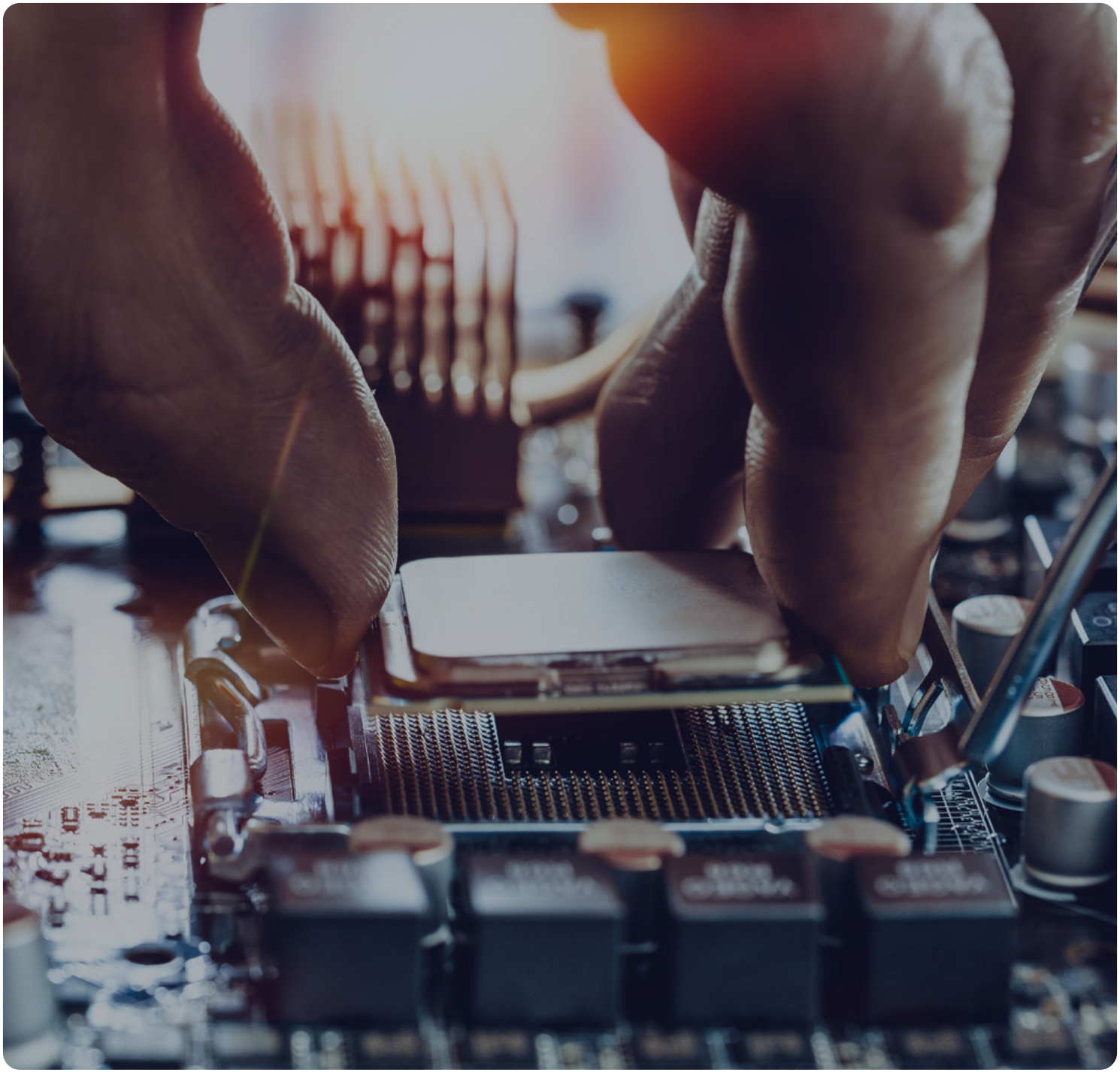
IONOS



The global chip shortage

Why semiconductors have become scarce and how the cloud can help in the short term.

IONOS

Content

IONOS

Summary

- As demand for chips continues to soar, supply chains will remain constrained at least for the foreseeable future. But almost nothing works without semiconductors these days. That means everyone – companies producing semiconductors and those relying on chips – should secure short-term workarounds while also aiming to build a more resilient supply chain.
- In order to be better prepared for the future, stakeholders need to re-examine short- and long-term strategies, factoring in lessons learned from chip shortages that have been disrupting manufacturing and beyond for months and years now.
- This paper will highlight factors and incidents that sparked the chip shortage and then tightened the bottleneck. This semiconductor shortage and the spillover into many sectors has raised fundamental questions about the health of the semiconductor supply chain and how to heal its shortcomings, with a focus on data centre server equipment and desktop workstations, in both the short and long term.
- Increased demand due to forced digitalisation and the impact of COVID-19, complicated by natural disasters, traffic interruption and lockdowns happening all at once have revealed fragilities in the supply chain. Market participants have to define what role they can play in strengthening the resilience of the supply chain, and find ways to minimise the effects of shortage at short notice, while not neglecting long-term planning.
- To identify the root causes of this shortage, businesses need to understand the dynamics within semiconductor manufacturing. What customers and markets are currently experiencing as a semiconductor shortage is in fact multiple shortages happening concurrently in different steps of the process, and supplier markets based on a multitude of dynamics and interdependencies. The interconnection between the underlying dynamics such as high barriers to entry, high geographic concentration, high fabrication utilisation and long manufacturing cycles is the reason why steeply increasing demand and external factors from natural disasters and human error to COVID-19-related lockdowns have disrupted the value chain since 2020.
- Consequently, none of the shortages in semiconductor manufacturing can be explained by just one cause. Most importantly, some of the underlying dynamics are unlikely to change in the short or medium term because they are deeply rooted in characteristics of semiconductor manufacturing.

- This white paper will explore and analyse the root causes of the global chip shortage by identifying key characteristics of semiconductor manufacturing. First and second order effects resulting from these characteristics reveal the lack of resilience within this vital value chain. Market mechanics, such as high market entry barriers, limited sources and extended manufacturing cycle times, will not change any time soon. There is no simple real short-term solution to this problem – it can only be addressed by long-term, strategic decisions. Nevertheless, an interim workaround is essential, and in fact, there is a possible solution.
- For many workloads the cloud could provide a way out of the problem. In choosing what might seem like an emergency exit, businesses have an opportunity to rethink and further develop their IT. So, the market needs to understand that strengthening the resilience of the global semiconductor value chain is a complex task for years to come, requiring structural changes, new business models and supplier relationships. And in the meantime the cloud may step in as more than just a stopgap measure. Acting as an IT infrastructure substitute is just one aspect. Collaboration enablement and digital cloud-based workplaces prolonging the life cycle of desktop hardware even further is another one.

