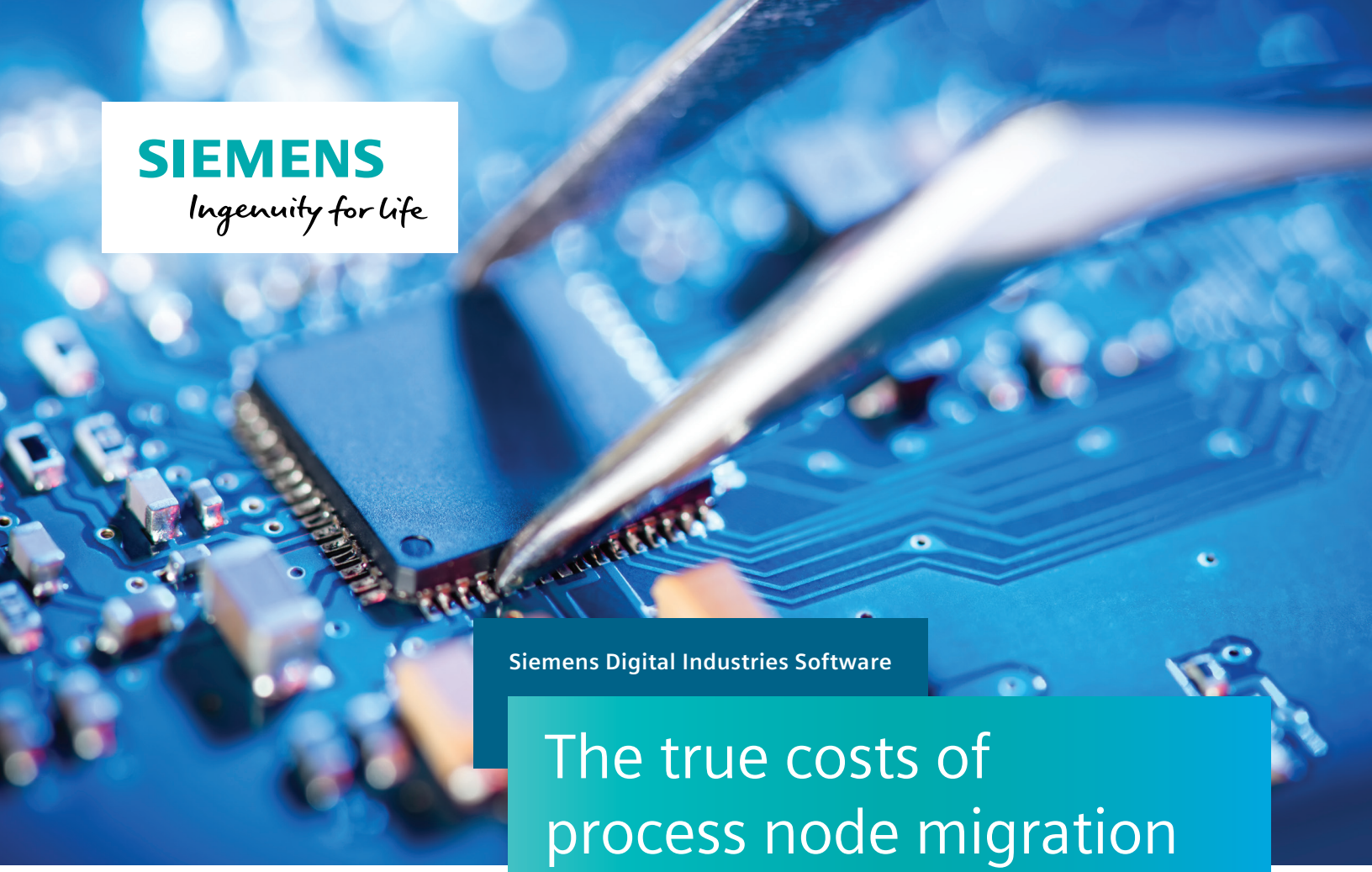

SIEMENS
Ingenuity for life

Siemens Digital Industries Software

The true costs of process node migration

Executive summary

Deciding when and how to make a process node transition is critical to business success. The solution that requires the least amount of total change – in the form of license configurations, required hardware resources, necessary tool qualifications, and adequate support infrastructure – will always be the most "inexpensive" option. Do you have all the information you need to make the right decision?

