From crunched to crushing it: How to handle the semiconductor shortage

IBM

Dr. Rami Ahola
Ingo Kobusch
Swadhin Pradhan
Matthieu van Bilsen

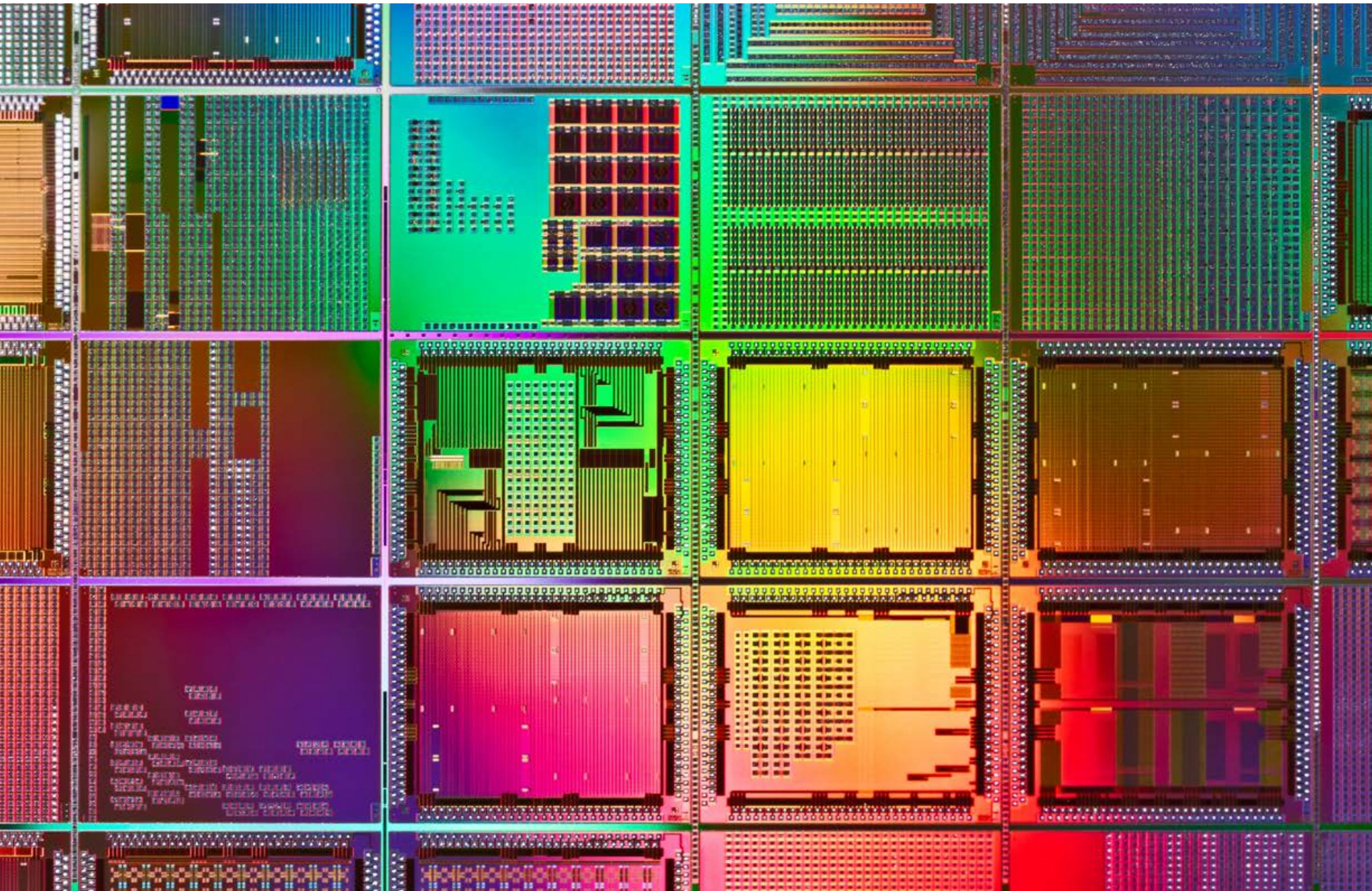


Table of contents

Executive Summary

The semiconductor industry is in turmoil, and for their customers – many of the companies integrating chips in their products – it's a huge disruption.

Semiconductors (or Integrated Circuits, "ICs" or "chips") are found in almost every product that comes to market today, and most industries have developed a huge dependency on chips. These ICs come in different levels of technology advancement, roughly put in 'Mature Nodes' based on older technology, with lower complexity and less advanced production process, and 'Advanced Nodes', the latest state-of-the-art technology (node refers to the physical size of the transistor).

The volume production of the 'Advanced Nodes' is very much concentrated in Taiwan and South Korea and the huge investments and the necessary skills make it very hard to set up production capacity somewhere else. Production of mature nodes is globally more distributed, but the older technologies can only support certain applications. The processor chip of the next smartphone however, will require an advanced manufacturing process. The combination of trade restrictions and the COVID-19 pandemic have caused a perfect storm for the semiconductor supply chain and its dependent customers.

