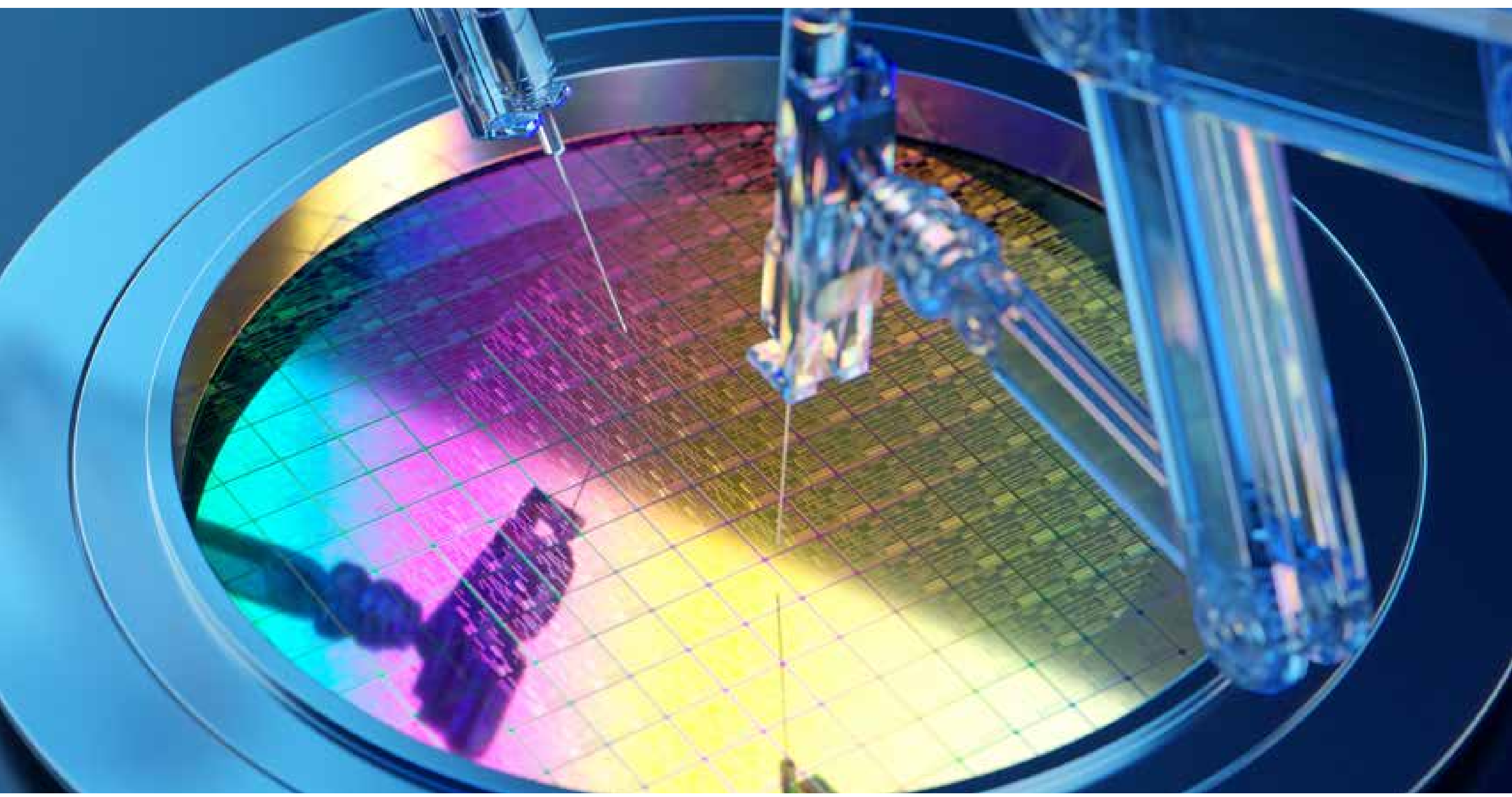
McKinsey & Company

Semiconductors Practice

Semiconductors have a big opportunity—but barriers to scale remain

Global semiconductor companies plan to invest roughly one trillion dollars in new plants through 2030. But first, the industry must overcome challenges.

This article is a collaborative effort by Bill Wiseman, Henry Marcil, and Marc de Jong, with Raphaela Wagner, Taylor Roundtree, and Teddy Stopford, representing views from McKinsey's Semiconductors Practice.



April 2025

Globally, semiconductor companies plan to invest about $1 trillion through 2030 in new fabrication plants (fabs),[1] and the global annual revenue of the industry is expected to reach more than $1 trillion by 2030. This does not include the significant potential revenue from the adoption of gen AI according to even modest upside scenarios.[2] Beyond satisfying market demand, these investments will also help regions increase supply resilience across the semiconductor value chain.

