# An economic analysis for optimal distributed computing resources for Mask Synthesis and Tape-out in production Environment

Chris Cork, Robert Lugg, Manoj Chacko
Synopsys Inc.
Synopsys Technology Park
2025 NW Cornelius Pass Rd.
Hillsboro, OR 97124

Shimon Levi
Tower Semiconductor Ltd.
Ramat Gavriel Industrial Area
P.O. Box 619
Migdal Haemek, Israel 23105

## ABSTRACT

With the exponential increase in output database size due to the aggressive optical proximity correction (OPC) and resolution enhancement technique (RET) required for deep sub-wavelength process nodes, the CPU time required for mask tape-out continues to increase significantly. For integrated device manufacturers (IDMs), this can impact the time-to-market for their products where even a few days delay could have a huge commercial impact and loss of market window opportunity. For foundries, a shorter turnaround time provides a competitive advantage in their demanding market, too slow could mean customers looking elsewhere for these services; while a fast turnaround may even command a higher price. With FAB turnaround of a mature, plain-vanilla CMOS process of around 20-30 days, a delay of several days in mask tapeout would contribute a significant fraction to the total time to deliver prototypes.

Unlike silicon processing, masks tape-out time can be decreased by simply purchasing extra computing resources and software licenses. Mask tape-out groups are taking advantage of the ever-decreasing hardware cost and increasing power of commodity processors. The significant distributability inherent in some commercial Mask Synthesis software can be leveraged to address this critical business issue.

Different implementations have different fractions of the code that cannot be parallelized and this affects the efficiency with which it scales, as is described by Amdahl's law. Very few are efficient enough to allow the effective use of 1000's of processors, enabling run times to drop from days to only minutes. What follows is a cost aware methodology to quantify the scalability of this class of software, and thus act as a guide to estimating the optimal investment in terms of hardware and software licenses.

Keywords: OPC, distributed processing, tape-out, economics, Amdahl's Law.

## 1. INTRODUCTION

The increasing market pressure for fast turnaround time of new products and the explosion in data processing for advanced technology nodes places the turnaround time of Mask Tape-out group under critical scrutiny from the people driving the business of those products. Fortunately, the availability of cheap commodity processors and Mask Synthesis software to leverage this, provides a way to buy back the increasing processing time by investing in extra hardware. This paper addresses how the Mask Tape-out group can assess where the sweet spot in terms of computing investment lies as function of the total business strategy. Even for a silicon Foundry, which is somewhat buffered from the economic situation of individual products, a faster tape-out to prototype turnaround time can be marketed as a competitive advantage, even perhaps realizing a premium price.

## 2. OPC CPU TIME SCALING WITH TECHNOLOGY NODE

Silicon Design rules have been and for the foreseeable future been following Moore's law [1] which means that every 2-3years the density of silicon chips double or their minimum features size roughly reduces by a factor of 1.5.



Photomask and Next-Generation Lithography Mask Technology XII, edited by Masanori Komuro,
Proceedings of SPIE Vol. 5853 (SPIE, Bellingham, WA, 2005) · 0277-786X/05/$15
doi: 10.1117/12.617132

Current process nodes for instance have included: 180nm, 130nm, 90nm, 65nm and 45nm the ratio of subsequent generations approximate R ~ 1.5. While the chip size remains roughly constant the number of polygons will increase such that the total length of edges should scale as R. As the number of segments that will need OPC applied to it increase, so will the required accuracy of the final correction. This will impact the number of other edges that need to be considered to determine that best correction i.e. edges within the proximity ambit, which will not scale as fast, leading to additional data flattening. The number of model kernels and iterations with which to achieve that correction may also increase. Moore's law also predicts a doubling of processor speed [2], and naturally correction recipes and engines will become more efficient with each node, so this will help to soften the CPU requirements. Nevertheless, Mask Tape-out groups must inevitably face the multiplicative effect of about 7 factors between each generation potentially leading to a significant increase in Mask Tape-out time for a new technology node run on hardware optimized for the previous node.

