Deniz Igan
deniz.igan@bis.org

Phurichai Rungcharoenkitkul
phurichai.rungcharoenkitkul@bis.org

Koji Takahashi
koji.takahashi@bis.org

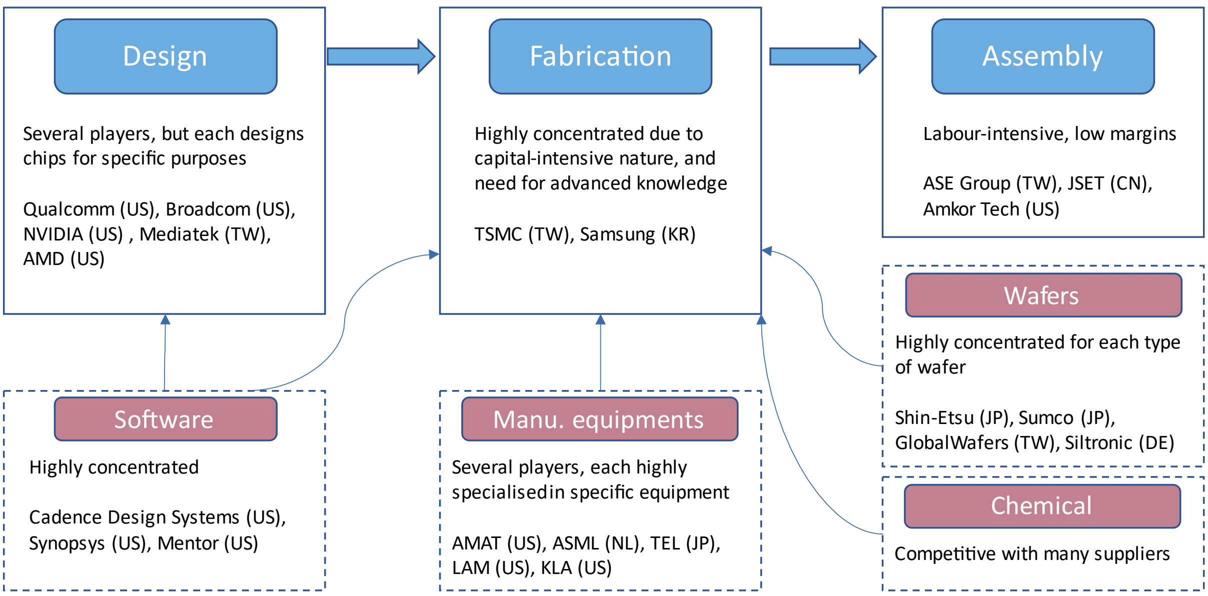
Online annex for Bulletin on "Global supply chain disruptions"

Part 1: The semiconductor value chain

Semiconductor shortages have had notable repercussions on downstream industries since the pandemic, notably for the automobile sector. Even when their value added to final goods is small, chips can be so crucial that their shortages disrupt the larger supply chain. Indeed, chip shortages continue to hamper manufacturing activity, albeit to a lesser extent than last year. Risks of near-term chip bottlenecks still remain, not least as the industry is becoming increasingly politicised, as it is now widely regarded as critical for national security. This annex provides a brief overview of the unique characteristics of this industry, which has become a key source of bottlenecks and volatility.

Semiconductor production involves three main stages: chip design, fabrication and assembly (Graph A1). Very few companies have the capacity to integrate all three – Intel and Samsung are such manufacturers – and even then, they can produce only certain types of chip. Dedicated chip designers, such as Qualcomm, Broadcom and NVIDIA, specialise in advanced and application-specific chips, working closely with producers at the fabrication ("fab") stage. Design is a difficult enterprise, which can cost more than half a billion dollars per new cutting-edge chip. Fabrication also requires a high degree of expertise and is extremely capital-intensive. The cost of building a new factory runs into tens of billions of dollars. TSMC and Samsung are major players. The last stage – assembly – is the most labour-intensive and competitive.

Semiconductor value chain: stylised network and key players — Graph A1



Source: BIS summary, based on Kleinhans and Baisakova (2020).

Of course, the entire length of the chip supply chain is far more complex than this stylised picture. Chip designers and fabs require specialised software, which only a few firms in the world are capable of writing. Setting up a new fab requires specialised equipment, sometimes produced by only one firm (eg the Dutch firm ASML is the only one with the extreme ultraviolet lithography technology needed to produce chips smaller than 7 nanometres). Learning how to use such highly customised equipment effectively requires years of close collaboration between fabs and machine manufacturers. Fabs also require silicon wafers as raw input to etch and create transistors, another concentrated sector, in which Japanese players dominate. Each of these upstream suppliers, in turn, maintains highly intricate supply chains of its own, sourcing unique materials from many more entities upstream.