These investments may substantially increase the world's semiconductor production footprint, but the road to fully realize the benefits from these investments will be rough. Beyond executing announced capital projects—which is proving to be a major challenge—five other barriers may impede advancements from newly invested capital over the long haul, particularly in the North American and European markets: underlying capital and operating cost dynamics, increasing material demands, offshore concentrations of raw materials and packaging, logistical and handling issues, and talent shortages. The semiconductor industry will need to confront these barriers if it expects to fully realize the benefits from announced investments and others that may come.

Underlying semiconductor capital and operating cost dynamics

Over the past several years, countries have been encouraged to strategically regionalize semiconductor manufacturing and the supply chain. In the United States, for example, the Bipartisan Infrastructure Law, the CHIPS Act, the Inflation Reduction Act, and state-specific incentives are enabling major opportunities. Similar incentives are available in other markets, including Europe, India, Japan, Mainland China, Southeast Asia, South Korea, and Taiwan. As a result, public funds that support semiconductors are enabling countries globally to build more fabs domestically. In the United States, however, these up-front capital and long-term operating costs for front-end capacity such as advanced logic fabs, as well as materials, packaging, and other segments, are inherently higher relative to other regions, such as Mainland China, Southeast Asia, and Taiwan. European markets have lower up-front capital and higher operating costs than the United States.

The ongoing change in cost structure is important considering the disparity in how certain industries are progressing: Automotive and industrials are seeing a downturn in investment, while AI and compute are excelling, so the shift in cost structure is expected to affect automotive and industrial-focused companies more severely. Meanwhile, AI has disproportionately driven investment and demand toward a few suppliers and distributors, which has boosted pricing power and profitability for these leading companies while the rest of the industry continues to struggle with recovery from recent downturns.

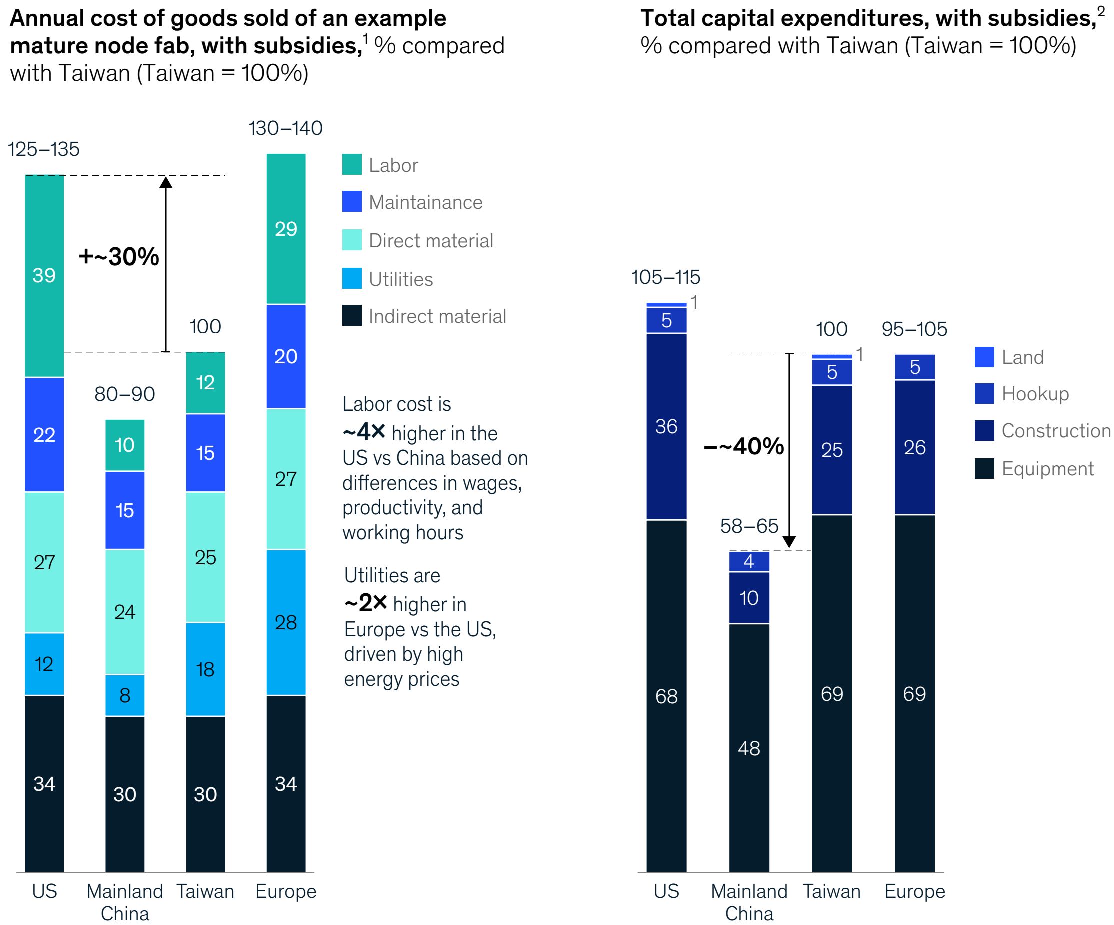
According to McKinsey analysis, with subsidies accounted for, a standard mature logic fab built in the United States will cost roughly 10 percent more to build and have up to 35 percent higher operating costs than a similar facility built in Taiwan. Europe's operating costs are roughly comparable to those of the United States because Europe's higher energy costs are offset by its lower labor costs. Mainland China has up to a 20 percent cost advantage in total subsidized operating expenses and up to a 40 percent advantage on subsidized capital expenses compared with the Taiwan market (Exhibit 1). This analysis does not fully account for differences in fab sizes, however; larger fabs, which are common in Taiwan, benefit more from economies of scale than the smaller fabs that are typically found in Europe, Mainland China, and the United States.