| | Parameter | Scaling Factor |
|---|---|---|
| 1 | Total length of Edges within chip area | R |
| 2 | Number of segments per unit length | R |
| 3 | Number of edges within ambit | ~R to $R^2$ |
| 4 | #Model Kernels to achieve required accuracy | ≥1 |
| 5 | Number of OPC iterations | ≥1 |
| 6 | Extra edges due to more aggressive RET | ≥1 |
| 7 | Data flattening due to larger relative ambit | ≥1 |

**Table 1. OPC CPU time scaling with technology node**

## 3. 1st SILICON CYCLE-TIME DICTATES TAPEOUT SPEED

A Fab Mask Tape-out group provides the service of taking a design and providing data in a format that can be sent to the mask shop. This is a time critical step for process development as well as the introduction of new products. For today's multi-billion dollar Fabs to be economic the cycle time in process steps must not change dramatically between process nodes. Many steps, such as deposition and etch, see little change in processing time, while those that may such as lithography equipment justify a significant hardware and software investment to maintain productivity levels.

There are two key timing considerations upon which having a faster tape-out can provide added value to a new product. The first is the "Setup time": typically a Fab will start the first silicon wafers for a new product when, say, the first three masks arrive at the Fab from the Mask Shop. Anything that enables the data for the mask or the masks themselves to be delivered quicker will reduce the time to working prototype. The second key event is cycle time, if the first silicon lot will need to wait at any photolithography step until the mask needed to process that level arrives this will again impact time to working prototype. The speed of a first silicon lot between subsequent Metal layers (Metal and Via) is typically 2-3 days. As advanced processes require multiple masks to be created often six or more, so if the cycle-time to do the data processing on these masks exceeds that of the Fab cycle time, there will be a delay on the lots which will be multiplied across the number of metal layers. Turnaround time improvement will give an economic benefit only up to the Fab cycle time. Faster turnaround times, however, will give no additional gain.