Extremely optimized industry supply chains (e.g., Just-in-Time (JIT) in automotive) were completely disrupted, which caused production stoppages and delays. It is almost impossible to solve this quickly because of the established architecture of the semiconductor industry. Although the nature of the challenges can differ by geography, industry policies, trade barriers and established industry practices, there are several long-term and short-term remedies that can be taken.

The short-term options

- Understand and leverage your full bill-of-materials (BOM), understanding in detail what's in your full hierarchical BOM (including that of your module suppliers) to gain negotiation leverage with your suppliers
- Weigh the cost of inventory versus the cost of shortage, reassessing your inventory management policies, accepting a high stock (of low-cost components) to avoid production stoppages or delays
- Think like your supplier – create a longer-term relationship with your supplier, understand what drives their success and how you can contribute to that in order to become a more attractive trading partner in this seller's market

The long-term options

- Build your own capacity, however only possible for those with very deep pockets and the stamina to build up skills and expertise
- Consume less, reducing the number and variety of semiconductor components by combining more functions onto a single chip
- Become more software defined using a general-purpose processor chip and implement the unique product functionality entirely through software.

A Global Situation

The global dependency on ICs

When you read this paper, you are most likely viewing it on an electronic device containing a number of Integrated Circuits (ICs) or 'chips'. These semiconductor components are critical building blocks of many devices that surround our everyday life, from the alarm clock in the morning, the coffee maker, the smartphone, to the vehicle, e-bike or airplane that takes us to work. The elevator and aircon controller in the office building, the video devices for the web meetings, the NFC payment for your second coffee – almost everything is directly or indirectly powered by ICs.

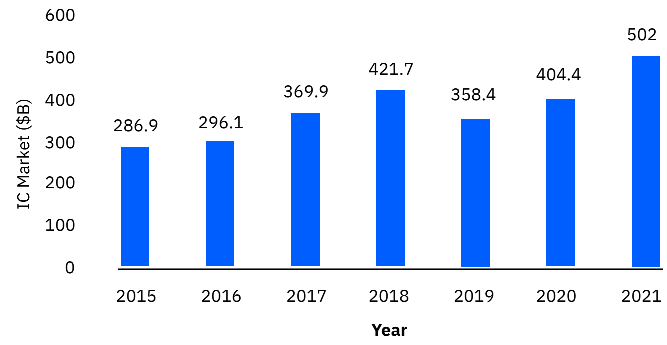
Already this list of examples gives an idea of the variety of ICs that are out there. They range from the most complex high-performance chips that power our indispensable smartphones to more simple sensors that check if the lift door is open or closed, or the crash sensor in a car that triggers the airbag.

Different types of semiconductor devices are classified by the technology node, basically a measure of the feature size of a chip. Smaller nodes (= smaller feature size, the latest in technology) are called advanced nodes (e.g.,14nm, 7nm), the 'older' larger sizes are called trailing edge or mature nodes.

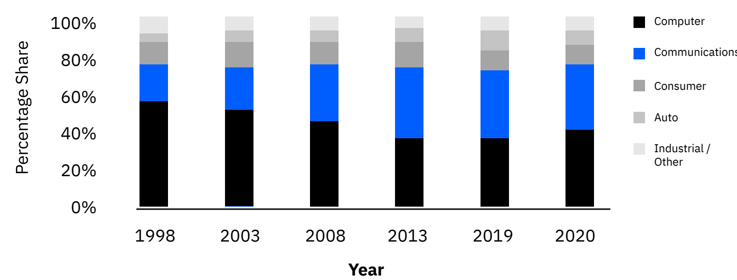
- Advanced node products are for example the processor chips in smartphones or the ones used for high performance computers like the IBM mainframe processors, game consoles, etc. Today, there are only two or three semiconductor companies left that can produce advanced node ICs. This is mainly because of the complexity of the processes and the huge (upwards of $10 billion) investments required to set up a single manufacturing facility and tools

- Mature nodes or trailing edge nodes are used for a variety of products from sensors to standard semiconductor products like amplifiers to simpler processors and controllers used in many automotive applications, for example

Worldwide Integrated Circuits (IC) Sales Growth



Integrated Circuit Market Share by System Type ($)



Source: IC Insights

A Global Situation

Why is there an IC shortage in the first place?

The global semiconductor industry was hit by a perfect storm of supply chain disruption caused by rising trade restrictions plus the novel corona virus. COVID-19 not only disrupted the supply chain due to travel restrictions and lockdowns that caused factories to close. At the same time, it caused an unprecedented spike in demand for computers, network equipment and communication devices to support the new way of remote working and the worldwide digitalization push.

