

INTERCONNECT OPTIMIZATION FOR RELIABILITY AND SCALABILITY

Rafi Saied¹ and Vikas Manra²

¹Senoir Staff Engineer, Intel Corp., Folsom, CA, USA

²Senior Design Engineer, Intel Corp., Folsom, CA, USA

ABSTRACT

As process nodes continue to shrink to improve transistor density and performance, it is causing an increase in resistance of interconnects. At higher voltages devices speed up even more, causing interconnect to become the frequency limiter. Relatively in every node the interconnect delays continue to increase. In High Speed designs timing optimization is done across several modes to coverage design for multiple PVT points. With growing demand for Internet of Things (IoT) and autonomous driving, the need for high speed designs to be reliable and dependable at extreme conditions of high voltage and temperature continues to increase. Due to this the path profile across multiple PVT have changed with special focus on high voltage due to interconnect dominance. While most tools are capable of multi corner optimization, the increase in number of modes has challenges and more enhancements are needed to address interconnect. One such area is Interconnect scaling with repeater optimization at Section and Full chip levels. While previous papers have looked at several techniques to improve timing and slope for a given corner [1], [2], [3], this paper describes a formula to design the repeater solution to optimize delay scaling across different corners such as High Voltage and Typical Voltage to maximize performance.

KEYWORDS

Repeater, interconnect, routing, optimization, multi corner, scaling, reliability.

1. INTRODUCTION

It is increasingly challenging to address interconnect scaling in deep sub-micron designs as process modes continue to scale transistors and interconnect density increases causing an increase in wire resistance. This increases the effort to solve performance and power and adds complexity to design [4]. There have been many papers that address optimizing interconnect using different strategies [4], [5], [6] and improving quality [8]. The Industry tools and methodology is designed to converge across multiple corners using well known formulas and routing algorithms to optimize interconnect and other optimization techniques to improve repeater performance and power [5], [10], [11], [15]. The repeater optimization is done by tools which are multi corner aware, that finds the best routing based on the maximum repeater distance (MRD). The driving fub is then required to size the output driver based on the topology of the route.

There are many aspects involved in performance verification of interconnect such as process technology, voltage, temperature, library design of repeater cells, routing. This paper describes a new way of optimizing interconnect through:

1. Repeater library cells Pruning
2. Balancing the RC to device delay ratio across voltage.
3. Placement analysis

We will show that this optimization not only improves scalability but also helps power by reducing size (Z) of total transistors for the route.

2. CURRENT METHODOLOGIES

Repeater optimization is done by tools which are multi corner aware, that finds the best routing based on the maximum repeater distance (MRD). Maximum repeater distance is the maximum length a given net can be before it needs a repeater. This is based on the process, the metal layer and the pitch [14]. For e.g., if MRD is 180um in a given Process for Metal X, then the net which is 500um long must have at least two repeaters. The driving fub is then required to size the output driver based on the topology of the section route. The typical standard of repeater insertion flow is to use the mathematical formulas such as Fig.1 to determine the placement of repeaters for the interconnect optimization.

$$D = k \left[0.7R_{inv} \left(\frac{C_{wire}}{k} + C_{inv} \right) + \frac{R_{wire}}{k} \left(0.4 \frac{C_{wire}}{k} + 0.7C_{inv} \right) \right]$$

Fig.1: Formula showing the delay of the segmented wire [7].

The above formula shows the delay of the segmented wire, where k is each segment and Rwire and Cwire are the resistance and capacitances of the whole wire, Cinv is the input capacitance of the repeater and Rinv is the effective output resistance [7]. Each repeater adds some device delay but reduces the resistive delay of interconnect, thereby improving overall delay and slope (slew) of the whole wire.

3. REPEATER CELL LIBRARY ANALYSIS

The first thing we analyzed is the repeater library cells. Library cells are the building blocks of any design and play a very important role in every aspect of the design, timing, quality, variation, noise and RV. Typically, the library exists and is considered production quality. There are two scenarios possible:

1. Production library on a mature process,
2. Production library on a new process

In both cases, since the library is expected to be production quality, only few fundamental checks are run which usually include, delay, arcs, caps and power checks. While these are standard checks that are usually done to ensure functionality and basic quality, but they might not be fully scrutinized to see if more optimization can be done. When starting a new design with a new library/process more effort must be spent on analyzing the library quality for the design. It is not possible to that all 100% of the library cells are optimized across all vectors of delays, cap, scaling and power all the time. This is because:

1. Library layout gets scaled from one process generation to next and may not be redrawn entirely, with some exceptions.
2. A few late Design Rule fixes on the library cell can cause monotonicity issues which might be too late to address in time for Design schedule.
3. Schedule pressures, and resources limitation might prevent completion all possible optimizations.