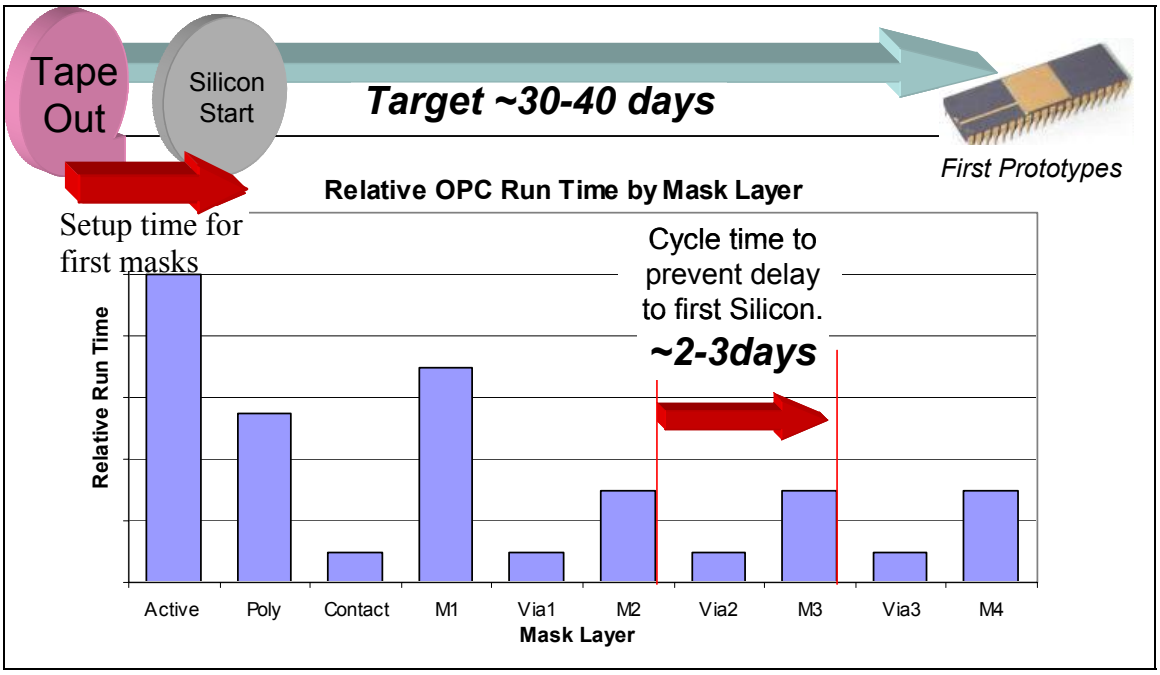



**Fig.1 How first silicon cycle time dictates tape-out time requirements.**

## 4. THE IMPLICATIONS OF AMDAHL'S LAW

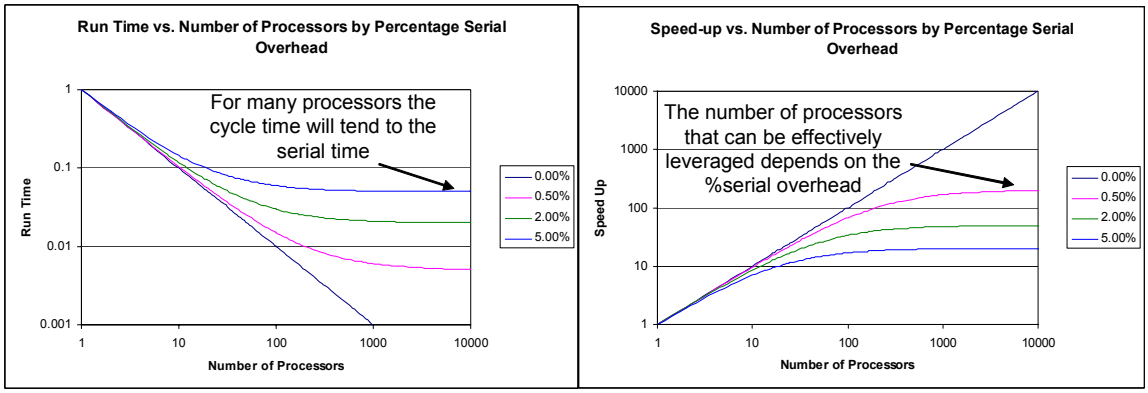
To be able to predict the scalability of any OPC correction recipe, one needs to be able to model how run time will vary with number of processors.

**Eqn. 1**

$$RunTime = s.T + \frac{(1-s).T}{N}$$

T = The amount of wall clock time the sequential version of the program takes to execute.
N = Number of processors dedicated to the parallel version of the program during execution.
s = Serial overhead. The fraction of the execution time T that must be spent in sequential mode

Amdahl's law [3] described how in a non-bandwidth limiting computing system, adding extra processors will reduces overall cycle time. This is done by considering the proportion of the code that can and cannot be parallelized. The ability to effectively leverage multiple CPUs for this task depends on the percentage serial overhead time of the recipe, as shown by Figs. 2 and 3 below. The equation shows that as the number of processors used tends to infinity the total wall clock time will tend to the serial time (Fig. 2), this means the practical limit on number of processors that can be effectively used is a function of the % serial overhead (Fig. 3).




**Figs. 2 & 3 – Shows how the advantage of adding extra processors is limited by the %serial overhead of the computation.**

For a serial overhead of 5%, tens of processors can be used effectively, for 0.5% hundreds can be used and thousands of CPU for even lower serial overhead values.

Fig. 4 shows Speed up vs. Number of Processors on an actual 90nm design for up to 100 CPUs using Proteus OPC. This graph shows an effective serial overhead of 0.2%. Others, using a recipe, presumably, with an even lower serial overhead, report near-linear performance using thousands of processors. [4]

