

Advanced Fly-By Routing Topology for Gbps DDR5 Systems

Shinyoung Park
Bufferchip
Rambus Inc.
 Seoul, Republic of Korea
 parks@rambus.com

Vinod Arjun Huddar
Bufferchip
Rambus Inc.
 Bangalore, India
 vhuddar@rambus.com

Abstract— From DDR3 and beyond, the fly-by has been widely used as it can support high data rate operations by providing smaller trace stubs and capacitive loadings. Even so, beyond a certain number of loadings, the fly-by also starts to have trouble in keeping up with high data rates. To accommodate larger loadings, either data rate needs to be reduced or advanced equalization techniques are required. To address this limitation of fly-by topology, we propose advanced fly-by topology routing.

Keywords— *Fly-by routing, advanced fly-by topology, parallel bus interface, DDR5*

I. INTRODUCTION

In parallel bus interfaces like DDR, there are two routing topologies well known i.e., T topology and fly-by topology. Conventional, low-speed DDR systems distribute signals to multiple receivers using a topology in which those signals propagate to all the receivers in the system at the same time using T-branches. In such systems, the propagation delays introduce timing skew into the system, thereby limiting the operating frequency of the bus and impacting the performance of high-speed memory systems. In addition, the performance of these topologies is also limited by capacitive loading [1]. Both the flight time skew problem and the capacitive loading issues discussed above can be solved using the fly-by architecture.

The fly-by topology routing also known as daisy chain topology, routes signals in a chain from single transmitter to multiple receivers. This methodology supports higher frequency operations by reducing the amount of trace stubs from T-topology, thus improving signal integrity. Typical routing topology for Gbps data rates for multi receiver parallel bus interfaces has been fly-by routing for multiple signal integrity reasons. Fly-by topology incurs less simultaneous switching noise and reduces the number via stubs. Because the arrival times of the signals at the receiver are distributed in time, the time at which the signals encounter the input capacitance of each of the receivers is similarly distributed, thus reducing the capacitive loading issues discussed above. The reduced capacitive loading enhances the signal integrity and enables higher data rate signaling [2]. Fly-by architectures enable subsystems that require operational data rates that are significantly greater than those achievable with conventional approaches. Using fly-by architectures allows designers to relax PCB trace length requirements allowing much simpler and more compact memory sub-system layouts. In addition, fly-by architectures

provide system benefits by enabling DDR systems to operate with GHz data rates. This superior DDR system performance results in improved performance in desktops, notebooks, enterprise servers and storage, HDTVs, gaming systems, and handheld portable devices for end users.

If we increase the number of receivers connected to single transmitter, beyond certain number, fly-by routing also starts having trouble in keeping up with high data rates seen in DDR5 interface i.e., meeting setup and hold requirements of each of receivers [3]. To accommodate higher number of receivers, data rate needs to be reduced or better equalization schemes at receiver needs to be implemented to meet minimum eye requirement. To address this limitation of fly-by topology, we propose advanced fly-by topology routing. Advanced fly-by routing topology can double the number of receivers as compared to fly-by with same data rate for Gbps speed DDR5 interface or improve eye with same higher loading conditions without need for better equalization schemes at receivers. Simulation results comparing fly-by and advance fly-by topology with clock and data patterns will be discussed.

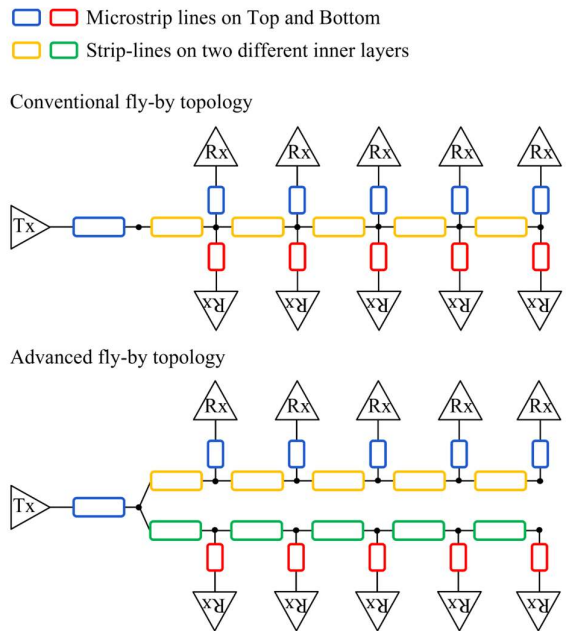


Fig. 1. Conventional fly-by topology vs proposed advanced fly-by topology

II. TIME-DOMAIN SIGNAL INTEGRITY ANALYSIS

A. Description of channel designs

The PCB (printed circuit board) channel designs used in the analysis are shown in Fig. 1 with 10 loadings per channel. In the conventional fly-by topology, the transmitter output signal travels a short microstrip line trace on the top layer for breaking out from package BGAs and then a strip-line on M-2 layer, where M stands for the total number of the layers of PCB. The stripline trace daisy-chains 5 loading sections, each with two receivers one on the top and the other on the bottom layer of PCB. We observed the signal integrity at the last receiver on the top layer since it showed the worst performance for its longest trace and via lengths.

In the advanced fly-by topology, after the first break-out trace, the routing branches off to two strip-line traces. The trace is designed on M-2 layer daisy-chaining 5 receivers on the top layer and the other trace is designed on 3rd layer daisy-chaining 5 receivers on the bottom layer. We observed the signal integrity at the last receiver on the top layer since it showed the worst performance for its longest trace. The last receivers on the top and the bottom showed similar performance as the layers on which traces are designed are symmetric to each other. All transmission lines are designed with 50 ohms characteristic impedance. For both topologies, the last two receivers are VDD-terminated to 80 ohms as default setup.

B. Experiment with clock pattern at 4.4 Gbps

The transmitter outputs clock pattern with 4.4 Gbps data rate. Fig. 2 (a) and (b) show the waveforms at the last receiver input with conventional fly-by topology and the proposed topology, respectively, both having the default 10 loadings and 80-ohm VDD-termination at the last 2 receivers. The signal swing at the receiver input on the proposed topology is twice and its arrival time is early by approximately 130 ps compared to that on the conventional indicating that the effective loading is smaller.

Fig. 2 (c) shows the waveform at the last receiver input with conventional fly-by topology with 5 receivers. The VDD-termination at the last receiver is 40-ohm, designed half of the default, for analysis in the same DC swing condition. The swing and arrival time of Fig. 2 (b) and (c) are similar indicating that the effective number of loadings with the proposed routing topology is 5, half of the actual number.

C. Experiment with data pattern at 3.6 Gbps

The transmitter outputs data pattern with 3.6 Gbps data rate. Infinite non-repeating data pattern was assumed for 1E-16 very low bit error rate (BER) simulation. Fig. 3 (a) and (b) show eye diagram at the last receiver input with conventional fly-by topology and the proposed topology, respectively, both having the default 10 loadings and 80-ohm VDD-termination at the last 2 receivers. Unlike the clock pattern result, the proposed routing topology show similar or worse eye margin compared to the conventional because it produces more reflection noise. The low frequency components in the data reflected at the receivers propagate to the receiver's daisy-chained by the trace on the other layer, and therefore the accumulated reflection noise becomes larger due to the longer stub length. Therefore, for broadband data transmission, terminating the channel that reduces reflection becomes important.

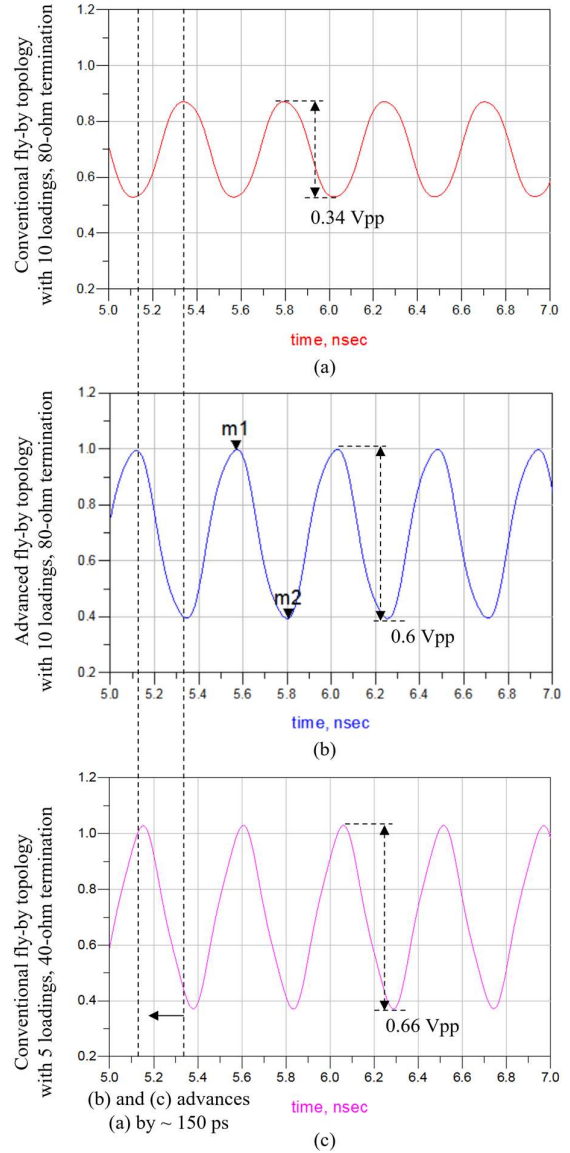


Fig. 2. Waveforms at the last receiver input of (a) the conventional fly-by and (b) the advanced fly-by with 10 loadings and 80-ohm termination at the last two receivers, and (c) the conventional fly-by with 5 loadings and 40-ohm termination at the last receiver.

Fig. 3 (c) shows eye diagram at the last receiver input with the proposed topology having 10 loadings and 50-ohm VDD-termination at the last 2 receivers. Despite the reduced DC swing due to the stronger termination, eye margin is larger because low frequency reflection noise is reduced by the termination matched to the channel impedance and high frequency signal arrived at the receiver is larger as the effective loading is smaller. The conventional fly-by having 10 loadings and 50-ohm VDD-termination at the last 2 receivers showed the worst eye margin. This is because of small DC swing due to stronger termination and large reflection noise. The effective termination seen is 25-ohm as the top and the bottom DRAMs are close to each other.

D. Experiment with data pattern at 2.4 Gbps

The transmitter outputs data pattern a 2.4 Gbps data rate. Infinite non-repeating data pattern was assumed for 1E-16

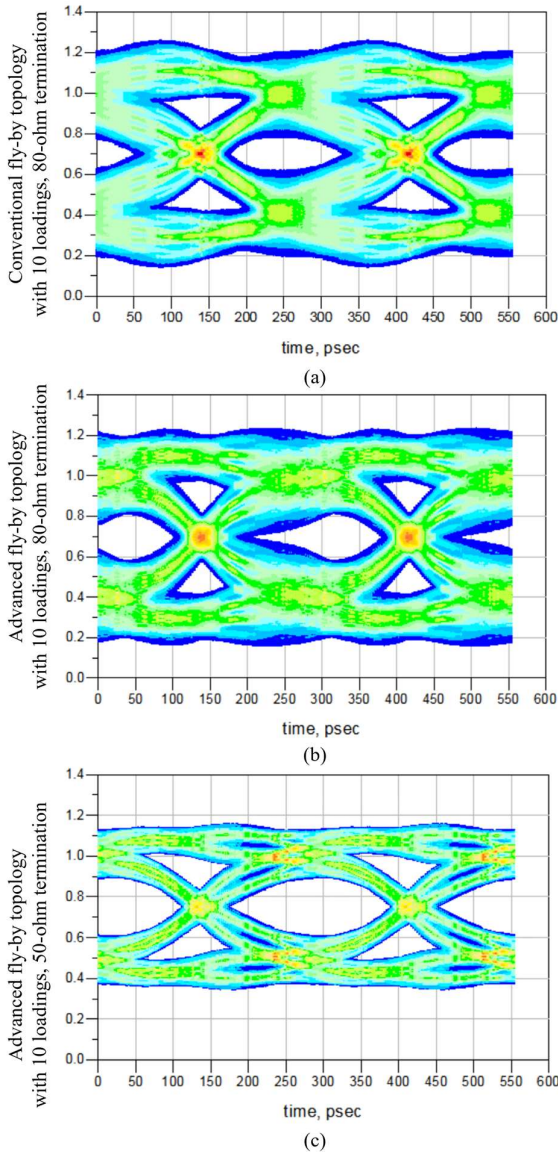


Fig. 3. 3.6 Gbps eye diagram at the last receiver input of (a) the conventional fly-by and (b) the advanced fly-by with 10 loadings and 80-ohm termination at the last two receivers, and (c) the advanced fly-by with 10 loadings and 50-ohm termination at the last two receivers.

very low BER simulation. Fig. 4 (a) and (b) show the waveforms at the last receiver input with conventional fly-by topology with 80-ohm VDD-termination at the last two receivers and the proposed topology, with 50-ohm respectively, both having 10 loadings. At 2.4 Gbps, the proposed topology shows smaller eye margin compared to the conventional even with the termination matched to the trace impedance unlike at the result at 3.6 Gbps. At low data rates, the high frequency signal loss due to the loading is less significant. Therefore, the conventional fly-by topology allows large DC margin and shorter stubs is preferred even though the effective loading is large. The conventional fly-by having 10 loadings and 50-ohm VDD-termination at the last 2 receivers showed the worst eye margin. This is because of small DC swing and large reflection noise as explained in the previous section.

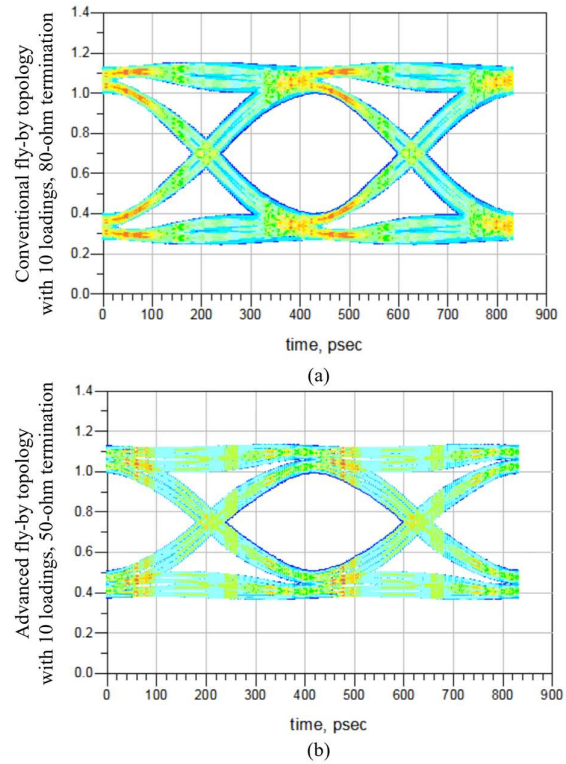


Fig. 4. 2.4 Gbps eye diagram at the last receiver input of (a) the conventional fly-by and (b) the advanced fly-by with 10 loadings and 50-ohm termination at the last two receivers.