John Ferguson
Siemens EDA

Node transitions

In the integrated circuit (IC) industry, Moore's Law has held largely true for over 50 years (figure 1). Historically, this truism was enabled through manufacturing process improvements that supported the creation of increasingly smaller transistor devices. The smallest possible transistor pitch for a given manufacturing process is commonly referred to as the process node. In that 50-year span, a new process node was established roughly every 18 to 24 months. With a smaller transistor size, designers were able to incorporate more transistors per die area, increasing total compute power node over node.

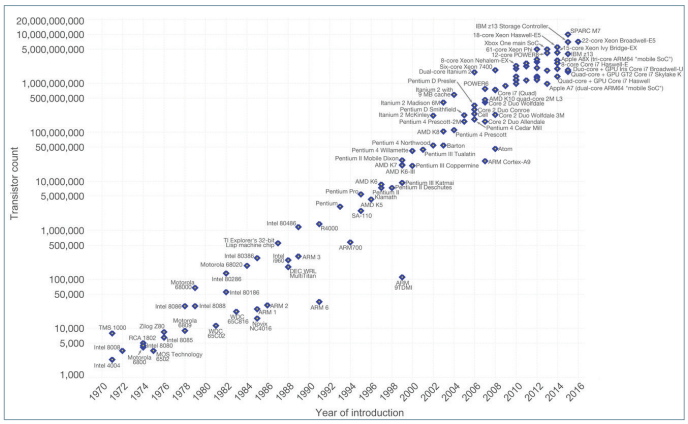

Figure 1: Moore's law has held relatively steady for over 50 years[1].

But Moore's Law is not exclusively related to just the number of transistors on a chip. As the transistor pitch is decreased, its channel length is reduced, ultimately leading to faster switching power. As a result, each new process node has also been associated with faster total chip performance. To build on this effect, other manufacturing improvements through the years have further extended the performance of the device. Examples include the switch from aluminum to copper for the primary interconnect between devices, smaller, low-k dielectric materials for the transistor gate region, and more.

Also implied by Moore's Law are the economic impacts. As mentioned, as the size of the transistor decreases, the total number of transistors per area can be increased. This relationship essentially reduces the total cost per transistor, and traditionally has made migration to newer process nodes particularly attractive. Additionally, as we travel down Moore's highway, we've also periodically seen increases in the total wafer size used in foundries (figure 2). This change has also had an impact on total manufacturing cost, since fitting more chips on each wafer brings down the aggregate costs. Lastly, as with the case for any new manufacturing process, over the lifetime of a product, the costs will come down – they may still be more than they were at the previous node, but not nearly as high as the initial entries.

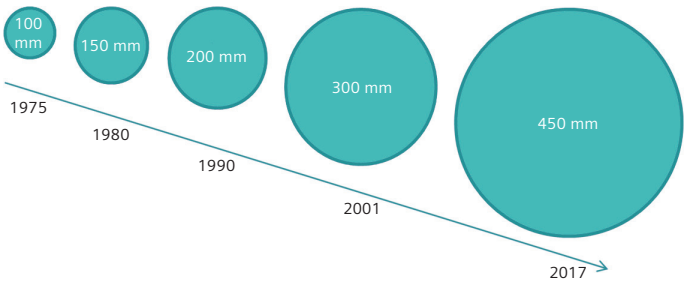

Figure 2: Increasing wafer size reduces the manufacturing cost per chip[2].

Of course, there are other mechanisms counteracting some of this economic gain. As we squeeze more and more transistors into the same area, we have less room to connect between them. This necessitates an increase in the total number of metallization layers required (figure 3), which in turn impacts both the economics (as more masks are now required to manufacture the full chip), and also the chip electrical behavior (because more power is required to push the signals through so many layers of metals and vias).