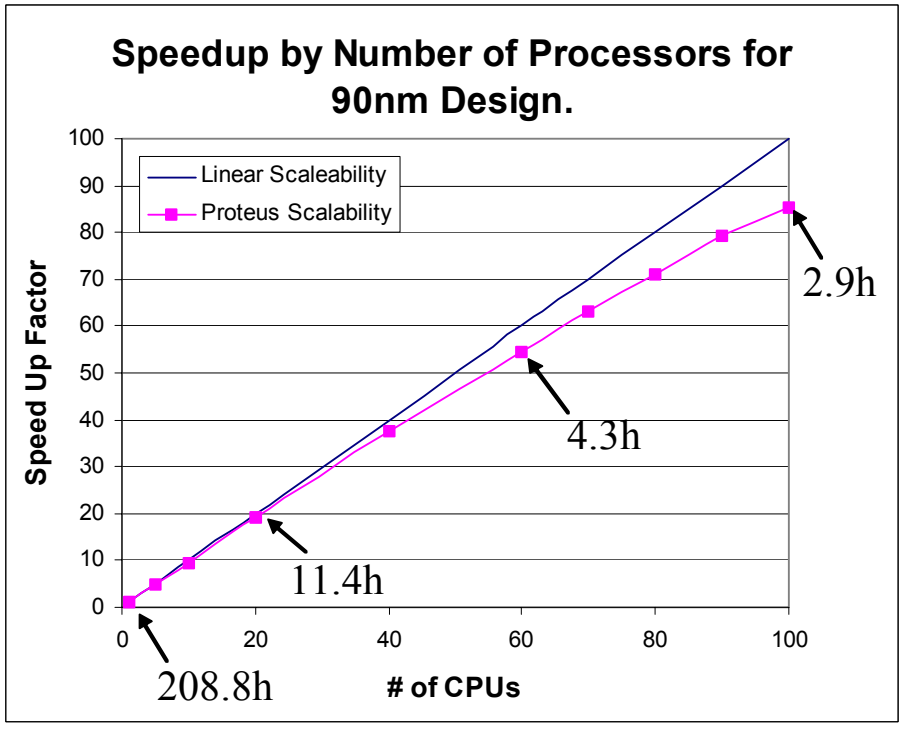



**Fig. 4 Speed up Factor vs. Number of CPUs for a 90nm OPC job.**

## 5. THE COST OF BEING LATE WITH A NEW PRODUCT

Fabs and especially foundries are judged on the turnaround time from receiving data to working prototypes. Mask tape-out is on the critical path for any new product introduction or new process development, any delay in mask tape-out will reflect on competitiveness of Fab. The increasing complexity of designs and the market pressure for

speed, mean design teams have to walk the fine line between racing towards tape out and cautiously avoiding errors [5]. For the more advanced products design re-spins are more common than not, and one re-spin will double the impact of mask tape-out delays. The global semiconductor market is estimated at $100 billion/year, of which ASICs take 10% with 10,000 designs a year [6]. The loss in revenue of chip designs due to delays in time to market has been estimated at between 1-5% per month depending on the product life-span [7]. The high mask costs of the advanced process nodes mean that high volume, high value ASICs, ASSPs and Microprocessors are the first OPC intensive products to use these nodes. The market for these products supplying for instance the cell-phone or DVD market have the shortest product lifetimes and thus are most sensitive to time to market. A loss in revenue due to delay of significantly more than 5% a month may be expected for such designs.

## 6. THE COST OF BEING LATE WITH A NEW PROCESS

The first availability of suitable Fab equipment usually dictates when process development for a new technology node can begin. Moore's law means there is a limited time within which to develop and put into manufacturing the new technology. For the mask tape-out group the process development or transfer phase can be the first indication of the step function increase in CPU requirements. If an appropriate investment has not been budgeted for CPU resources there will be a corresponding step function increase in Mask tape-out delay. For the latest technology nodes successful OPC and RET and its feedback to suitable design styles can be the area requiring the most significant yield learning [8]. The impact of even a few percentage improvements in yield on a multi billion dollar Fab, can have a very significant effect on Fab profitability and competitiveness [9]. This translates directly to purchasing that extra percentage of the enormously expensive Fab equipment and materials to supply the customer. The profitability of the Fab is therefore enhanced if more yield learning cycles can be achieved within the limited time available for process development [10,11]. This in turn is determined by the cycle time from tape-out to silicon. If no investment is made in new hardware at this crucial stage, the typical one day mask tape-out turn around time can spiral to a week for the first mask, and the latter masks will add additional days of delay. While investing in new hardware can not only enable extra design cycles, but new algorithms on how to improve design for manufacturability could be implemented within the OPC software by the user, at no extra cost to cycle time, if further processors are bought to take this load.