So some effort must be spent on analyzing the library in more detail, to identify such un-optimized cells and prune them from the usage list based on design priority. Power, area, timing,

robustness and RV reliability are some of the indicators for creating a prune list. Sometimes it is a few cells of different drive strengths that can be pruned and sometimes it whole cell family type itself. The next section will explain the idea through the pruning studies with examples.

Figure 2, shows the delay in pico seconds of different repeater buffers of same family but different drive strengths. Each of these repeaters are driving 30% of their Cmax loads with same slew rate. Here size x5 is not very optimal from delay point of view, it is off by 3% relative to others and preferred to be added to the prune list.

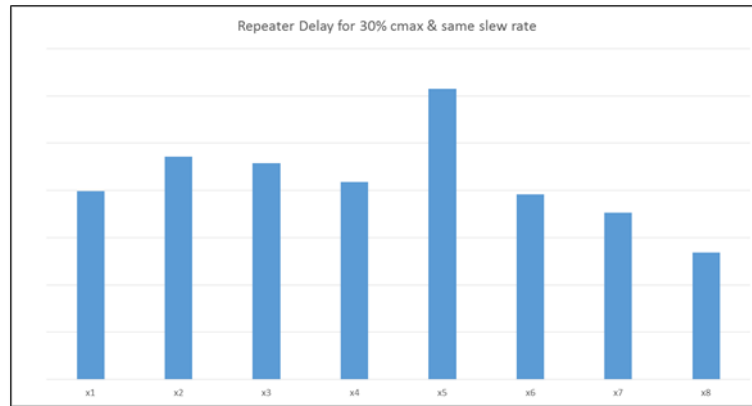


Figure 2. Delay (ps) of repeater cell of different drive strength driving 30% of Cmax

In the next level of pruning we looked at cells that didn't scale well from one revision of the node to another. This production library was generated based on a new revision of the process. Based on the way this library was generated not all cells saw the intended scaling. In Table 1 below the intended average scaling was -6%. As you can see there are some issues with type3 repeater family of drive strengths M, H and I.

Rpt Size	type1	type2	type3	type4	type5
D			-7.00%	-5.60%	-5.60%
E			-7.90%		
F			-7.90%	-6.10%	-5.90%
G			-6.50%	-5.90%	-5.70%
H			-0.50%	-6.20%	
I	-5.10%	-6.10%	-0.50%	-6.30%	-6.50%
J	-5.10%	-6.40%	-6.20%	-5.80%	-6.70%
K	-7.70%		-5.90%	-7.70%	
L	-5.50%	-5.80%		-5.80%	-6.70%
M	-4.10%		9.60%	-5.00%	
N	-5.20%	-5.30%		-4.70%	-7.00%
O	-2.70%			-6.80%	

Table 1. Scaling % of different repeater cell types of different drive strength relative to previous process revision.

Those cells were added to the prune or "don't use" list. Similarly we also added some gates to prune list based on Reliability analysis. We reviewed RV results of few blocks and identified cells that had the most violations due to p/n ratio and other criteria removed those cells (we called them unbalanced cells) from our usage list. There was no impact to timing and our analysis showed that just by removing 8 unbalanced cells, 5%-10% RV effort was reduced. This is an example of limiting usage based on RV criteria, by blocking a very small subset of RV "expensive" cells.

4. BALANCING THE RC DELAY TO DEVICE RATIO

For our study, we analyzed the repeater nets from an existing design that had gone through repeater optimization using standard tools and traditional methods. The current repeater optimizations techniques tend to find best solution to balance across both typical and high voltage PVTs and tend to have an average RC to Device ratio of 0.4. Which means for 100ps of total repeater net delay about 40% is device and the remaining is interconnect.