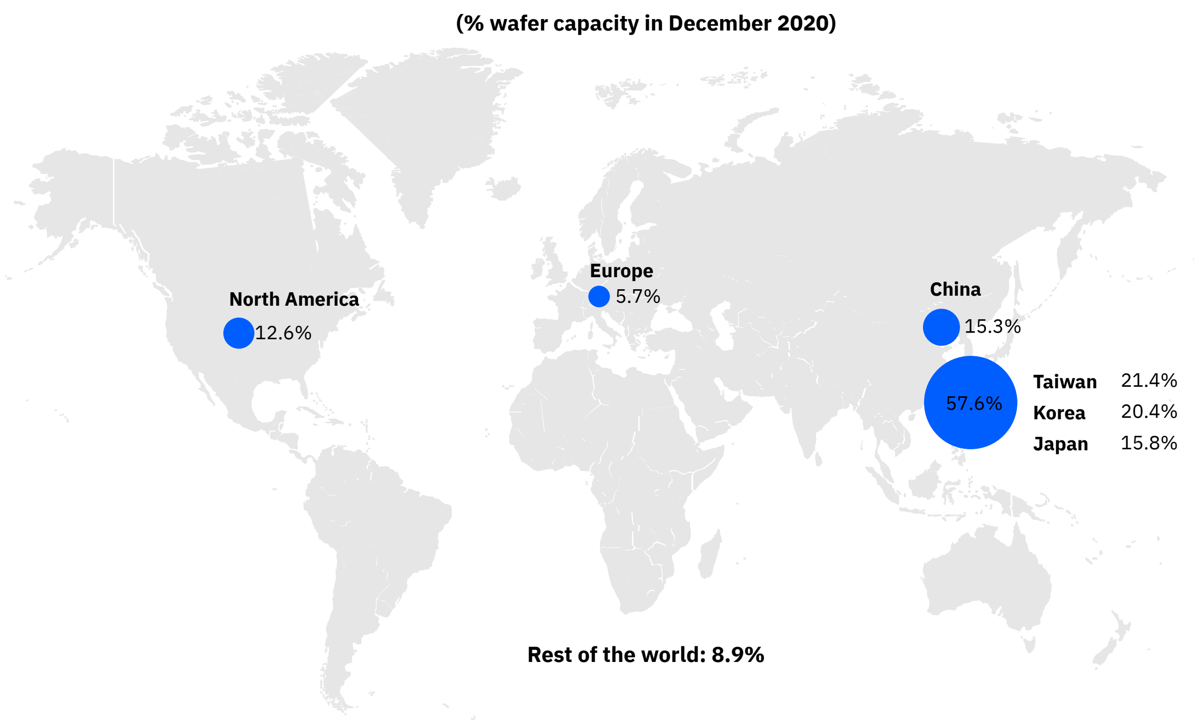
Inventory optimization processes have worked well for many industries during "normal" times, when supply was virtually unlimited and delivery time estimations were very accurate. In the COVID-19 situation, these same processes caused line stoppages in the automotive industry and delays in smartphone delivery. Obviously, the growing importance of ICs for more and more products and services, i.e. the growing demand itself is a driving factor of the shortage.

Why can't the suppliers just build more capacity?

The production lead time for an advanced node IC can easily exceed three months, and even for the more commodity types of ICs it is still over a month. The IC supply chain consists of three main steps: design, frontend fabrication, and backend assembly and test. Most semiconductor companies have specialized in one of the steps and outsource the other two steps.

Over the years, a geographical consolidation has happened for two of the three steps. While design is still globally distributed, advanced node frontend production is almost exclusively located in Taiwan and South Korea. Frontend production for mature nodes is more geographically diversified in Japan, China, Europe and the UK, and North America. Backend assembly and test is mostly consolidated in Taiwan, Malaysia, Philippines, and China.