IONOS

1 Introduction

As the world grapples with a massive chip shortage[1], which is causing global delays in the manufacture of cars, televisions, laptops, smartphones, and other electronic devices, including desktop computers and data centre servers and equipment, there's one thing that can be counted on. That is the reliability of cloud computing when it is needed. The chip shortage is just as global in its reasons as it is in its failure to satisfy a customer demand that has been bouncing back after a year plagued by the coronavirus pandemic. Many industries worldwide are still affected by the current semiconductor or chip bottleneck.

Almost every industry is dependent on semiconductors to some extent; Goldman Sachs estimates[2] for instance that around 169 industries globally are impacted by these chip shortages. The recent lack of semiconductors and their spillover damage to many other industries also explain governments' increased interest in the semiconductor industry and in a strengthened resilience of these critical value chains.[3] Meanwhile, semiconductor spending has climbed to unprecedented levels.[4]

The effects of global chip shortages may be reduced by cloud computing. Rather than one distinct shortage, we are currently facing multiple shortages at different processing steps and input streams within the supply chain – and these are in dire need of mitigation.

1. Duberstein, B. (2021): Is the Chip Shortage Over? Not So Fast, https://www.fool.com/investing/2021/07/26/is-the-chip-shortage-over-not-so-fast/.
2. Howley, D. (2021): These 169 industries are being hit by the global chip shortage, https://finance.yahoo.com/news/these-industries-are-hit-hardest-by-the-global-chip-shortage-122854251.html.
3. European Commission (2021): COMMISSION STAFF WORKING DOCUMENT - Strategic dependencies and capacities, https://ec.europa.eu/info/sites/default/files/swd-strategic-dependencies-capacities_en.pdf.
4. Gain, V. (2021): Semiconductor spending at an all-time high amid chip shortage, https://www.siliconrepublic.com/machines/semiconductor-spending-global-chip-shortage.

IONOS

2 How has the global semiconductor value chain become disrupted?

Disruption did not happen out of the blue. It's important to keep in mind that consumer electronics such as laptops, smartphones, tablets and PCs for home or office use make up the lion's share of global semiconductor demand in terms of quantity. Demand for these goods depends considerably on the general economic situation.

> The first and most important factor that disrupted the global semiconductor value chain was the sharply increasing demand for processor cores due to the COVID-19 pandemic and the US-China technology dispute.[5]

The switch to remote ways of working and learning is also a contributing factor. Since the first half of 2020, working from home and homeschooling have been the new normal in many countries. Many companies initially lacked the necessary equipment and infrastructure to enable remote working; as a consequence many PCs and laptops were bought in a very short space of time. As a result of remote work, video calls became the standard form of communication throughout society, creating high demand for data centre and server equipment. In addition, the obligation to stay at home due to curfews and lockdowns meant that many people purchased gaming consoles and other electronic devices.

Another reason for the increasing demand seems to be the US-China technology rivalry. When the US placed export bans on Huawei in 2019, some Chinese companies started hoarding chips[6] out of fear of facing similar challenges if put on the U.S. Entity List, thereby narrowing global availability.

Global semiconductor sales were 18% higher in 1Q21 than in 1Q20, 29% higher in 2Q21 than in 2Q20 and 28% higher in 3Q21 than in 3Q20.[7] Forecasts also predict that 11% more semiconductor units will be sold in 2022[8] after a rise of 25% in 2021 as a whole.

5. Hall, M. (2021): Surging Chip Demand, Digital Transformation and COVID-19 – Insights from Wells Fargo, https://www.semi.org/en/blogs/technology-trends/chip-demand-digital-transformation-covid19-insights-Wells-Fargo.

6. Campbell, C. (2021): Inside the Taiwan Firm That Makes the World's Tech Run, https://time.com/6102879/semiconductor-chip-shortage-tsmc/.

7. Semiconductor Industry Association (2022): Global Semiconductor Sales Increase 23.5% Year-to-Year in November; Industry Establishes Annual Record for Number of Semiconductors Sold, https://www.semiconductors.org/global-semiconductor-sales-increase-23-5-year-to-year-in-november-industry-establishes-annual-record-for-number-of-semiconductors-sold/.

8. IC Insights (2022): 2022 Semiconductor Sales to Grow 11% After Surging 25% in 2021, https://www.icinsights.com/news/bulletins/2022-Semiconductor-Sales-To-Grow-11-After-Surging-25-In-2021/.

In addition to the steep increases in demand for semiconductors, many external shocks have put a further strain on the global semiconductor value chain in the last two years. In some cases, **government-ordered lockdowns**[9] have forced several plants to shut down entirely.

Natural disasters and power outages have brought further disruptions to an already strained supply chain.

9. Wu, D. et al. (2021): Chip Shortage Set to Worsen as Covid Rampages Through Malaysia, https://www.bloomberg.com/news/articles/2021-08-23/chip-shortage-set-to-worsen-as-covid-rampages-through-malaysia.

3 Semiconductor value chain characteristics

Characteristics of the semiconductor value chain
There are six characteristics defining the inner fabric of semiconductor manufacturing:

- Considerably high division of labour
- Very high capital intensity
- Very high knowledge intensity
- Relatively long manufacturing cycle times
- Pronounced transnationality and
- Distinctive lock-in effects

Division of labour: The semiconductor industry's high level of innovation and efficiency is rooted in a highly specialised and interdependent ecosystem, where a high division of labour is distinctive not just in supplier markets, but across the stages of production:

- Design
- Wafer fabrication and assembly
- Testing
- Packaging

Modern chip production involves thousands of highly specialised companies. The first process step, chip design, relies on access to third-party intellectual property vendors and electronic design automation tools. The two process steps that follow, front-end and back-end manufacturing, depend on a variety of chemical suppliers, manufacturing equipment vendors, cleanrooms, and process automation, to name just a few. There is a reason that large semiconductor manufacturers such as Intel and TSMC are recognised as outstanding suppliers each year – the advanced level of labour comes with economic pressure to constantly innovate.[10]

Capital intensity: Semiconductor manufacturing is highly capital intensive, especially cutting-edge wafer fabrication. Building a modern semiconductor plant (for 5nm chips for instance) requires $20 billion in capital expenditure[11], and a single cutting-edge lithography machine from ASML costs $175 million.

10. McKinsey&Company (2011): McKinsey on Semiconductors, https://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/semiconductors/pdfs/mosc_1_revised.ashx

11. LD Investments (2021): ASML: Riding The Chip Wave, https://seekingalpha.com/article/4422340-asml-riding-chip-wave.

Large fabs (as semiconductor plants are called by insiders) need around 20 of them. These extremely high capital expenditures for cutting-edge manufacturing are one reason the market has been heavily consolidated over the past 20 years. The only three companies that still operate cutting-edge fabs (TSMC, Samsung and Intel) accounted for more than 50% ($59.4 billion) of global semiconductor capital spending in 2020.