While investing in more-expensive fabs locally may provide meaningful benefits—such as resilience, lead-time reduction, and employment opportunities—the underlying cost dynamics could effectively create higher costs that direct consumers of chips and end customers may ultimately absorb. For example, public funding in the United States is largely offsetting up-front capital

[1] "Exploring new regions: The greenfield opportunity in semiconductors," McKinsey, January 29, 2024.
[2] Ondrej Burkacky, Julia Dragon, and Nikolaus Lehmann, "The semiconductor decade: A trillion-dollar industry," McKinsey, April 1, 2022.

Exhibit 1

Including subsidies, the cost of semiconductor fabs is higher in the US and European markets than in the Taiwan and Mainland China markets.



[1]Considers wafer starts per month for a node fab for 28-nanometer nodes, with 28,000 wafer starts per month fab capacity. Assumes utilization of 85% for all regions, with subsidies from state and federal funds for capital expenditure reductions, labor benefits, tax credits, electricity reduction, and land reduction. Includes an uncertainty of approximately five percentage points, reflecting variations in individual cost categories.
[2]Accounts for a 5% uncertainty margin, derived from alternative capital expenditure estimation methodologies.

McKinsey & Company

expenditures, but compared with other regions, there are limited incentives to offset long-term operational costs of these facilities.

This build-out could also increase varied cyclicality across the semiconductor supply chain. As incentives spur the investment of increasing fab capacity across geographies, uncertainties in market demand (including the speed at which AI may affect demand, the recovery of automotive semiconductor sales, and personal electronics) could lead to lower capital utilization across the industry's supply chain. Because much of the industry and its supply chain depend on high utilization (typically more than 75 percent) for economics to be favorable, the industry could experience further boom and bust cycles.

One of the major factors in the fab construction cost difference is variations in labor cost.

Up-front capital cost structures
Building a new fab in the United States is significantly more expensive than building a fab in Asia, mostly because of the cost of construction, and the cost of building a fab in Europe is roughly similar to in Taiwan but significantly higher than in Mainland China. Otherwise, the costs to procure tools for the fab, up-front freight, and logistics are relatively comparable across regions because of the global and consolidated nature of the supply chain. However, China benefits significantly from government-backed equipment leasing programs: The government purchases equipment and leases it to fabs at no cost, providing another form of subsidy that reduces up-front capital requirements

One of the major factors in the construction cost difference is variations in labor cost. Constructing a fab is extremely labor-intensive, requiring input from a wide variety of white- and blue-collar trades, and labor contributes a large portion of the up-front cost of building a fab. In the United States, labor costs for these trades are typically four or five times more expensive on a per-hour basis than in Asia and are apparent across the entire capital expenditure cycle, from initial site preparation to tool hookup and installation.[3] In Europe, labor can be two to three times more expensive than in Asia. Labor shortages across many industrial sectors also heighten labor costs.

Moreover, significant new fab construction has not occurred in the United States or Europe at such a scale in decades, meaning there have been fewer qualified workers and lower productivity, which increases the total cost of a project. With the increased demand for skilled craft workers experienced in building semiconductor fabs, a growing number of construction projects have to use a workforce with less fab construction experience, exacerbating productivity issues and increasing the time it takes to complete projects. Extended timelines are more apparent in the United States and Europe than in Asia. In East Asia, fabs have been completed and have achieved volume production 28 to 32 months after construction started. Conversely, due to permitting and construction delays, timelines for some US fabs have been pushed out to more than 50 months to achieve the same results. In Europe, timelines for fabs are typically 40 to 50 months.

Long-term operating cost dynamics
Regional operating costs for a new fab in the United States and Europe compared with Asia are more expensive than up-front capital costs. The core factors driving operating cost differences are labor (both direct labor and maintenance) and utilities, especially for Europe. Costs for direct and indirect materials, such as up-front equipment, generally experience cost parity across regions if the impact of subsidies is not considered.