Inelasticity of chip supply

The intricate value chain, with high concentration in most stages of production, explains why chips are particularly prone to supply disruption. Production capacity is highly inflexible, as each producer works with the same trusted partners with the required expertise rather than with multiple entities that otherwise might have offered more resilience. In addition, chips are often purpose-specific, limiting their substitutability. This is indeed one reason why auto producers have faced particularly acute bottleneck problems, as they could not replace the older-generation chips needed to meet safety standards by the newer ones that are in greater supply.[1,2]

The chip industry's unique complexity and concentration also stands in the way of efforts to reshore. Not only is investing in new production facilities very costly, but building new working relationships with qualified suppliers is a daunting task. Attaining greater supply chain resiliency in the longer term, at least in the semiconductor industry, could come at a substantial cost during the transition.

Demand factors

Meanwhile, the semiconductor industry is facing robust growth in demand associated with the Internet of Things. More consumer electronics now have built-in microchips, as technology has paved the way for greater computational power and increasing data flows. This in turn has created demand for larger data centres that require more chips (eg cloud servers). As of 2020, the global semiconductor markets totalled around USD 450 billion, and are projected by some industry experts to grow 40% over the next five years.

Strong secular demand, inelastic supply and concentrated GVC together mean imbalances can easily develop and spill over to other supply chains. Indeed, current chip shortages are not so much an aggregate phenomenon – global semiconductor sales rose steadily over the past two years. The problem is instead one of supply-demand mismatches, as fast-growing sectors during the pandemic such as personal computer and home-office equipment have crowded out others such as the auto industry. Continued strong demand from producers of various products could keep competition for semiconductors intense, highlighting the risk of more protracted bottlenecks at least in some segments.

1 Another reason is the sheer number of chips required – a modern car can embody than 3,000 chips.

2 Overcoming limited substitutability requires technological adaptations, eg Tesla re-wrote its software and substituted available chips for those in short supply. See https://bit.ly/3poh8BT.

Part 2: Supply chain pressures and inflation – a VAR analysis

This annex describes a simple VAR model used to estimate the contribution of bottlenecks to inflation (as presented in Graph 4, second and third panels).

The VAR model includes two endogenous variables, the growth rate of real GDP and CPI inflation (PCE inflation only for the United States). The Global Supply Chain Pressure Index (GSCPI) from the New York Fed is included in the VAR as an exogenous variable. To control for oil and food price inflation, year-on-year changes in the WTI crude price and the IMF's global food price index are included as exogenous variables. To cater for possible pandemic-induced structural changes, we allow the coefficients on the exogenous variables to change after Q2 2020. The VAR model is described as follows:

$$Y_t = \sum_{i=1}^{4} \beta_i Y_{t-i} + \alpha X_t + \gamma X_t * DT_{Q2,2020} + \epsilon_t,$$

where Y_t and X_t are vectors of the endogenous and exogenous variables, respectively. $DT_{Q2,2020}$ is a dummy variable that takes value one after Q2 2022 and ϵ_t indicates residuals.

The quarterly two-variable VAR model is estimated using the data from Q1 1998 to Q2 2022. The contribution of GSCPI to CPI inflation is calculated based on the estimated VAR model for each country by using a standard recursive computation (namely the effect of changes in GSCPI on CPI inflation is cumulated over time as implied by the dependence of CPI inflation on the lagged variables).

In the sensitivity analysis of the third panel of Graph 4, we take a two-step approach. First, we recover the implied VAR residuals required to match the estimated VAR forecasts to the September 2022 FOMC median projections of PCE inflation and GDP growth. This first step assumes the GSCPI follows an estimated AR(1) process and the oil price evolves in line with the future prices while the food price index continues to decline to the level of January 2020 at the same pace as observed recently. In the second step, we calculate the counterfactual paths of inflation and GDP growth by assuming that the GSCPI remains unchanged from Q2 2022 until the end of simulation period while keeping the VAR residuals calculated in the first step. For the oil and food prices, we assume that they follow the same paths as the first step. This analysis shows the impact of the change in the path of GSCPI on inflation when other shocks are kept identical to the first step.

There are several caveats to this analysis. First, the GSCPI might be capturing the amplification mechanism through the GVCs rather than bottlenecks. Second, the analysis might overestimate the impact of the GSCPI if there are omitted variables correlated with the GSCPI that affect inflation (eg if GDP growth inadequately controls for demand factors). Third, it abstracts from the feedback effect from GDP and inflation to the GSCPI, which could also lead to an overestimation of the contribution of the GSCPI. Finally, the graphs show results based on point estimates, although the estimation errors are ineligible.