Knowledge intensity: Across all industries, semiconductor companies have one of the highest research and development expenditures. In 2020, the semiconductor industry spent more than 14% of revenue on R&D. Fabless chip design companies that outsource manufacturing, such as Nvidia, AMD and MediaTek, typically invest around 20–25% of their revenue in R&D. However, semiconductor manufacturing also relies on extensive process knowledge based on decades of experience and skilled workers. To develop future manufacturing processes, foundries (a synonym for fabs) and integrated device manufacturers (IDM) have R&D collaborations with research and technology organisations, equipment and chemical suppliers and their fabless customers. Historically, US companies and institutions have had the highest share of global semiconductor R&D, but South Korean, Taiwanese and Chinese companies have become important R&D partners in recent years.

Manufacturing cycle times: Producing a single chip requires up to 1,500 steps[12], each based on hundreds of variables. Some process steps during wafer fabrication, such as oxidation and coating, lithography, etching and doping, are repeated hundreds of times, depending on the specific chip. Thus, wafer fabrication from start to finish takes, on average, 12 weeks but can last up to 20 weeks. Then, the wafers are delivered to back-end manufacturers for assembly, testing and packaging. In total, producing a semiconductor can take more than six months. Consequently, the industry is characterised by long-term planning with customers placing their orders well in advance.

Transnationality: The US, Japan, South Korea, Taiwan, the EU, China and several Southeast Asian countries play critical roles within the semiconductor value chain. No region is able to source all necessary inputs and perform every process step domestically.

Lock-in effects: In this transnational value chain, having close connections within the ecosystem is essential to develop competitive products. However, this in turn creates strong lock-in effects between companies, making it harder to switch suppliers or manufacturers. One example is the close business relationships between chip design companies and foundries for contract manufacturing.

12. University Wafer: How do you make Silicon Wafers into PC Chips, https://www.universitywafer.com/how-to-make-silicon-wafers-into-computer-chips.html.

Choosing a foundry's process node for a new chip design is a long-term, strategic decision as chip design companies cannot simply switch nodes once the chip has been developed (i.e. from manufacturer 1's 5nm process to manufacturer 2's 5nm process).[13] A chip design, especially for cutting-edge chips, is always designed for and therefore dependent on a fab's specific process node. Another example is lock-in between fabs, manufacturing equipment and chemicals. Manufacturing equipment might work best with chemicals from a specific vendor because of an longstanding R&D collaboration, making it unlikely that fabs will switch equipment vendors for fear of disrupting the production process. These functional interactions create strong lock-in effects across the complete value chain.

The interdependence of these six characteristics of the global semiconductor manufacturing industry have led to further dynamics within the value chain over the past decades. These dynamics are **increasingly high market entry barriers**, the need for high fab utilisation due to economic pressure that in turn leads to conservative capacity investments, dependence on limited sources for inputs and manufacturing, and a high geographic concentration for certain production steps.

13. Bauer, H. et al. (2020): Semiconductor design and manufacturing: Achieving leading-edge capabilities, https://www.mckinsey.com/industries/advanced-electronics/our-insights/semiconductor-design-and-manufacturing-achieving-leading-edge-capabilities.

IONOS

4 Sharply increasing demand and external factors

The semiconductor value chain encounters challenges such as limited agility and resilience[14], which is why sudden demand surges and certain types of external shocks, such as natural disasters or corona-related lockdowns, have potentially highly disruptive effects with huge spillover damages. The reasons for that lack of agility and resilience are specific dynamics that stem from the interplay of the value chain's circumstances.

Why did skyrocketing demand disrupt the value chain?

Sudden increases in demand caught the semiconductor value chain off balance because of high market entry barriers, high fab utilisation and limited sources.

High market entry barriers

The value chain did not cope well with the surge in demand because the high market entry barriers in semiconductor manufacturing make it impossible for any company from outside the ecosystem to fill in if demand exceeds supply. The high market entry barriers result from the high capital intensity and high knowledge intensity.

The barriers for taking up production of semiconductor manufacturing including complementary supplier markets such as manufacturing equipment and chemicals mean that even in the mid-term, the value chain can only rely on existing companies.

High fab utilisation

Another reason why rapidly growing demand can be highly disruptive to the value chain is the need for constant high fab utilisation rates in semiconductor manufacturing. As fabrication owners need to invest substantial amounts of money in equipping their fabs, these huge capital investments are profitable only if the fab plants operate 24/7 with utilisation rates of 80% or above. So, the operational goal of high fab utilisation is a direct result of the high capital intensity of semiconductor manufacturing. Operating close to full capacity is the only way to amortise the high investment costs. However, this means that the market has very limited spare fabrication capacity and fabs

What is a fab?

A complete new fab build takes at least three years and requires a lot of CAPEX – up to $10 billion in investment.

14. Alam, S. (2021): Chip shortages impact for supply chain resiliency, https://www.accenture.com/us-en/blogs/high-tech/chip-shortages-impact-for-supply-chain-resiliency.

are quickly booked out if there is an unexpected increase in demand. In 3Q20, when shortages started to materialise in earnest, fab utilisation rate was already at 95% or above.[15]

The business goal of high fab utilisation rates in a market with fluctuating demand leads to another second-order effect: conservative capacity investments.

Even as fab owners such as Samsung, TSMC, Intel and many others announced substantial capacity investments in the coming years, Intel poured $20 billion into two facilities in the US[16], wanting to avoid overcapacity at all costs. In the future, semiconductor manufacturing will also be defined by periods of oversupply and undersupply. Fab owners have a strong economic incentive to utilise as much existing capacity as possible before investing in new fabs that cost up to $20 billion for a cutting-edge production facility.

Inside a fab: take a closer look

A semiconductor fabrication plant, commonly known as a fab, is a factory where devices such as integrated circuit (IC) chips are manufactured. These are the chips we find in everyday electrical and electronic devices. The making of a semiconductor device involves a multiple-step lithography sequence to create a pattern in the photoresist, as well as chemical processing, during which electronic circuits are gradually created on a silicon wafer. These steps include etching, ion implantation, deposition and photoresist coating.

ICs are made of layers, from about 0.005 to 0.1 mm thick, that are built on the semiconductor substrate one layer at a time, with perhaps 50 or more layers in a final chip. After adding a layer, so-called deposition, the layer is etched, using lines and geometric shapes in the exact locations where the material is deposited.

The entire manufacturing process, from start to packaged chips ready for shipment, takes six to eight weeks. All fabrication takes place inside the cleanrooms of these fabs. In more advanced semiconductor devices, such as modern 7 nm nodes, fabrication can take between 11-13 weeks on average.