Semiconductor manufacturing capacity by geographic region



A Global Situation

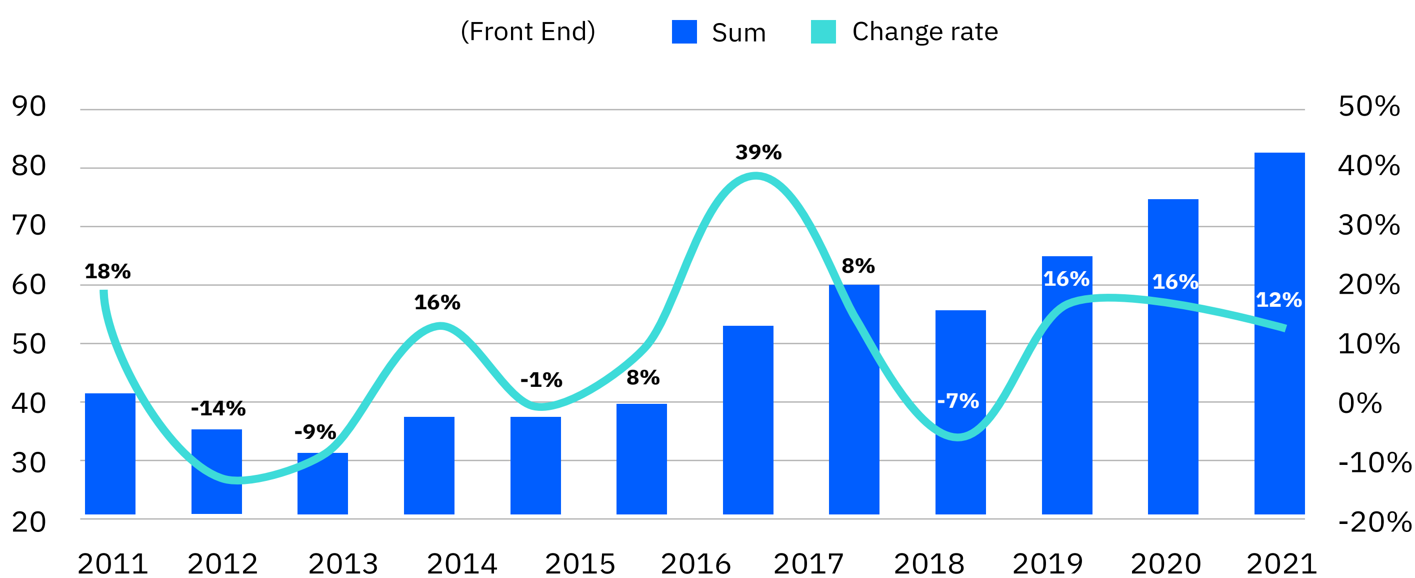
This geographical clustering of each function can render multi-sourcing strategies of IC customers useless, because the IC suppliers might have outsourced some of the steps to the same service provider (e.g., same outsourced semiconductor assembly and test – OSAT – provider) creating a single point of failure in the supply chain, which is exactly what the customer tried to prevent with multi-sourcing in the first place.

Building a chip production facility – called a "fab" – requires investments between a billion dollars for a mature node fab and upwards of $10B for an advanced node fab and will take several years. It also requires a stable infrastructure, in particular a steady and large supply of electricity and water, and of course the employees with the skills to run the fab. The extremely high investment makes new entrants into the IC manufacturing sector rare, except for China, where the government is investing a fortune to boost the local semiconductor industry. For the few remaining companies and countries that satisfy all the requirements, the next challenge is that there is also a shortage of the equipment required for the semiconductor manufacturing process.

For mature nodes, China's goal to become self-sufficient for semiconductor manufacturing by 2025 has led to a shortage of the required equipment – an issue further exacerbated by trade restrictions that led semiconductor equipment manufacturers to reduce production capacity.

For the most advanced nodes, the problem is even worse: for example, there is only a single company in the world that can produce the extreme ultraviolet (EUV) lithography tools required. Even though these machines cost well in excess of $100M each, the sole manufacturer of them has been running at maximum production capacity for years.

Fab Equipment Spending



Source: World Fab Forecast Report, 1Q21 Update, Published by SEMI

Industry impact

Which industries and which geographies are impacted the most?

The global chip shortage has impacted most industries, and across a broad spectrum of products within those industries. The automotive industry is the most impacted despite being among the smallest revenue generators for most established chip manufacturers such as TSMC.

Other industries ranging from consumer electronics (smartphones, smart TVs, white goods, etc.) to power (solar cells and wind turbines) to computing (HPC and IoT) are equally facing challenges around chip shortages.

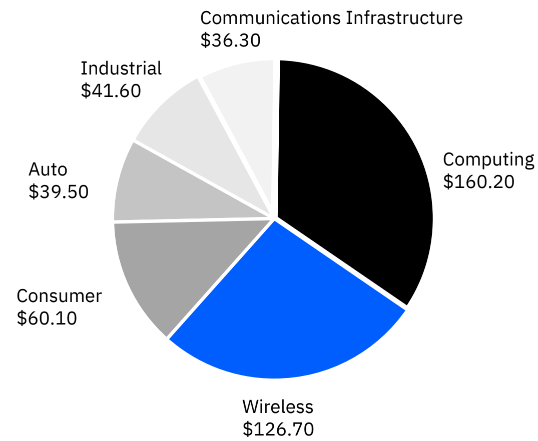
At $39.5 billion, the automotive industry makes up less than 9 percent of chip demand by revenue, according to IDC. The chip shortage will lead to lower vehicle production in 2021 and 2022. The shortage is expected to cost the global automotive industry $110 billion in revenue in 2021, according to AlixPartners.

The primary reason for the auto industry being hit the hardest is their reliance on very mature technology nodes (40 nanometers and older), which not many top suppliers are prioritizing. Moreover, the auto companies have been behind when it comes to securing inventory or entering into long term contracts with the chip suppliers.

For the leading chip manufacturers that are capable of producing the most advanced nodes, supplying chips for the computing, communications and consumer electronics markets (smartphones, gaming consoles, etc.) will invariably be more attractive than the automotive market because the chips are more complex (and hence more expensive), and the unit volumes are an order of magnitude larger. And even in the more attractive markets, gaming console companies continue to struggle fulfilling orders for their consoles and forecast supply issues continuing at least until the end of 2021.