III. CONCLUSION

In this paper, we addressed the limitation of conventional fly-by routing topology and proposed advanced fly-by that can double the number of receivers as compared to fly-by with same data rate for Gbps speed DDR5 interface or improve eye with same higher loading conditions. We discussed signal integrity of the proposed topology by comparing time domain simulation results with clock and data patterns. The proposed topology can improve signal integrity at high data rates by reducing the effective loadings but only under the condition of proper termination that gives good impedance matching to transmission lines. In addition, at low data rates, the conventional fly-by is preferred because the high frequency signal loss due to the loading is less significant.

REFERENCES

- [1] W. Feng, L. Kai and G. Ze, "Investigation of DDR T-Topology Port Resistance," *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, 2020, pp. 98-108, doi: 10.1109/ITAIC49862.2020.9338836.
- [2] N. Na, J. Wang, S. Long, C. Su, T. To and Y. Wang, "Exploring DDR4 Address Bus Design for High Speed Memory Interface," *2017 IEEE 67th Electronic Components and Technology Conference (ECTC)*, 2017, pp. 1843-1848, doi: 10.1109/ECTC.2017.247.
- [3] C. -C. Chiu, K. -Y. Yang, Y. -H. Lin, W. -S. Wang, T. -Y. Wu and R. -B. Wu, "A Novel Dual-Sided Fly-By Topology for 1-8 DDR With Optimized Signal Integrity by EBG Design," in *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 8, no. 10, pp. 1823-1829, Oct. 2018, doi: 10.1109/TCPMT.2018.2810239.

Differential Via Optimization for PCIe Gen5 Channel based on Particle Swarm Optimization Algorithm

Chulhee Cho, Kwangho Kim, Manho Lee, Jaeyoung Shin, Sungjin Yoon, Youngjae Lee, Chayoung Song, Wooshin Choi, Myoungbo Kwak, Youngdon Choi, Jung-Hwan Choi, and Hyungjong Ko
Samsung Electronics Co. Ltd, Hwaseong, South Korea
 E-mail: chulhee1.cho@samsung.com

Abstract— In high-speed SerDes channels, it becomes more important to reduce impedance mismatches to minimize signal return. Most of the mismatches are due to the differential via on PCB which is essential component to make up the PCIe Gen 5 system, and this mismatch should be reduced for the high-speed signal quality. To effectively minimize the mismatch, this paper presents an equation based TDR estimation model of the differential via, and the model is verified to commercial model of the coupled transmission line. And this paper also proposes a method for optimizing the design parameters of the differential via by applying a reward based on TDR impedance to PSO algorithm. The optimization procedure is then applied to one of the actual PCB designs to verify the optimized design parameters.

Keywords—PCIe, Differential via, Signal integrity, Design parameter optimization, Time-domain reflectometry (TDR), Printed circuit board (PCB), Particle Swarm Optimization (PSO)

I. INTRODUCTION

For SerDes (Serializer/Deserializer), its data rate has been going up with their generation, and especially PCIe Gen5 have to reach their speed up to NRZ (Non-Return Zero) 32 Gbps. To achieve this speed, the interconnection channel is required to guarantee their bandwidth at least fundamental frequency, 16 GHz, and even more up to 3rd, 5th harmonics of the fundamental. In this situation, when it comes to the return loss, there is no other equalization techniques, so technical approach to reduce the signal return has become more important. And the most of the return loss is caused by the impedance discontinuity at the differential via on PCB which is essential component to make up the PCIe Gen5 system. In these days, discontinuity at the via is considered one of the main factors for the low signal quality in high-speed, and efforts to reduce this discontinuity are being continued [1-2].

In an effort to reduce discontinuity at differential via, some optimization methods were studied in previous studies. In [3], the optimization was performed through analysis of design parameters of via such as stub resonance and anti-pad size effect, and in [4-6], statistical techniques, Taguchi's method and analysis of variation were used respectively. These approaches focused on reducing the simulation cases, since they need 3D electromagnetic (EM) simulation based on 3D modeling (e.g. HFSS). And it takes long time to get S -parameter results for the complicate structure such as the case of differential via which has many parameters to be designed.

This paper presents an equation based estimation of TDR (Time Domain Reflectometry) characteristics for the differential via to overcome this time-consuming problem. In addition, since differential via has many parameters to be tuned, making it difficult to calculate all results, the PSO (Particle Swarm Optimization) algorithm is selected and a reward model based on TDR impedance is suggested. And then, the suggested procedure is applied to one of the actual PCB designs to verify the effect of the optimized design parameters by comparing its TDR, S -parameters and BER (Bit Error Rate) contour at 1E-12.

II. MODELING AND OPTIMIZATION OF DIFFERENTIAL VIA

A. Equation based TDR Modeling for the Differential Via

A single via structure has similar to the coaxial structure, and the differential via can be considered as twin-rod. Through the capacitance and inductance characteristics of twin-rod, the impedance of the differential via Z_{via} and the effective dielectric constant DK_{eff} can be derived as defined in [7]. But, like in [7], because it is not in common that the horizontal length t and the vertical length w of via have different value each other, t and w can be considered as the same value equals to $2r$. In this condition, the equation about Z_{via} and DK_{eff} can be simplified as follows:

$$Z_{via} \cong \frac{60}{\sqrt{DK_{eff}}} \times \ln \left(\frac{s}{2r} + \sqrt{\left(\frac{s}{2r}\right)^2 - 1} \right) \times \ln \left(\frac{W+b}{4r} \right) \quad (1)$$

$$DK_{eff} = DK_{avg} \times \frac{\ln \left(\frac{s}{2r} + \sqrt{\left(\frac{s}{2r}\right)^2 - 1} \right)}{\ln \left(\frac{W+b}{4r} \right)} \quad (2)$$

where s is via-to-via pitch, r is radius of via barrel equals to the drill diameter and DK_{avg} is $(DK_{xy} + DK_z)/2$ [7] as shown in Fig. 1(a). If the via pitch s is smaller than the anti-pad size W , there exists overlapped anti-pad area between the via as shown in Fig. 1(b). To consider this case, the horizontal length of anti-pad, b , was divided into the following two cases:

$$b = \begin{cases} W, & S > W \\ \frac{s}{2} + \frac{W}{2}, & S \leq W \end{cases} \quad (3)$$

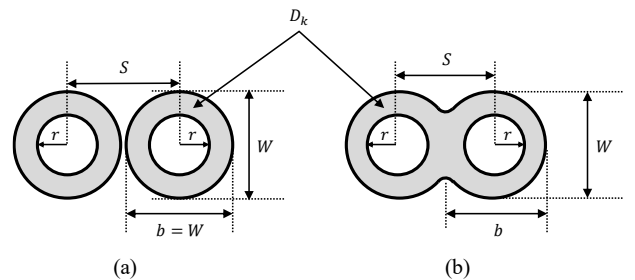


Fig. 1. Dimension of differential via on top view : (a) completely separated anti-pad case, and (b) overlapped anti-pad case.

Also, in a symmetric transmission line case, the mixed-mode S -parameters can be found in [8]:

$$\begin{bmatrix} S_{d1d1} & S_{d1d2} & S_{d1c1} & S_{d1c2} \\ S_{d2d1} & S_{d2d2} & S_{d2c1} & S_{d2c2} \\ S_{c1d1} & S_{c1d2} & S_{c1c1} & S_{c1c2} \\ S_{c2d1} & S_{c2d2} & S_{c2c1} & S_{c2c2} \end{bmatrix} = \begin{bmatrix} P_o & Q_o & 0 & 0 \\ Q_o & P_o & 0 & 0 \\ 0 & 0 & P_e & Q_e \\ 0 & 0 & Q_e & P_e \end{bmatrix} \quad (4)$$

$$P_{e(o)} = \frac{(Z_{0e(o)}^2 - Z_0^2) \sinh \gamma_{e(o)} l}{2Z_{0e(o)} Z_0 \cosh \gamma_{e(o)} l + (Z_{0e(o)}^2 + Z_0^2) \sinh \gamma_{e(o)} l} \quad (4a)$$

$$Q_{e(o)} = \frac{2Z_{0e(o)} Z_0}{2Z_{0e(o)} Z_0 \cosh \gamma_{e(o)} l + (Z_{0e(o)}^2 + Z_0^2) \sinh \gamma_{e(o)} l} \quad (4b)$$

The propagation constant γ is a complex number and it has attenuation coefficient (α) and phase delay coefficient (β). The former is associated with signal loss, which is set to almost zero in this paper, and the latter is associated with angular velocity which is the same as the speed of light over square root of the effective dielectric constant.

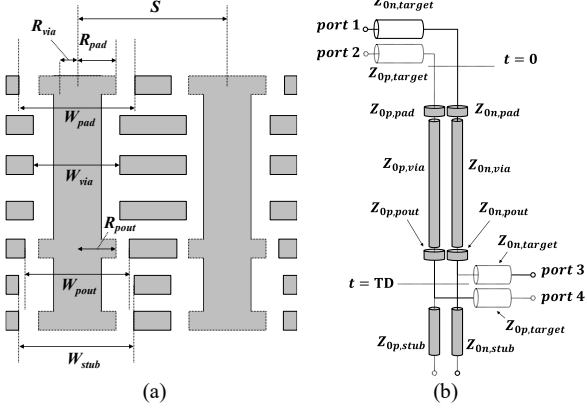


Fig. 2. (a) Side view of differential via implemented in the multi-layer PCB. (b) Four-port network connection for TDR estimation of the differential via.

In multi-layer PCB, the differential via has usually 8 design parameters such as center to center via pitch s , break-in pad radius R_{pad} , break-in pad anti-pad size W_{pad} , via radius R_{via} , via anti-pad size W_{via} , break-out pad radius R_{pout} , break-out pad anti-pad size W_{pout} , and stub anti-pad size W_{stub} as depicted in Fig. 2(a). If the signal is assumed to propagate from top to fifth layer, the lower side of conductor could be modeled as stub as shown lower side of Fig. 2(b). To get a total characteristics of the differential via, each mixed mode S -parameters was converted to standard S -parameters by using M matrix in [8], and connected properly using cascading formula in [9]. At the both end side of the via, the ideal transmission lines were connected to include discontinuity at $t = 0$ and $t = TD$, and, in this paper, the delay and single-ended target characteristic impedance were selected as 0.5 ns, and 42.5Ω respectively. By converting the connected standard S -parameters to the mixed mode again, total via network for the differential mode was simplified into two ports. Then the TDR impedance can be easily derived from Inverse Fourier Transform of S_{d1d1} , denoted as v_r , and input source v_s in time-domain as follow:

$$Z_{TDR} = Z_{0,d} \times \frac{v_r}{v_s - v_r} \quad (5)$$

where $Z_{0,d}$ is the differential mode reference impedance and 100Ω is used in this paper.

B. Validation of the Coupled Transmission Line Model

The accuracy of the Z_{via} and DK_{eff} was well proven in [7] based on ADS simulations. In this paper, because we did not select ADS tool but the equation based model, only the accuracy of the S -parameter for the coupled transmission line was compared to that of ADS CLINP. The input parameters of CLINP are as follows; odd mode impedance is 25Ω , length is 100 mm, odd mode dielectric coefficient is 6.15 and odd mode attenuation coefficient is 0.0001 dB/mm. The differential mode insertion loss and return loss from the equation based model and ADS CLINP show good correlation for the same parameters until 20 GHz as shown in Fig. 3.

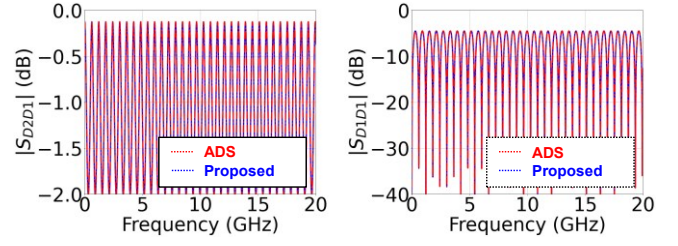


Fig. 3. Comparison in the mixed mode insertion loss S_{D2D1} (left), and return loss S_{D1D1} (right) between the proposed equation based model with the coupled transmission line model of ADS.

C. Hyper Parameters and Reward Modeling for PSO

PSO algorithm is a widely-used optimization algorithm and it has iteration process [7]. For a given value of iteration, each time the local best $pbest$ and global best $gbest$ are calculated and updated through (6) and (7), and the final $gbest$ value gives the optimized values. In the process, hyper parameters such as pop_size (particle size), w_{min} , w_{max} (inertia factors), c_1 (self confidence) and c_2 (swarm confidence) are needed. In this paper, the values were considered as 50, 1.0, 1.5, 0.2 and 0.6 respectively. Experimentally, most of $pbest$ and $gbest$ were converged within 20 iterations.

$$v_i^{k+1} = wv_i^k + \frac{c_1 R(pbest_i - s_i^k)}{\Delta t} + \frac{c_2 R(gbest_i - s_i^k)}{\Delta t} \quad (6)$$

$$s_i^{k+1} = s_i^k + v_i^{k+1} \Delta t \quad (7)$$

where

$$w = \frac{total\ epoch - current\ epoch}{total\ epoch} \times (w_{max} - w_{min}) + w_{min};$$

R is a random value, v and s indicate the velocity and the location of each particle respectively.

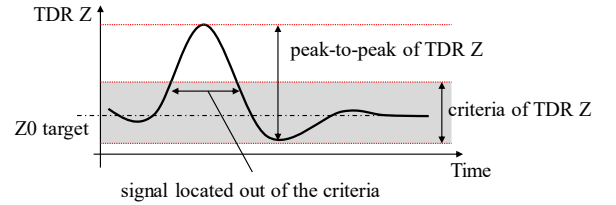


Fig. 4. Reward parameters in TDR impedance waveform.