The heart of a fab is the cleanroom, an area where the environment is controlled to eliminate dust on a nanoscale. All fabrication steps take place here. It also houses the lithography system and other machinery required for IC production. Under the desk floor is the socalled sub fab, which contains auxiliary equipment such as the drive laser. The utility fab – where the pumping and abatement systems for vacuum and cooling are located – is usually found one floor below this.

Information provided by ASML

15. Business Wire (2021): UMC Reports Fourth Quarter 2020 Results, https://www.businesswire.com/news/home/20210127005391/en/UMC-Reports-Fourth-Quarter-2020-Results.

16. Bobrowsky, M. (2022): Intel to Invest at Least $20 Billion in Ohio Chip-Making Facility, https://www.wsj.com/articles/intel-to-invest-at-least-20-billion-in-ohio-chip-making-facility-11642750760.

Limited sources

Another factor that contributes to supply constraints within the value chain is the reliance on limited or single sources. There is no abundance of suppliers for a particular process step, type of equipment or chemical because of the industry's high knowledge intensity, high division of labour and strong lock-in effects. The interplay of these three characteristics means that quasi-monopolies are very common throughout the value chain because companies have to specialise to stay competitive.

- TSMC and Samsung are increasingly becoming cutting-edge foundries.[17]
- ASML is said to be the only supplier for very advanced lithography equipment.[18]
- Tokyo Electron has 90% of the global market for other types of equipment such as coaters and developers.
- Specialty chemicals and wafers are also regularly single-sourced, simply because of lack of competition or highly customised processes.

17. Shilov, A. (2021): TSMC and Samsung Foundry Becoming Dominant Makers of Advanced Chips, https://www.tomshardware.com/news/tsmc-and-samsung-foundry-becoming-dominant-makers-of-advanced-chips.

18. Shead, S. (2021): TECH Investors are going wild over a Dutch chip firm. And you've probably never heard of it, https://www.cnbc.com/2021/11/24/asml-the-biggest-company-in-europe-youve-probably-never-heard-of.html.

5 Why did increasing demand overstretch the value chain?

The semiconductor value chain was disrupted not only by unforeseen increases in demand, but also by a series of natural disasters, human error and corona-related lockdowns often resulting in temporary production loss. The global semiconductor value chain is susceptible to these types of external shocks not merely because of the long manufacturing cycle times but also because of two additional factors – high geographic concentration and limited sources.

Extended manufacturing cycle times

Simply because wafer fabrication alone takes at least 12 weeks or more, a single power outage in a fab might lead to 12 weeks of severely hampered or lost production.[19]

High geographic concentration

The semiconductor value chain is spread across several geographic regions and jurisdictions. However, chip manufacturing is relatively concentrated, especially cutting-edge wafer fabrication and back-end capacities such as assembly, testing and packaging. There are several reasons for this, including government incentives and outsourcing of labour-intensive production steps, and they derive from decades of specialisation through a high division of labour in a transnational value chain. For example, Samsung and TSMC account for 75% of foundry production capacity globally.[20]

East Asia is also the most important region for back-end manufacturing. This high geographic concentration increases the risk of supply chain disruptions in the event of disasters or lockdowns.

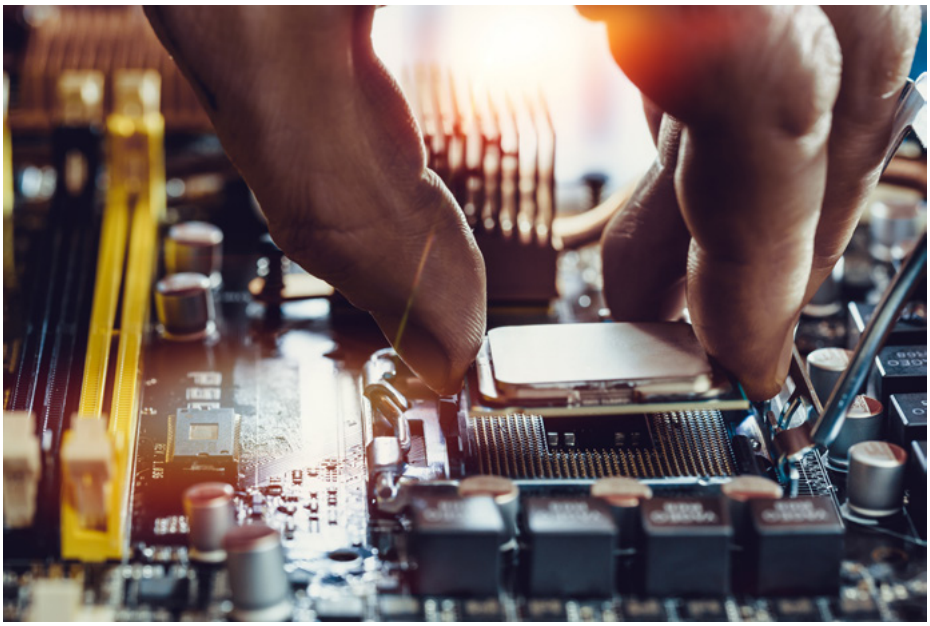
19. Bradshaw, T. / White, E. (2021): Texas winter storm blackouts hit chip production, https://www.ft.com/content/ec2f93ad-d23c-4ff4-867a-59385d1cc8a6.

20. Deloitte (2020): Rise of the "Big 4" The semiconductor industry in Asia Pacific, https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/technology-media-telecommunications/cn-tmt-rise-of-the-big-4-en-082820.pdf.

Limited sources in niches

The inability to quickly switch to second or third sources is a challenge not only during periods of skyrocketing demand but also in the event of an external shock. The high knowledge intensity, high division of labour and strong lock-in effects created specialised companies that have become indispensable. That is why external shocks can be especially disruptive for fabs, chemicals and wafer suppliers.

In summary, a deluge in demand and external shocks were highly disruptive because of the underlying dynamics within the semiconductor value chain. High market entry barriers, the economic need for high fab utilisation rates and limited sources throughout the value chain resulted in substantial disruption to the value chain when demand surged. In addition, external shocks throughout 2020 and 2021, from natural disasters to human error and lockdowns, further disrupted the value chain because of its – once again – long manufacturing cycle times, high geographic concentration and limited sources or inability to easily switch to a second or third source. Therefore, what customers and markets are currently experiencing as a semiconductor shortage are, in fact, multiple shortages happening simultaneously in different process steps and supplier markets based on a multitude of dynamics and dependencies. Most importantly, some of these underlying dynamics are unlikely to change in the future because they are inherent characteristics of this value chain.



6 Why did external factors disrupt the value chain?

6.1 Special cases of semiconductor shortages

Load factor in fab utilisation
Foundries' as well as IDMs' production capacity was stretched to the limit as early as 4Q20 when automotive suppliers ran out of chips.[21] The existing wafer capacity was already completely reserved by other customers, and there was simply no overcapacity available to accommodate car manufacturers. Additionally, with a global market share of less than 12%, automotive semiconductors used to play a smaller role in the market compared to consumer electronics or telecommunication – a position the automotive industry may have misjudged. At the end of 2020, lead times for automotive microcontrollers had already extended to 14 weeks and continued to rise constantly.

