PCIe Gen-5 Design Challenges of High-Speed Servers

Mallikarjun Vasa, Chun-Lin Liao, Sanjay Kumar, Ching-Huei Chen, Bhyrav Mutnury,
DellEMC Infrastructure Solutions Group.
{Mallikarjun_Vasa, Chun-Lin_Liao, Sanjay_Kumar22, Carol_Chen2, Bhyrav_Mutnury}@dell.com

***Abstract*— Supporting PCIe Gen-5 in high-speed servers has become a challenge to designers. Parasitic effects that were benign at PCIe Gen-4 speeds are impacting PCIe Gen-5 operations adversely. Optimizing signal and ground via placement, anti-pad dimensions, AC capacitor placement and trace routing in dense pin area are becoming important as sensitivity to loss, impedance and crosstalk is high at 32 Gbps. In this paper, few approaches to minimize channel impedance discontinuity and near-end/far-end crosstalk (NEXT/FEXT) are discussed. Two channels with optimal design practices and regular design practices are compared and contrasted.**

Keywords—PCIe Gen 5, Via Design, Crosstalk

I. INTRODUCTION

With the increasing demand for bandwidth in server designs, PCI Express (PCIe) has been extended from fourth generation (Gen-4, 16 Gbps) to fifth generation (Gen-5, 32 Gbps). PCIe Gen-5 has 31.25 ps of unit interval (UI) and uses 85-ohms channel impedance as its previous generation. The end-to-end channel loss is 36 dB @ 16 GHz. In order to meet the channel loss requirement, better PCB material(s) and/or cable(s) are needed along with optimal stack-up and layout design practices [1].

To meet the small UI requirement, better impedance control and crosstalk minimization is required on every component such as vias and connectors [2][3]. In the past, sideband termination on PCIe connectors are studied on [4]. Via patterns and optimization methods are presented in [5]-[7]. Most of the work was focused on optimizations on connector or add-in-card (AIC). In this paper, the focus is on the mother board design.

In this study, PCIe Gen-5 design challenges on server mother board channels are introduced in section II. Several channel parameters are analyzed and optimized in section III. Two channels with optimal design and regular design practices are compared and contrasted in section IV. Conclusions are summarized in section V.

II. SERVER DESIGN CHALLENGE

High-speed servers offer feature rich, configurable and customizable PCIe configurations to the customer. As a result of this, high-speed server designs are usually complex. In rack and blade servers, PCIe topologies typically have multiple boards and cables in the path to connect the root-complex to the end-device. Server PCIe topologies are also complicated with many of them having multiple connectors and/or cables in their path. Besides, PCIe channel design in servers needs to reserve channel loss budget for CPU package and AIC loss. On PCIe Gen-4, CPU package loss was typically around 5 dB with only 15 dB channel loss budget reserved for server PCIe channel. On PCIe Gen-5, CPU package loss is at 9 dB and there is only 17.5 dB channel loss budget allotted for server PCIe channel. From PCIe Gen-4 to Gen-5, the Nyquist frequency got doubled from 8 GHz to 16 GHz, but the loss budget for server design increased from 15 dB to 17.5 dB. This server budget should cover PCB trace loss, via loss, connector loss, cable loss and AC capacitor loss. Channel loss requirement is so restrictive that every channel parameter needs to be optimized to minimize their loss.

III. DESIGN OPTIMIZATION

In this study, several channel parameters are analyzed to optimize PCIe Gen-5 channel. A 20-layer stack-up with 8-signal layers is used. The core and prepreg thickness are 5 and 6 mil, respectively. The total board thickness is 120 mil. Except for a very short trace around breakout area on top layer, all PCIe Gen-5 signals are routed on inner layers and are shielded with two adjacent ground (GND) layers. A via stub of 10 mil is maintained on all channels.

