ROADMAPPING VS. S-CURVES: HOW TO SWITCH TO THE NEXT S-CURVE

Analyzed using the example of the semiconductor industry

Gerd Grau, Ph.D.[1]

[1]Dept. of Electrical Engineering and Computer Sciences, University of California, Berkeley, United States of America, grau@eecs.berkeley.edu

Abstract

There are two very important tools in managing technology: roadmapping and s-curves. Roadmapping can be a very effective tool to evolutionarily advance an existing technology. Conversely the idea of s-curves aids managers in the decision to make a revolutionary change to a new technology. The semiconductor industry is a prime example of a very successful roadmapping exercise. However, with continued scaling traditional microfabrication based on top-down lithography techniques becomes exceedingly expensive and complex. Many academic and industrial researchers work on alternative technologies to switch to the next s-curve. This work examines a first order approach to analyze such new technologies. The case of bottom-up nano-assembly is used as an illustration. Its merits are contrasted with current technology to come to a first assessment of its viability as the next s-curve. However, this is only a starting step to guide managers and technologists into the right direction when investigating new technologies.

JEL Classification: C44, D81, G11

Keywords: decision theory, managing technology

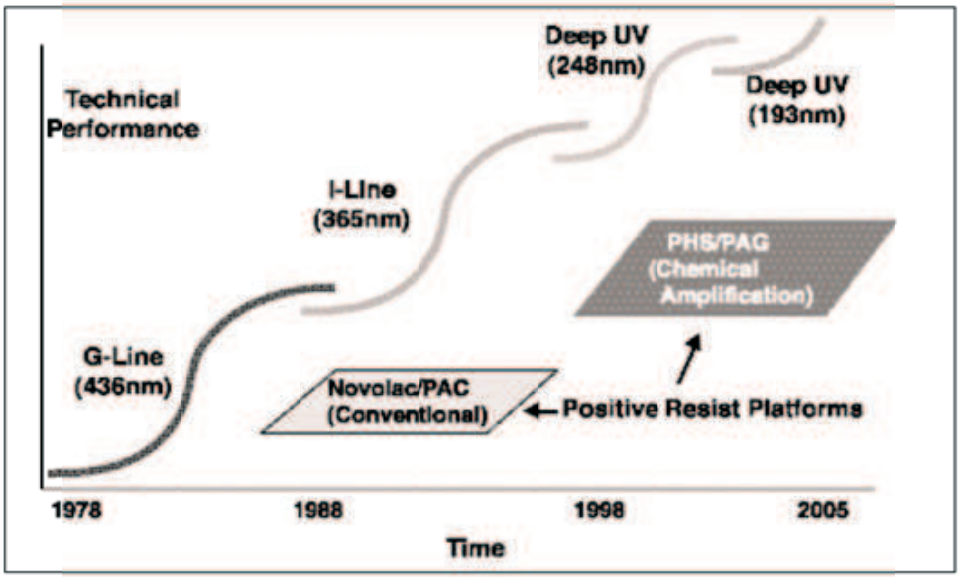
1 Introduction

Over the last decades the semiconductor industry has been one of the most successful industries to ever undertake a roadmapping exercise. It has been able to maintain Moore's Law (the exponential increase of chip performance over time (Moore 1965)) despite of many technological challenges that had to be overcome.

It has been predicted many times in the past that this trend will come to an end; however, so far the industry has always managed to overcome these roadblocks. The International Technology Roadmap for Semiconductors ("International Technology Roadmap for Semiconductors (ITRS) 2011 Edition" 2012) has given researchers and managers goals to strive for and security that the industry as a whole, including equipment suppliers, would move in the same forward direction. The ITRS has been crucial in shaping the way the semiconductor industry has become a 226 billion dollar sales industry in 2009 (Semiconductor Industry Association 2010).

An alternate view of looking at technological advancement is by considering s-curves. S-curves describe the effect that the performance of a technology often starts to improve slowly before some rapid improvement, which will eventually level off (see Figure 1). It is important to identify when a technology has reached the point at which its development starts to level off and one has to consider alternate technologies even if their initial performance is worse.