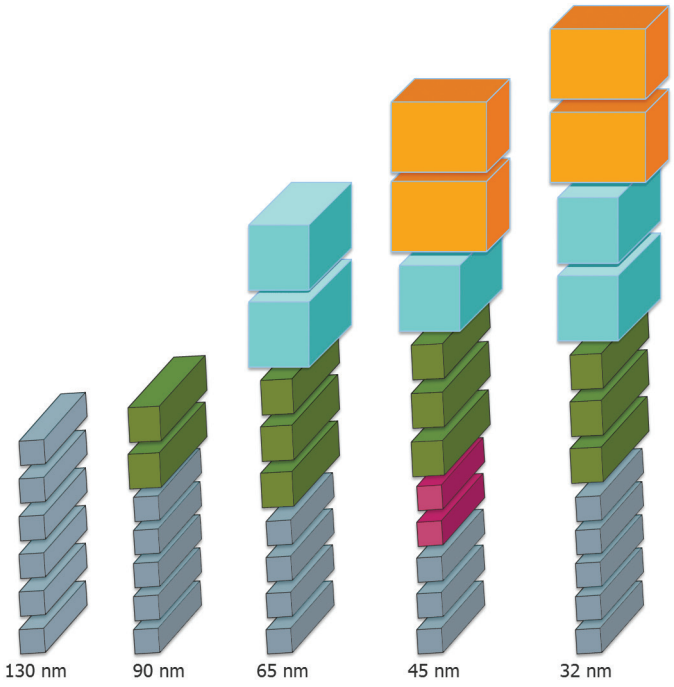

Figure 3: The increase in metal layers increases manufacturing cost and complexity, and impacts chip electrical behavior[3].

In addition, starting roughly during the transition from 28 nm processes to 20 nm processes, some of the implications of Moore's Law no longer hold. With the switch from CMOS transistors to FinFET transistors (figure 4), what most foundries refer to as their process node is no longer truly about the minimum transistor gate pitch, but rather a relative comparison of the transistor performance to previous nodes.

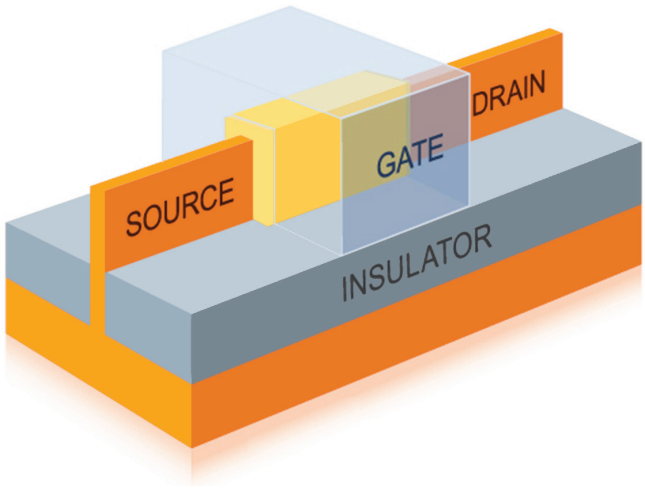

Figure 4: The introduction of finFET transistors changed the definition of node scaling (source: GLOBALFOUNDRIES. Used by permission).

If you were to review, for example, a 20 nm process, you would be hard-pressed to find any device geometry pitch that is actually 20 nm between components. In fact, from the 28 nm to 20 nm node, minimum transistor pitch change is not nearly as large as the expected 8nm. While the number of transistors that can be placed in a given area continues to improve for each full node transition, the cost to achieve that overall area reduction is growing at a much higher rate than previously.

At the most advanced process nodes (20 nm and down), the ability to accurately print the finFET device fins and other fine-grained structures requires overcoming some very challenging lithographic limitations. This manufacturing issue was first addressed with the introduction of double patterning, and we now have multiple multi-patterning techniques to reduce the lithographic impacts of printing in dense regions by splitting layout geometries between two or more masks. Of course, each mask comes at a considerable cost to both the design and manufacturing processes (figure 5). Again, though, prices typically decline after the initial introduction, so taking a long-term view can help a company assess the true cost/benefit over the expected lifetime of that node. All of these factors taken together may lead some companies to more carefully consider the long-term cost/benefit analysis of a node transition.