The heightening costs of labor also raise fab operating costs. Direct labor accounts for about 30 percent of a US fab's total costs, and maintenance accounts for 20 percent of overall fab costs. In Europe, direct labor accounts for 20 percent of a fab's total costs, and maintenance accounts for roughly 15 percent. US labor costs are two to four times higher than in Asia and account for up to 20 percentage points

[3] Kevin Xu, "The cost of deglobalization," *Noema Magazine*, February 23, 2023.

more of the total operating costs than estimated fab labor in Asia. Similarly, compared with Asia, labor costs in Europe are two to three times higher and account for roughly ten percentage points more of the total operating costs. Maintenance costs, heavily influenced by labor rates and with high demand for overtime, are up to 50 percent and 30 percent higher in the United States and Europe, respectively, than fab operational expenses in East Asia.

Utility costs vary significantly across regions. The United States benefits from low baseline energy costs, while energy prices in Europe are two to three times higher. Both regions offer large-volume consumption discounts of about 10 percent, which is modest compared with the subsidies provided to fabs operating in the Mainland China market (which reach up to 70 percent) and in the Taiwan market (which are at 30 percent).

The difference in utilities costs between regions could change over time if lower-carbon chips experience higher demand. Low-carbon energy sources in Asia tend to cost more than in the United States because Asia has to import green energy in more circumstances due to space constraints and its historical aversion to nuclear energy.

Increasing material demands for advanced node manufacturing and packaging

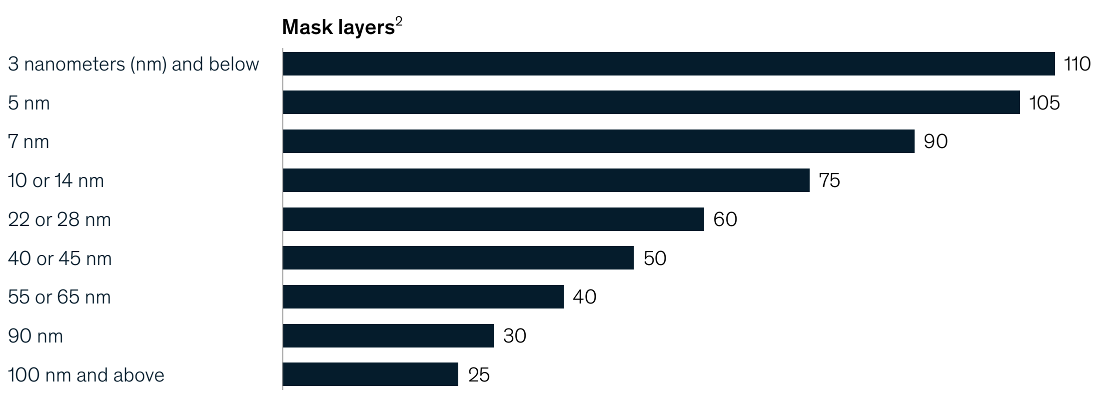
Leading-edge chips—especially those less than ten nanometers (nm)—and advanced packaging (AP) require increasingly complex manufacturing methods, which increases the demand for the materials that enable these processes and create the finished product. In the United States and Europe, investments have largely been focused on establishing more advanced-manufacturing capacity, while only a fraction of investments announced have been for semiconductor materials. As a result, the United States and Europe may increase their reliance on international capacity to fulfill needs for basic and specialty materials, many of which are currently manufactured in Asia.

Specifically, as semiconductor technology progresses to use smaller nodes, the number of mask layers—used as part of the process that defines and fills conductive channels in the chip—for producing a single wafer increases disproportionally. For example, while a 65-nm process node wafer may require about 40 mask layers to complete, a leading-edge five- or three-nm process node requires up to 110 mask layers (Exhibit 2).

Exhibit 2

Materials consumption increases dramatically with smaller nodes because advanced chips have more mask layers.

Number of mask layers in chip tech node bands[1]



[1]Data demonstrated on table is representative of most integrated circuits and wafers while understanding that numbers may change based on the tools used and the building process for a chip.
[2]Mask layer count is based on approximate feature size.

McKinsey & Company

Beyond node size, AP is also increasing material consumption. Many AP technologies use interposers or base dies to connect multiple chips, require carrier wafers as part of the manufacturing process, or include an increasing number of interconnects (through-silicon via, for example). All these incremental process steps require more materials than mature technologies do.

Through 2030, about 26 percent of the new capacity added in the United States and 11 percent of the new capacity in Europe is likely to be for nodes sized seven nm or less, compared with 15 percent in the United States and 6 percent in Europe today. The shift to using more mask layers and consuming additional materials as part of AP processes could increase the United States' total US material consumption by as much as 60 percent and Europe's by 65 percent—and many of these materials may need to come from international markets. This 60 percent increase in material consumption for the United States is roughly a third higher than the projected 45 to 47 percent increase in wafer start capacity, and the 65 percent increase in consumption for Europe is roughly 20 percent higher than the projected 50 to 55 percent increase in wafer start capacity.