Fig. 4 shows the parameters for the reward in the TDR impedance of a differential via. The reward consists of positive reward and negative reward denoted as $R_{positive}$ and $R_{negative}$ respectively. The positive reward becomes larger than 1 if the TDR peak-to-peak smaller than impedance criteria (gray colored area). To distinguish a case that absolute value of TDR impedance is out of the criteria, the negative reward was defined and given. Total reward is represented as follow:

$$Reward = R_{positive} - R_{negative} \quad (8)$$

where

$$R_{positive} = \frac{criteria\ of\ TDR\ Z}{peak-to-peak\ of\ TDR\ Z};$$

$$R_{negative} = \frac{length\ of\ signal\ located\ out\ of\ the\ criteria}{length\ of\ total\ signal}.$$

III. SIMULATION RESULTS

The differential via optimization was applied to one of PCIe Gen5 PCB design in which channel has two transition points by via. Table I shows normalized values of each design parameters of differential via comparing typical design parameters with optimized design parameters. In Fig. 5(a), as calculation of TDR using proposed method in the previous section, the peak of impedance of before and after optimization are 107 Ω and 85 Ω , and it could be confirm that the level of those results were quite similar to that of via in whole channel simulation from HFSS, as shown Fig. 5(b).

TABLE I. NORMALIZED DESIGN PARAMETERS

Design Parameters	Description	Original values	Optimized values
s	Center to center via pitch	1.00	0.50
R_{pad}	Radius of break-in Pad	1.00	0.64
W_{pad}	Anti-pad size of break-in pad	1.00	0.61
R_{via}	Radius of via	1.00	1.13
W_{via}	Anti-pad size of via	1.00	1.50
R_{pout}	Radius of break-out pad	1.00	0.50
W_{pout}	Anti-pad size of break-out pad	1.00	1.34
W_{stub}	Anti-pad size of via stub	1.00	1.50

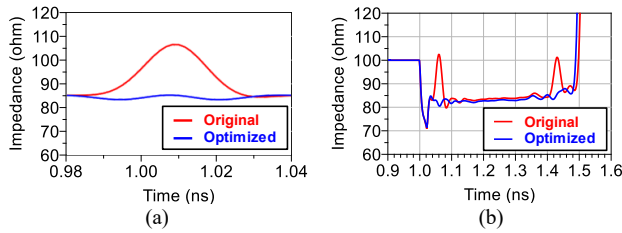


Fig. 5. TDR impedance comparison between the optimized via design PCB with the original design : (a) differential mode for single-ended four-port network of differential via described in Fig. 2(b), and (b) design parameter applied and analyzed for whole channel with HFSS.

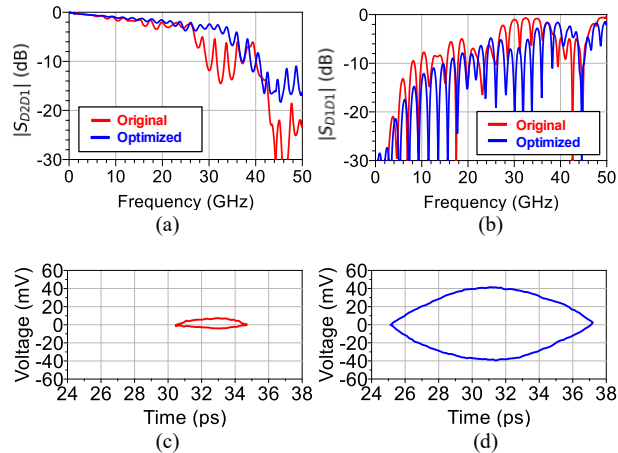


Fig. 6. Comparison with an original design (red) to a via optimized design (blue) PCB in S -parameter (a) insertion loss and (b) return loss. And BER contour at 1E-12 of 32 Gbps speed simulation in (c) typical design, and (d) via optimized design.

Fig. 6(a) and (b) show the insertion loss and the return loss of the typical and optimized differential via system, and in the optimized case, those two parameters were improved from 2.66 dB to 1.43 dB, and from 8.4 dB to 10.5 dB at 16 GHz respectively. Also, as depicted in Fig. 6(c) and (d), the eye height and width of BER 1E-12 contour for 32 Gbps speed without equalizer were improved from 11 mV to 81 mV and from 4.38 ps to 12.19 ps, respectively.

IV. CONCLUSION

In high speed SerDes channels, it has become more important to reduce signal return, and one of the main cause of the signal return is the impedance mismatch of the differential via on PCB. To minimize this mismatch, the design parameters should be optimized, but the previous studies used 3D EM modeling have had disadvantages to take time to optimize complicate structure of via. In this paper, the equation based TDR estimation model of the differential via was suggested and the PSO algorithm based on the rewards defined from the TDR impedance was selected to the optimization method. Then the suggested method was applied to the differential via on PCIe Gen5 PCB and 8 design parameters were optimized. As a result, the optimized results showed good improvement in signal quality. In the future, the measurement results will be confirmed by PCB fabrication. Since it is important to consider the effects caused by not only signal via but ground via, further research considering the ground via effects is also planned to be conducted later.

REFERENCES

- [1] J. Zhang, J. Lim, W. Yao, K. Qiu and R. Brooks, "PCB via to trace return loss optimization for >25Gbps serial links," *2014 IEEE International Symposium on Electromagnetic Compatibility (EMC)*, 2014, pp. 619-624.
- [2] L. W. Chew, C. Y. Tan, M. D. Chai and Y. R. Lim, "PCB Channel Optimization Techniques for High-Speed Differential Interconnects," *2022 International Conference on Electronics Packaging (ICEP)*, 2022, pp. 189-190.
- [3] B.-G. Kang, Hyun Kim, Hee-do Kang and J.-G. Yook, "Optimization of via structure in multilayer PCB for high speed signal transmission," *2008 Electrical Design of Advanced Packaging and Systems Symposium*, 2008, pp. 105-108.
- [4] A. Zenteno, D. Reina and G. Regalado, "Optimization of PCB via design considering its physical length parameters," *2010 53rd IEEE International Midwest Symposium on Circuits and Systems*, 2010, pp. 256-259.
- [5] S. Zhou and G. Lu, "Robust optimization of PCB differential-via for signal integrity," *2013 Proceedings of the International Symposium on Antennas & Propagation*, 2013, pp. 1336-1339.
- [6] Yalin Guan and Shilei Zhou, "Electromagnetic optimization of PCB differential-via for signal integrity using analysis of variance," *2013 IEEE INTERNATIONAL CONFERENCE ON MICROWAVE TECHNOLOGY & COMPUTATIONAL ELECTROMAGNETICS*, 2013, pp. 22-25.
- [7] L. Simonovich, E. Bogatin and Y. Cao, "Differential Via Modeling Methodology," in *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 1, no. 5, pp. 722-730, May 2011.
- [8] Jeonhyeon Cho, Eakhwan Song, Heegon Kim, Seungyoung Ahn, Jun So Pak, Jiseong Kim, and Joungho Kim, "Mixed-Mode ABCD Parameters: Theory and Application to Signal Integrity Analysis of PCB-Level Differential Interconnects," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 53, No. 3, Aug 2011.
- [9] G. R. Simpson, "A Generalized n-Port Cascade Connection," *1981 IEEE MTT-S International Microwave Symposium Digest*, 1981, pp. 507-509.
- [10] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, 1995, pp. 39-43.

Hardware Verification of Via Crosstalk Cancellation for Differential BGA-to-BGA Links

Katharina Scharff¹, Xiaomin Duan¹, Matteo Cocchini², Hung Nguyen², Nicole Selezinski³, Dierk Kaller¹, Hubert Harrer¹

¹ IBM Deutschland Research & Development GmbH, Germany

² IBM Systems, IBM Corporation, USA

³ Hamburg University of Technology, Germany

Abstract—Coupling in dense via pinfields is known as a major contributor to crosstalk. In this work we present and verify a novel wiring strategy for links with a symmetrical pinfield-to-pinfield topology that reduces crosstalk significantly and requires no extra shielding vias. By twisting the orientation of the p and n wires of the differential stripline pair, a cancellation of the crosstalk between the differential via pairs is achieved. It can be implemented without any additional space requirements. The proposed design is fabricated and the crosstalk is measured with a vector network analyzer. It is shown that small changes in the orientation of the wiring in the pinfield region can reduce far-end crosstalk significantly.

Index Terms—Via, Crosstalk, Cancellation

I. INTRODUCTION

The data rates of high-speed signals are increasing further, while, simultaneously, the density of the wired traces on a printed circuit board (PCB) becomes larger as well. This requires a lot of effort from designers to ensure adequate signal integrity margins and reliable connections on PCBs. Crosstalk is one of the main disturbances that impact signaling on a PCB. It causes significant signal deteriorations and is one of the major noise sources. Especially channels using PAM4 are susceptible to noise [1].

Vias and via pinfields are major contributors to crosstalk. To prevent high crosstalk, either the distance between aggressor and victim needs to be large or a sufficient number of shielding vias need to be added. Both solutions require additional space on the PCB, hence reducing the IO density in the pinfield. If this is not possible due to spacing constraints, changes to the layout of the pinfield can make a large impact on the crosstalk behavior. It has been shown in [2]–[5] that it is possible to reduce or cancel out part of the crosstalk.

We propose a novel design of the wiring in the pinfield area where far-end crosstalk (FEXT) can be reduced or canceled out completely. In principle, differential via crosstalk can happen in two polarities as illustrated in Fig. 1. Links that consist of a symmetrical ball grid array (BGA-to-BGA) topology, where the via coupling occurs twice along the line (as seen in Fig. 2), can take advantage of this effect. It is based on the observation that the orientation of the p and n wires of a differential pair can change the sign of the coupled crosstalk pulse. Depending on the coupling polarity and wiring distance between the two BGAs, the FEXT pulses from both via fields can either add up or cancel out each other. By carefully designing the

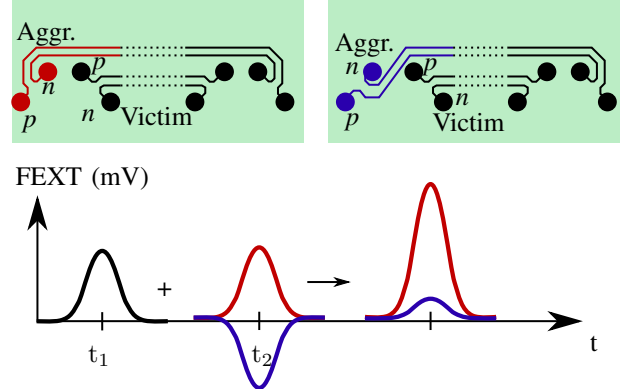


Fig. 1. Two examples of via coupling with different polarities of the crosstalk pulse. By shifting the order of the p and n wires, the FEXT pulses from both via pairs can either add up or cancel out each other.

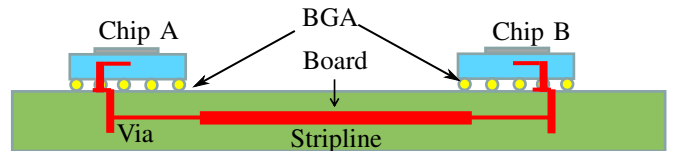


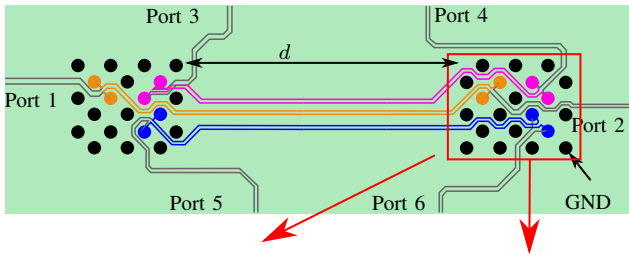
Fig. 2. Example of a BGA-to-BGA link. Two chips are connected with via BGAs and a PCB.

fanout and wiring length, it is possible to minimize the FEXT contributions from vias with no additional space requirement. Two test vehicles with two pinfield layouts were fabricated, one without any cancellation effect and one with the crosstalk cancellation effect. The S-parameters of both structures were measured with a vector network analyzer (VNA). We show that the cancellation structure leads to a significant reduction of crosstalk.

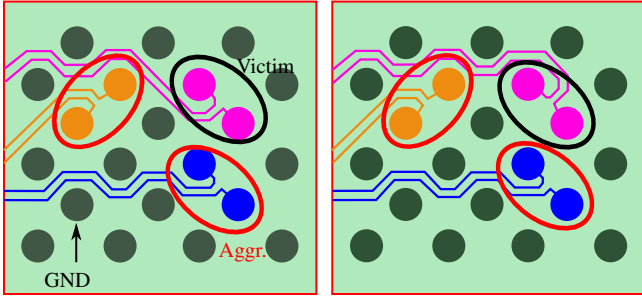
II. FAR-END CROSSTALK CANCELLATION EFFECT

The effect of crosstalk cancellation is well-known and widely used in e.g. twisted wires and conductors [6]. Efforts exist to apply it on the PCB wires, for example, in [4]. Here, we present its application on differential vias.

Unlike single-ended vias, the polarity of differential via coupling can be manipulated by swapping the position of the p and n vias and their associated wiring. This is usually



(a)



(b)

(c)

Fig. 3. PCB test vehicle with two different implementations of the right pinfield. The grey striplines are positioned in the top signal layer and used to connect to the coaxial ports. The via pitch is 1.5 mm and the distance d between the pinfields is 30.46 mm. (a) Complete model. (b) Test vehicle without cancellation effect (TV1). (c) Test vehicle with cancellation effect (TV2). Only the striplines in the lowest layer are shown in (b) and (c).

disregarded in a link simulation since the crosstalk phase is mostly considered random. However, certain link topology, e.g. a BGA-to-BGA communication link shown in Fig. 2, consists of two via coupling locations. In this type of setup, it is possible to manipulate the polarities of the two coupling instances to enable their cancellation.

As shown in Fig. 1, there are two possible fanout wiring orientations associated with the differential via that results in opposite coupling polarity. The cancellation effect is a result of superposition of the two via crosstalk instances with controlled timing. This can be achieved by twisting the orientation of the p and n wires of a differential via pair with attached striplines. The coupling between two differential pairs is continuous along the entire length of the pairs and the total FEXT at victim is the net results of all couplings along the line. By twisting parts of the victim pair, i.e. escaping the via in opposite direction, the polarity of the crosstalk pulse is reversed. Ideally, if it happens at the same time, the crosstalk cancels exactly and will not appear at the victim. This requires the distance between the two via couplings of the victim and aggressor lines to be identical. Furthermore, the attenuation along the line will impact the cancellation.

