elevate other considerations such as environmental impact or labor conditions through a mix of regulatory pressure and private actions. The approaches outlined above may offer a pathway to do the same for supply resiliency. In addition, they may also help reduce the risk of moving from today's fragile and overly specialized production network to a new one that is fragmented, encumbered by regional overcapacity, and propped up by subsidies. Instead, expanding the policy toolbox could spur movement toward a smarter, more nimble globalized production network.

__Abigail Berger__ received her master's degree from Carnegie Mellon University in engineering and public policy. __Hassan Khan__ is an executive fellow at Semiconductors and Supply Chains for the National Network for Critical Technology Assessment. __Andrew Schrank__ is the Olive C. Watson Professor of Sociology and International and Public Affairs at Brown University and a CIFAR fellow in Innovation, Equity, and the Future of Prosperity. __Erica R. H. Fuchs__ is a professor in the Department of Engineering and Public Policy at Carnegie Mellon University, director of the National Network for Critical Technology Assessment, and a research associate with the National Bureau of Economic Research.