A. Insertion Loss for Different Via Transion Lengths

As per the introduction, the loss budget of the PCIe Gen-5 channel did not double as the data rate. So, designers should intelligently select the routing layer so that overall channel and via loss is within the loss budget. For a 20-layer stack-up with mid-loss material (Dk-3.8 Df-0.015), the loss per via can vary by 0.3 dB or more depending on routing layer. It can be seen from Fig. 1 that inner1 via loss is 0.1 dB and inner8 via loss is ~0.4 dB at 16 GHz.

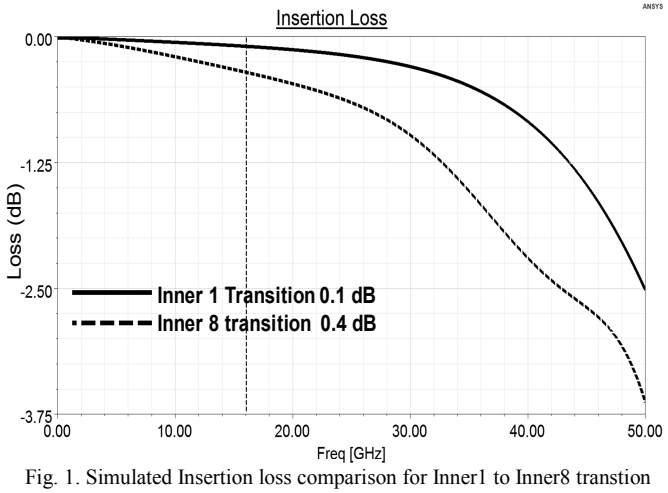


Fig. 1. Simulated Insertion loss comparison for Inner1 to Inner8 transtion

B. Via Anti-Pad Dimensions

Anti-pad refers to the area between signal vias and the ground/power plane, this is necessary to ensure signal via does not short other non-signal layers. The size and shape of the antipad would determine the impedance discontinuity. At high

978-1-7281-6161-7/20/$31.00 ©2020 IEEE

data rates, one common anti-pad cannot be followed for all the via transitions as each layer experiences a different capacitance due to planes, and their impedance are different accordingly, as shown in Fig. 2(B). In Table. 1 it is shown that optimized impedance is the combined effort of anti-pad (Ap), signal to ground distance (Sg), via pitch (Vp) and material property as shown in Fig. 2(A).

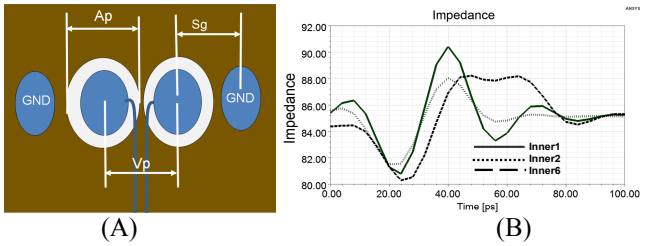


Fig. 2. (A) Top view of differential pair via (B) simulated Impedance plot for different layer transitions

TABLE 1. IMPEDANCE VARIATION ON DIFFERENT SIGNAL LAYERS

Layer	Via_Pitch (Vp)	Anti_pad (Ap)	Signal to Ground pitch (Sg)	Impedance
1	38 mil	38 mil	33 mil	(81-91) ohm
2	38 mil	36 mil	33 mil	(82-88) ohm
3	38 mil	34 mil	33 mil	(81-88) ohm
6	38 mil	32 mil	33 mil	(81-91) ohm
8	38 mil	32 mil	33 mil	(80-92) ohm

C. *GND Vias*

Ground vias are needed to be placed adjacent to PCIe signal transition vias to have good signal return path and crosstalk shielding. On PCIe Gen-4, two symmetric ground vias are typically enough. In this study, two and four adjacent ground vias structures are analyzed to see their crosstalk shielding performance. Three structures are shown in Fig. 3 and their FEXT comparisons are shown in Fig. 4. Structures (A) and (B) have 2 ground vias adjacent to PCIe transition vias, but their ground via locations are different. Since structure (A) has two ground vias between DIFF1 and DIFF3 pairs, it has better crosstalk immunity between DIFF1 and DIFF3. On the other side, structure (B) has better crosstalk performance between DIFF1 and DIFF2. However, four ground vias structure (C) has the best crosstalk performance compared to two ground via structures.