"The semiconductor crisis, from everything I see and I'm not sure I can see everything, is going to drag into '22 easy because I don't see enough signs that additional production from the Asian sourcing points is going to come to the West in the near future."
Carlos Tavares, CEO of Stellantis

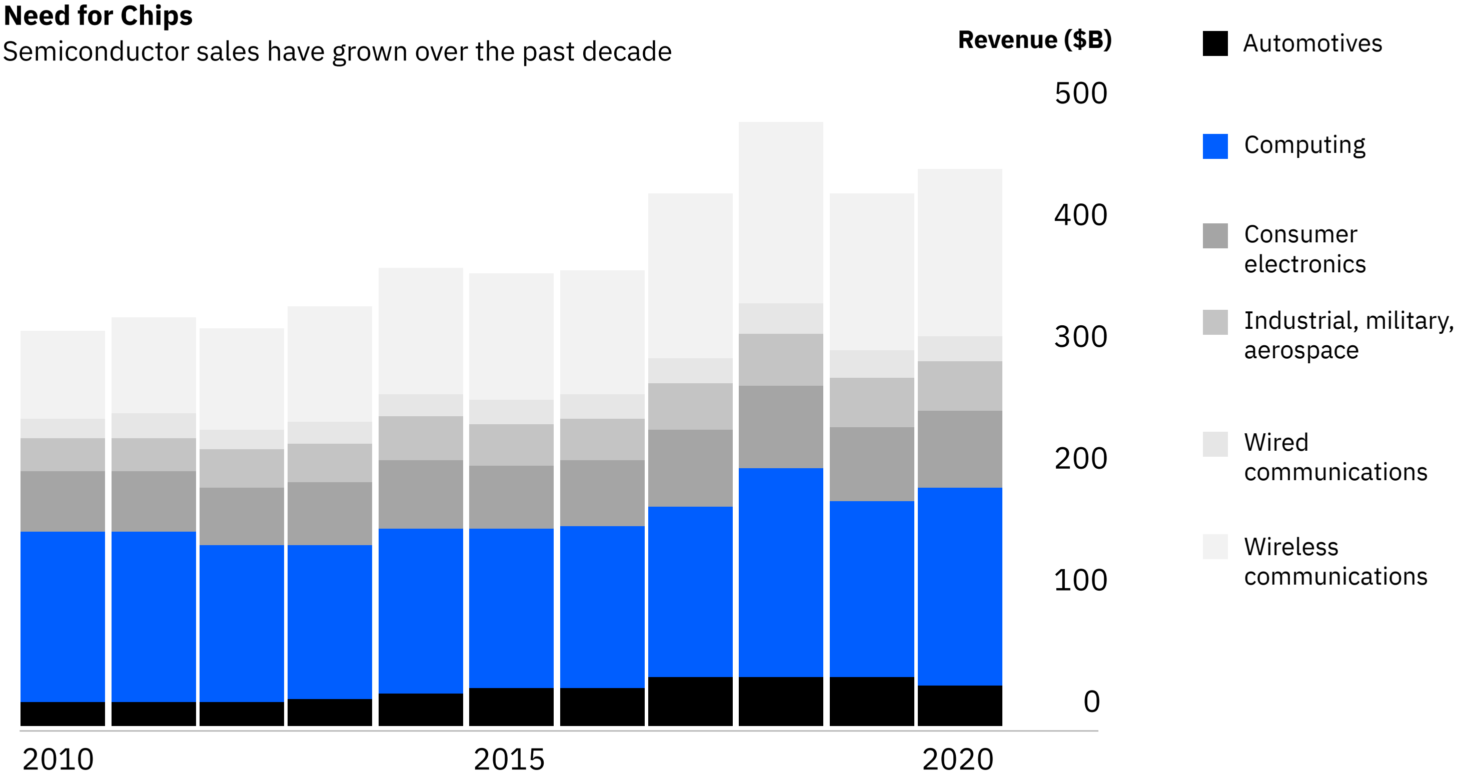
Chip Demand by Revenue ($B)



Source: IDC

Industry impact

Need for Chips
Semiconductor sales have grown over the past decade



Note: Data does not include foundry-only businesses such as TSMC or Globalfoundries.
Source: IDC

In terms of geography, the global chip shortage recovery will also depend on the ongoing COVID-19 situation. Possible COVID outbreaks in supply regions such as Taiwan or South Korea may impact production.

Many companies in these regions have announced higher capex numbers for the years to come. In May 2021, South Korea announced $450 billion of support to help expand domestic chip manufacturing, and Taiwan's TSMC plans to invest $100 billion over three years to meet similar goals. However, it will take a number of years for the new manufacturing capacity to come online.