Figure 1: S-curves for different lithography technologies in semiconductor processing (Bowden 2004)



The most successful manager of technology has to be able to work with both tools: roadmapping and s-curves. Roadmapping is a crucial tool in achieving the high growth at the center of the s-curve and in maintaining this growth for as long a time as possible to maximize the return on initial investments. But strategic

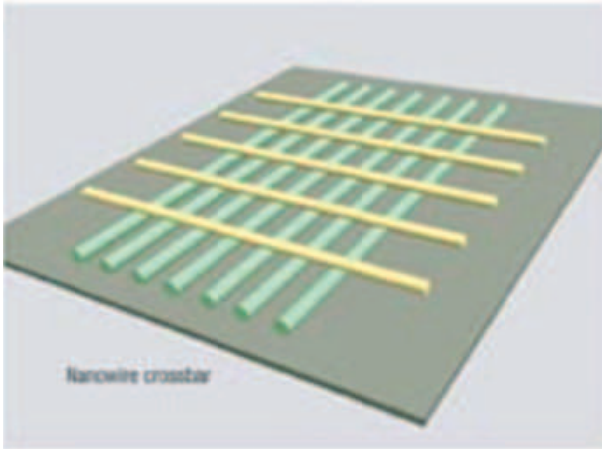
thinking in terms of s-curves is just as crucial in the long term to ensure that new trends are not overlooked, which could ultimately lead to losing one's competitive advantage to other companies.

2 Bottom-up as the potential next s-curve

As device dimensions shrink further and further, classical microfabrication becomes increasingly challenging, complex and costly. The classical approach to microfabrication uses top-down techniques to deposit and etch materials and to define critical feature dimensions. The dominant patterning technique in the past and nowadays is photolithography. Photolithography has the great advantage of exposing whole wafers at once offering unparalleled throughput. However, the resolution of optical lithography is fundamentally limited by the wavelength of the light that is being used. Tricks such as phase-shifting masks have extended the use of optical lithography way past what was thought possible in the past and it is unclear at which point a hard limit will be reached. The other problem with photolithography is the cost of masks increasing rapidly with decreasing feature size. There are other top-down patterning techniques that try to tackle these problems differently such as electron beam lithography or even single atom manipulation using scanning tunneling microscopy (STM). These techniques offer higher resolution and don't require expensive mask sets. However, there is a general trade-off between resolution and throughput. These high resolution techniques suffer from very limited throughput. Since throughput translates directly into cost per wafer, no other top-down technique has yet been able to replace photolithography.

This is one of the main motivations for many researchers to pursue a different paradigm of microfabrication: bottom-up nanotechnology – the potential next s-curve. In contrast to top-down manufacturing critical dimensions are now not defined explicitly with direct patterning techniques such as photolithography anymore. Instead nanostructures assemble themselves into the desired geometry. Mechanisms such as surface, electrostatic or chemical forces are utilized to guide components to their desired location. This requires a detailed understanding of the forces at play. Especially complex structures with many degrees of freedom will be very difficult to design because there is no direct way of manipulating the structure. This is the reason why it is projected that the first bottom-up structures that make it into production are likely going to be simple two-dimensional arrays of nanowires (NW) (Lu and Lieber 2007).

Figure 2: Schematic of nanowire crossbar array e.g. for memory application (Lu and Lieber 2007)



These nanowires can be based on a range of different materials including carbon nanotubes (Chen et al. 2003) (Rueckes 2000), silicon nanowires (Dong et al. 2008) (Cui 2001) or more exotic materials such as $Ge_2Sb_2Te_5$ (Jung et al. 2006) or GeTe (Lee et al. 2006). The actual memory element can be based on various different working principles including resistance switching (Chen et al. 2003), field effect transistors (Huang 2001) or even mechanical switching (Rueckes 2000). Such crossbar structures are of limited complexity for the case of regular memory arrays. More complex arrays such as logic circuits will require a considerably larger amount of effort to make the process controllable. But even simple memory arrays will suffer from problems with yield. Because the fabrication, placement and alignment of nano-components will rely on very small forces that cannot be controlled directly, there will be more statistical fluctuations reducing device and system yield. (A. DeHon, Lincoln, and Savage 2003) This will require more sophisticated error correction techniques and redundancy built into the system than in current circuits to keep chip yields up even if a number of devices on any chip are faulty. This, however, will impact the density of the system, which of course is one of the main reasons to pursue bottom-up nanotechnology in the first place. Another factor that might limit the density of nano-circuits is the interfacing with higher level circuits and the outside world. At some point nano-devices and interconnects will have to interface with metal lines whose dimensions are defined by classical photolithography. As the understanding of bottom-up assembly gets better, more and more of the interconnects will also be manufactured using bottom-up techniques. How-