An illustration of the effect is shown in Fig. 1. In both structures the differential via pairs have the same orientation. The striplines however are launched in opposite directions. At t_1 the coupling occurs between the two differential via pairs. Since their orientation is the same the direction of the pulse is identical for both structures. Assuming aggressor and victim stripline are of equal length, the induced crosstalk pulse on

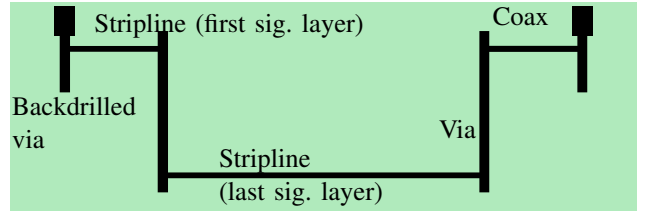


Fig. 4. Position of striplines in cross-section. The via diameter is 12mil. GND planes and the other signal layers are not shown.

the aggressor line and the actual signal pulse on the victim line arrive at the second via pairs at the same time t_2 . At t_2 the coupling of the second via pairs is added to the FEXT that propagated from the first via pair. The direction of this second coupling depends on the orientation of the p and n wires. If the pulses have opposite directions it will lead to the cancellation of the FEXT at the receiver.

III. LAYOUT

To investigate the previously discussed cancellation effect, two test structures were developed and fabricated on a PCB. The fabricated structure uses a cross-section with 6 signal layers. The cross-section is made of standard FR4 layers and ultra low loss layers, where the signal traces are placed. One complete structure is shown in Fig. 3. Each structure is composed of two pinfields that are connected with differential striplines. They are located in the lowest signal layer (see Fig. 4). Each differential signal via inside the pinfield is connected with another differential line segment on the top signal layer. They lead to single-ended ports for measurements (not shown in Fig. 3). Each structure includes one victim and two far-end aggressors. The left pinfield of the model is identical for both structures. The layout of the right pinfield is different, in order to investigate the cancellation effect. These two variations are shown in Fig. 3. In TV2 (see Fig. 3 (c)) the order of p and n for port 4 is swapped, which should lead to a smaller FEXT in structure 2 compared to TV1. The FEXT is measured for victim port 4 and aggressors 1 and 5.

IV. MEASUREMENT RESULTS

The two structures that were introduced previously were fabricated on a PCB. The FEXT S-parameters were measured up to 40 GHz with a VNA. TV1 uses the routing shown in Fig. 3 (b) without cancellation and TV2 uses the pinfield with cancellation (Fig. 3 (c)). Since only 4-port measurement equipment was available, the unused ports were terminated with 50Ω resistors to avoid reflections from unterminated ports. A picture of the fabricated structure with attached resistors is shown in Fig. 5.

Fig. 6 shows the differential FEXT for TV1 and TV2 with two different aggressors, port 1 and 5. The victim in both cases is port 4. For aggressor 1 in Fig. 6 (a), the crosstalk is indeed considerably smaller for TV2. For aggressor 5, the crosstalk in TV2 is smaller up to 25 GHz, at higher frequencies however, it is higher than for TV1. This could be due to coupling in

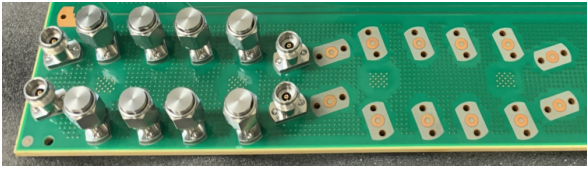


Fig. 5. Fabricated PCB structure with test vehicle 1(TV1) on the left and test vehicle 2 (TV2) on the right. 4 connectors are attached. The other ports are terminated with 50Ω resistors.

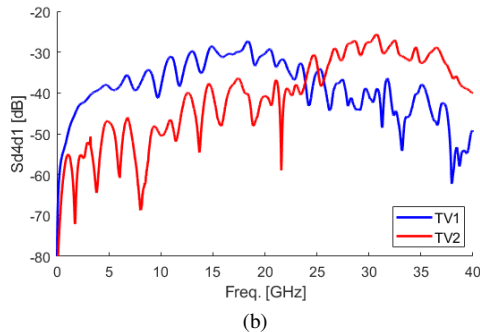
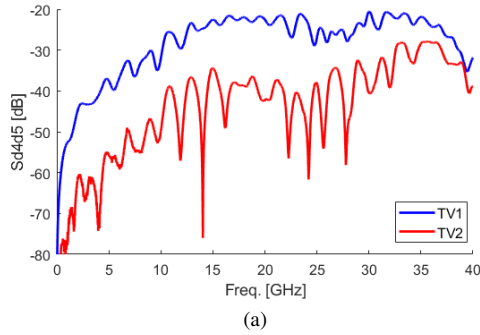
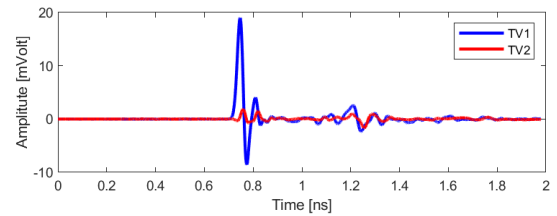


Fig. 6. S-parameters of aggressors 1 and 5 for both test vehicles. (a) S_{d4d5} . (b) S_{d4d1} .

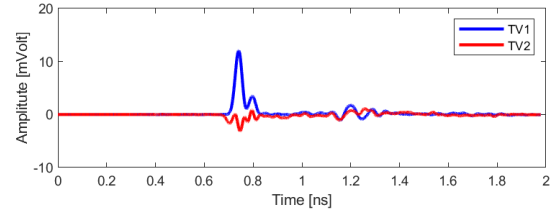
the stripline area as well as second order effects. This is only relevant if the Nyquist frequency of the data rate is in this frequency region or higher.

The crosstalk can also be observed in time domain similar to a TDR (time domain reflectometry) analysis. Fig. 7 shows the step responses of the FEXT for an excitation with a 15 ps rise time. A waveform with this rise time would have a bitrate of 13.33 Gbps if the bitrate is 20% of the unit interval. As discussed in Sec. II, under ideal timing conditions the contributions of the coupling inside the via arrays cancel out and reduce the crosstalk to almost zero. TV1 shows a significantly higher FEXT pulse for both aggressors, whereas the crosstalk for TV2 is very small. The peak of S_{d4d5} is at 18.99 mV for TV1 and 1.86 mV for TV2, respectively. For S_{d4d1} the peaks are at 12.04 mV for TV1 and 1.12 mV for TV2, respectively.

It can also be observed that the crosstalk of TV2 is not driven to zero. The remaining crosstalk is mainly due to stripline coupling in the pin area, that is significantly lower in its contribution to the total FEXT compared to via coupling.



(a)



(b)

Fig. 7. Crosstalk pulses in time domain of S_{d4d5} and S_{d4d1} for both test vehicles. The exciting step has a rise time of 15 ps. (a) Step response of aggressor 5. (b) Step response of aggressor 1.

V. CONCLUSION

We introduce a cancellation effect for differential FEXT applicable to via crosstalk in a BGA-to-BGA configuration and show that with no additional space requirement the cancellation effect can be implemented into a PCB design. The effectiveness of the cancellation scheme is validated with measurements. In the future, the impact of the routing length will be investigated.

REFERENCES

- [1] D. D. Sharma, "Pci express® 6.0 specification at 64.0 gt/s with pam-4 signaling: a low latency, high bandwidth, high reliability and cost-effective interconnect," in *IEEE Symposium on High-Performance Interconnects (HOTI)*. IEEE, Aug. 2020.
- [2] Q.-M. Cai, L. Zhu, X.-B. Yu, L. Zhang, C. Zhang, Y.-Y. Zhu, and X. Cao, "Far-end crosstalk mitigation using homogeneous dielectric substrate in DDR5," in *12th International Workshop on the Electromagnetic Compatibility of Integrated Circuits (EMC Compo)*. IEEE, Oct. 2019.
- [3] B. Chen, S. Pan, J. Wang, S. Yong, M. Ouyang, and J. Fan, "Differential crosstalk mitigation in the pin field area of SerDes channel with trace routing guidance," *IEEE Transactions on Electromagnetic Compatibility*, vol. 61, no. 4, pp. 1385–1394, Aug. 2019.
- [4] D. G. Kam, H. Lee, J. Kim, and J. Kim, "A new twisted differential line structure on high-speed printed circuit boards to enhance immunity to crosstalk and external noise," *IEEE Microwave and Wireless Components Letters*, vol. 13, no. 9, pp. 411–413, Sep. 2003.
- [5] J. Tang, X. Yang, J. A. Hejase, M. Bohra, Y. Zhang, X. Duan, D. Kaller, W. D. Becker, and D. M. Dreps, "Far end crosstalk mitigation of differential high speed interconnects within printed circuit board via fields," in *IEEE 30th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*. IEEE, Oct. 2021.
- [6] R. Voelker, "Transposing conductors in signal buses to reduce nearest-neighbor crosstalk," *IEEE Transactions on Microwave Theory and Techniques*, vol. 43, no. 5, pp. 1095–1099, May 1995.

Inverse Design of Embedded Inductor with Hierarchical Invertible Neural Transport Net

Oluwaseyi Akinwande*, Osama Waqar Bhatti*, Madhavan Swaminathan*[†]

*[†] School of Electrical and Computer Engineering [†] School of Material Science and Engineering

*[†]3D Systems Packaging Research Center, Georgia Institute of Technology, Atlanta, USA

Abstract—Heterogeneous integration of voltage regulators in power delivery networks is a growing trend that employs embedded inductor as a key component in significantly improving the power distribution. In this work, we propose a neural network framework called the hierarchical invertible neural transport for the inverse design of an embedded inductor. With this invertible method, we obtain the probability distributions of the parameters of the embedded inductor design space that most likely satisfy the desired specifications. We also learn the impedance response for free in the forward design. In the forward design, our results show a 2.14% normalized mean square error when compared with the output response of a fullwave EM simulator.

Index Terms—power delivery, inverse design, emdedded inductor, neural networks, transport maps

I. INTRODUCTION

Early-stage prototyping in electronic design automation (EDA) is often an herculean task due to a large amount of variables that are explored in the design space. The design cycle is often plagued with iterating through series of designs in an attempt to find the optimal parameters that satisfy the target specifications. These sorts of evaluations are usually compute- and time-intensive. Optimization methods and surrogate modeling have been proposed to tackle this task effectively [1]. However, the best solution may still not be obtained, or several possibilities may be ignored.

In recent times, machine learning (ML) methods have been employed to model the forward and inverse mappings for a set of inputs and outputs. Consider a design space X of a parameterized system, as illustrated in Fig. 1, that forms the input of the ML model, with corresponding output response Y . This mapping relationship can be represented as:

$$\mathcal{F} : X \rightarrow Y, \quad (1)$$

where \mathcal{F} is the forward mapping. The forward model learns the input-output relationship and predicts the output response given the input parameters. To estimate the best set of input parameters that satisfies the desired target, we find the inverse mapping:

$$\mathcal{F}^{-1} : Y \rightarrow X. \quad (2)$$

Inverse problems are often ill-posed and intractable because they fail the existence and uniqueness tests. Existence verifies if the inverse exists, and uniqueness determines the ambiguity brought by the one-to-many mapping in the inverse direction. The problem of invertibility is not novel, and several methods have been proposed to address it. In the domain of artificial neural networks, state-of-the-art generative models such as the

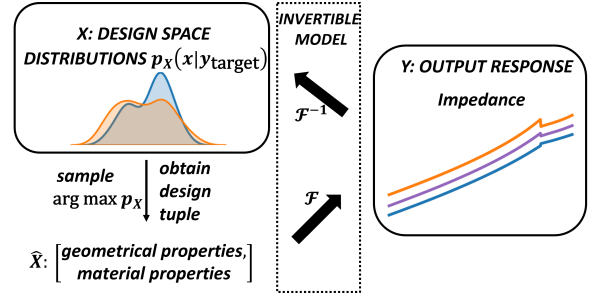


Fig. 1. Model-based invertible framework that offers a custom solution.

generative adversarial network (GAN) [2], variational auto-encoder (VAE) [3], and invertible neural network (INN) [4] address this issue by generating conditional posterior distributions rather than point-estimates. In particular, the INN has some merits which include efficient computation of forward and inverse mappings, and direct modeling of its likelihood function [4].

In this paper, we propose an inverse system modeling, design and identification using hierarchical invertible neural transport net for embedded inductor in integrated voltage regulator. With inverse design, the design parameters can be straightforwardly obtained from the output objectives. It reduces design cycle time and related costs by increasing the overall efficiency of the design process. We investigate both the inverse mapping to learn the probability distributions that satisfy a desired specification, and the forward mapping to verify the solution.

II. HIERARCHICAL INVERTIBLE NEURAL TRANSPORT NETWORK (HINT)

A. Theory of Transport Maps

The theory of transport maps is based on the concept of constructing a coupling, i.e., a transport map, between a complex target probability measure ν and a simpler source probability measure μ [5]. We seek a transport map $T(x)$ such that μ is supported on X and ν is supported on Y , i.e.,

$$T : X \rightarrow Y \quad (3)$$

We can generate samples of ν by pushing forward the μ through the map. In order to conserve mass, we require [5]:

$$\mu(T^{-1}(A)) = \nu(A) \quad \forall A \subset Y. \quad (4)$$

$\mu(T^{-1}(A))$ is called the push forward of μ through T , denoted $T_{\#}\mu$. Therefore, a compact form for (4) is given as [5]:

$$T_{\#}\mu = \nu. \quad (5)$$

Choosing such a transport map involves minimizing the cost c of moving a unit of mass from x to $T(x)$. This leads to the Monge formulation of optimal transport, given as [5]:

$$\min \left\{ \int_X c(x, T(x)) d\mu(x) \mid T_{\#}\mu = \nu \right\}. \quad (6)$$

The solution to this constrained optimization problem is the optimal transport map. To address this problem, we pose the following questions: (1) Does the minimizer T^* exist? (2) If it exists, is it unique? (3) Is it feasible?