D. *GND vias on SMT GND pads*

Surface mount (SMT) connector is better for PCIe Gen-5 channel because of its better loss performance. It is recommended to add two ground vias on "heel" and "toe" sides of all ground connector pins as shown in Fig. 5. However, due to routing space limitation, some ground pins could not have two ground vias, and some ground vias could not be placed close to connector ground pins. This section uses simulation data to show the SI risk of several different ground via patterns. Fig. 5 shows the different ground via patterns analyzed in this paper. The patterns include (a) single ground via on different locations (b) two ground vias on different locations and (c) four ground vias structure. With different ground via numbers and locations, the insertion of one PCIe pair and FEXT between two adjacent PCIe pairs are compared on Fig. 6. From the observation, single ground via the resonant peak on FEXT will happen around 30 GHz. On the two ground via structures, if the ground vias could not be placed close to connector pad, it also could not get good performance. The best pattern is placing 4 GND vias on 10 mil/40 mil away from two sides of connector ground pads.

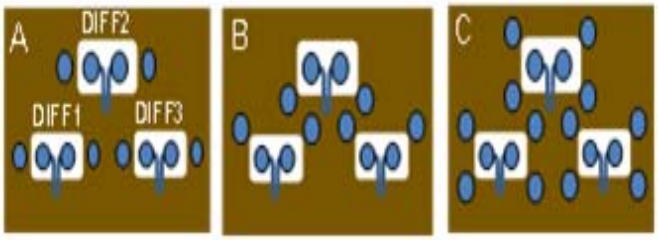


Fig. 3. Comparison of different transition via patterns.

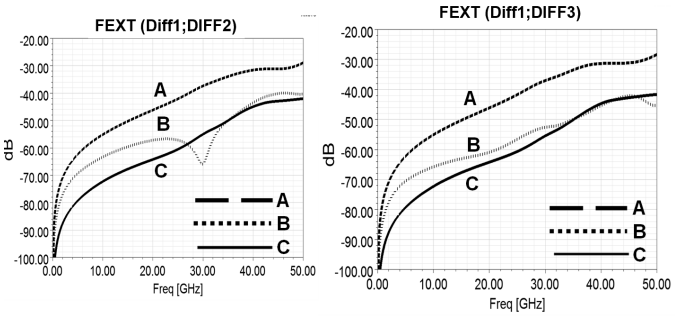


Fig. 4. Comparison of Simulated far end crosstalk (FEXT) between different transition via patterns.

E. *Crosstalk in Dense Pin Area*

If PCIe pinouts of the chip are not close to the edge of chip but in the inner area, then those PCIe signal needs to penetrate through dense pin area to the outside open area. In the dense pin area, PCIe signal traces should not go through another different pair vias. In Fig. 7, two yellow circles are P/N pins of one PCIe pair (DIFF1) and another PCIe pair (DIFF2 of blue lines) passes through P/N pins of DIFF1. Another pair (DIFF3) do not go through DIFF1 P/N pins. It can be expected that the crosstalk between the two differential pairs (DIFF1 and DIFF2) will be much higher than crosstalk between (DIFF1 and DIFF3). Simulation results are shown in Fig. 8 and it is seen that the crosstalk is ~5 dB higher between DIFF1-DIFF2 compared to DIFF1-DIFF3. In cases when signals penetrate through dense BGA pin area, crosstalk can be reduced by proper layer selection so the via crosstalk can be reduced and selecting a GND pattern around BGA signal vias.