ever, at least for the first generation of bottom-up systems it is very likely that only the most critical active dimensions will be fabricated bottom-up and all interfacing and interconnects will be manufactured top-down. For instance first generation nanowire memory arrays would have metal interconnects right up to the edge of the nanowire array with no nanoscale interconnects. This will reduce the effective bit density that can be achieved with such systems. However, it is unclear whether these two factors – redundancy due to limited yield and interfacing with classical metal lines – will limit the density and thus usefulness of bottom-up manufactured systems.

3 First order analysis of simplest bottom-up memory array case

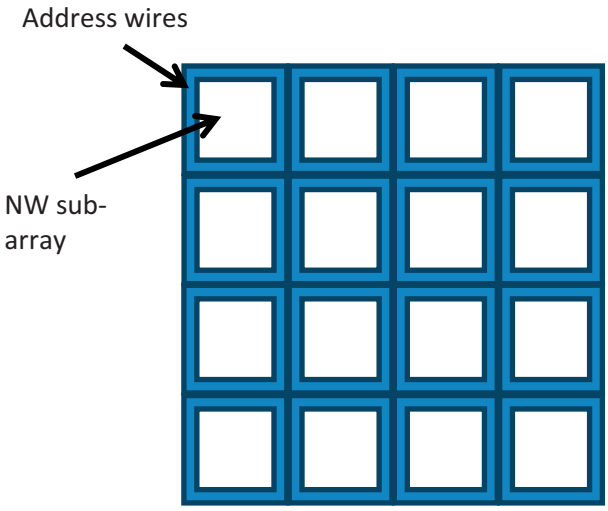
3.1 Introduction to analysis

When switching to the next s-curve the first generation of the new technology will usually be the simplest possible case before more complex versions will be available, which will lead to rapid improvements later. It is demonstrated here how managers can perform a simple first-order analysis to investigate the implications of moving to the new technology without going into too much of the technical details. These details will of course be crucial when finalizing plans. However, it is often important to have a simple model to guide efforts at the early stage.

One can assume a generic nanowire crossbar memory array. The actual device working principle is secondary because the geometrical considerations that go into the analysis are universal for all crossbar arrays.

The aim is to relate the effective bit density of a nanowire memory array to yield, nanowire dimensions and the size of addressing lines. The impact of individual devices and of whole nanowires failing is taken into account. Due to the fact that longer nanowires will be more likely to have a defect somewhere along their length the memory array is split up into smaller blocks, which improves yield but also increases the number of addressing lines needed (see Figure 3). This leads to a trade-off.

Figure 3: Nanowire array architecture assumed in this model with metal address wires and sub-arrays



In order to study the effect of yield on density the simplest error correction method possible is used as a limiting case. It is assumed that after fabrication all devices and nanowires are tested to detect non-functioning devices and wires. These memory locations are stored and simply not used during operation as a memory array. This reduces effective bit density simply because the number of functioning bits per unit area is smaller than the nominal number of crossbar nodes. More sophisticated error correction codes could of course be used. For instance if there is a break in a wire in this model, it will be disregarded completely. A more sophisticated system could potentially route signals around such a fault. But since it is unclear how large the overhead for such a system would be and how great the gains would be, it will not be considered here where the simplest case is taken as the limiting case.

The second factor reducing effective bit density is the overhead due to address wires. The pitch of address wires will be considerably larger than that of the nanowire array because metal address wires are patterned by standard photolithography. The redundancy necessary to achieve a set of unique nanowire codes also contributes to the addressing overhead. Because it is assumed that whole nanowires are more likely to be faulty with increasing length, the large memory array is split up into a number of smaller sub-arrays or blocks. This increases the overall yield. How-

ever, the overhead due to metal addressing wires also increases. This leads to the conclusion that there will be an optimal sub-array size for a given nanowire yield, nanowire pitch and metal pitch that will balance nanowire yield versus addressing overhead.