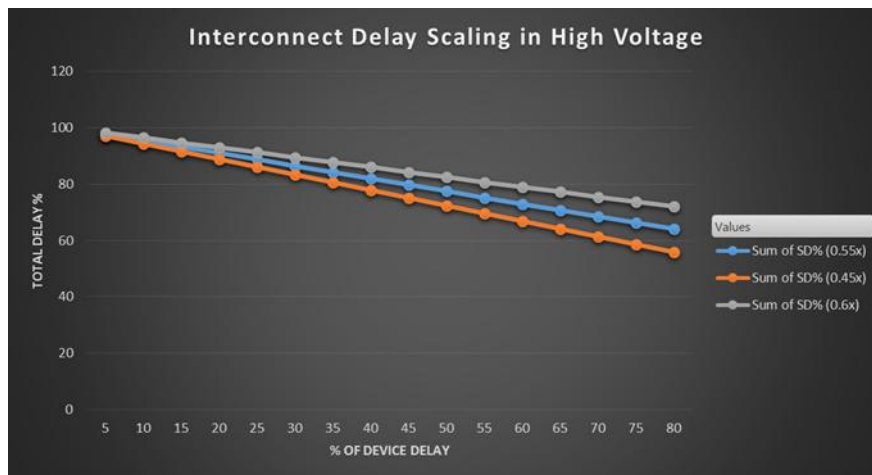


Figure 3. Total Delay scaling in High voltage for different RC to Device delay ratios.

Figure 3 above shows for various percentage of Device delay of the total interconnect delay, how the total net delays scales in high voltage. This graph highlights the well-known fact that as devices speedup in high voltage, interconnect does not [9]. So higher the percentage of device delay in your total net delay, the better the scaling will be. This paper looks at the approach to increase the % of device delay by small percentages more than what the traditional well known interconnect optimization techniques provide. Because these techniques optimize for the best delay in typical voltage, our approach calls for slight degradation in the typical voltage to get better scaling in high voltage where the most demand for performance is.

While this ratio of 0.4-0.5 is helpful for typical corner and also optimizes the number of repeaters required, the average size of the repeaters used is big. The new idea focuses on increasing the device ratio to interconnect by additional 5% to 10% depending on the topology of the net and other criteria, and at the same time optimizing the driver sizes by limiting the upper drive strength cells used.

Let's take the example of the net in Figure 4 below. This net has an RC of 103 and repeater delay of 72ps. This has a device ratio of 0.42. After our optimization the RC delay improved by 15%, and repeater delay degraded by 13%, but the total delay in both typical and High voltage corner improved. Here not only did the High voltage delay improve by 7%, but since we limited the upper sizes, and used more devices the fub output driver size was cut in half and the total Z on the net was reduced by 10%.

The way we achieve the result is to add a repeater at the first available gas station, rather than waiting for the prescribed MRD. For our study we picked nets in section that are using a fub

output driver greater than drive strength. Then using the new idea, we added the first repeater at the closest repeater bay and then added more as per the MRD. We saw that this improved almost all metal layers and even for distances as low as 170um. Additionally due to the pruning of high drive strength repeater cells, we reduced overall Reliability verification effort by 10% and reduced IR drop violations by 5%. This approach doesn't have to be all or nothing. It can be used on targeted nets that belong to interconnect dominated paths in the design.

5. RESULTS

In Figure 4, below we show a conservative approach where we increased the device ratio by a small percentage. Here we focused on power reduction by reducing the Z (the sum of device widths used). The light green and blue columns show the total delay of net: Light green is the RC and blue is the Device delay. Post our optimization, (yellow + dark green), total delay is same or less, but the device component of delay (dark green) has increased and RC component (yellow) has reduced. We have shown for various metal layers and different length of the total wires. Figures 5 and 6 show similar data for both the typical and high voltages.