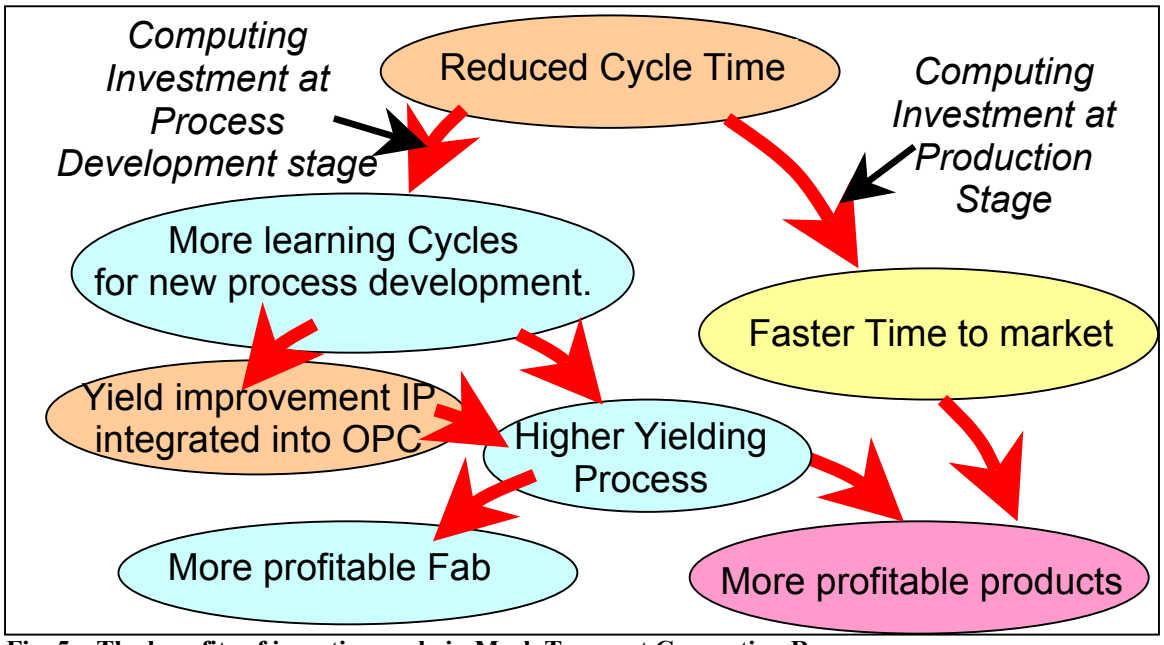



**Fig. 5 – The benefits of investing early in Mask Tape-out Computing Resources**

## 7. MARGINAL COST AS A METRIC FOR DETERMINING OPTIMUM HARDWARE INVESTMENT

In principle hardware purchasing decisions can be based on margin cost. This is the point at which the cost of adding extra processing capabilities equals the extra revenue realizable from an incremental cycle time improvement. Although revenue improvement numbers are hard to determine, their values can be considerable as has already been discussed.

From Amdahl's law for a given serial overhead there is a diminishing return on adding extra processors.
Cost benefits may be hard to quantify but will include faster product introduction and better yield.
Incremental Time saving from using N+1 compared with N processors $\approx (1-s_i)T_i/N_i^2$
Cost benefit of saving time over 1 year = $\sum CostOfDelay_i(1-s_i)T_i/N_i^2$
Incremental yearly depreciation Cost of node (includes hardware cost, maintenance and licenses)

$= Depreciation,$

Optimum number of processor based on an average run:

$$N = \sqrt{No.ofTapeouts \times CostOfDelay\ (1-s\ )T/Depreciation}$$

## 8. GRAPHICAL REPRESENTATION OF OPTIMAL NUMBER OF PROCESSORS BASED ON MARGINAL COST

Taking an average product cost of delay and cycle time allows for a graphical representation for the benefit in reducing setup time only (see Fig. 6):

$$N = \sqrt{[(1-s)T].[Costof\,Delay.NumberOfTapeOuts/Depreciation]} \qquad \textbf{Eqn. 2}$$