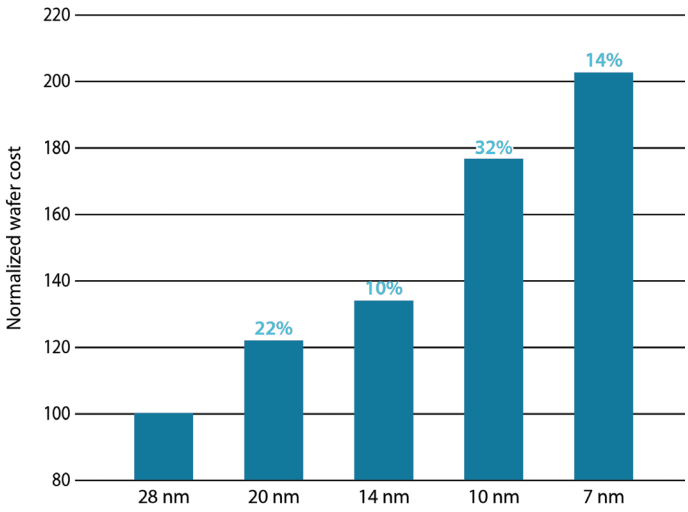

Figure 5: Normalized wafer cost trend – node to node evolution[5].

These manufacturing changes, however, are not the only source of economic disruptions through the process of migrating to newer nodes. Less commonly acknowledged are the increased costs to design at a new node. With each new process node, the new

manufacturing approaches manifest into new design constraints (figure 6). Existing EDA software tools may be impacted or new tools/flows may be required. For example, more design rules are needed to accurately account for the parasitic impacts of metallization when characterizing a chip's performance, or to analyze device context to capture stress impacts upon the behavior of each transistor, or properly split a dense layer to two or more masks. More design for manufacturing (DFM) analysis is required to predict the impact of manufacturing limitations in processes like lithography and chemical-mechanical polishing (CMP).

These new or tighter constraints impact many parts of the design flow, and often require new tooling and training for the designers. They also typically increase the total time required to take a design from concept to tape-out at a new node, as well as requiring a significantly larger amount of computation, often reflected in costs in terms of increased hardware and the corresponding increase in power and cooling requirements.

We can get a sense of the significance of these impacts by looking at the relatively straightforward case of design rule checking (DRC) changes across the process nodes (figure 7). With each new target process node, the total number of design rules required for sign-off increases dramatically. Equally noticeable is the increase in the number of DRC operations required to implement these rules, a reflection of their increased complexity.

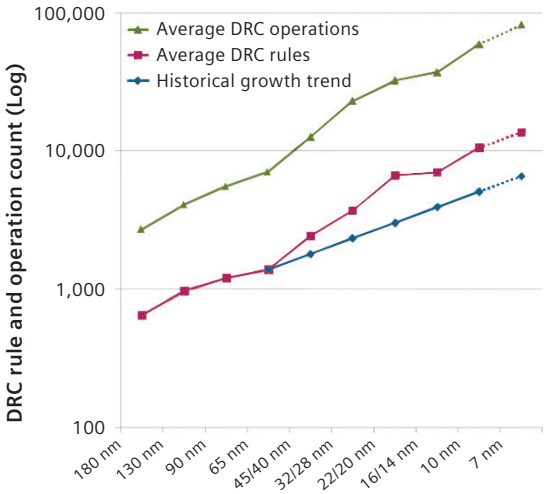

Figure 7: Average growth in design rules and DRC operations over node progressions.