3.2 Results of analysis

It is clear that the effective bit density will increase both with wire and device yield. The effective cell pitch will converge to the nanowire pitch as yield gets close to 100%. As one shrinks the nanowire pitch, the yield per device will become a less important factor giving only relatively small gains compared to the density gains that can be achieved from shrinking nanowire pitch. Physical nanowire pitch will have a much larger effect on overall density than metal address line pitch simply because the address lines make up a smaller portion of the total memory array than the nanowires, which actually provide the active devices. The main benefit from scaling metal lines will be that smaller sub-arrays become favourable further reducing the need for long, high yield nanowires.

One can perform a simple calculation to see that even a device yield of 75% and a total nanowire yield of 78% (for an optimized block size of 545 devices per nanowire with nanowire pitch 10nm and metal pitch 50nm) are sufficient to achieve an effective memory cell pitch of less than 2.5 times the nanowire pitch. These yield numbers are reasonable for nanowire arrays (Andre DeHon and Wilson 2004) and still very far low compared to what can be achieved in modern top-down microfabrication. It should also be noted that this calculation is a limiting case in terms of worst case performance. The metal pitch was assumed to be 50nm, which might be lower with state-of-the-art lithography tools. Also more sophisticated error correction codes and re-routing techniques might be implemented in a real system to save some of the devices if only part of a nanowire is corrupted and the rest can be used by re-routing around the broken parts.

The more critical factor than yield seems to be actual physical nanowire density and to a lesser extent metal addressing lines density. Scaling of the metal lines mainly affects the dependency on length dependent yield due to the reduced addressing overhead for smaller arrays. Scaling of the nanowire pitch sets the floor for minimum cell pitch or maximum bit density that can be achieved even with perfect yield and highly scaled address lines. The Langmuir-Blodgett technique can indeed give a very small nanowire pitch on the order of 6nm (Acharya et al. 2006).

This shows that it might be possible to get to very high density memory arrays that are not critically dependent on yield if other problems such as misalignment or abruptness of dopant modulation during growth can be controlled. This could be a new paradigm of micro/nanofabrication based on high density, low yield and very simple error correction.

3.3 Comparing the results with the incumbent technology

After analyzing the potential for a new technology the next step is a comparison with the incumbent technology. In this case one of the main competitors of any new nano-memory technology will be Flash as the currently leading technology. The first question that can be asked regarding Flash is whether the scheme outlined above – low yield, high density and simple error correction – could also work for Flash. This is very likely to be uneconomical. As explained above this scheme relies heavily on scaling of the physical cell pitch. This is achieved relatively easily using bottom-up techniques. However, Flash being a top-down technique requires very high resolution photolithography to get down to 10nm or below. Since every Flash chip requires a sizable number of masks, cost would explode if one were to try to implement a high density, low yield strategy. Instead it seems most likely that the scaling of Flash will continue as before at a steady pace focusing on high yield.

The two strengths that any bottom-up technique will have are fabrication cost and density. The simple fabrication using self-assembly should always outperform classical microfabrication in terms of cost even if additional effort is required for error correction. However, the advantage in terms of density is not quite so clear. According to the 2011 ITRS multi-level cell (MLC) Flash will have a density of about 10^{12}bits/cm^2 in 2020 ("International Technology Roadmap for Semiconductors (ITRS) 2011 Edition" 2012). This is equivalent to a cell pitch of 10nm and thus very close to what might be achievable with bottom-up techniques. It will thus be a very close race between bottom-up techniques and Flash. A potential scenario could be that Flash hits a technological or economic brick wall and will be abandoned at some point. Or alternatively scaling of Flash will go on for quite some while before alternative technologies are considered. It will be interesting to see which technology will win. Due to the inertia of the industry and past investment it seems most likely, though, that the scaling of Flash will continue as long as possible. But depending on what type of nanowire array is used, it might have

other beneficial properties for certain applications such as retention time, speed, mechanical flexibility or cost.