In Europe, a key demand center, the imbalance between the needs of production and the ability to secure chip supply is quite evident. As an example, many of the large auto makers who have raised red flags are from the continent, but only a small fraction of the global chip manufacturing capacity is in Europe. The current situation is also helping distribute the global chip manufacturing footprint more evenly around the world again.

For example, the US today accounts for approximately 12% of global semiconductor manufacturing capacity, down from 37% in 1990. This situation is influencing the US government to refocus on domestic chip manufacturing and companies are responding positively. For example, Intel and TSMC have announced plans to invest $20B and $12B, respectively, into new manufacturing capacity in Arizona.

Potential remedies

As chip manufacturers can currently sell everything they can produce, most of them are obviously scrambling to build additional capacity as quickly as possible. This doesn't help in the short term, as getting a new fab online takes time. However, analysts are starting to predict an overcapacity situation as soon as the second half of 2023. In the meantime, there are several possible approaches that consumers of ICs can take to address the shortage problem. There are several short-term possibilities, and taking a sufficiently long-time horizon opens some additional options.

Long term options

Build your own

An obvious answer to uncertain supply would be to develop your own capability to manufacture the chips you need. While this may sound like a far-fetched idea, it may be worth looking into for some companies. As discussed earlier, the investment and the skills required to build a leading edge fab are far beyond the means of most semiconductor companies, let alone a medium size product company. But not all chips are leading edge, and the investment required for an older technology node is substantially smaller. Even that investment is certainly out of reach for most companies, but for a large company with a high demand of analog chips for the foreseeable future – an automotive company for example – building a fab or acquiring one might be worth considering as an option to secure supply. Industry giants such as Tata have reportedly been evaluating such options already. Whether building or buying, this is a long lead time solution that is unlikely to help in the short term.

Consume less

The second longer-term option to address the shortage would be to reduce your dependency on semiconductors. The world is not going back to mechanical switches and relays any time soon. However, there can be several thousand chips in a car. It would definitely be possible to move to a slightly less distributed architecture and centralize more functions onto a single chip.

In the automotive industry, specifically, there are two main arguments against centralization: First, the industry's supply chains have been organized around very independent modules – the rear view mirror, for instance, is a self-sufficient module that contains the intelligence required for dimming, etc. – and reducing the number of chips would require increasing dependencies between modules; Second, a significantly more centralized architecture would also mean significantly more wiring in the car, increasing the weight of the car quite substantially.

It is worth mentioning, though, that Tesla is taking a somewhat different direction compared to the traditional auto companies, including developing their own specialized chips. Similar good reasons for distributed architectures exist in other industries as well, and going from 3,000 chips to 10 chips is not a feasible idea. However, going from 3,000 chips to 2,400 chips may be much more practical, and already a considerable reduction in semiconductor dependency. Changing the product architecture will not happen overnight, though, i.e. this is not a short-term solution either.