Offshore concentrations of critical precursor raw materials and packaging

One of the primary challenges in manufacturing semiconductor materials lies not only in producing the final materials but also in sourcing the critical raw materials and components at the beginning of the process. Many materials used in the manufacturing process are at least somewhat available globally, except for some made primarily in Japan, but essential raw materials such as gallium, germanium, copper, and tungsten are concentrated in a few locations around the world, which increases intermediary chemical producers' dependence on these locations despite some being geopolitically sensitive.

While the semiconductor industry has a relatively smaller market demand on the global supply of these key raw materials, the concentration in only a few countries and regions can potentially disrupt the semiconductor industry if export restrictions are imposed. Countries and industry players outside of concentrated regions will depend heavily on their material imports.

Materials such as tungsten (a precursor for tungsten hexafluoride, ultimately used to deposit tungsten metal in semiconductors), germanium (used in production of silicon germanium wafers), and cobalt have niche availability, meaning that more than 70 percent of such materials' market share is controlled by single countries, which could exacerbate supply chain constraints (Exhibit 3). Unlike wafer manufacturing locations, the availability of these critical raw materials is fixed by geography: Countries either have these materials or they don't. This characteristic makes the supply chain vulnerable to disruptions or necessitates the sourcing of more-expensive alternatives.

Beyond the materials used in semiconductor manufacturing, there is also a global imbalance in standard and advanced assembly, testing, and packaging (ATP). Traditional packaging is concentrated in low-cost markets, with Mainland China, Southeast Asia, and Taiwan, accounting for approximately 75 percent of global supply. In comparison, the Europe and Middle East area and the United States each account for less than 5 percent of global supply of traditional ATP.

This trend is even more extreme for AP. Western markets exhibit an extremely low share of the AP manufacturing capacity. Logic AP is concentrated in the Taiwan market, which hosts more than 70 percent of global capacity, and high-bandwidth-memory AP is concentrated in South Korea, which hosts about 85 percent of global capacity. This reality is even starker in terms of the localization of AP capacity—the Taiwan market accounts for essentially the entire remainder of global high-bandwidth-memory capacity, and Korea accounts for an additional 13 percent or more of global logic AP capacity. Conversely, the United States accounts for 1 percent of global logic AP capacity and 0 percent of global high-bandwidth-memory AP.

Exhibit 3

Critical raw materials are concentrated among a few subtier component suppliers.

Concentration of materials in top 3 producers, %

No. 1 producer | No. 2 producer | No. 3 producer | Critical concentration[1]

		No. 1 producer	No. 2 producer	No. 3 producer	Total	Top 3 producers	Semiconductor industry share of material production, %
Commodities	Aluminum	58	6	6	70	China, India, Russia	<1
	Titanium	35	16	13	64	China, Mozambique, South Africa	<1
	Copper	24	11	8	43	Chile, Democratic Republic (DR) of Congo, Peru	20
	Tungsten	76	13	4	93	China, Vietnam, Russia	<1
	Silica	47	11	7	65	United States, Italy, France	15
	Gold	11	10	9	30	China, Russia, Australia	5–10
	Tin	23	19	14	56	China, Indonesia, Myanmar	20
Specialties	Palladium	40	38	11	89	Russia, South Africa, Canada	5
	Fluorspar	62	20	7	89	China, Mexico, South Africa	<1
	Cobalt	70	6	5	81	DR Congo, Indonesia, Russia	5
	Zirconium	34	22	10	66	Australia, South Africa, China	2–5
	Rare earths	71	14	8	93	China, United States, Australia	10–15
	Tantalum	32	20	11	63	DR Congo, Brazil, Rwanda	40
	Noble gases	50	20	15	85	Ukraine, United States, China	80–90[2]
	Gallium	98	1	1	100	China, Russia, Japan	98
	Germanium	94	4	1	99	China, Russia, Japan	10–15
	Niobium	93	6		99	Brazil, Canada, Russia	<5

[1]Critical concentration is defined as a single producer accounting for greater than 50% of material production in a market in which semiconductor consumption accounts for at least 10–15% of the material's consumption.
[2]Reference for neon because it has the highest share of all noble gases.
Source: C. Reichl and M. Schatz, *World mining data 2024*, Austria Federal Ministry of Finance, 2024; World Population Review

McKinsey & Company

Logistical and handling issues due to the shift in production and increased demand of materials

There are multiple risks to the broader supply chain beyond material precursor availability and the capacity allocation of ATP. Logistics, the state of national infrastructure, and the financial stability of players across the supply chain all factor into the ongoing health of the industry. A recent assessment on supply chain resilience for key material transportation positions wet chemicals, bulk gases, and nitrogen trifluoride specialty gas as the parts of the supply chain that would be the best candidates for onshoring or nearshoring (Exhibit 4).