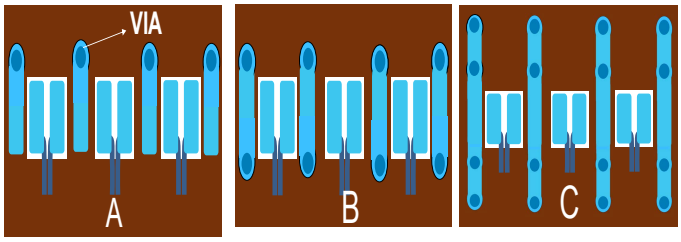


Fig. 5. (A) Single GND via placed Inner Side, (B) Two GND via placed either side, (C) Four GND via placed on either side.

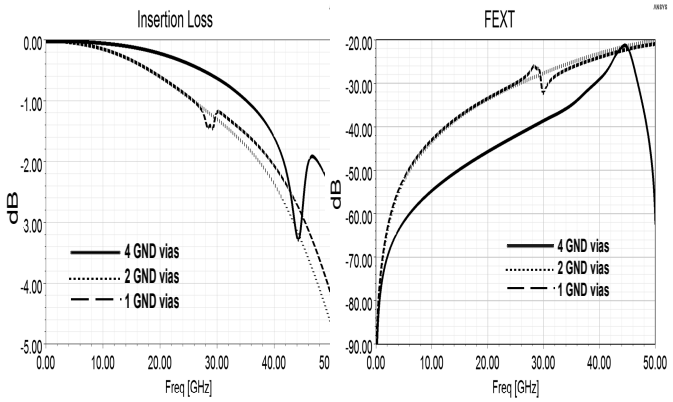


Fig. 6. Simulated Insertion loss and FEXT comparison of three structures of Fig. 5.

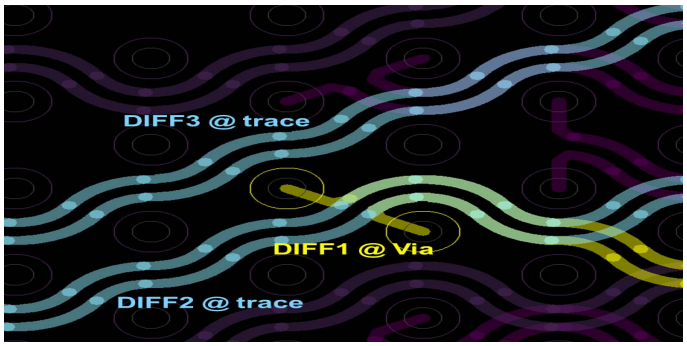


Fig. 7. Differential pairs in pin field area.

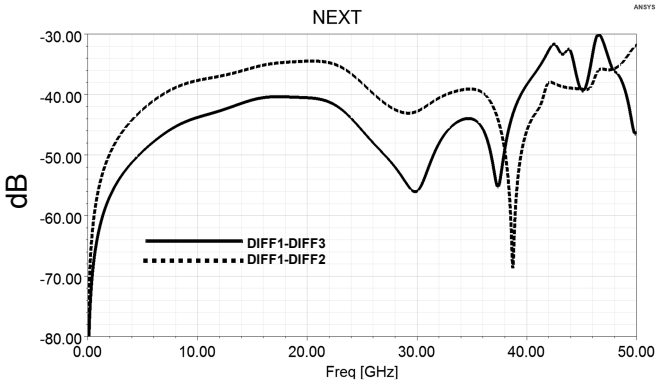


Fig. 8. Simulated NEXT for traces in pin field area.

IV. Optimal versus Regular Channels

This section uses two PCIe channels to show how much difference in eye opening can be achieved from optimal and regular channel designs. The optimal channel uses all the best design practices from section III, the regular channel on the other hand uses regular design practices. Individual channel models are cascaded to result in a channel model as shown in Fig. 9. This channel is a PCIe Gen-5 Tx channel, which includes one adapter, one AC cap and two PCIe connectors. Trace length on planar and adapters are 4" and 2", respectively. Optimal channel uses Inner1 transition vias, 5 mil core and 6 mil prepreg. Connector and transition vias have 4 adjacent GND vias. Regular channel uses Inner8 routing, 3 mil core and 4 mil prepreg. Connector and transition vias have only 2 adjacent GND vias. Trace length of both the channels are same. The cascaded model showed that the optimal channel loss is 3 dB lower and the FEXT and NEXT are 10 dB lower compared to regular channel. Eye diagram of the two channels are shown in Fig. 10, optimal channel has (EH/EW) = (37.0 mV/15.8 ps) and the regular channel (7.0 mV/7.8 ps).