B. The Recursive Coupling Block

HINT uses normalizing flow based on the change-of-variables law of probabilities to model complex distributions from a simple one. The normalizing flow pipeline T is a composition of recursive coupling blocks f_i that are invertible, given as [6]:

$$T = f_{C1} \circ f_{Q1} \circ \dots \circ f_{CN} \circ f_{QN}, \quad (7)$$

where each pair $f_{C_i} \circ f_{Q_i}$ in T is a composition of a triangular map and an orthogonal transformation, respectively, where the former is known as a Knothe-Rosenblatt rearrangement in transport maps [5]. Fig. 2 shows one such composition of a recursive affine coupling block and an orthogonal transformation. A parameterized flow model $f_{\theta_i} = f_{C_i} \circ f_{Q_i}$ splits the input vector x into $[x_1, x_2]$ and transforms them by an affine function with coefficients e^s and t with element-wise operations as [4]:

$$x'_1 = x_1, \quad x'_2 = x_2 \circ e^{s(x_1)} + t(x_1). \quad (8)$$

It is easy to see that (8) is trivially invertible. This way, the inverse flow composition is easily computed as

$$T^{-1} = f_{QN}^{-1} \circ f_{CN}^{-1} \circ \dots \circ f_{Q1}^{-1} \circ f_{C1}^{-1}. \quad (9)$$

T in (7) is used to transport the data distribution $p_X(x)$ to a standard normal latent distribution $p_Z = \mathcal{N}(0, I)$, while T^{-1} in (9) can then be used to draw a sample $z \sim p_Z$ in the latent space to obtain p_X .

To infer the inverse solution, we can turn any z sampled from the latent space into a corresponding x conditioned on y as [6]:

$$p_X(x|y) = p_Z(f(x, y)|y) \cdot |\nabla f(x)|. \quad (10)$$

This invertible model has the benefits of efficient computation of the conditional posterior probabilities because of the dense Jacobian determinant ∇f using the recursive design, and a simpler training objective where the parameters of the flow f_{θ} can be learned from (10) via a maximum likelihood loss [6]:

$$\mathcal{L} = \frac{1}{2} \|f(x, y|\theta)\|_2^2 - \log |\nabla f(x|\theta)|. \quad (11)$$

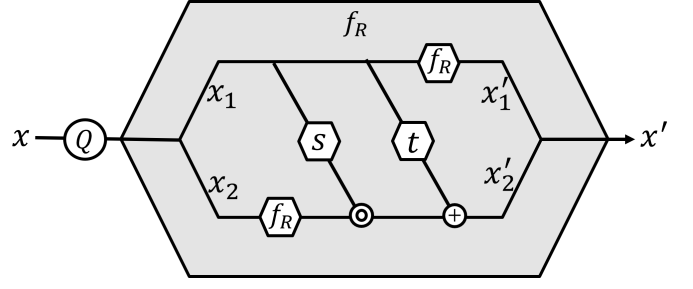


Fig. 2. A recursive affine coupling block. The inner functions f_R perform the same sequence of operations as the outer gray block, repeated until the maximum hierarchy depth is reached [6].

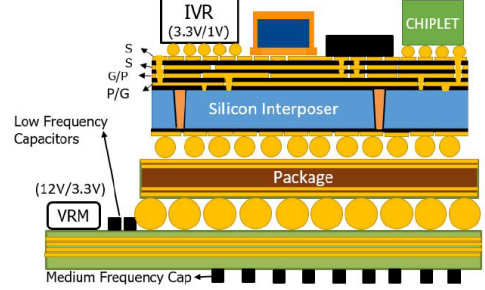


Fig. 3. Stack-up of the considered 2.5D integrated system [7].

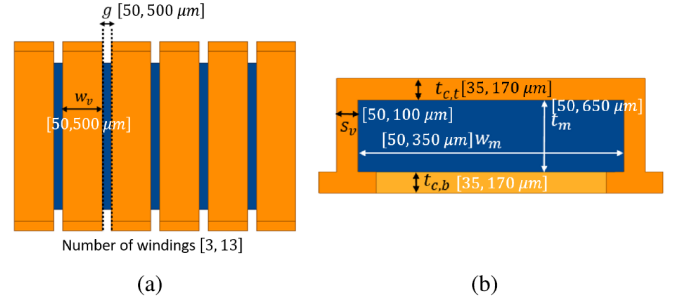


Fig. 4. Solenoid inductor [7]. (a) Top view. (b) Side view.

III. APPLICATION: INVERSE DESIGN OF EMBEDDED INDUCTOR

For a demonstration of the proposed method, we apply the HINT method for the inverse design of an embedded solenoidal inductor. The inductor is made up of a Nickel-Zinc (NiZn) ferrite magnetic core and it is integrated on the top metal layer of a silicon-interposer-based 2.5D heterogeneously integrated system as in Fig. 3 [7]. The output response of the solenoidal inductor is the impedance $Z = ESR + j\omega L$, where L is its inductance and ESR is its equivalent series resistance. The objective here is to (1) obtain the solenoidal inductor design parameters that correspond to a given specification of impedance response Z , and (2) validate the design parameters through a forward evaluation.

A. Model Setup

The design parameters of the solenoidal inductor and their range of values are shown in Fig. 4. The target characteristic

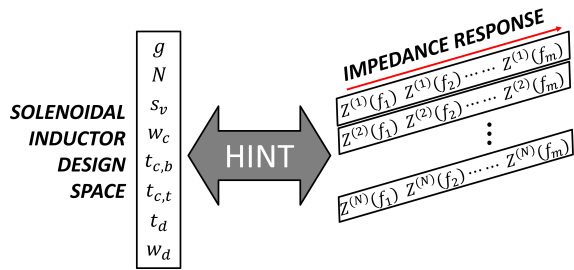


Fig. 5. Proposed Hierarchical Invertible Neural Transport (HINT) model setup for embedded inductor.

investigated is the impedance response Z with the frequency being swept from 15–98 MHz for 34 frequency points. Using Latin Hypercube Sampling, we determine 949 samples and extract their inductance and ESR with Ansys HFSS. We split the data into train and test sets.

The objective here is to determine an invertible mapping between the design space X and output response Y . The proposed model setup is shown in Fig. 5. The HINT model is constructed using 2 recursive coupling blocks with 3 recursion levels. Each recursive block contains the scale s and shift t networks which are constructed with fully connected neural networks with one hidden layer of 256 neurons, Rectified Linear Unit (ReLU) activation functions, and batch normalization layers. On the input side of the model setup, there are 8 inductor design parameters. The output variable is the impedance response Z . The HINT model is trained for 500 epochs.

B. Results

During the inference process, we choose a random response y_{target} from the test set and we obtain the inverse solution using (10). The HINT model generates rich conditional posterior distributions of the embedded inductor design parameters as shown in Fig. 6. Next, we obtain an inverse tuple from these distributions by sampling the points with the highest densities in the embedded inductor design space. The tuple obtained is $\{128.7 \mu\text{m}, 499 \mu\text{m}, 109.8 \mu\text{m}, 145.2 \mu\text{m}, 303 \mu\text{m}, 10.7 \text{ mil}, 8.2 \text{ mil}, 9\}$. We take this tuple and perform a forward evaluation with the HINT model to obtain the impedance response shown in Fig. 7. In the forward design, we achieve a 2.14% normalized mean square error, averaged over 10 inference runs.

IV. CONCLUSION

We present the HINT method for the inverse design of embedded inductor for integrated voltage regulator. We applied this method to obtain the probability distributions of the inverse solutions that satisfy target specifications.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. CNS 16-2137259 - Center for Advanced Electronics through Machine Learning (CAEML).

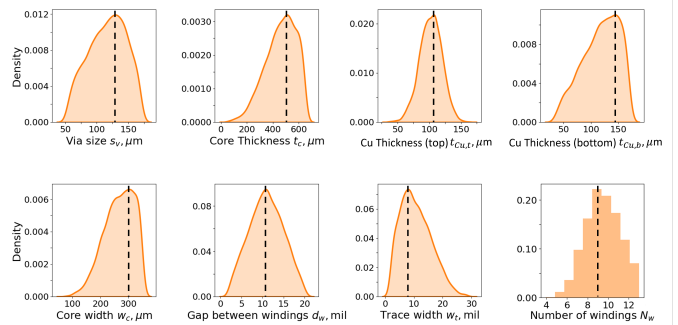


Fig. 6. Predicted conditional posterior distributions $p(x|y_{\text{target}})$ of embedded inductor design parameters. Black vertical dashed lines indicate the points with the highest densities. When the points corresponding to the highest densities are sampled for the design parameters, the tuple obtained is $\{128.7 \mu\text{m}, 499 \mu\text{m}, 109.8 \mu\text{m}, 145.2 \mu\text{m}, 303 \mu\text{m}, 10.7 \text{ mil}, 8.2 \text{ mil}, 9\}$.

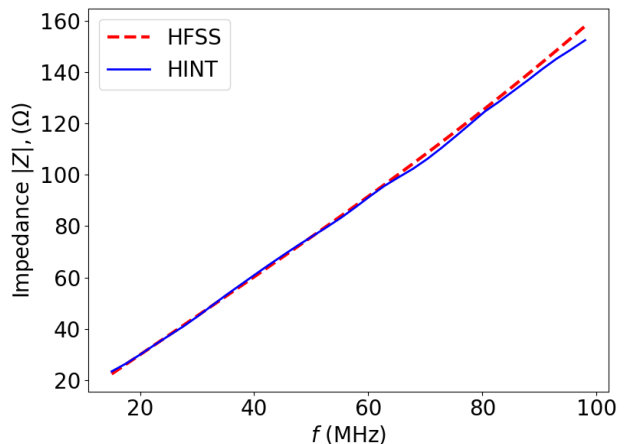


Fig. 7. Forward evaluation, showing impedance response for the embedded inductor, with HFSS and the trained HINT model for generated tuple $\hat{x} \sim p(x|y_{\text{target}})$.

REFERENCES

- [1] M. Swaminathan, H. M. Torun, H. Yu, J. A. Hejase and W. D. Becker, "Demystifying Machine Learning for Signal and Power Integrity Problems in Packaging," in IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 10, no. 8, pp. 1276-1295, Aug. 2020, doi: 10.1109/TCPMT.2020.3011910.
- [2] Ian Goodfellow et al., "Generative Adversarial Nets", in Advances in Neural Information Processing Systems 27 (NeurIPS) 2014.
- [3] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114, 2013
- [4] Lynton Ardizzone, Jakob Kruse, Carsten Rother, Ullrich Köthe, "Analyzing Inverse Problems with Invertible Neural Networks", International Conference on Learning Representations (ICLR) 2019.
- [5] Y. Marzouk, et al., "Sampling via measure transport: An introduction". In R. Ghanem, D. Higdon, and H. Owhadi, eds., "Handbook of Uncertainty Quantification", 1–41, Springer, 2016.
- [6] Kruse, Jakob et al., "HINT: Hierarchical Invertible Neural Transport for Density Estimation and Bayesian Inference," arXiv, 2021, doi:10.48550/ARXIV.1905.10687.
- [7] H. M. Torun et al., "A Spectral Convolutional Net for Co-Optimization of Integrated Voltage Regulators and Embedded Inductors," 2019 IEEE/ACM International Conference on Computer-Aided Design.

Towards Accelerated Transient Solvers for Full System Power Integrity Verification

Antonio Carlucci*, Stefano Grivet-Talocia*, Scott Mongrain[§], Sid Kulasekaran[§], Kaladhar Radhakrishnan[§]

*Dept. Electronics and Telecommunications, Politecnico di Torino, Italy

[§]Intel Corporation, Chandler, AZ, USA

antonio.carlucci@polito.it

Abstract—This paper proposes a novel framework for power integrity verification of multicore systems, including voltage stabilization provided by multiple integrated voltage regulators at the core interfaces. The proposed framework adopts a two-stage macromodeling strategy to derive a compact representation of the full system dynamics as observed from each core. These dynamics are parameterized by the time-varying duty cycle provided by dedicated feedback controllers to each voltage regulator, here implemented through an averaged model. We show that the proposed simulation framework has the potential to outperform direct transient analysis based on SPICE engines.

I. INTRODUCTION AND PROBLEM STATEMENT

This paper addresses system level power integrity verification for multi-core microprocessors with integrated voltage regulators providing per-core voltage domain granularity. Our objective is to simulate an entire power delivery network structured as a cascade of multiple stages between the main supply and the compute domains inside the microprocessor [1]. In particular, we address the transient solution of the complete power delivery, including voltage regulation effects provided by Fully Integrated Voltage Regulators (FIVR) [2] as well as the coupling from one core to another due to the shared input network. The large-scale nature of this simulation problem, both in terms of expected dynamic order and number of ports/signals to be evaluated, combined with the nonlinear FIVR circuitry and the associated feedback regulation loops, make this problem particularly challenging. A brute-force SPICE transient simulation based on suitable models for all system parts is feasible only for small-scale low-core platforms, but it does not have potential to scale to higher complexities, as demanded by state of the art HPC or AI many-core processors.

The reference structure is schematically depicted in Fig. 1. At the motherboard level, the output of an on-board Voltage Regulator Module (VRM) is routed through the PCB power planes and the power pins of the microprocessor package. The entire board and package power distribution structures are collectively denoted here as *input network*, which also includes a linear model of the VRM and the contribution of all board and package decoupling capacitors. The latter are usually optimized to meet specific target impedance requirements [3], [4], but once this optimization is performed they can be considered as integral parts of the input network, which in turn

978-1-6654-5075-1/22/\$31.00 ©2022 IEEE

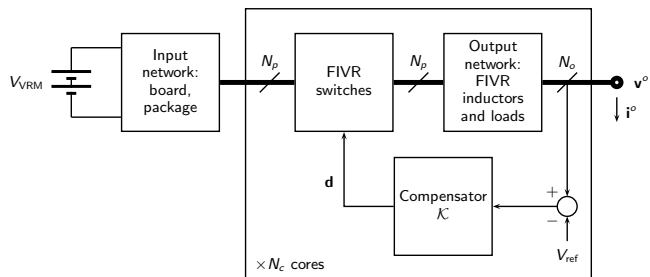


Fig. 1. Schematic illustration of the power distribution system under investigation, including N_c cores whose voltage is regulated by N_p -phase FIVRs.

can be collectively represented a large-scale distributed Linear and Time-Invariant (LTI) multiport described by a transfer matrix $\mathbf{H}_b(s)$.