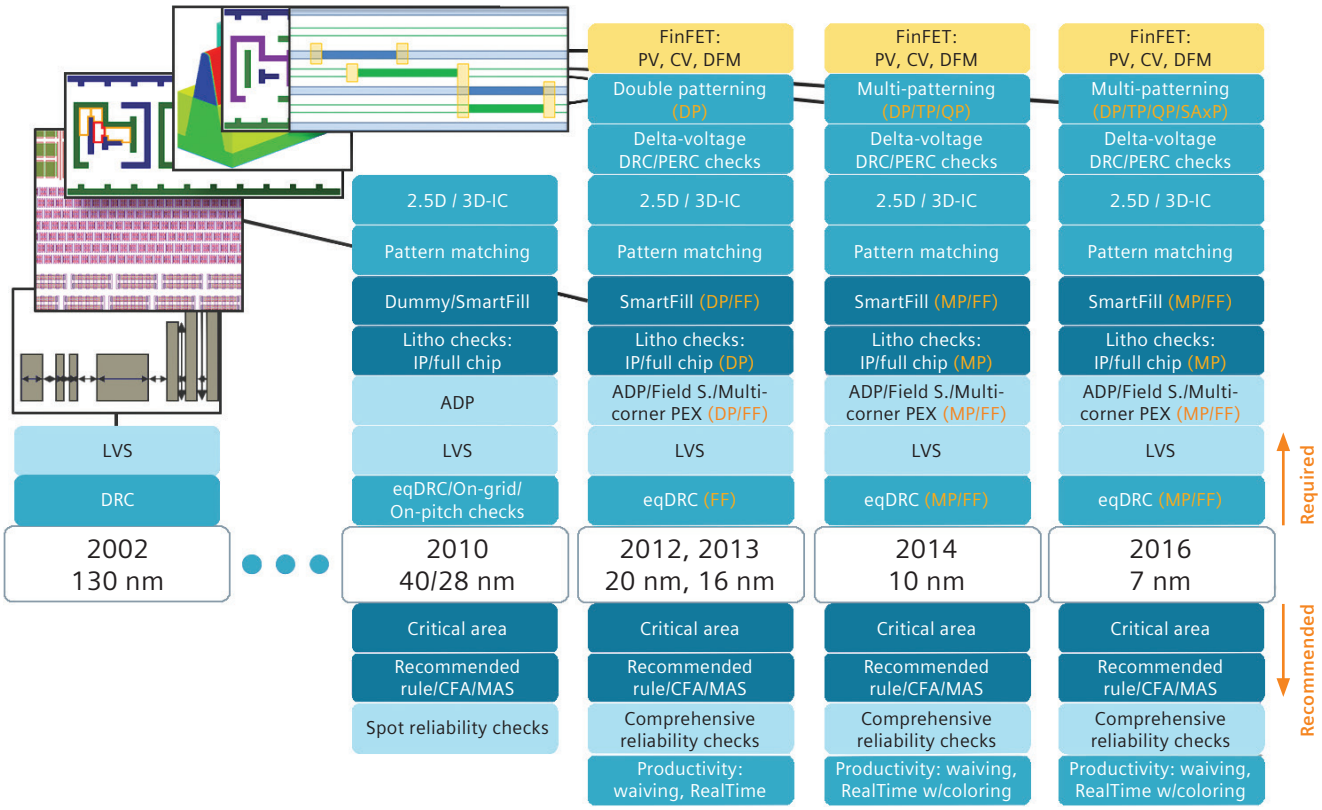

Figure 6: Design requirements have increased significantly node over node for physical and circuit verification, as well as DFM optimization.

When these increases are coupled with the fact that the number of layers, total physical sizes of the chips, and total geometry data are also increasing, there is a compound effect on the computing requirements for each node (figure 8).