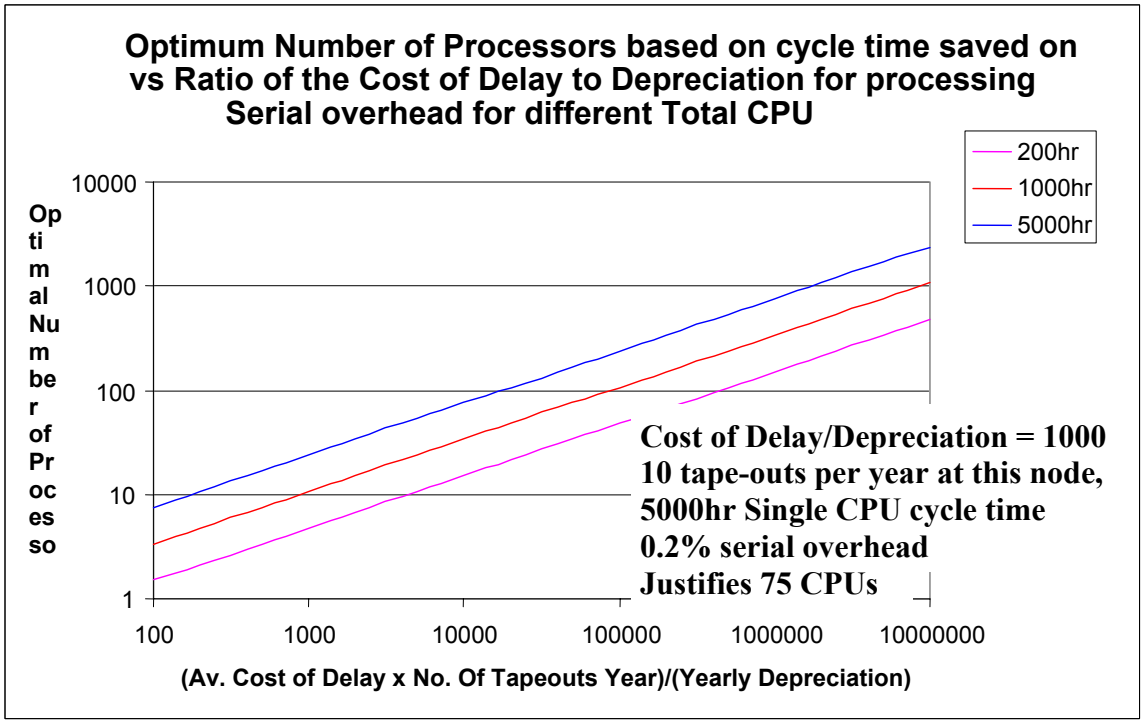



**Fig. 6 – Using marginal cost to determine optimum number of CPUs for OPC as a function of Cost of Delay to Depreciation for delay in receiving the first masks only.**

For a typical process there are 20-30masks, avoiding any cumulative delay across all of these will dominate the cost calculation for extra CPUs. The calculation becomes more complex as now the value of adding extra CPUs is no longer a smooth function, but either zero for times faster than the cycle time or a function similar to the "setup time" case. In Fig. 7 we consider adding an addition NM metal/via layers all the same single processor CPU time. Now the function is similar to the "setup time" case but the slope of the axis is increased by a factor (2.NM+1) up until the Fab metal cycle time is reached. Beyond that the curve is flat until a significantly better Cost of Delay to Depreciation ratio can justify purchasing extra CPUs to giving a turnaround time improvements to the "Setup Time" alone.

**Eqn. 3**

$$Depreciation = CostOfDelay \times NoOfTapeouts\{2NM\begin{bmatrix} 1 & T(s+(1-s)/N) > 1day \\ 0 & T(s+(1-s)/N < 1day \end{bmatrix} + 1\}(1-s)T/N^2$$