Even if the broader supply chain remains financially healthy and can expand at the pace needed to meet projected demand, logistical infrastructure is critical to the on-time delivery of materials that enable ongoing operations (see sidebar, "Altman Z-scores and semiconductor supply status").

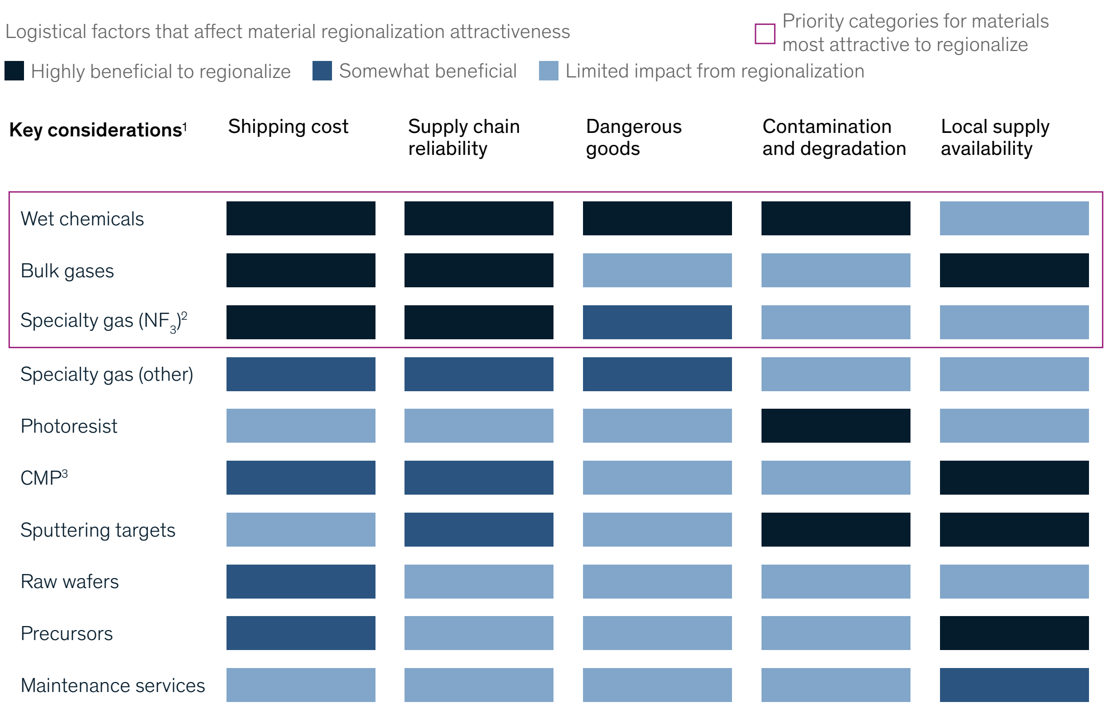
Currently, delivery issues are exacerbated by the lack of top-ranking maritime ports in the United States and Europe. These markets collectively have only five of the top 20 ports globally—two in the United States and three in Europe. In contrast, 14 of the top 20 ports are in Asia, reflecting the region's dominance in global shipping infrastructure.[4] North America's highest-ranked port is 53rd[5] globally and

[4] "World ports ranking: The leaders in container traffic," CERL, April 11, 2023.
[5] *The Container Ports Performance Index: A comparable assessment of performance based on vessel time in port*, World Bank Group, July 18, 2024.

Exhibit 4

Logistical factors make regionalization attractive for certain segments of supply.

Considerations for regionalization



[1]For key considerations, "contamination and degradation" refers to impurity contamination and chemical degradation during shipping; "dangerous goods" refers to materials that are on the International Maritime Dangerous Goods list; "shipping cost" refers to high relative shipping costs due to volume, diluted material, etc; "local supply availability" refers to when suppliers of the commodity exist in the US; "supply chain reliability" refers to reliability of ports and other logistics related to shipping affecting timely delivery.
[2]Nitrogen trifluoride.
[3]Chemical mechanical planarization.

McKinsey & Company

located in Charleston, South Carolina; Europe's highest-ranked port is tenth globally and located in Algeciras, Spain. Additionally, the total volume handled by the ten largest US ports is equal to approximately 79 percent of the volume managed by the ten largest European ports and only 27 percent of the ten largest ports in China, highlighting a significant capacity gap. These disparities affect cost control measures and extend shipping times for US and European markets, further intensifying global supply chain inefficiencies.

Without improving maritime infrastructure in the United States and Latin America, the growing demand of materials and components that need to be imported to these areas could be delayed significantly or cause broader problems to the industry. On top of lacking maritime infrastructure, ports are experiencing other disruptions that have delayed supply chains further. The semiconductor industry's lengthy supply chains are vulnerable to disruptions, making lane and mode changes challenging.

Further, rail and ground routes are seeing increased demand, but without sufficient infrastructure improvements, additional issues could ensue. Increasing costs force businesses to raise prices or seek other transportation modes—which are usually more expensive than rail—and switching rail providers is difficult due to location constraints.