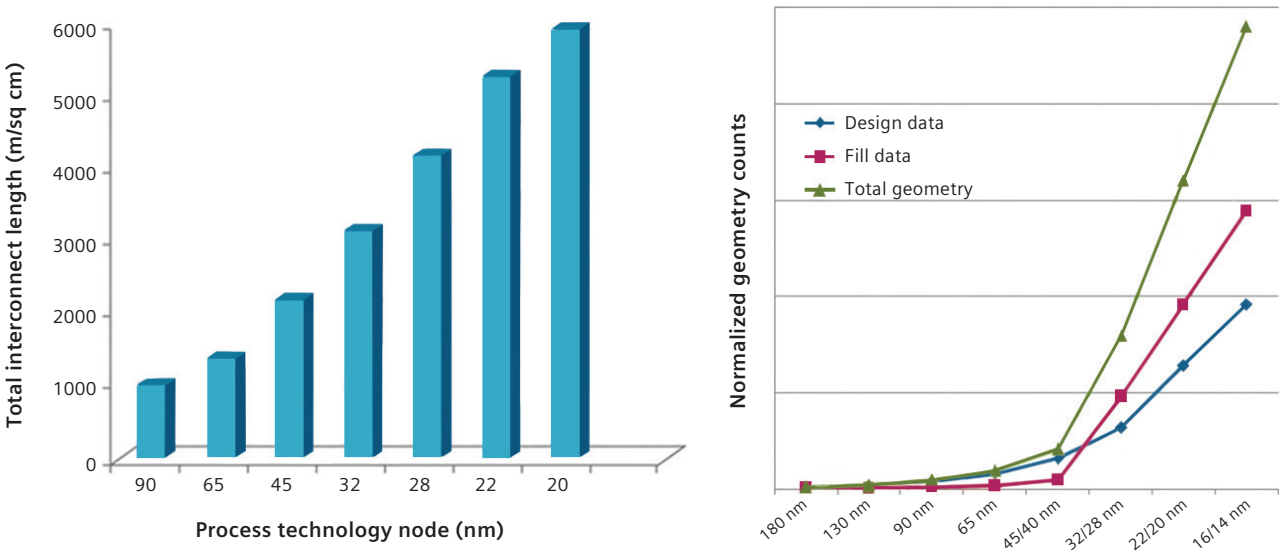

Figure 8: Multiple growth factors impact the time and costs of node progression (left graph[4]).

Faster processing of design rules on faster hardware is essential, but even that can be insufficient to achieve acceptable runtimes unless also combined with efficient hardware scaling and distributed processing on multiple CPUs.

Growth in the effort per design team isn't just limited to computation, either. With everincreasing complex designs, the number of designers needed for both analog/custom and digital implementation design projects grows when moving to advanced nodes. In analog/custom designs, staffing need is being driven by the rapidly growing number of analog and mixed-signal blocks in a design, an increasing number of complex circuits in nm-node analog, more power-saving modes, etc. Over the last ~5 years, it hasn't been uncommon to hear of 50% node-over-node IP designer staffing increases.

With all these factors, how do you make the best decisions on the software and hardware you use?

Tool performance

The costs associated with physical verification tools are often measured solely on the basis of running full-chip verification in what the user considers a "reasonable runtime," which varies from company to company, and design to design. However runtime is determined, continuous performance enhancement coupled with efficient hardware scaling (more resources) is critical to achieving runtime goals at all nodes, but especially at advanced nodes.

Industry-leading tools like the Calibre® nmDRC™ and Calibre nmLVS™ platforms continuously improve the speed and efficiency of their engines, enhance their verification flows, and partner with the foundries to optimize the foundry-written rule decks for efficient processing, but with the exponential node-overnode increase in compute demands, scaling is the only remaining means to maintain constant turnaround time (figure 9).

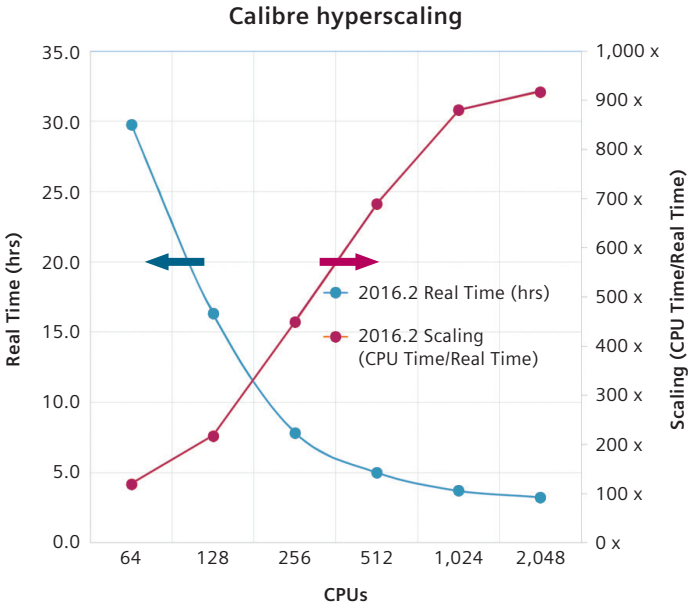

Figure 9: The Calibre nmDRC tool operates efficiently by scaling up to thousands of cores, depending on design size and runtime targets.

Another consideration is how you access advanced verification tool functionality. For example, multi-patterning verification may be required at 22/20 nm and below, while pattern matching technology is now frequently used at advanced nodes for identifying complex layout configurations that typically result in lithographic hotspots. However, fabless users may be reluctant to adopt such new software if it is not mandated by their foundry, whereas foundries are reluctant to implement rules that may require their customers to obtain new software.