Extended manufacturing cycle times
When the supply is stable, manufacturing a chip takes four to six months.[22] These long production times are incompatible with the highly complex and non-transparent just-in-time supply chain that's typical of the automotive sector.

Limited sources
Automotive chips have stringent safety requirements such as weather resistance, fault tolerance, redundancy, etc. that must be certified along with the production process. This limits the number of fabs that automotive chip suppliers can rely on, putting further stress on an already strained supply chain during times of scarcity.

21. Design & Reuse (2020): Total Revenue of Top 10 Foundries Expected to Increase by 18% YoY in 4Q20 While UMC Overtakes GlobalFoundries for Third Place, Says TrendForce, https://www.design-reuse.com/news/49112/revenue-ranking-to-10-foundries-4q20.html.

22. ASML: How microchips are made, https://www.asml.com/en/technology/all-about-microchips/how-microchips-are-made.

6.2 Chemical shortages

Front-end and back-end manufacturing relies on hundreds of different chemicals and materials. The lack of one type of chemical can have a domino effect throughout the entire value chain and can interrupt the whole manufacturing process, as was the case with Ajinomoto Build-up Film (ABF) substrates. ABF substrates are an unexciting, but essential part of every chip that uses laminated packaging. Functioning as a layer that connects different components within a chip, ABF substrates are widely used in chips for graphics cards, servers, smartphones and laptops, just to name a few.

In addition to the ABF substrate supply constraints caused by spiking demand for games consoles and graphic cards, two fires at a major substrate supplier, Unimicron, in October 2020 and again in February 2021, and problems with comparatively low yield (<70%) at 3 different suppliers (Unimicron, Nan Ya, Kinsus) led to further shortages. Large customers such as AMD, TSMC, Samsung and Intel are planning strategic investments and partnerships with suppliers such as Unimicron and Ibiden to secure their ABF substrate supplies. Three familiar dynamics led to the shortages of this chemical substrate:

- Conservative capacity investments
- Limited sources
- Geographic concentration

Conservative capacity investments
As substrates are a low-margin business, substrate suppliers have been hesitant to expand their production capacity. This, in turn, led to many years of underinvestment in additional capacities. When the market encountered external shocks and skyrocketing demand at the same time, suppliers had no room to produce more, as they were already operating near to or at full capacity. Some analysts predict that demand will decrease after the pandemic ends in 2022/2023 and then increase again or just stabilise before growing again in 2025. It is not unreasonable to expect that these variations in utilisation could be expensive for manufacturers. Beyond this, new competitors could have entered the market in a few years' time. These uncertainties might be another reason why producers are pursuing a conservative investment strategy.

Constrained sources and external shocks
As mentioned above, Unimicron experienced two fires at its plants, which led some customers to switch to a smaller supplier, Nan Ya (6% of global market share).[23] Consequently, Nan Ya could not compensate for the demand of all the customers that usually source their ABF substrates from Unimicron. In reaction to the tight supply, many customers, such as Nvidia, now plan to diversify their supplier network for ABF substrates.

[23] Nanya Technology Corporation (2021): 2020 Annual Report, https://www.nanya.com/en/Activity?Action=Get_IRannualreport_FileName&Id=20.

Geographic concentration
The leading ABF substrate suppliers are based in Taiwan (Unimicron Technology, Kinsus Interconnect Technology, Na Ya) and Japan (Ibiden, Shinko Electric). This high geographic concentration poses risks during natural disasters or pandemic-related lockdowns in these regions.

6.3 (Back-end) equipment shortages

Expanding capacity in existing fabs can be done much more quickly than building new fabs (18 months as opposed to three years). However, supply constraints for certain types of manufacturing equipment pose a challenge to short-term capacity expansions. One example is wire-bonders, which are often used for packaging as one process step in back-end manufacturing of trailing-edge components such as microcontrollers. ASE Group, the largest chip packaging company, reported that wire-bonding makes up 80% of their chip packaging processes and that lead times for wire-bond equipment, for example, from market leader Kulicke & Soffa, rose to 40–50 weeks in 1Q21.[24] So, short-term back-end capacity expansion will take longer due to equipment shortages stemming from the interplay of two dynamics: limited sources and conservative investments in capacity.

Limited sources
As the demand for trailing-edge chips gained traction during the beginning of the shortage, Kulicke & Soffa was already running at full capacity. Consequently, chip packaging companies such as ASE Technology found that their wire-bonding capacity was 30–40% below demand. Packaging companies and their equipment suppliers maintain close relationships on the one hand but strong lock-in effects on the other, thereby making it infeasible to quickly source equipment elsewhere.

Conservative capacity investments
Mature nodes, front-end as well as back-end, have seen very limited capacity investments in recent years. Thus, equipment suppliers have increasingly focused on equipment for modern fabs (300mm wafers) instead of equipment for older fabs (200mm wafers). The sudden demand for mature nodes cannot be met due to the lack of node equipment at the same level of maturity. As long as equipment suppliers are not able to meet the demand for chip making machines, foundries and packaging companies cannot expand their capacity.

24. CSIMarket (2022): KLIC Sales vs. its Competitors Q4 2021, https://csimarket.com/stocks/compet_glance.php?code=KLIC.

6.4 Disruptions in wafer fabrication

In February 2021, Samsung, NXP and Infineon had to temporarily suspend plant operations for several weeks due to winter storms causing power outages in Texas,[25] resulting in lost production and hundreds of millions of dollars in lost revenue. The power outage damaged not only manufacturing equipment but also components in the facilities' infrastructure that had been expected to last the life of the facility. The power outage exacerbated disruptions in an already strained supply network. External shocks, such as power outages, disrupt not just wafer fabrication but also the entire value chain mainly because of two dynamics: limited sources and long manufacturing cycle times.

6.5 Limited sources

Customers of the Samsung foundry were not able to simply move their production to a different foundry because a chip design is always based on a dedicated process node from a specific company. As wafer fabrication, on average, takes up to three months, a considerable amount of production is lost during an external shock and lead times quickly increase.

25. BBC News (2021): Texas freeze shuts chip factories amid shortages, https://www.bbc.com/news/technology-56114503.

IONOS

7 Why strengthening the resilience of the semiconductor value chain is an ambitious undertaking

Challenges of adapting to fluctuating demand

Reviewing four cases in more depth has illustrated the interplay of the many dynamics that led to a variety of shortages at different process steps and inputs. The global semiconductor value chain cannot quickly adapt to sudden increases in demand mainly because of the interaction of three dynamic factors all rooted in fundamental characteristics of semiconductor manufacturing: high market entry barriers, high fab utilisation and limited sources.