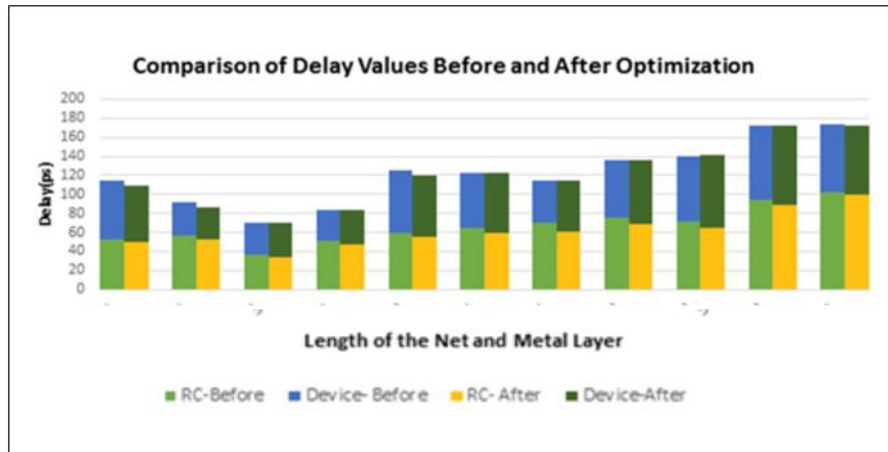


Figure 4. Delay distribution before and after optimization.

We saw similar results for nets of different lengths and metals as evident from the figures 5 and 6 below.

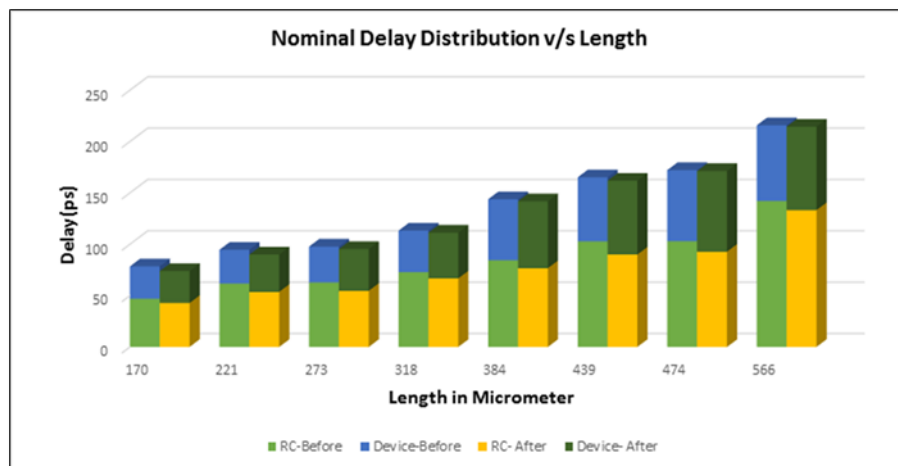


Figure 5. Delay distribution for different lengths of net in typical voltage corner.

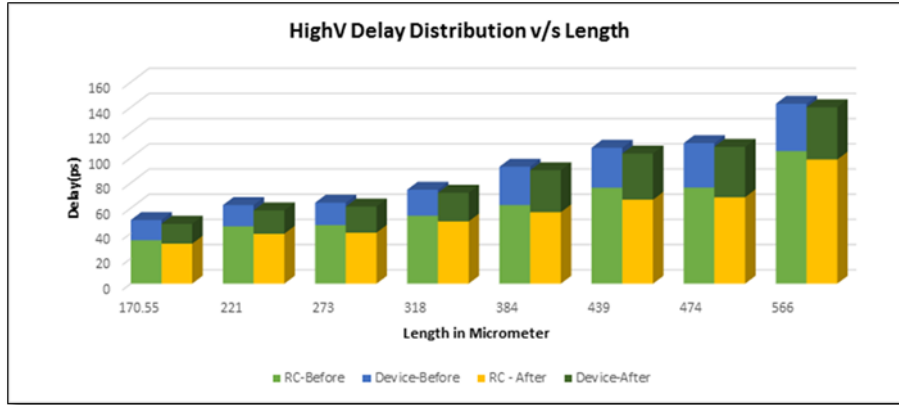


Figure 6. Delay distribution for different lengths of net in High Voltage corner.

6. CONCLUSIONS

In all examples shown above, overall timing is same or better and we have reduced the RC component of the total delay. Reducing the RC component of the net gives better scalability from typical to high voltage, as devices improve with higher voltage but RCs don't [13]. In summary we see that by improving the RC to device ratio by even a small percentage, we see benefits not only in timing and Power (through Z reduction), but we will also see better scalability as we are reducing the RC component, that doesn't scale with voltage.

Table 2. Overall gains from this approach.