Because of these factors, the Calibre nmPlatform is structured to provide common verification functionality at all nodes. We work closely with the foundries to understand the technological requirements of each node to ensure the core verification functions work with the decks for that node. In most cases, this process is made easier by the fact that all the major foundries develop their design rules and decks using Calibre tools as their plan of record tools. This helps ensure that customers can confidently port their existing Calibre installations to new versions that will have the necessary functionality and performance for the new node, without waiting for tool qualification.

Advanced functionality is available on an as-needed basis, so that customers can acquire only the functionality they need at each node. Multi-patterning, pattern matching, advanced fill technology, advanced reliability checking – all of these and more can be readily adopted and implemented when and as they are needed or required.

Of course, software costs alone are not sufficient to understand the total cost impact to the design house of a node migration. The total hardware costs must also be considered. A true understanding of the total cost of physical verification from a hardware point of view must address the maximum number of processors needed over the course of a design flow. Once the number of processors required for each tool is understood, then the total cost for each tool can be calculated.

It is critical to consider not just the price per processor, but also additional one-off and recurring hardware costs, as described below.

1. Total cost to purchase appropriate hardware. This cost can vary dramatically depending on specific requirements, such as:
   - Number of processor nodes
   - CPUs or cores per node and corresponding availability of virtual cores
   - RAM required per node
   - Disk storage required
   - Network configuration required
   - Rack housing architecture to hold the nodes
2. The recurring cost of supporting the hardware, including:
   - Hardware vendor support cost
   - Internal IT support cost
3. Licensing and support costs associated with grid queuing and allocation software
4. Recurring electrical costs to power the processor farm configuration
5. Fixed and recurring costs associated with cooling the processor farm
6. Real-estate costs to house a processor farm

The fixed cost of the hardware can vary greatly, depending upon the specific requirements of the verification software in question. In the past few years, the cost of processors has fallen significantly, and this trend appears to be continuing. It is now economically feasible to purchase a single machine with > 64 total physical cores. Such an approach can help to reduce total housing and related costs. In addition, for advanced node, large, full-chip designs, physical verification runs may require significant RAM. The total amount of RAM that can be housed on a node is associated with the number of available dual in-line memory modules (DIMMs) in the system. Having access to such large devices also ensures a large number of DIMMs to better ensure sufficient memory is available for the job (figure 10).

Also, effective use of a compute farm doesn't have to be (and shouldn't be) limited to one function, like physical verification. To make the most effective use of multiple processors, it is important to consider all the design and corporate software functions that the farm can support. Users should work with qualified hardware vendors who have the expertise to ensure that they take best advantage of all their resources.

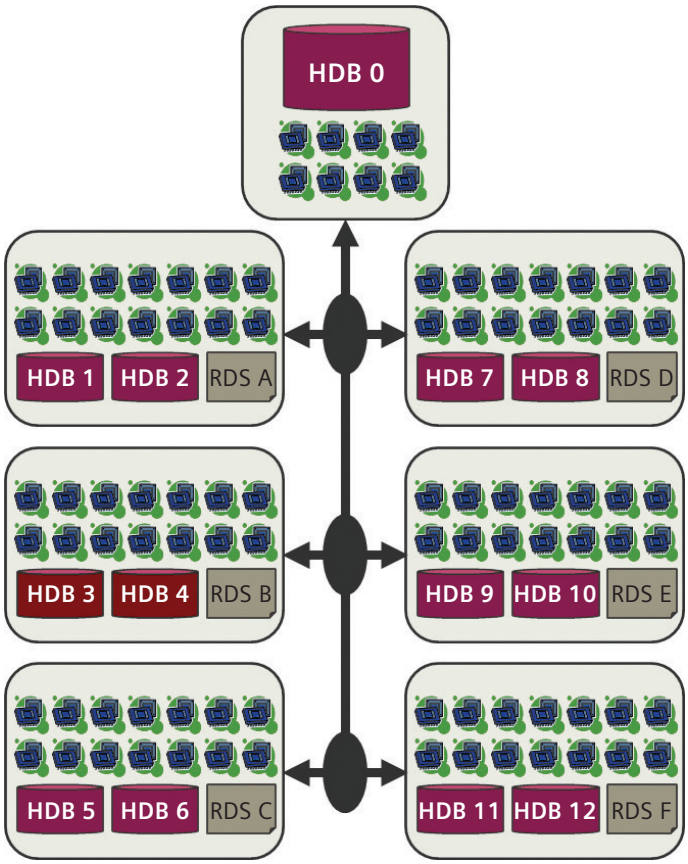

Figure 10: The Calibre nm Platform uses a variety of hardware and software innovations to ensure fast verification runtimes for all designs at all nodes.