Inside the chip, a second voltage regulation stage is implemented through FIVRs, consisting of multi-phase switching power supplies (e.g., buck converters). Power transistors, switching control circuits, and the output decoupling for these FIVRs are fabricated on-die, while the inductors are placed in the package. As a result, the FIVR output is a filtered and regulated voltage that is distributed through the die power rails to reach logic devices in their respective power domains. In this work, the switching banks of all FIVRs are represented by nonlinear and time-varying circuit blocks. The FIVR switches are controlled through feedback loops that sense the output voltage of each core and translate the tracking error with respect to a reference voltage V_{ref} into the appropriate duty cycle. This operation is attained through dedicated per-core controllers, denoted as \mathcal{K} in Fig. 1. The figure also shows blocks denoted as *output network* which include a circuit model of the PDN of each core, including integrated MIM capacitors which provide the output decoupling, plus a detailed electromagnetic model of the integrated inductors that complete the topology of the FIVRs. This output network can be represented as a LTI system with a transfer matrix $\mathbf{H}_c(s)$.

In order to set notation, we consider a system with N_c identical cores, whereas the number of phases of each FIVR is N_p . As a result, the input network $\mathbf{H}_b(s)$ has $N_c N_p + 1$ ports, each core k is represented by a transfer function $\mathbf{H}_{c,k}(s)$ with N_p ports interfaced to the switches and N_o output ports

where the transient voltage is to be computed, so that the overall output network $\mathbf{H}_c(s)$ has a total of $N_c(N_p + N_o)$ ports. The time-varying duty cycles of all cores are collected in the vector $\mathbf{d}(t) \in [0, 1]^{N_c}$. The main objective of this work is to compute efficiently the transient voltages $v_{k,n}^o(t)$ at all $n = 0, \dots, N_o - 1$ ports of each core $k = 0, \dots, N_c - 1$, excited by predefined current load signals $i_{k,n}^o(t)$ acting concurrently. All results in this work refer to a model of a 11th Generation Intel[®] Core[™] microprocessor, for which $N_c = 4$, $N_p = 4$ and $N_o = 36$. The specific models that are used in this work include only partial output decoupling, as will be evident from AC and transient results.

II. FORMULATION

The initial step in our problem setup is to properly characterize both the input network and the output network. Since these are viewed as LTI subsystems, we follow the standard practice of performing a set of full-wave electromagnetic analyses of the distributed interconnects (board+package) and components (e.g. the FIVR inductor banks), obtaining sampled scattering responses at the ports of interest. These are combined with any lumped terminations (e.g. decoupling capacitors), and the resulting assembled scattering samples are processed by a rational macromodeling engine based on Vector Fitting (VF) with passivity enforcement [5], [6], so that both $\mathbf{H}_b(s)$ and $\mathbf{H}_{c,k}(s)$ are available as a set of linear state-space equations and the associated synthesized SPICE realization. Such macromodels enable a direct (reference) SPICE simulation of the complete system, once complemented with circuit models of the switches and the compensators. This will provide the solution that we will use as reference, both in terms of accuracy and runtime.

One of the key aspects of proposed formulation is the adopted representation for the switches, here represented through averaged models. For each core k and phase j , the corresponding set of FIVR switches is represented by an ideal transformer with turn ratio $1 : d_k(t)$, where $d_k(t)$ is the duty cycle signal resulting from the compensator \mathcal{K}_k of core k . This assumption has its own limitations but is known to be accurate when the buck converters operate in Continuous Conduction Mode (CCM). In particular, if each duty cycle signal $d_k(t)$ is “frozen” and considered as a fixed parameter (e.g. by disconnecting the controllers and opening the feedback loops), the entire structure becomes a large-scale LTI system, which can be fully characterized by the output impedance matrix $\mathbf{Z}(s, \mathbf{d})$ relating the output voltages to the output excitation currents through $\mathbf{V}^o(s) = \mathbf{Z}(s, \mathbf{d})\mathbf{I}^o(s)$.

Figure 2 depicts a sweep over frequency and duty cycle of two representative output impedance elements. The particular structure of such responses and their dependence on the duty cycles \mathbf{d} enable a second layer of model order reduction through a second rational fitting stage, where common poles p_ν are used to represent all impedance entries, and where the

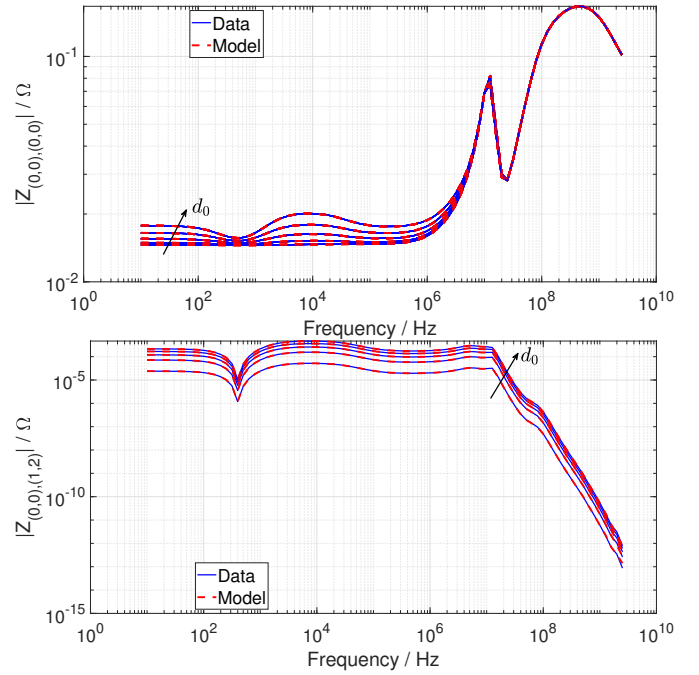


Fig. 2. Open-loop output impedance responses $Z_{(k_1, \ell_1)(k_2, \ell_2)}(j\omega, \mathbf{d})$, at port ℓ_1 of core k_1 while exciting port ℓ_2 of core k_2 , plotted over a sweep of duty cycle values d_0 at fixed $d_1 = d_2 = d_3 = 0.1$. The macromodel responses (dashed lines) are compared to reference AC sweeps from HSPICE (thin solid lines).

associate residues \mathbf{R}_ν are parameterized

$$\mathbf{Z}(s, \mathbf{d}) = \sum_{\nu=1}^{\bar{\nu}} \frac{\mathbf{R}_\nu(\mathbf{d})}{s - p_\nu}. \quad (1)$$

A closed-form parameterization of the residues is obtained by a low-order polynomial interpolation. A state-space realization of the impedance (1), including also the contribution of the constant input V_{VRM} , can be obtained as

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}_o(\mathbf{d})\mathbf{i}^o(t) + \mathbf{B}_i(\mathbf{d})V_{\text{VRM}} \\ \mathbf{v}^o(t) = \mathbf{C}\mathbf{x}(t) \end{cases} \quad (2)$$

where the dependence on the duty cycle is not affecting the open-loop dynamics (matrix \mathbf{A} collecting the poles p_ν is constant) but only the input-state mappings $\mathbf{B}_{i,o}$.

The final step in proposed framework is to reintroduce the closed loop control of the duty cycle signals. This is achieved through the following system of ODEs

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}_o(\mathbf{d}(t))\mathbf{i}^o(t) + \mathbf{B}_i(\mathbf{d}(t))V_{\text{VRM}} \\ \mathbf{v}^o(t) = \mathbf{C}\mathbf{x}(t) \\ \dot{\mathbf{w}}(t) = \mathbf{A}_\mathcal{K}\mathbf{w}(t) + \mathbf{B}_\mathcal{K}\mathbf{e}(t) \\ \mathbf{d}(t) = \mathbf{C}_\mathcal{K}\mathbf{w}(t) + \mathbf{D}_\mathcal{K}\mathbf{e}(t) \end{cases} \quad (3)$$

where the dynamics of all compensators are represented as a (vectorized) state-space system (subscript \mathcal{K}). The vector of error signals $\mathbf{e}(t) \in \mathbb{R}^{N_c}$ feeding the compensators is defined as $\mathbf{e}(t) = \mathbf{N}\mathbf{v}^o(t) - \mathbf{V}_{\text{ref}}$, where \mathbf{N} is a constant selector matrix and \mathbf{V}_{ref} collects the reference voltages. Note

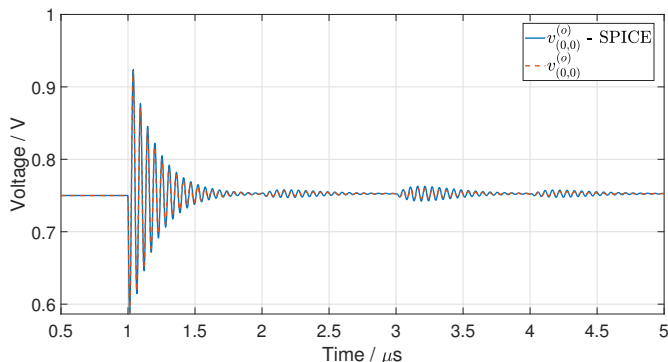


Fig. 3. Transient response of the regulated voltage $v_{k,\ell}^o(t)$ at core $k = 0$, port $\ell = 0$ induced by a sequential transient current step (10 A / 5 ns) excitation per core. The proposed solver response (dashed line) is compared to the reference HSPICE response (thin solid line).

that the explicit time-dependence of all signals is highlighted in (3), so that it becomes evident that the system is not represented as a standard LTI but rather as a nonlinear system in Linear Parameter Varying (LPV) form with feedback. The nonlinearity is however only algebraic (polynomial).

In order to integrate the above system of ODEs numerically, we consider a uniform time step Δt with $t_n = n\Delta t$ for $n = 0, 1, \dots, N_t$, and we initialize all system states at t_0 with their DC solution (no output current excitation and all voltages equal to the V_{ref}). Denoting the approximation induced by discretization as $\hat{\mathbf{x}}_n \approx \mathbf{x}(t_n)$, we introduce an additional relaxation assuming that $\mathbf{d}(t)$ and $\mathbf{e}(t)$ are piecewise constant in each sub-interval, that is $\hat{\mathbf{d}}_n \approx \mathbf{d}(t_n) \approx \mathbf{d}(t)$ for all $t \in [t_n, t_{n+1}]$, and similarly for $\mathbf{e}(t)$. This approximation enables the application of recursive convolutions to integrate equations (3) from t_n to t_{n+1} , where the ODEs are locally linear and the PDN system is decoupled from \mathcal{K} . Differently from standard rational macromodel transient simulation approaches [5], the coefficients of each recursive convolution are time-dependent and updated at each time step. In a compact form, the proposed discretized solution reads

$$\begin{aligned} \hat{\mathbf{x}}_{n+1} &= e^{\mathbf{A}\Delta t} \hat{\mathbf{x}}_n + \\ &\int_{t_n}^{t_{n+1}} e^{\mathbf{A}(t_{n+1}-\tau)} \left[\mathbf{B}_o(\hat{\mathbf{d}}_n) \mathbf{i}^o(\tau) + \mathbf{B}_i(\hat{\mathbf{d}}_n) V_{\text{VRM}} \right] d\tau \\ \hat{\mathbf{v}}_n^o &= \mathbf{C} \hat{\mathbf{x}}_n \\ \hat{\mathbf{w}}_{n+1} &= e^{\mathbf{A}_{\mathcal{K}}\Delta t} \hat{\mathbf{w}}_n + \int_{t_k}^{t_{n+1}} e^{\mathbf{A}_{\mathcal{K}}(t_{n+1}-\tau)} \mathbf{B}_{\mathcal{K}} \hat{\mathbf{e}}_n d\tau \\ \hat{\mathbf{d}}_n &= \mathbf{C}_{\mathcal{K}} \hat{\mathbf{w}}_n + \mathbf{D}_{\mathcal{K}} \hat{\mathbf{e}}_n \end{aligned}$$

III. NUMERICAL EXPERIMENTS

The scheme proposed here has proven effective for transient simulation of the power delivery network of a 4-core microprocessor. In this real-world test case, we are able to show that a prototypal, non-optimized and non-parallel Matlab

implementation of this solver is already about $10\times$ faster than commercial circuit solvers like HSPICE, while still providing accurate results for the purposes of power integrity verification.

Each of the four cores in the test case is a FIVR domain with $N_o = 36$ ports on the die side, totalling $P = 144$ output ports. The obtained reduced-order PDN model (1) has dynamic order $\bar{\nu} = 24$ and is parameterized in terms of the duty cycle \mathbf{d} with polynomial degrees $\rho_o = 2$ and $\rho_i = 1$ for the corresponding two terms in (2). The raw data used to build this macromodel are the PDN Z-parameters sampled for 625 values of \mathbf{d} arranged on a uniform grid in the parameter space, resulting from parametrically-swept AC analyses performed in HSPICE. The macromodel accuracy is demonstrated in Fig. 2.

The PDN system, initially at steady-state, is excited with a 10 A step per core with rise time 5 ns, uniformly distributed among all ports of each core. The individual cores are activated sequentially at $\{1, 2, 3, 4\} \mu\text{s}$, and a transient simulation is performed up to $T = 5 \mu\text{s}$ with a fixed time step $\Delta t = 0.1 \text{ ns}$. In order to perform a fair comparison, the total number of time steps is the same as in the reference HSPICE simulation. The transient results at one representative output port of proposed solver are reported in Fig. 3 where also the reference HSPICE solution is depicted for comparison. The RMS error in the output voltage with respect to the HSPICE transient simulation is 3.9 mV, corresponding to a relative 0.5% cumulative RMS error. In terms of runtime, proposed solver completed the transient analysis in 209 s whereas HSPICE required 1920 s, with a corresponding speedup factor of about $9.6\times$.

IV. CONCLUSIONS

This paper provided a proof of concept of a macromodel-based transient solver for full-system power integrity verification, including core voltage stabilization through averaged models of integrated regulators. The proposed approach has been validated on a model of the power distribution network of an Intel-based 4-core microprocessor, showing excellent accuracy with respect to reference SPICE simulation and confirming a good potential for a dramatic speedup. Future developments will be dedicated to code optimization and parallelization, as well as scalability to higher core counts.