4 Conclusions

This work highlights the importance of both roadmapping and s-curve analysis as tools for the management of technology. The example of bottom-up nanowire memories is taken to illustrate the first order process in determining whether a new technology has the potential of taking over from the current technology, which is equivalent to switching to the next s-curve and creating a new roadmap or incorporating it into a new version of the current roadmap (e.g. ITRS). It is hypothesized that bottom-up nano memories could lead to a new paradigm of high density, low yield nanofabrication. Whether this shift will really occur depends on the severity of the problems that will be encountered with top-down traditional microfabrication in the future and whether bottom-up technology will really be able to live up to the expectations. However, such a first order technology analysis can be the first step in guiding both mangers and technologists to look into the right direction for a new technology and to ask the right questions.

5 Appendix

References

Acharya, S., A. B. Panda, N. Belman, S. Efrima, and Y. Golan. 2006. "A Semiconductor-Nanowire Assembly of Ultrahigh Junction Density by the Langmuir-Blodgett Technique." *Advanced Materials* 18 (2) (January 19): 210–213. doi:10.1002/adma.200501234.

Bowden, Murrae J. 2004. "Moore's Law and the Technology S-Curve." *Current Issues in Technology Management, Stevens Alliance for Technology Management* 8 (1).

Chen, Yong, Gun-Young Jung, Douglas A A Ohlberg, Xuema Li, Duncan R Stewart, Jan O Jeppesen, Kent A Nielsen, J Fraser Stoddart, and R Stanley Williams. 2003. "Nanoscale Molecular-switch Crossbar Circuits." *Nanotechnology* 14 (4) (April 1): 462–468. doi:10.1088/0957-4484/14/4/311.

Cui, Y. 2001. "Functional Nanoscale Electronic Devices Assembled Using Silicon Nanowire Building Blocks." *Science* 291 (5505) (February 2): 851–853. doi:10.1126/science.291.5505.851.

DeHon, A., P. Lincoln, and J.E. Savage. 2003. "Stochastic Assembly of Sublithographic Nanoscale Interfaces." *IEEE Transactions On Nanotechnology* 2 (3) (September): 165–174. doi:10.1109/TNANO.2003.816658.

DeHon, Andre, and Michael J. Wilson. 2004. "Nanowire-based Sublithographic Programmable Logic Arrays." In , 123. ACM Press. doi:10.1145/968280.968299. http://portal.acm.org/citation.cfm?doid=968280.968299.

Dong, Yajie, Guihua Yu, Michael C. McAlpine, Wei Lu, and Charles M. Lieber. 2008. "Si/a-Si Core/Shell Nanowires as Nonvolatile Crossbar Switches." *Nano Letters* 8 (2) (February): 386–391. doi:10.1021/nl073224p.

Huang, Y. 2001. "Logic Gates and Computation from Assembled Nanowire Building Blocks." *Science* 294 (5545) (November 9): 1313–1317. doi:10.1126/science.1066192.

"International Technology Roadmap for Semiconductors (ITRS) 2011 Edition." 2012. Accessed May 8. http://www.itrs.net/Links/2011ITRS/Home2011.htm.

Jung, Yeonwoong, Se-Ho Lee, Dong-Kyun Ko, and Ritesh Agarwal. 2006. "Synthesis and Characterization of Ge2Sb2Te5 Nanowires with Memory Switching Effect." *J. Am. Chem. Soc.* 128 (43): 14026–14027. doi:10.1021/ja065938s.

Lee, Se-Ho, Dong-Kyun Ko, Yeonwoong Jung, and Ritesh Agarwal. 2006. "Size-dependent Phase Transition Memory Switching Behavior and Low Writing Currents in GeTe Nanowires." *Applied Physics Letters* 89 (22): 223116. doi:10.1063/1.2397558.

Lu, Wei, and Charles M. Lieber. 2007. "Nanoelectronics from the Bottom Up." *Nat Mater* 6 (11) (November): 841–850. doi:10.1038/nmat2028.

Moore, G.E. 1965. "Cramming More Components onto Integrated Circuits." *Electronics* 38 (8) (April 19): 114–117; reproduced in Proceedings of the IEEE, 86, 82–85 (1998).

Rueckes, T. 2000. "Carbon Nanotube-Based Nonvolatile Random Access Memory for Molecular Computing." *Science* 289 (5476) (July 7): 94–97. doi:10.1126/science.289.5476.94.

Semiconductor Industry Association. 2010. "Industry Fact Sheet." http://www.sia-online.org/cs/industry_resources/industry_fact_sheet.