Market entry barriers – in particular high capital intensity and high knowledge intensity – and the challenge of limited sources (high knowledge intensity and high division of labour as well as strong lock-in effects) will not change any time soon.

However, the operational goal of high fab utilisation and the resulting conservative capacity expansions due to fluctuating and uncertain demand are not set in stone, although they are hard to change. The conflicting priorities between high fab utilisation and the ability to cope with quickly changing demand have led to many boom-and-bust cycles in the semiconductor market. Investments in additional capacity or in a new fab are made only if an expansion is economically viable – when high utilisation rates can be achieved quickly. As a consequence, new fabs are built when demand for chips is larger than the supply produced by fab capacity, and shortages and stockpiling are already occurring. Fabs make more money in times of scarcity and their customers so far have not had incentives to pay for spare capacity.

> Initial government subsidies, short-lived guarantees and simply building more fabrication plants will not fundamentally change this dynamic because future fabs will also have economic pressure to achieve high utilisation rates.

Because it takes at least a year to expand an existing fab and around three years to build and ramp up a new fab, demand visibility is crucial for semiconductor manufacturing. The current shortages have the potential to change the business relationships between fabs and their customers to improve demand visibility and make the value chain more resilient. **Some foundries require long-term agreements and exact prepayments from their customers for future fabs in exchange for guaranteed wafer capacity per customer.**[26] Another development is non-cancellable, non-refundable chip orders.

---

[26] Shilov, A. (2021): TSMC Gets Billions in Pre-Payments for Fab Capacity, https://www.tomshardware.com/news/tsmc-collects-huge-prepayments.

The inherent characteristics and dynamics of the semiconductor value chain show that adding capacity alone is not a successful strategy for making the supply chain more resilient and agile in terms of sudden demand increases. Achieving more resilience is deeply rooted in the question of how to incentivise overcapacity.

8 Resilience against future impacts

The examples explored above show that the global semiconductor value chain struggles to cope with external shocks such as natural disasters, human error and lockdowns, mainly due to single or limited sources, high geographic concentration and long manufacturing cycle times. Long manufacturing cycle times are a structural feature of the complex process of semiconductor manufacturing, and nothing can be done to change that.The same is not necessarily true for high geographic concentration, primarily because of the high division of labour and the use of single or limited sources throughout the value chain. Diversification, especially as natural disasters are becoming more frequent due to climate change, increasingly becomes unavoidable. As pointed out in the section on back-end equipment shortages, external shocks cannot only be narrowed down to disruptions in manufacturing processes.

Identifying bottlenecks in the value chain where companies are indispensable due to the high division of labour and lock-in effects, leaving customers dependent on single or limited sources, may be the first step. Consequently, possible alternative sources can be explored, at least in the long term, or diversification of those quasi-monopolies may be incentivised. Similarly, if a region accounts for the lion's share of a certain production step such as Taiwan for cutting-edge wafer fabrication, or the provision of a critical input, it is highly likely that most companies based in that region will have production outages when external shocks occur. The snow storm in the US, the lockdown measures in Malaysia and the earthquake in Taiwan have already demonstrated the disruptive potential of the interplay between external shocks and geographic concentration.

As diversification is not always possible, even in the long term, semiconductor customers, such as automobile manufacturers, must prepare better for chip supply disruptions. A first step is increased transparency of the value chain but also closer relationships with suppliers and strategic inventories for production-critical chips – measures that seem to have helped Toyota keep its car production running much longer than most of its competitors despite chip shortages.

9 Conclusion

As we have seen, the global semiconductor value chain is not in good shape. It is highly efficient and innovative, but prone to disruption by external shock. It does not adapt well to rapid increases in demand from the fabs and their customers, who often have divergent long-term business interests. This is not new to the semiconductor market; in fact, this is neither the first nor the last boom-and-bust cycle.

However, chips play a much more critical role in almost every sector of today's industry than 10 or 20 years ago. To better cope with demand surges, fabs need an economic incentive for overcapacity instead of striving for 85% and higher fab utilisation rates. Similarly, making the semiconductor value chain more resilient to external shocks involves more transparency as a first step. But any supply chain that depends on chips, such as the automotive supply chain, will also need to invest in substantial inventory on their own and better supplier relationships.

Governments can certainly provide the right incentives to lessen the high geographic concentration in the long term, for example in cutting-edge wafer fabrication and back-end manufacturing, at least to some extent. They may also be able to push industry toward increased supply chain transparency, a better flow of information and strategic inventories. But some of the key challenges within the global semiconductor value chain come down to business models and supplier relationships that are hard to change from the outside.

The global chip shortage continues to wreak havoc around the world, and no easy solution is in sight for those awaiting vital components or equipment. Tech leaders have had to take a flexible approach to the crisis, and for many working in IT, the big changes and rapid digital transformation of the coronavirus pandemic have been kind to them, strengthening their role within companies. With a shortage of key equipment set to continue for some months, strong relationships with vendors will be key for IT teams.

Buffer capacities might play a role too. Capacities which are built up by service providers to be let in times of surging demand can bridge demand overhang on a short-term basis. In the long run, those service providers could act as a more predictable commercial bulk purchaser for chip manufacturers.

IONOS

10 How the cloud can help in the short term

The way out: Moving IT to the cloud

Functioning IT is essential for almost all companies and administrative bodies. While some components are now more readily available than in 2021, others remain scarce. "The semiconductor industry is massive and diverse," asserts a leading analyst at Gartner. "So although some parts of the industry are doing better, other parts aren't, and the end effect from a purchasing perspective is the same: so the microprocessor part of the business is doing okay at the moment, but the power control part isn't, and you can't make PCs or laptops without power components."

For some IT departments, this has led to trouble getting hold of the technology needed to operate on-premises data centres according to GlobalData. Data centres have not been hit as hard by the shortage of core CPUs – these chips are being produced with high priority because they are profitable product lines. Cloud service providers have well established procurement relationships with manufacturers and can leverage these sources relatively easily. But cores are not the only component of a data centre, and lengthening lead times for network switches or power chips and resistors have had an effect.

The shortage has also coincided with a spike in memory prices.

How have tech leaders responded to the chip shortage?

For other tech leaders, the chip shortage has meant making some adjustments to the way they work. The shortage of equipment for data centres has exacerbated the trend to move services to the cloud. "One option some enterprises are resorting to is the public cloud as an alternative to struggling to build or extend their own custom data farms within a feasible time frame."

As Forrester has it:

"Run to the cloud. Speaking of cloud providers, this could be your right move. They have plenty of capacity to serve your needs, and you can let them worry about the chips. It's not always the right option, but it is often a good one."

