

# The Economic Limit to Moore's Law

By KARL RUPP, *Graduate Student Member IEEE*

SIEGFRIED SELBERHERR, *Fellow IEEE*

*Institute for Microelectronics, TU Wien  
1040 Wien, Austria*



Back in 1965 Gordon E. Moore observed in his article [1] that “the complexity for minimum component costs has increased at a rate of roughly a factor of two per year (...).” In 1975 Moore refined his component count estimation to a doubling every two years, thus a reduced exponential growth compared to his initial estimation. He argued that between 1965 and 1975 the additional factor cleverness, which is for example the reduction of the number of components needed for storing a single bit, had lead to a faster growth in that period, which was now exhausted. This (adjusted) prediction known as *Moore's Law* has become a business-dictum for the whole semiconductor industry.

The higher component density has lead to a decrease in end-consumer prices. However, the costs for producers follow a converse trend: R&D, manufacture, and tests become more and more expensive with each new generation. This observation is coined *Moore's Second Law*: facility costs also follow an exponential growth. Despite this exponential growth of facility costs,

the cost per shipped unit decreases at an exponential rate. Moore himself already observed in 1995 that the semiconductor industry cannot continue its fast exponential growth indefinitely, since it would exceed the gross world product (GWP) at some time [2].

## I. GROWTH LIMITED BY ECONOMY?

Moore's Law states that the number of transistors per chip doubles every year, which is mathematically expressed as

$$n(t) = n_{2009} \exp(\alpha t), \quad (1)$$

where  $n(t)$  is the number of transistors per chip,  $t$  is the time measured in years from the beginning of 2009 (to ease the formulation of initial conditions),  $n_{2009} = 10^9$  is the number of transistors per chip in the beginning of 2009 and  $\alpha = \ln(2)/2 = 0.34$  is the growth parameter in accordance with Moore's Law.

On the other hand the exponential increase in fabrication facility costs  $c(t)$  over time can according to Moore's Second Law be written as

$$c(t) = c_{2009} \exp(\beta t). \quad (2)$$

The growth parameter  $\beta$  is determined from fabrication facility (fab) costs in the past decades and indicates a growth of a factor 100 within

30 years, thus  $\beta = \ln(100)/30 = 0.13$ , which corresponds to an average growth of fab costs of 14% p.a. [3]. The constant  $c_{2009}$  is equal to today's fab costs of approximately five billion dollars.

A natural barrier for fab costs is in principle given by the total revenue of the semiconductor industry. Since this revenue has experienced a tremendous growth in the past century itself, we better formulate a bound for fabrication costs in terms of a fixed share  $\varepsilon$  of the GWP  $g(t)$ :

$$c(t) = \min\{\varepsilon g(t), c_{2009} \exp(\beta t)\}, \quad (3)$$

where  $\varepsilon = 0.02\%$  appears to be plausible considering that the semiconductor business revenue has been about 0.5% of the GWP over the past ten years, several fabs are built each year and  $\varepsilon = 0.02\%$  corresponds to four percent of the total semiconductor industries' revenue per fab. Gordon Moore himself wrote [2] "I do not know how much of the GWP we can be, but much over one percent would certainly surprise me."

The GWP  $g(t)$  also grows exponentially, but slower than the semiconductor business in the past:

$$g(t) = g_{2009} \exp(pt) \quad (4)$$

We assume an average GWP growth of  $p = 3\%$  p.a., which is estimated from growth rates over the past decades [5].

Fab costs are essentially determined by two factors: process complexity and capacity. The former is a direct measure for R&D output, while the latter increases with each technology generation in order to achieve competitive costs per wafer (*economy of scale*) as has already been addressed in Moore's article [1]. While fab costs were more or less free to grow in the past, limited by the currently available technology only, economic limits sooner or later come into play. Consequently, the growth parameter  $\beta$  in (2) has to drop by a factor of four from 0.13 to

0.03 in order to match the GWP growth rate  $p$  in (4). However, a reduction of the growth rate of fab costs by a factor  $p/\beta$  also results in a reduction of the growth of money (directly or indirectly) spent on R&D, thus a reduction of the growth of transistor counts with time,  $dn/dt$ , is to be expected. A mathematical formulation including such economic limitations is

$$\frac{dn}{dt} = \min\{C_{\text{Moore}} \exp(\alpha t), C_{\text{GWP}} \exp(\alpha(p/\beta)t)\}, \quad (5)$$

where  $C_{\text{Moore}}$  and  $C_{\text{GWP}}$  are appropriate constants. If the first term is the limiting factor, then the number of components grows according to Moore's Law in (1). If economics determine the growth by means of the second term in (5), the factor  $p/\beta$  shows the economic limit with an asymptotic growth parameter of  $\alpha(p/\beta) \approx \alpha/4$  depicted in Fig. 1: This results in a reduced asymptotic growth of component counts per chip by a factor of two every eight years only. The GWP share parameter  $\varepsilon$  does not influence the reduced growth rate due to economy, it only influences the time at which we run into economic limitations.