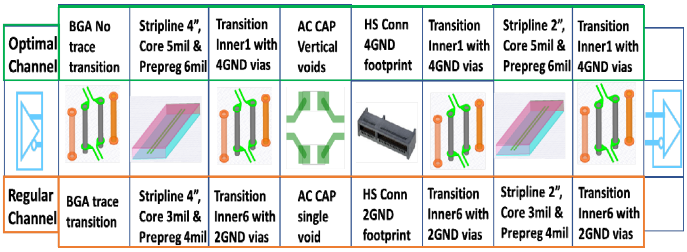
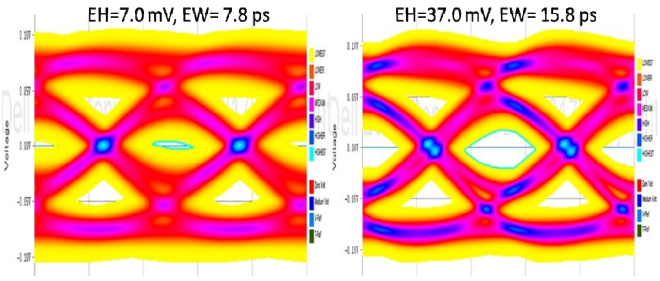


Fig. 9. PCIe Gen-5 channel topology.



(A) (B)

Fig. 10. Comparison between (A) regular channel and (B) optimal channel

V. Conclusions

In this paper, several channel parameters are optimized to reduce their loss, reflections and crosstalk. At PCIe Gen-5 data rate, any channel imperfection can begin to have a huge impact on SI robustness. Optimizing all aspects of the PCIe channel becomes critical to a successful design. In this paper, the importance of the proposed optimizations is shown using two example channels and it is evident that optimized design practices are needed at PCIe Gen-5 speeds.

References

[1] H. F. Xue *et al.*, "PCI Express 5.0 Solution Extending Study Through PCB Stackup and Geometry Optimization," DesignCon 2020.

[2] X. M. Gao, P. Z. Yang, J. X. Zhu, and G. Yang, "Modular Platform Design & Optimization for PCIe 5.0 IPs Validation," DesignCon 2020.

[3] L. Shan, D. Freidman, C. Kennedy, W. Persak and K. Lau, "Backward Compatible Connectors for Next Generation PCIe Electrical I/O," 2018 IEEE 68th Electronic Components and Technology Conference (ECTC), San Diego, CA, 2018, pp. 1798-1804.

[4] Y. Zhou, W. Shi and S. Sudhakaran, "A new approach to mitigate PCI express Gen4 crosstalk from sideband signals in connectors," 2017 IEEE 26th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS), San Jose, CA, 2017, pp. 1-3.

[5] C. Ye, X. Ye and E. L. Miralrio, "Via pattern design and optimization for differential signaling 25Gbps and above," 2016 IEEE International Symposium on Electromagnetic Compatibility (EMC), Ottawa, ON, 2016, pp. 312-317.

[6] S. Chen, C. Chen, C. Liao, J. Chen, T. Wu and B. Mutnury, "Via optimization for next generation speeds," 2017 IEEE 26th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS), San Jose, CA, 2017, pp. 1-3.

[7] M. Vasa, A. C. Reddy, B. Mutnury, S. Kumar and Vasanth RD, "High speed interconnect optimization," 2015 Asia-Pacific Symposium on Electromagnetic Compatibility (APEMC), Taipei, 2015, pp. 8-11.

978-1-7281-6161-7/20/$31.00 ©2020 IEEE