Resorting to the public cloud can be a sound way to ensure a decent time-to-market for businesses' IT workloads. Cloud providers have enough size and negotiating power to mitigate the demand issues; while a typical IT organisation purchases 10 to 50 servers a year, cloud providers are buying them by the hundreds or thousands.

Luckily, IT departments have avoided some of the worst effects of the chip shortage because their deployments tend to be planned. This contrasts with

the automotive industry, which has suffered big problems due to its reliance on just-in-time supply chains. Estimates on when the chip shortage might ease vary from the later months of 2022 as Gartner has it or even up to 2023 on some components.[27] Many chip companies are investing heavily in capacity with new chip foundries being constructed around the world, but most of these will take at least another year to come online. It looks like the shortage will ease sometime in the first half of 2023 at the earliest. Some lagged effects could drag on well beyond that. So, turning to the cloud is not the worst of ideas.

Opportunities for digital transformation

Since necessity is the mother of invention, those leaders have an opportunity to re-architect their IT structure by moving from hardware-based inhouse solutions to cloud-based ones. Businesses are in a sink-or-swim situation, but there's an opportunity here to move away from hardware-based equipment to more cloud-based solutions. This trend had already been kick-started by the pandemic; even as lockdowns lifted, many organisations around the globe recognised that remote and hybrid work was here to stay. At the same time entire business models have moved online (e.g. online retail because people still wanted or needed to consume). So, all the things people would need in daily work had to be moved and made operational in their homes and not out of home any more. Services delivering just that had to rapidly scale processes, staff and IT.

It may seem likely that the cloud would be similarly affected by the chip shortage, but that is not the case. The shortage is having less of an impact on public cloud service providers because of their buying power. They purchased a great deal of infrastructure years in advance of demand and are busy adding additional capacity all the time.

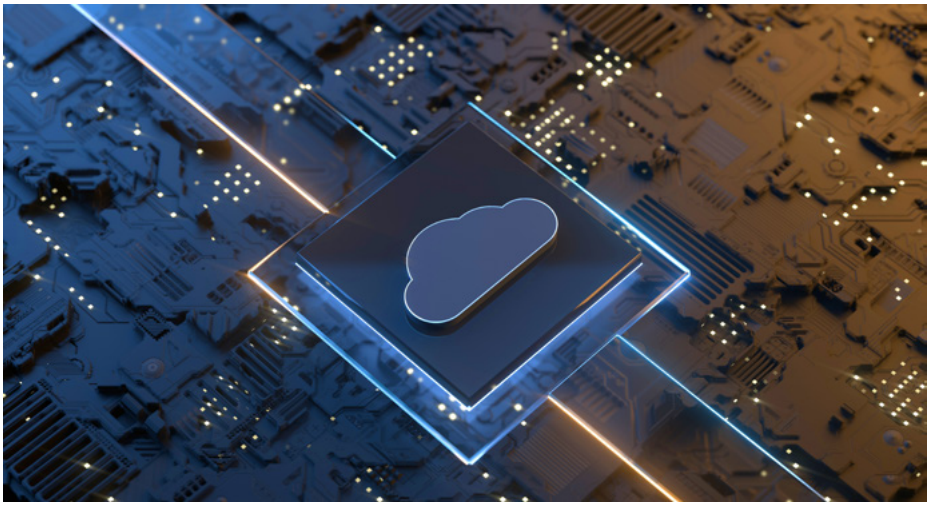
First, the shortage impacts traditional enterprise data centres more than cloud providers. The good news for cloud providers, or those who use cloud providers, is that they are less sensitive to chip price and availability issues compared to private data centre owners. And there are reasons for this:

- Cloud providers do a much better job of sharing chip-based resources, given that they leverage virtualised and multi-tenant systems. The typical data centre will not be as efficient at sharing chip-based resources, no matter if they are virtualised or not.

27. Shead, S. (2021): The global chip shortage could last until 2023, https://www.cnbc.com/2021/05/12/the-global-chip-shortage-could-last-until-2023-.html.

- Cloud providers can keep prices lower per processing cycle because they have a much longer-term view of pricing and its effects. It is to their advantage to keep usage prices low since the number of customers they acquire translates directly into long-term recurring income. For the standard data centre, it is just sunk costs that will not be fully utilised for many years.

Second, cloud service providers now drive more innovations of chips used in cloud computing systems. Some cloud providers are inventing, producing and leveraging their own chipsets. Because many large cloud providers now control all steps of the chip development process and the chips are optimised for their specific requirements, these providers no longer rely on the chip producers for their innovations or their chip costs and power optimisations. Other cloud providers seek close cooperation with chip manufacturers in order to align on long-term roadmaps with chip production planning. Some if not all private data centre owners cannot profitably operate and innovate at the same levels as cloud providers' experience curve deliver.



IONOS

11 Multi-cloud for greater independence

When you think about making use of the cloud as the means of overcoming semiconductor and network equipment shortages with a mid- to long-term perspective, cloud architecture design might better be orientated to more than one public cloud in parallel. Aligning IT workload requirements with the opportunities of different cloud service offerings will give you even greater independence and strategic leeway against future bottlenecks for IT fittings and infrastructure originating from the cloud provider itself.

Digital workplaces in the cloud

Beyond that, waiting times for laptops and other desktop equipment continue to drag on due to the chip shortage, so cloud desktops are becoming more popular. Instead of waiting, organisations can apply the concept of right-sizing to their business's digital needs and swap hardware for software and cloud-based solutions.

Worldwide end-user spending on public cloud services was predicted to grow 18.4% in 2020 alone, to total $304.9 billion, according to Gartner.[28] And this has really created a need for cloud-hosted desktops, which offer many clear business benefits – especially in light of the chip shortage.

Advantages of cloud desktops

There are five primary advantages to using cloud desktops:

- Unlimited performance without become obsolete
- Greater business agility
- Better security
- Comprehensive observability
- Flexible pricing models

28. Gartner (2021): Gartner Forecasts Worldwide Public Cloud End-User Spending to Grow 18% in 2021, https://www.gartner.com/en/newsroom/press-releases/2020-11-17-gartner-forecasts-worldwide-public-cloud-end-user-spending-to-grow-18-percent-in-2021#:~:text=Worldwide%20end%2Duser%20spending%20on,research%20vice%20president%20at%20Gartner.

IONOS

12 Why the cloud isn't just a short-term solution

Roadmap to business modernisation

Cloud computing providers are not the only solution, but they do have the upper hand when you consider the chip shortage from all angles. They will feel some of the effects of the technology shortages, but the current crunch is much less likely to affect their operations or prices.

If anything, scarcity will continue to drive independent innovation among cloud providers or will drive operation effectiveness amongst cloud companies.

Global spending on cloud services reached a new record in Q3 2021, at $49.4 billion,[29] however the impact of the global chip shortage on the data centre industry is imminent.