In bulk road transport, mass chemical shipments can be challenging. For example, many carriers avoid transporting peroxide, which is widely used in fabs as a cleaning agent, because it requires extra driver training, can cause equipment damage, and requires additional cleaning needs. These issues highlight a need to localize semiconductor operations where possible.

Talent shortages

The semiconductor industry faces a pervasive talent shortage. From 2018 to 2022, job postings for technical roles in the United States and Europe grew at a CAGR of more than 75 percent. The shortage stems from high attrition, an aging workforce nearing retirement,[6] high demand for talent due to industry growth, and a lack of supply from an insufficient number of training programs or graduates opting for other careers. The talent shortage is a global issue, however. Countries such as India, Saudi Arabia, and the United Arab Emirates are also striving to build semiconductor ecosystems and secure investments[7] and are experiencing similar labor shortages. Semiconductor companies need new strategies to address these gaps.

For example, talent clusters—areas with a significant concentration of semiconductor presence and skilled professionals—foster innovation through proximity of professionals to

[6] "How semiconductor companies can fill the expanding talent gap," McKinsey, February 2, 2024,
[7] "Exploring new regions: The greenfield opportunity in semiconductors," McKinsey, January 29, 2024.

The semiconductor industry faces a pervasive talent shortage. The shortage stems from high attrition, an aging workforce, high demand for talent, and a lack of supply.

Altman Z-scores and semiconductor supply status

The Altman Z-score is a financial metric used to predict the likelihood of a company going bankrupt within the next two years. It combines five financial ratios—which are calculated from working capital, retained earnings, EBIT, market value of equity and value of total liabilities, and revenue—and uses a weighted sum of total assets to produce a single score that indicates the probability of bankruptcy.

For the purposes of this article, an Altman Z-score assessment details the status of OEMs and suppliers in the semiconductor industry as they have changed since the COVID-19 pandemic, comparing the 2017–19 and 2020–23 periods. The assessment showed four results:

— *Top performers:* back-end processing, sputtering targets, and metrology

— *Greatest improvements:* photolithography, substrates, and photoresists

— *Largest deterioration:* wafer processing, metrology, and automation and handling

— *Worst performers:* wafers and polysilicon, photoresists, and bulk and specialty chemicals

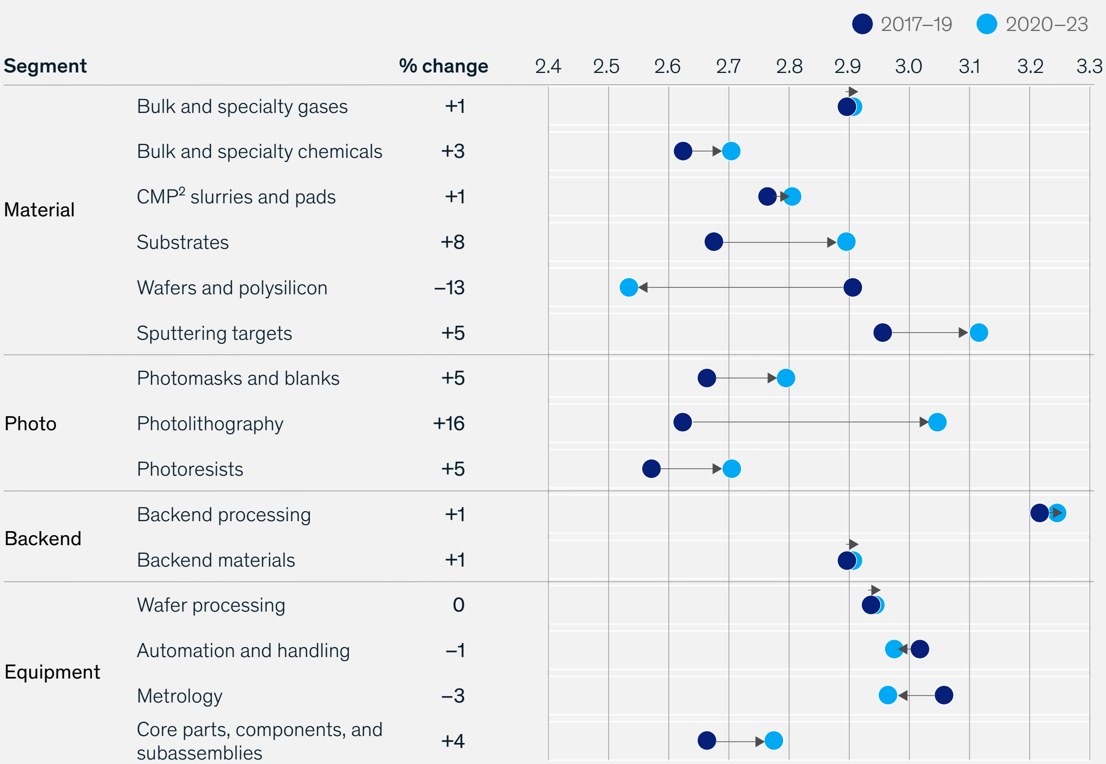
The semiconductor supply chain was financially stable before COVID-19 but struggled during the 2020–23 chip shortage. By 2024, it showed financial improvement overall, though certain subtier sectors still face challenges today. For example, substrates, once moderately healthy, became a bottleneck due to high demand and long lead times for new facilities. From 2014 to 2019, Ajinomoto Build-up Film substrate supply exceeded demand by up to 20 percent, but during the pandemic, demand outstripped supply by 20 to 30 percent annually due to rising demand and long lead times for new facility construction.