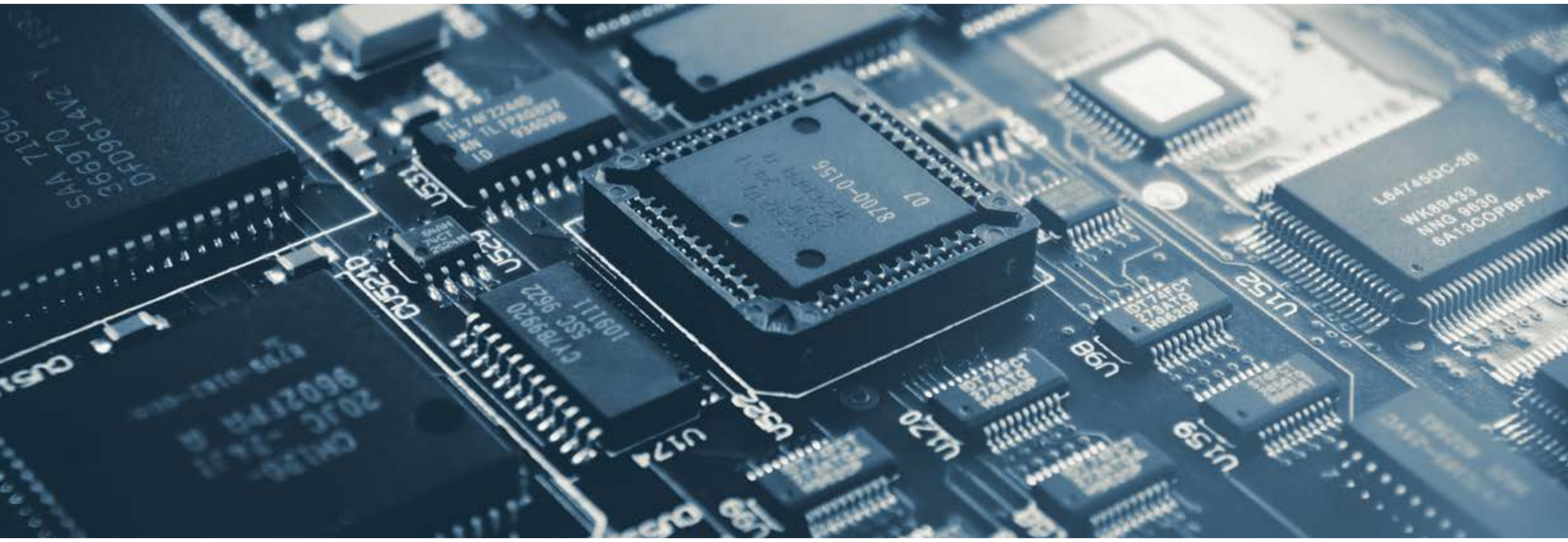
Become more software-defined

Another possible way to become less dependent on specific semiconductor manufacturers is to drive towards increasingly software-defined products. Today, many electronics products still require one or more custom chips that perform specific functions in the product. Producing a custom chip just for your product most likely means that you are dependent on a single supplier, or two at best. Most products simply don't have the sales volumes to economically distribute their chip sourcing over a larger number of suppliers.

Using a custom chip is in most cases a deliberate design choice that was made when doing the system design for the product. The other choice would have been using a general-purpose processor chip running software that creates the desired functionality. Typically, the design choices are driven by factors such as speed (you need a custom chip to perform the functionality fast enough) and cost (at your volumes, a fast enough processor chip would cost you more than a custom chip).

Moore's law means in practice that the speed of chips will go up and the cost will go down, and fast. The design choice you made five years ago may not be justified any more. With faster and cheaper general purpose processor chips, fewer and fewer product functions require (or can commercially justify) custom chips anymore. Your design choice today could very well be to use a general-purpose processor and implement your unique product functionality entirely through software. General purpose chips are available from many more sources, significantly reducing your dependency on any single supplier.

Potential remedies



Short term options

Understand and leverage your full bill-of-materials

Larger purchase volumes almost always mean better negotiation leverage with your suppliers. In today's multi-tiered supply chain architectures, however, many companies don't even know exactly which components are in their full bill-of-materials (BOM), and where they are coming from.

The opaque hierarchy is exactly what was intended when originally designing the supply chain, as it goes a long way in reducing the complexity seen from the top of the hierarchy. In today's situation, however, shortage of a single component anywhere in the hierarchy may halt the production of the entire system. For instance, when that small chip that's responsible for rear view mirror dimming is not available, the entire $50,000 automobile isn't coming out of the manufacturing line.

Now, the company making the rear view mirror module may only be buying that one component from the semiconductor company, and therefore has negligible leverage when negotiating for scarce supply. If, however, the automotive company looked throughout their entire hierarchical BOM, they may find that they are buying 500 or 1,000 components per car from the same semiconductor company. One thousand components per car gives you a lot more negotiation power than one component per rear view mirror. The lesson? Understand what exactly is in your full hierarchical BOM, understand where all of it is coming from, and use the full BOM to gain negotiation leverage with your suppliers.

Weigh the cost of inventory versus the cost of shortage

Many industries have moved towards just-in-time (JIT) manufacturing. The automotive industry may have taken JIT to its extreme, but most other industries have worked hard for a long time to reduce their inventories as well. The natural consequence of reduced inventory is of course a reduced tolerance for any disruptions to component supply. As discussed earlier, the automotive industry is a prime example: having spent years taking inventories down to practically zero, they were hit hard and fast when the semiconductor supply was disrupted. The current crisis may well be the end of across-the-board JIT.

The cost of holding a $0.50 semiconductor part in inventory is negligible compared to the cost of stopping the production and delivery of a $1M piece of medical equipment because you ran out of that $0.50 part. Companies should reassess their supply chain practices going forward. Instead of simply minimizing inventory, you should be comparing the cost of holding inventory to the risk and cost of component shortages.

For components that have a limited number of alternative suppliers or other systemic supply risks, it may well be less costly to hold some amount of buffer inventory than to risk bringing your production to a halt.