The prediction of the time at which we run into economic limitations is very sensitive with respect to the GWP share parameter  $\varepsilon$ : Choosing  $\varepsilon = 0.03\%$ , a growth reduction is

predicted around 2025, whereas the choice  $\varepsilon = 0.01\%$  shows first signs of reduced growth already in 2015. Thus, with joint funding of large fabs an economic growth capping can be shifted many years into the future so that we eventually face limitations imposed by physics first.

## II. FROM FREE MARKET TO OLIGOPOLY?

Considering the market structure of the semiconductor business, integrated device manufacturers (IDMs), who both design and manufacture, will only be able to afford fab costs up to an individual  $\varepsilon_{\text{IDM}}g(t)$ , simply because they have to cover both fab expenses and investments into R&D. In contrast, foundry companies can fully focus on device manufacture, potentially being able to afford  $\varepsilon_{\text{foundry}}g(t)$  with  $\varepsilon_{\text{foundry}} > \varepsilon_{\text{IDM}}$ . As soon as fab costs exceed  $\varepsilon_{\text{IDM}}g(t)$ , the business is thus subjected to a restructuring process: IDMs are unable to afford the fabrication of the latest generation of devices any longer. The only way to stay competitive for IDMs in mass markets is to become a fabless company then, fully focus on chip design, and rely on foundry companies for fabrication. Such a transition can already be observed today and is to a certain degree self-energizing: With each IDM becoming a fabless company, another foundry company grows

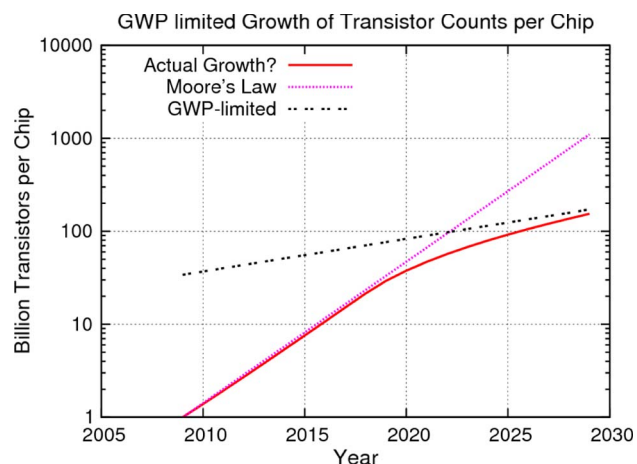


Fig. 1. Possible future transistor counts per chip due to economic limitations.

and pushes its individual GWP share parameter  $\varepsilon_{\text{foundry}}$  to a larger value.

However, the number of semiconductor foundries can only be small, if the fab costs keep increasing at a rate higher than the GWP: The large, cost-effective fabs provide best and most transistors per dollar, pushing smaller competitors out of the market. Thus, the free market of (leading edge) chip manufacturing may turn into an oligopoly. The consequences of such an oligopoly can be manifold: Research and development could be

focused and accelerated, while syndicates may lead to a considerable slowdown of innovations, because the remaining companies will not be interested in a technology race. Since a successful market entry for new companies is almost impossible, existing fabs could be sufficient to skim the market within such an oligopoly.

### III. CONCLUSION

There have been numerous papers and discussions about the lives and

deaths of Moore's Law, all of them dealing with several technological questions. Just as Moore observed in 1975 an exhaustion of the factor cleverness, we may soon experience an exhaustion of the economically unlimited growth. With such a reduced growth of transistor counts, it will take some additional years until we finally hit a fundamental barrier for the MOS layout imposed by physics as discussed at this place some time ago [5]. ■

### REFERENCES

- [1] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, no. 8, Apr. 1965.
- [2] G. E. Moore, "Lithography and the future of Moore's law," *Proc. SPIE*, vol. 2437, no. 8, pp. 2–17, 1995.
- [3] Y. Nishi and R. Doering, Eds., *Handbook of Semiconductor Manufacturing Technology*, 2nd ed. Boca Raton, FL: CRC, Jul. 2007.
- [4] World Development Indicators Database, World Bank. [Online]. Available: <http://devdata.worldbank.org/data-query/>
- [5] J. R. Powell, "The quantum limit to Moore's law," *Proc. IEEE*, vol. 96, no. 8, pp. 1247–1248, Aug. 2008.