Gains	
Speedup (freq)	1.37% Speed up at High Voltage
Driver size Z reduction	28%
Total Z reduction	4%
RC Delay Improvement	11%
Device Delay Increase	12% at Typical Voltage

One caveat of this approach is it will increase the overall number of repeaters and we might not have enough room to add repeaters for all nets, but we can prioritize and do this on the nets that have high RC to device ratio and work downwards until we have room for repeaters. Additionally we were also able to make the design more robust by removing high drive strength repeaters that can cause IR drop at high frequencies and reduce the reliability verification effort. This approach can be used to prioritize or optimize specifically what the design needs: improve scalability, improve reliability and power. The table 2 above summarizes the overall gains.

ACKNOWLEDGEMENTS

Special thanks to Suraj Kashyap and Zi Wen for helping with the analysis.

REFERENCES

- [1] P P, Vinay Kumar. "Efficient Cluster Based Repeater Insertion for Slope Improvement," DTTC 2011
- [2] Mishra, V. "Repeater insertion methodology for SkyLake GT," DTTC 2013
- [3] Aguilar, J. "Montecito Repeater Insertion Methodology," DTTC 2005

- [4] Vaidyanathan, Kaushik. "Overcoming interconnect scaling challenges using novel process and design solutions to improve both high-speed and low-power computing modes", IEEE International Electron Devices Meeting (IEDM) 2017.
- [5] Tang, Min. "Optimization of Global Interconnect in High Performance VLSI Circuits", 19th international conference on VLSI Design, 2006.
- [6] Koh, Hong Yeow. "Efficient Static Timing Analysis Method for Accurate FPGA Interconnect Performance", SNUG 2018.
- [7] Moiseev, Konstantin "Multi-Net Optimization of VLSI Interconnect", pgs.45-46.
- [8] Damanpreet Kaur and V.Sulochana, "Crosstalk Minimization for Coupled RLC Interconnects Using Bidirectional Buffer and Shield Insertion", International journal of VLSI design & Communication Systems, Vol. 4, No.3, June 2013.
- [9] J.-P. Schoellkopf, "Impact of interconnect performances on circuit design", Proceedings of the IEEE International Interconnect Technology Conference, 1998.
- [10] Laxminath Tripathy and Chitta Ranjan Tripathy, "A New Interconnection Topology For Network On Chip", International Journal of Computer Networks & Communications (IJCNC) Vol.10, No.4, July 2018.
- [11] S. Rajendar, P. Chandrasekhar, M. Asha Rani, Ambati Divya, "A Novel Low Power High Dynamic Threshold Swing Limited Repeater Insertion for On-Chip Interconnects", International Journal of VLSI design & Communication Systems (VLSICS) Vol.5, No.6, December 2014.
- [12] S. Kobayashi, M. Eda, Y. Hayashi, "New interconnect structure design methodology by Layout-design-based Interconnect Structure Optimization System", Proceedings of the IEEE 2000 International Interconnect Technology Conference, June 2000.
- [13] James S. Clarke*, Christopher George, Christopher Jezewski, Arantxa Maestre Caro, David Michalak, Jessica Torres, "Process Technology Scaling in an Increasingly Interconnect Dominated World", Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers 2014, Pgs 1-2
- [14] K. Yamashita, S. Oadaka, "Interconnect scaling scenario using a chip level interconnect model", IEEE Transactions on Electron Devices, 2000, Vol. 47, Issue: 1, Pgs. 90 – 96
- [15] N.S. Nagaraj, W.R. Hunter, P.R. Chidambaram, T.Y. Garibay, U. Narasimha, A. Hill, H. Shichijo, "Impact of interconnect technology scaling on SOC design methodologies", Proceedings of the IEEE International Interconnect Technology Conference, 2005, Pgs. 71 - 73

Authors

Rafi M Saied is a Senior Staff Engineer at Intel Corp. in Folsom, CA. He holds a VMaster's Degree in Electrical Engineering from Arizona State University. He is driving Performance verification methodologies and Reliability Verification for High Speed Designs for microprocessors.



Vikas Manra is a Senior Design Engineer at Intel, Folsom. He has a Master's degree in Electrical and Computer Engineering from Southern Illinois University. Currently working as Timing convergence lead for Big Core Projects with expertise in Static Timing Analysis, Floor planning, high frequency/low power digital circuit design.