REFERENCES

- [1] K. Radhakrishnan, M. Swaminathan, and B. K. Bhattacharyya, "Power delivery for high-performance microprocessors—challenges, solutions, and future trends," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 11, no. 4, pp. 655–671, 2021.
- [2] E. A. Burton, G. Schrom, F. Paillet, J. Douglas, W. J. Lambert, K. Radhakrishnan, and M. J. Hill, "FIVR — Fully integrated voltage regulators on 4th generation Intel® Core™ SoCs," in *2014 IEEE Applied Power Electronics Conference and Exposition - APEC 2014*, Mar. 2014, pp. 432–439.
- [3] M. Swaminathan and A. E. Engin, *Power integrity modeling and design for semiconductors and systems*. Prentice Hall Press, 2007.
- [4] I. Erdin and R. Achar, "Mcb-dpo: Multiport constrained barrier method-based decoupling capacitor placement optimization on irregularly shaped planes," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 12, no. 4, pp. 665–675, 2022.
- [5] S. Grivet-Talocia and B. Gustavsen, *Passive macromodeling: Theory and applications*. John Wiley & Sons, 2015.
- [6] "IdEM R2018, Dassault Systèmes." [Online]. Available: www.3ds.com/products-services/simulia/products/idem/

Reinforcement Learning for the Optimization of Power Plane Designs in Power Delivery Networks

Seunghyup Han*, Osama Waqar Bhatti[†], and Madhavan Swaminathan[‡]

*seunghyup@gatech.edu, [†]osamawaqarbhatti@gatech.edu, [‡]madhavan.swaminathan@ece.gatech.edu

*[†][‡]School of Electrical and Computer Engineering, [‡]School of Material Science and Engineering
3D Systems Packaging Research Center (PRC), Georgia Institute of Technology, Atlanta, GA, USA

Abstract—This paper proposes a deep deterministic policy gradient (DDPG) based method to optimize the power plane in power delivery networks (PDNs). The proposed method considers the degrees of freedom of a plane design in a layer, determining the parameters for creating a power plane. The results show that the proposed method can provide an optimized power plane design even in a plane layer with a restricted region.

Index Terms—Deep deterministic policy gradient (DDPG), Power plane, Power delivery network, Reinforcement learning (RL)

I. INTRODUCTION

Due to the increase in operating frequencies and current loads in current high-performance computing systems, a robust power delivery network (PDN) design that provides a stable power supply to ICs has become increasingly challenging. The current PDN design for high-speed systems utilizes a pair of power and ground planes. By securing a large power plane, the plane capacitance and low DC resistance decrease the PDN impedance below the target impedance in the frequency range of interest, guaranteeing that the power supply noise is below the threshold level. However, since an IC requires several power rails with different voltage levels, designing multiple power planes to meet each target impedance specification is becoming a critical optimization problem because of the limited area and space constraints, as shown in Fig. 1.

As machine learning has shown rapid development in the field of signal and power integrity [1], various reinforcement learning (RL)-based methods have been proposed for PDN optimization problems [2], [3]. A Deep Q network-based technique for designing the plane structure of a PDN is proposed [4]. However, due to the nature of the value-based RL method, the number of possible unit cells that can be assigned as plane structures is limited, making it difficult to design a plane with various shapes and areas.

In this paper, we propose a deep deterministic policy gradient (DDPG) RL-based method that applies a continuous action space to determine the power plane creation parameters. The size and shape of the plane can be modified with continuous values at each step, thereby obtaining the PDN with the power plane satisfying the target impedance. The proposed method can provide the optimized power plane design with the minimum area and desired shape for the given specifications, even in a plane layer with a restricted region.

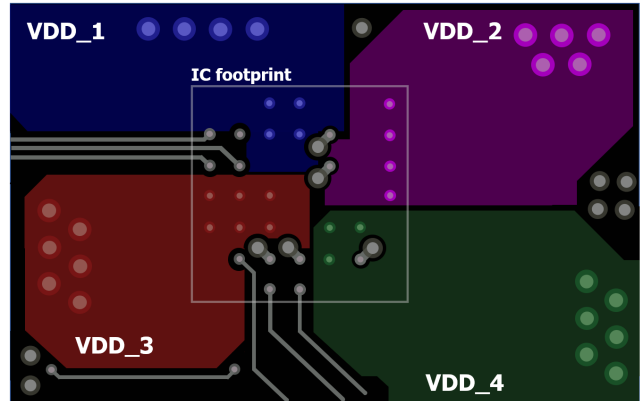


Fig. 1: Layout of the multiple power planes with different voltage levels

II. PROPOSED METHOD

The problem of power plane optimization in PDNs can be expressed as a Markov decision process (MDP) consisting of a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$, where \mathcal{S} , \mathcal{A} , \mathcal{P} , and \mathcal{R} represent a set of states, set of actions, state transition probability, and reward, respectively. Fig. 2 shows the details of the MDP defined in this problem. A state, s_t , which can define a power plane at each time step, is a multi-dimensional vector given by:

$$s_t = [d_1, d_2, d_3, \dots, d_n] \in \mathcal{S} \quad (1)$$

where d is the distance between each vertex of the power plane and the center point at which the IC port is located. An action a_t , determined by the RL agent based on the current state, is the increment of each distance d , given as:

$$a_t = [\Delta d_1, \Delta d_2, \Delta d_3, \dots, \Delta d_n] \in \mathcal{A} \quad (2)$$

A reward r_t is the feedback from the PDN environment after the state transition by the action. In this optimization problem, the reward is given only when the impedance profile below the target impedance in the frequency range of interest is achieved. Based on this MDP, our goal is to determine a policy $\pi(a_t|s_t)$ that maximizes the expected cumulative reward, thereby designing a power plane with a minimized area while satisfying the target impedance. To achieve this optimal policy, we use DDPG, an off-policy actor-critic RL algorithm dealing with a continuous action space [5].

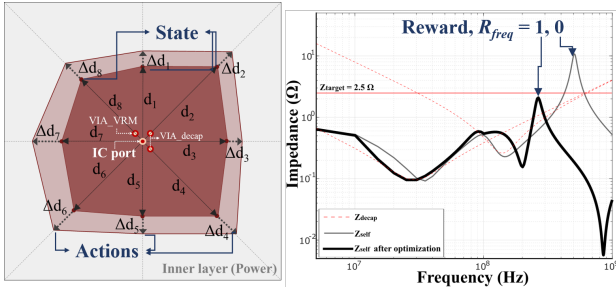


Fig. 2: Details of MDP defined in the design problem of the power plane

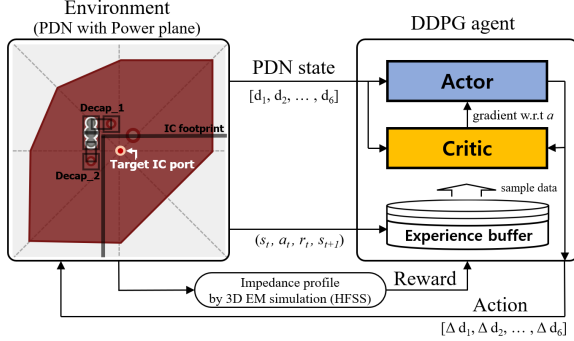


Fig. 3: Overall framework of the optimization of the power plane layout using the proposed DDPG based method

A. DDPG Algorithm

An overview of the DDPG framework applied to the proposed method is shown in Fig. 3. At time step t , the environment produces a state vector of the power plane and sends it to the DDPG agent. The agent then chooses an action for each distance of the plane based on its policy $\pi(a_t|s_t)$. After the actions are taken to the PDN, the environment generates the next state vector s_{t+1} and the reward based on the defined function given as:

$$r_t = R * \left(\frac{1}{A_t} - Var(s_t) \right) \quad (3)$$

where A_t and R are the area of the power plane at step t and the reward given when the target impedance is satisfied, respectively. In (3), the subtracted variance term leads to the exclusion of impractical power plane designs with heavily jagged shapes from the optimization results. The tuple (s_t, a_t, r_t, s_{t+1}) generated in each step is stored in the experience buffer.

The DDPG agent has two networks: the actor, $\mu(s|\theta^\mu)$, that maps states to actions, and the critic, $Q(s, a|\theta^Q)$, that outputs a state-action value where θ^μ and θ^Q are their respective network parameters. [5]. For stable training of the networks, they each have a target network and are updated using a random minibatch of N experience tuples sampled from the buffer. Since the expected reward J needs to be maximized,

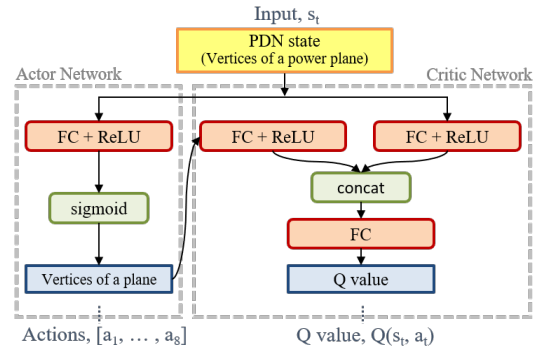


Fig. 4: Detailed structure of the proposed DDPG agent

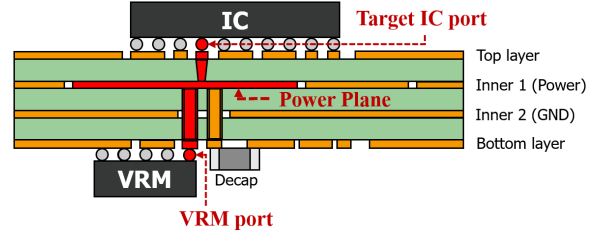


Fig. 5: The test PDN model with 4-layer PCB

the actor is updated using the policy gradient, given as:

$$\nabla_{\theta^\mu} J = \frac{1}{N} \sum_{i=1} \nabla_a Q(s, a|\theta^Q)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu} \mu(s|\theta^\mu)|_{s=s_i} \quad (4)$$

The critic is updated to minimize the TD error defined as:

$$L = \frac{1}{N} \sum_i (y_i - Q(s_i, a_i|\theta^Q))^2 \quad (5)$$

where y_i is the estimated return at the next time step obtained by the target network.

B. Details of the proposed DDPG architecture

Fig. 4 shows the details of the proposed DDPG architecture. A vector form of the state is fed to the actor and critic networks as an input. The actor network has FC layers with the ReLU activation function, where the last layer is the sigmoid function that outputs continuous action values in the range of $[0, 1]$. The critical network uses both action and state values as inputs. The last hidden layers of the network for each input are concatenated, and the last FC layer outputs the Q value that evaluates the current state and action.

III. POWER PLANE OPTIMIZATION USING THE PROPOSED DDPG BASED METHOD

We apply the proposed method to optimize the power plane in PDNs. Fig. 5 shows the test PDN model. The power plane to be optimized is in the second layer. The VRM and decaps placed on the bottom layer and the target IC on the top layer are connected to the plane with vias. The via for the power supply and those connected to the two pre-assigned decaps

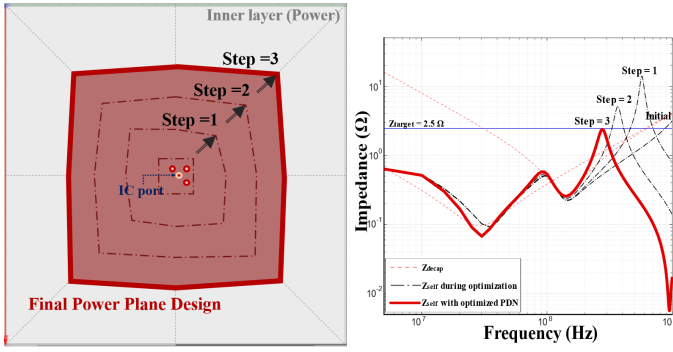


Fig. 6: Optimized power plane design satisfying the target impedance at the target IC port

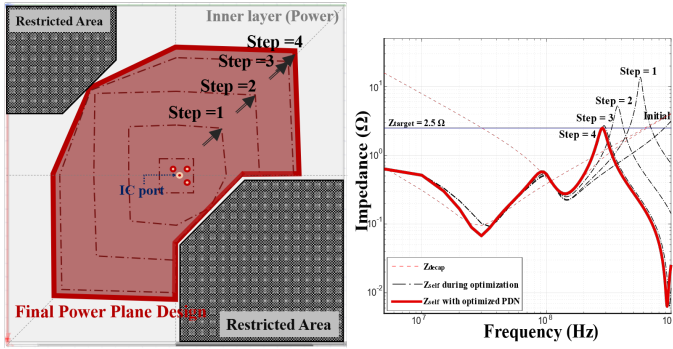
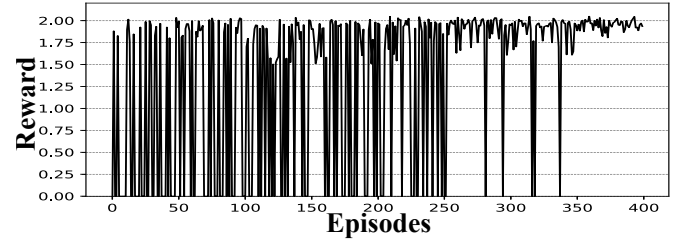


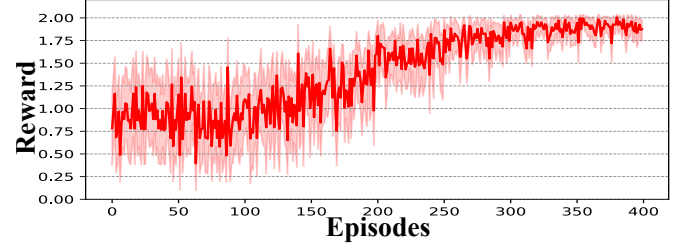
Fig. 7: Optimized power plane design in the PDN with a plane-restricted region

are located close to the IC port. We simulate the PDN using Ansys HFSS [6] to obtain the impedance profile at the target IC port. The entire area where the power plane can be extended is divided into eight equal parts centered on the IC port, and the vertices of the polygon forming the power plane are located on the boundary of each area. In the initial state, the distance between each vertex and the center point is 10% of the total length of each boundary line. This distance is increased by 10%-20% depending on the action of each step, thereby increasing the area of the power plane with an arbitrary shape.

Using the proposed method, we obtain the optimized power plane design and its impedance profile at the target IC port, as shown in Fig. 6. As the distance between each vertex and the center point is increased in each step, the size and shape of the power plane area are modified, thereby obtaining the final power plane design (polygon formed with thick red lines) that satisfies the target impedance in the frequency range of interest (50MHz-1GHz). Fig. 8 shows the convergence property of the proposed method. The reward successfully converges to the maximum number as the DDPG agent is trained, indicating that a minimized area of the power plane with a uniform shape is achieved. We also optimize the test PDN, where there is a plane-restricted region for vias or other purposes. As shown in Fig.7, at the end of the training, the power plane design with a minimized area that meets the target impedance can be



(a)



(b)

Fig. 8: (a) Convergence of the reward calculated based on the impedance profile and the area and shape of the power plane. (b) Average reward during the training .

achieved using the proposed method without relying on human design experience.