Substrate health has since improved, with a 50 percent increase in Altman Z-scores from the 2017–19 to the 2020–23 period. Based on the Altman Z-score analysis, suppliers across the semiconductor supply chain have strong financial health across most categories, but certain core supplier categories (such as wafers and polysilicon, automation and handling, and metrology) have had a decline in Altman Z-score performance between the two assessed time period, although metrology and automation and handling are still top performers (exhibit).

Exhibit

Financial health of suppliers across component categories is generally strong, although some core suppliers have degraded.

Semiconductor supplier health, Altman Z-score,[1] 2017–19 vs 2020–23, average rating per segment



[1]Interpretation: Z>2.99 = low risk of bankruptcy. Z<2.99 but >1.81 = moderate risk and uncertain status. Z<1.81 = high risk of bankruptcy.
[2]Chemical mechanical planarization.
Source: Calculated Altman Z-scores from company financial statements

relevant organizations. Close proximity facilitates knowledge sharing and collaboration. These clusters also attract public and private investment more effectively thanks to the concentrated talent pool and established infrastructure, reducing risks for start-ups and new ventures.

The path forward for the semiconductor industry

While companies are currently grappling with high capital expenditures, the long-term impact of fabs in the United States and Europe, with operations that cost up to 35 percent more than those in Asia, remains uncertain. Companies building facilities in the United States and Europe will need to adjust their cost structures, accepted margins, or portfolios to make these assets viable, especially without systemic financial assistance to equalize regional cost differences. Continued investments in the United States and Europe will heighten the urgency to address these issues.

Manufacturers must find ways to manage rising costs, which are largely structural and difficult to change without significant shifts in cost composition. Innovations such as engineering design overhauls could reduce labor costs but require substantial R&D investment. The rising demand for labor, technology, and materials is driving up prices, making cost reduction difficult, especially given varying negotiation power for manufacturers. Manufacturers can use forward contracting and cost contingencies, such as indexing material costs to commodity prices, to manage often unpredictable price fluctuations.

US and European semiconductor stakeholders will want to look at how to increase material consumption needs to ensure the expansion of supply chain capacity is commensurate with the level of expansion in wafer manufacturing. Companies will also need to manage supply chains for growing consumption while recognizing that raw material production may be concentrated in a few countries, creating potential bottlenecks.

With respect to ATP, the CHIPS Act set aside $3 billion for the National Advanced Packaging Manufacturing Program and $500 million for International Technology Security and Innovation, which is focused on driving resilience in ATP. When coupled with the signaled willingness of ATP players to invest in facilities in other markets, there could be a rebalancing of global capacity similar to what is taking place in cutting-edge front-end fabs.

For new innovations, it will be essential to expand existing manufacturing hubs or greenfield ecosystems to build a new talent ecosystem. Regional areas of semiconductor expertise can help provide a larger labor pool. Some large semiconductor companies are developing this expertise by training workers in existing centers and bringing knowledge back to new hubs. Improving job desirability in the semiconductor industry requires rethinking career paths, upskilling the workforce, and reaching untapped talent pools, including women and older adults. Retaining specialized talent will also necessitate prioritizing cultural elements such as workplace flexibility, skill-sharing resources, and inclusivity.

As global effort intensifies on growing the semiconductor industry, be it for economic demand or national security, the sector faces impending change. Industry players must navigate shifting cost structures, supply chain and logistical challenges, rising material consumption amid geopolitical tensions, a relative lack of abundance of materials based on announced expansions, and a historic talent deficit. How companies manage these issues will directly affect the industry's ability to reach the projected $1 trillion or more in revenue by 2030.[8]

[8] Ondrej Burkacky, Julia Dragon, and Nikolaus Lehmann, "The semiconductor decade: A trillion-dollar industry," McKinsey, April 1, 2022.

Bill Wiseman is a senior partner in McKinsey's Seattle office, **Henry Marcil** is a partner in the Bay Area office, **Marc de Jong** is a senior partner in the Amsterdam office, **Raphaela Wagner** is a consultant in the Zurich office, **Taylor Roundtree** is an associate partner in the Atlanta office, and **Teddy Stopford** is a consultant in the Washington, DC, office.