Potential remedies

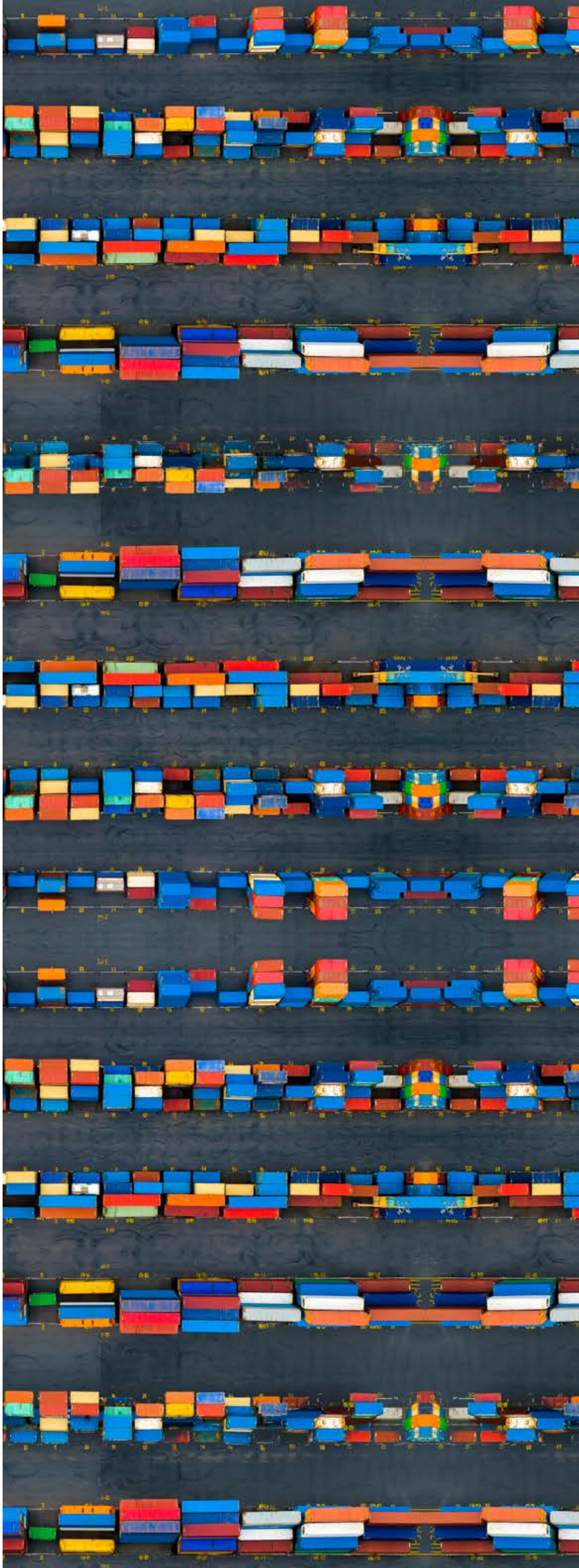
Think like your supplier

Most semiconductor companies are rational actors, attempting to maximize their profits over some time frame. Time frames may vary, but most companies are striving to find a balance between short term profits and long-term prospects. To improve your negotiation position with your semiconductor supplier, it can be helpful to put yourself in their shoes. First, try to understand which other customers are competing with you for the same supply. Then consider what factors would make the supplier want to prioritize your business over the other customers. Hard commercial factors include the price you are willing to pay, the purchase volumes you are committing to, and the length of purchase commitments.

It can be helpful to analyze the supplier's financials in some detail to understand how they would likely be weighing short term versus long term: If the supplier urgently needs cash, the price you pay today is likely the most important factor; if the supplier is building for the long term, then today's price will be less important than the length of your commitment to them.

In addition to the commercial aspects, there are several softer factors that can tilt the scales in your favor. You could be doing some joint development with your supplier, increasing their growth prospects in the future. You could be sharing data from your manufacturing line and field service that could help the supplier improve their manufacturing yields over time. Or you could be giving them favorable public visibility. And the list goes on. Consider how you can add value to your supplier, thereby making you a more attractive trading partner. It is a seller's market right now.

The current situation will eventually pass, as semiconductor companies build more manufacturing capacity. However, given the critical importance of semiconductors to practically every industry today, it is likely that shortages and supply crunches will happen in the future as well. Winners will be those that enter the next crisis better prepared.



Sources & References

Authors

Dr. Rami Ahola
Partner and Global Leader, Electronics Industry Center of Competence,
IBM Consulting
Member, IBM Industry Academy
rami.ahola@fi.ibm.com

Ingo Kobusch
Associate Partner
Electronics Industry Center of Competence
IBM Consulting
kobusch@de.ibm.com

Swadhin Pradhan
Advisor
Industrial Sector
IBM Market Insights
swapradh@in.ibm.com

Matthieu van Bilsen
Director Electronics
IBM Global Industry
Member, IBM Industry Academy
bilsen@nl.ibm.com

Sources

1. SEMI.org: "Global fab equipment spending poised to log three straight years of record highs",March 2021
2. IEEE Spectrum and IDC: "How and When the Chip Shortage Will End, in 4 Charts", June 2021
3. CNBC: "Chip shortage expected to cost auto industry $110 billion in revenue in 2021", May 2021
4. Reuters: "Global chip shortage will easily drag into 2022 -Stellantis CEO", July 2021
5. AGWEB: "John Phipps: The Impacts and Implications of the Great Microchip Shortage (Part 2)", April 2021
6. EIU: "Asia trade brief: the global chip shortage", June 2021
7. SIA: "TURNING THE TIDE FOR SEMICONDUCTOR MANUFACTURING IN THE U.S.", 2021
8. CNBC: "Intel is spending $20 billion to build two new chip plants in Arizona", March 2021
9. The Times of India: "After 5G, Tatas gear up for chip manufacturing", August 2021
10. The New York Times: "A Tiny Part's Big Ripple: Global Chip Shortage Hobbles the Auto Industry", April 2021