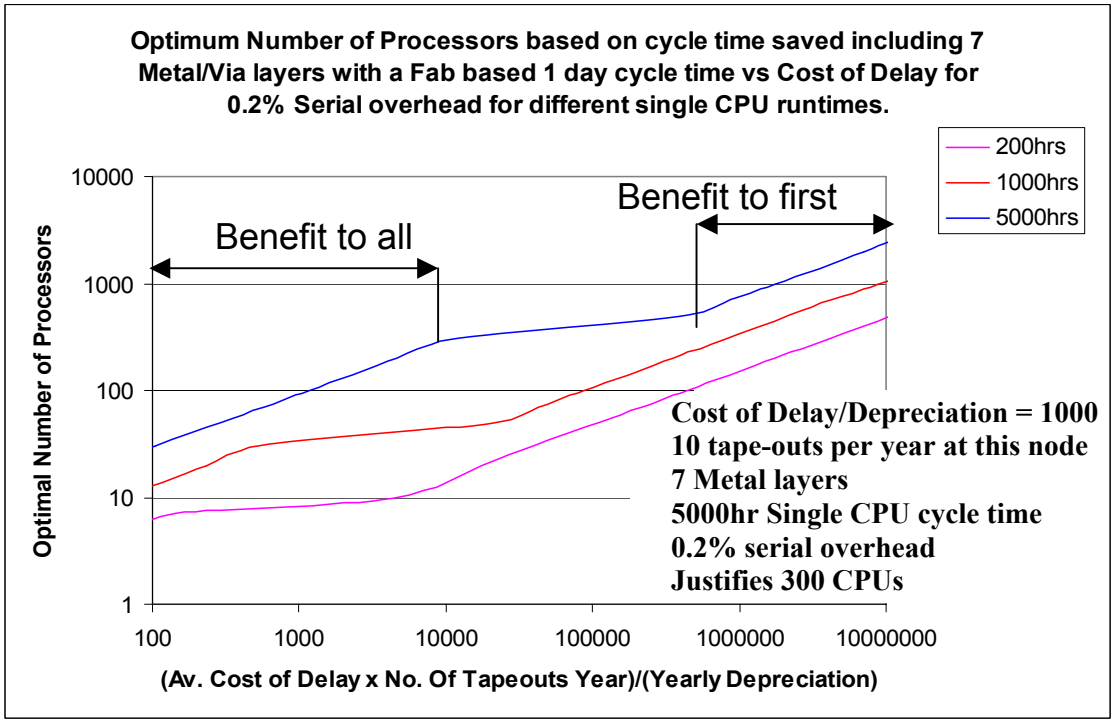



**Fig. 7– Using marginal cost to determine optimum number of CPUs for OPC as a function of Cost of Delay to Depreciation for delay in receiving multiple masks.**

## 9. CONCLUSIONS

The requirements in terms of CPU computation for OPC model evaluation between successive technology nodes scale much faster than the CPU speed with which to calculate it. Progressively more CPUs must be used to provide a mask tape-out turn around time which does not significantly impact the turnaround time of the first silicon in the Fab. Significant benefits can be gained by leveraging the massive parallelism in the latest OPC tools, including faster time to market, faster yield learning and smarter yield aware correction. Marginal cost can be used as a metric in combination with Amdahl's law to determine the optimum number of CPUs for mask tape-out such that the additional profit that can be accrued from reduced cycle time by adding one extra CPU balances the depreciation on that computing node.

## REFERENCES

[1] Cramming more components onto Integrated Circuits. G.Moore. Electronics 1965
[2] Microprocessors circa 2000.Gelsinger, Gargini, IEEE Spectrum Oct 1989.
[3] Robert Lugg, Mathias Boman, Jim Burdorf, Michael Rieger, "An Effective Distributed Architecture for OPC & RET Applications", Proc SPIE, p. 903-908 Vol. 4889. 2002
[4] 23 Feb 2004 Press Release: Synopsys' Proteus Optical Proximity Correction Software Delivers Near Linear Performance Increase Using 1000 Intel® Xeon™ Processors
[5] FPGAs Expand Their Roles as Best ASIC Replacement http://www.xilinx.com/company/success/asic.htm
[6] What is the next Implementation Fabric? IEEE Design and Test of Computers 0740-7475 2003
[7] Paradigm Shift in ASIC Technology http://www.easic.com/technolgy/Paradigm_Shift_in_ASIC_Technology.pdf
[8] Redefining Design for Yield. Srinivas Raghvendra - Compiler Magazine.
http://www.synopsys.com/news/pubs/compiler/art2_redefi-jan04.html
[9] Yield: The Key to Nanometer Profits. T.Lecklider Evaluation Engineering March 2005
[10] Understanding Fab Economics by Robert C. Leachman, John Plummer, and Nancy Sato-Misawa (June 1999) - Technical Report ESRC 99-05, Engineering Systems Research Center, UC Berkeley, June 1999.
[11] Measuring Dynamic Capabilities Practice and Performance in Semiconductor Manufacturing. JT Macher, DC Mowery. Organization Science 2004