One approach that is becoming more and more interesting to many users is the use of cloud computing to replace in-house hardware. Particularly for off-peak cloud hours, the costs can often be essentially equivalent to the costs of self-owned server farms, but without the hassles associated with the internal infrastructures of housing, powering and cooling. On the other hand, users must be able to count on access to sufficient cloud resources at the time they need them, not just during low-demand or off-peak times. In addition, cloud computing is still considered risky by many IC design houses due to uncertainty of the security levels protecting their data in a third party cloud.

Conclusion

When migrating from one node to a smaller node, there is always an increase in total cost, due to the need to run more complex tasks, often requiring more licenses, newer licenses, and more hardware. When migrating to the most advanced process nodes, the cost often increases dramatically due to the increased cost per mask, the increased total number of masks required, a significant bump in required hardware and tool licenses, and the fact that the reduction in die areas has been lost. In these cases, the migration to such nodes is primarily predicated on the need for faster performance and/or lower power. Interestingly enough, markets where these characteristics are attractive are still plentiful, with many design houses targeting process nodes at 16nm and below.

In the end, the solution that requires the least amount of total change – in the form of license configurations, required hardware resources, necessary tool qualifications, and adequate support infrastructure – will always be the most "inexpensive" option. The Calibre nmPlatform is continuously re-engineered to provide industry-leading performance at every node, while ensuring that customers can advance from node to node with the confidence that the Calibre tools they use for today's designs will be equally accurate, efficient, and comprehensive when addressing tomorrow's requirements.

References

1. Max Roser (2016). "Technological Progress." Licensed under CC-BY-SA by Max Roser. Published online at: https://ourworldindata.org/technological-progress.
2. "Does size matter? Understanding wafer size," anysilicon, Oct. 23, 2012. http://anysilicon.com/does-size-matter-understanding-wafer-size.
3. Adapted from C. Alpert and G. Tellez. "The importance of routing congestion analysis," IBM Research. DAC Knowledge Center. https://dac.com/knowledge-center/back-end-topics/physical-design-design-closure/importance-routing-congestion.
4. Adapted from International Technology Roadmap for Semiconductors (ITRS), 2013. http://www.itrs2.net.
5. Arindam Mallik, Naoto Horiguchi, Jürgen Bömmels, et al. "The economic impact of EUV lithography on critical process modules," in Proc. SPIE 9048, Extreme Ultraviolet (EUV) Lithography V, 90481R (April 17, 2014); doi: 10.1117/12.2046310.

Siemens Digital Industries Software

Headquarters
Granite Park One
5800 Granite Parkway
Suite 600
Plano, TX 75024
USA
+1 972 987 3000

Americas
Granite Park One
5800 Granite Parkway
Suite 600
Plano, TX 75024
USA
+1 314 264 8499

Europe
Stephenson House
Sir William Siemens Square
Frimley, Camberley
Surrey, GU16 8QD
+44 (0) 1276 413200

Asia-Pacific
Unit 901-902, 9/F
Tower B, Manulife Financial Centre
223-231 Wai Yip Street, Kwun Tong
Kowloon, Hong Kong
+852 2230 3333

About Siemens Digital Industries Software

Siemens Digital Industries Software is driving transformation to enable a digital enterprise where engineering, manufacturing and electronics design meet tomorrow. Our solutions help companies of all sizes create and leverage digital twins that provide organizations with new insights, opportunities and levels of automation to drive innovation. For more information on Siemens Digital Industries Software products and services, visit siemens.com/software or follow us on LinkedIn, Twitter, Facebook and Instagram. Siemens Digital Industries Software – Where today meets tomorrow.

siemens.com/eda

82034-C2 8/17 H