The findings, published by Canalys and Synergy Research Group, showed that growth was driven by hybrid and remote work as well as the growing use of industry-specific cloud applications. And while these latest figures reinforced the 35% annual growth Canalys and Synergy Research Group reported for Q1, the party could soon be coming to an end as data centre component providers see longer lead times and higher prices that will inevitably be passed on to the largest providers.

Besides managing supply chains to the best of their abilities, the providers building an advantage are focused on developing their go-to-market channels along with their product portfolios to catch up with the increasingly wide variety of customer use cases that has fueled demand since the start of the pandemic.

Companies who are looking to purchase physical servers are quickly realising that the wait time is unsustainable, and being forced to rethink their approach to infrastructure. Unlike most enterprises, many cloud providers have been able to buy ahead of time and in bulk, making them less affected by the manufacturing scarcity. This confluence of factors is contributing to a surge in cloud services, offering a faster, more efficient solution to an unruly amount of new data.

[29] Canalys (2021): Global cloud services spend hits record US$49.4 billion in Q3 2021, https://www.canalys.com/newsroom/global-cloud-services-q3-2021.

Cloud infrastructure helps free up server capacity, which in turn takes the pressure off organisations that would otherwise be waiting for new hardware. But the move to the cloud is not one size fits all – first taking the time to identify workload needs will enable companies to create an environment that best fits their business. The cloud solution of choice, whether it be the public cloud, the private cloud, or the hybrid cloud, depends on a variety of factors, use cases and limitations. When choosing the type of cloud, decision makers should consider the unique benefits of each option:

Public clouds are fully hosted by a third party and shared across organisations. These solutions offer:

- Lower costs
- Less Capex, fair Opex
- Extended flexibility
- Increased scalability
- Access to next-generation technologies
- Ease of use
- No lock-in

While supply chain issues have caused endless headaches, they have allowed many businesses to forge a new path for continued growth. With the right strategy and policies in place, cloud migration can lead to long-term value creation for businesses of any size.

As data centre hardware lead times stretch to 52 weeks, companies face the stark choice of either pausing projects or finding an alternative.

When the pandemic forced people to work, learn and play from home, sales of laptops, smartphones/tablets and gaming systems – and the semiconductors that power them – soared, along with streaming and other cloud services.

In fact, IDC reports that worldwide semiconductor sales grew to $464 billion in 2020, an increase of almost 11% over 2019. The market will grow to 12.5% in 2021, they say, reaching $522 billion – growth driven largely by consumers, who will continue to purchase digital products, consume data and adopt cloud services at unprecedented rates.[30]

Rather than wait on hardware – network switch manufacturers are reporting lead times of 52 weeks – many organisations are turning to the cloud to accelerate business-critical projects, increase agility and enhance business resiliency.

30. IDC (2021): Worldwide Semiconductor Revenue Grew 10.8% in 2020 to $464 Billion, Growth Will Accelerate This Year Despite Market Shortages, According to IDC, https://www.idc.com/getdoc.jsp?containerId=prUS47664821.

The result is that the shortage of available data centre hardware is directly affecting organisations with on-premises IT that need to add capacity for new projects or workloads. Businesses are therefore left with a stark choice: pause programs, perhaps indefinitely, or move to the public cloud.

A move to the cloud as a move to business modernisation

The current semiconductor shortage is a solvable issue. During previous shortages, the solution was to bring forward purchases of servers and semiconductor-reliant hardware, either with existing capital or through financing. Nowadays, when there is simply no hardware obtainable and capital bound to investments negatively influences solvency, renting it for time of actual use might be a sound approach. The most recently released IDC figures indicate that the public cloud is capable of meeting the recent increase in demand.[31]

Cloud might be the long-term fix

But the cloud also provides a faster, more cost-effective and long-term fix. With security, maintenance and scalability built in, the cloud is ideal for organisations that cannot get the hardware they need. And it offers plenty of data centre capacity. Simply spin up servers with your cloud provider and scale them back when no longer needed.

It is important to recognise that moving to the cloud does not have to be an all or nothing approach.[32] However, it should be acknowledged that there are also complexities to cloud deployments, especially large ones, at both the transitional and operational stages. These mean that organisations should ideally partner with an expert who can first help manage the move to the cloud and then offer a continued expert managed service provision.

An expert partner like IONOS Cloud can help assess your existing environment, determine costs and outcomes and provide digital transformation and managed cloud services to ensure success, backed by lock-in averse open-source technology and without the software licence hassle.

Trade wars, supply problems, sanctions and other factors related to the semiconductor shortage are unlikely to go away any time soon. A move to the cloud can help you overcome on-going challenges and stay ahead of competitors who are likely facing the same issues.

---

31. Marshall, D. (2022): Cloud Infrastructure Spending Increased in Third Quarter of 2021 with Overall Growth Expected for 2021, According to IDC, https://vmblog.com/archive/2022/01/14/cloud-infrastructure-spending-increased-in-third-quarter-of-2021-with-overall-growth-expected-for-2021-according-to-idc.aspx#.YitnfZPMK3K.

32. Clarke, A. (2020): Moving to the cloud doesn't have to be all or nothing, https://blogs.opentext.com/moving-to-the-cloud-doesnt-have-to-be-all-or-nothing/.

IONOS

About IONOS

With more than eight million customer contracts, IONOS is the leading European provider of cloud infrastructure, cloud services and hosting services. The product portfolio provides everything that companies need to be successful in the cloud: from domains to traditional websites and do-it-yourself solutions, from online marketing tools to fully fledged servers and an IaaS solution. The offer is aimed at freelancers, tradespeople and consumers and at corporate customers with complex IT requirements.

IONOS Cloud is the European cloud alternative and part of IONOS. With the Cloud Compute Engine, our product portfolio includes an IaaS compute engine with its own code stack for virtualisation, managed Kubernetes for container applications, a private cloud powered by VMware and S3 object storage. With our offer, we provide established large and medium-sized companies, regulated industries, the digital economy and the public sector with all the services that are necessary to be successful in and with the cloud.

IONOS was emerged in 2018 from the merger of 1&1 Internet and the Berlin-based IaaS provider ProfitBricks. IONOS is part of the publicly listed United Internet AG (ISIN DE0005089031). The IONOS brand family includes STRATO, Arsys, Fasthosts, home.pl, InterNetX, SEDO, United Domains and World4You.

More information is available at cloud.ionos.co.uk

Imprint

IONOS Cloud Ltd.
Discovery House 154 Southgate Street
Gloucester GL1 2EX
United Kingdom

IONOS Cloud Contact

Phone +44 333 336 2984
E-mail product@cloud.ionos.co.uk
Website https://cloud.ionos.co.uk/

Management Board

Hüseyin Dogan, Dr. Martin Endreß, Claudia Frese, Hans-Henning Kettler, Arthur Mai, Britta Schmitt, Achim Weiß

Chairman of the Board

Markus Kadelke

Copyright