IV. CONCLUSION

In this paper, we leverage the DDPG RL algorithm to optimize the power plane in PDNs to satisfy the target impedance with a minimized plane area. The implemented continuous action space allows the design parameter values of the power plane to be determined with continuous values, resulting in a power plane design with the desired shape. The results show that the proposed method achieves an optimized power plane with a minimized area even in the case of a power plane layer with a limited space.

REFERENCES

- [1] M. Swaminathan, H. M. Torun, H. Yu, J. A. Hejase, and W. D. Becker, "Demystifying machine learning for signal and power integrity problems in packaging," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 10, no. 8, pp. 1276–1295, 2020.
- [2] H. Park, J. Park, S. Kim, K. Cho, D. Lho, S. Jeong, S. Park, G. Park, B. Sim, S. Kim *et al.*, "Deep reinforcement learning-based optimal decoupling capacitor design method for silicon interposer-based 2.5-d/3-d ics," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 10, no. 3, pp. 467–478, 2020.
- [3] K. Son, M. Kim, H. Park, D. Lho, K. Son, K. Kim, S. Lee, S. Jeong, S. Park, S. Hong *et al.*, "Reinforcement-learning-based signal integrity optimization and analysis of a scalable 3-d x-point array structure," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 12, no. 1, pp. 100–110, 2021.
- [4] S. Lee, H. Kim, K. Song, J. Kim, D. Park, J. Ahn, K. Kim, and S. Ahn, "Deep reinforcement learning-based power distribution network structure design optimization method for high bandwidth memory interposer," in *2021 IEEE 30th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*. IEEE, 2021, pp. 1–3.
- [5] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.
- [6] Ansys-HFSS. [Online]. Available: <https://www.ansys.com/products/electronics/ansys-hfss>

Network Model Compensation For Single-Point Measurements Of Multi-Pin Devices When Using Non-Invasive Current Estimation

Chad M. Smutzer, Jordan R. Keuseman, Clifton R. Haider and Barry K. Gilbert
 Mayo Clinic Special Purpose Processor Development Group (SPPDG)
 Rochester, MN, USA

Email: smutzer.chad@mayo.edu, keuseman.jordan@mayo.edu, haider.clifton@mayo.edu, gilbert.barry@mayo.edu

Abstract—Power delivery network (PDN) model development is often simplified using superports or pin-groups for high pin count devices. This approach significantly reduces model complexity but can compromise accuracy in holistic time- and frequency-domain analyses. In the context of the non-invasive current estimation (NICE) technique for packaged, high-performance integrated circuits (ICs), this paper describes the limitations of using pin-group PDN models in distributed impedance applications. DC error calibration factors are calculated, and tuned AC error compensation networks are proposed. These fundamental techniques for working with overly simplified models in highly distributed power integrity applications are derived and demonstrated through simulation and measurement on exemplar hardware.

Keywords—Load Current Calculation, Power Delivery, PDN Modeling, Distributed Network Compensation

I. INTRODUCTION

The measurement of load current profiles for high-power electronic devices is problematic when traditional techniques cannot be employed. The Non-Invasive Current Estimation (NICE) methodology [1][2] was developed to enable indirect current measurement in applications where series sense resistors and inductive coil probes are prohibitive. Using voltage measurements at the source $V_1(t)$ and load $V_2(t)$, in conjunction with the admittance parameters $Y(s)$ for the power-delivery network (PDN), circuit currents can be calculated from two-port network theory.

As previously reported and partially resolved [3], the practical application of NICE has presented some challenges. The accuracy of the voltage measurements and PDN model is paramount when applying the technique. We demonstrated that oscilloscope ground isolation is required to prevent undesirable current paths in the instrumentation environment thereby reducing voltage measurement errors. Despite this improvement, overall correlation between the applied and calculated load current continued to be unsatisfactory. Attention was turned to the PDN model and the known limitations of two-port network usage in distributed applications [4].

The simplicity of NICE is dependent upon the well-established practice of grouping pins at the source and load devices in PDN modeling [5]. For many designs, the resultant two-port model is most often used to evaluate broadband PDN impedance $Z(f)$ against a target goal [6]. When employed in NICE, the model simplification also reduces the matrix

mathematics. However, grouping pins in electromagnetic simulator tools inherently assigns an equipotential boundary condition to all device pins within the group. This grouping has the unwanted effect of negating the distributed qualities of the PDN R/L/C within the pin-field of a large ASIC or FPGA.

Modern high-power electronics have thousands of pins distributed within a packaged area for placement on the surface of printed circuit boards (PCB). The pin distribution may be symmetrical in some cases, but the PDN impedance is often asymmetrical when decoupling capacitors or PCB effects are considered. The image in Fig. 1 (a) depicts a sample footprint of a typical multi-stage voltage regulator (VR) and a large multi-pin load device with a planar shape representing the PCB PDN.

The current calculated by NICE is highly dependent upon the voltage measurement probe location within the load pin-field, as evident in the plots of Fig. 1 (b). The selection of which two pairs of pins to use in NICE can result in over- or underestimated steady-state (DC) current as well as load transition errors in the form of over/undershoot (AC). The problem is exacerbated by next-generation electronic devices with increasingly larger packages and pin arrays.

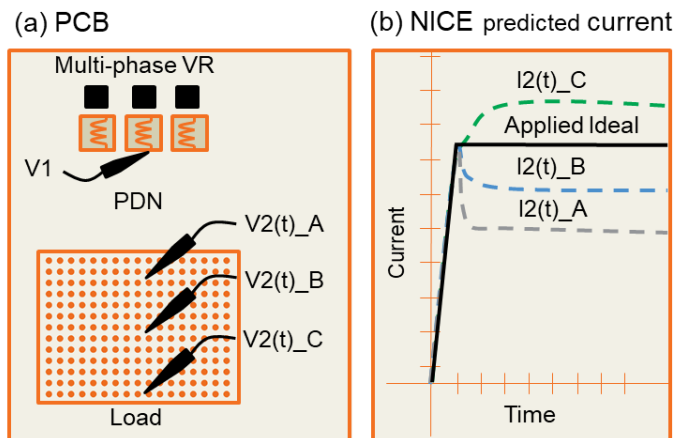


Fig. 1 – NICE Current Prediction Errors $I_2(t)$ A, B, C Resulting From Selected Voltage Probe Locations $V_2(t)$ A, B, C

In this paper, we address the impact of using grouped pins for PDN network modeling as applied in NICE. An explanation is provided for the specific AC and DC error terms that arise in highly distributed pin-field designs. Solutions are suggested for either mathematical correction or PDN network compensation,

which have been validated through simulations. Lab results are presented from a device under test (DUT) consisting of a high-performance VR and a high pin-count programmable load attached to a PCB. Finally, we apply the corrections to the NICE technique and demonstrate the improved current calculation results in the test hardware.

II. STEADY STATE COMPENSATION

The deleterious effect of pin-grouping is best visualized in voltage gradient plots from DCIR analyses. For the examples presented in Fig. 2, a DC current load was applied to a multi-pin device using individual sink elements on each voltage pin (a) and using a single sink with the voltage pins grouped (b) – with the same total current in each case. Unsurprisingly, the grouped pin condition nullifies the distributed impedance and does not accurately capture voltage gradient within the pin-field. This underlying issue is also apparent in broadband two-port PDN models necessary for NICE.

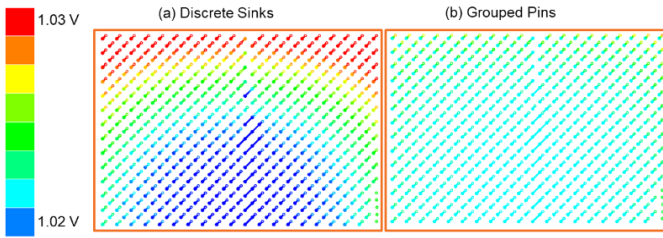


Fig. 2 – Pin-field Voltage Gradient With Distributed And Grouped Sinks

The simplified PDN network shown in Fig. 3 (a) is used to further illustrate the problem. In this sample circuit, nodes PA and PB represent two pins of a multi-pin load device connected to a VR through PDN impedances ZA and ZB, where $Z_A < Z_B$. Transient simulation was performed with inter-pin impedance ZAB included. While applying a time-aligned current step to each IA and IB, voltages at V(PA) and V(PB) were recorded to emulate the hardware test environment where voltage can be measured at any discrete pin. Subsequently, a two-port PDN model was extracted with one port at the VR site and the second port grouping pins PA and PB at the load. This pin grouping has the illustrated effect of nullifying ZAB.

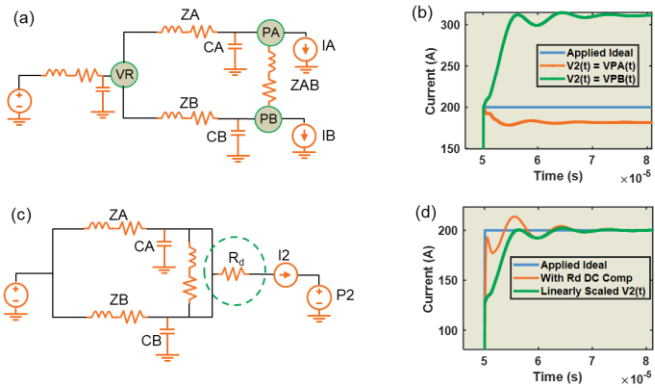


Fig. 3 – Simplified Example PDN Network Demonstrating the Effects of DC Compensation on NICE Current Prediction

With $I_A = I_B = 100A$ step, a total load current of 200A is expected. However, using NICE with the two-port network and each voltage V(PA) or V(PB) at the load, the calculated current is either over- or under-estimated, as depicted in Fig. 3(b). Significant current errors can be observed even when voltage deviations are small. This example demonstrates the spatial significance of voltage gradient and the impact of negating inter-pin impedance ZAB within a distributed pin-field.

The center region of a pin-field is typically where IR drop is maximal due to the PDN impedance from the VR being higher for inner pins. As demonstrated using V(PB) in the sample circuit, measured voltage at this location will result in an over-estimation of current using NICE. For this case, the addition of series resistor R_d at the grouped load port, as depicted in Fig. 3 (c), will compensate for the distributed resistance nullified by the pin-group. This approach effectively reconstructs the DC voltage drop expected at the measurement location which is otherwise ignored in the two-port PDN model. This is akin to extending the model reference plane to the voltage probe location. One way to calculate the compensation resistance is through a calibration process of applying known DC load currents and tuning R_d until the predicted steady state current matches the known applied current.

An alternative to adding R_d in the NICE PDN model is to linearly scale the measured V2(t) voltage using a correction factor also calculated from a calibration process. This technique does an equivalent job of accurately predicting the steady state current; however as observed in Fig. 3(d), the distortion in the corners of the predicted load current edges is worse when using the voltage correction factor scaling approach.

III. AC COMPENSATION NETWORK

The addition of the DC compensation resistor adequately corrects for steady-state errors in the NICE calculations; however, appreciable amounts of time-based distortion and ripple are still present in the predicted current. Akin to the previously described resistive gradient in the pin-fields, distributed inductance and capacitance is also erroneously nullified by grouping pins. Additionally, the spatial effects from non-uniform decoupling capacitor placement are not accurately captured by the two port PDN model.

Using the same simplified circuit described above, an AC compensation network, Z_{comp} , was added to the load port, as depicted in Fig. 4 (a). In this example, two complex shunt impedances are added to null the remaining ringing current after DC compensation. The exact Z_{comp} network architecture will vary depending upon the PDN application, but these two complex elements were suitable for the demonstration circuit.

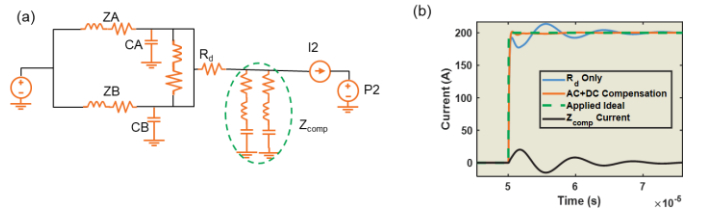


Fig. 4 – AC Compensation Network and Final Simulated NICE Results

With the addition of both the AC and DC compensation elements to the two-port PDN network in the simplified circuit, the current predicted by NICE matches the ideally applied 200A step, as shown in Fig. 4 (b). The current specific to the Z_{comp} network is also plotted (“ Z_{comp} Current”) to illustrate the time-aligned subtractive effect on the curve with only DC compensation (“ R_d Only”).

The Z_{comp} values can be obtained empirically through a measurement calibration process with known step loads in programmable test hardware (e.g. [7], [8]) and circuit element optimization in simulation. Once the entire compensation network is defined from a single step response and steady-state calibration, any current load profile signature can be accurately predicted by NICE including sinusoids, periodic pulses and arbitrary or unpredictable activity typical in ASIC/FPGA applications. The compensation network is a fixed error correction for spatial current flow in grouped pin-fields, at a single measurement location.

In applications where the PDN impedance from the VR to the load is much greater than the inter-pin spatial impedance within a pin-field, network compensation may be minor or unnecessary altogether. However, sometimes the compensation element values may be large or additional components are necessary if the pin-to-pin spatial impedance has a dominant influence on the PDN behavior.

IV. APPLICATION OF COMPENSATION IN MEASURED RESULTS

The same DUT described in [3] was employed to validate the grouped pin compensation strategy. For this hardware, the multi-pin load device was programmed to step from 100A to 200A as depicted by the “Applied Ideal” curve in Fig. 5 (b). Using only the extracted two-port admittance parameters for the PDN, the “Uncompensated NICE” curve contains both steady state and distortion errors. After applying the compensation network of Fig. 5 (a), the NICE calculation matched the programmed step very well.

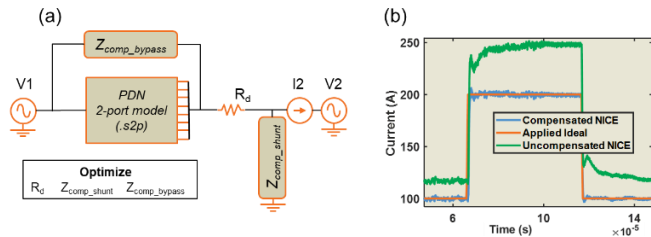


Fig. 5 – Test Hardware Compensation and Final Measured NICE Results

In this well-designed test hardware, the PDN impedance was intentionally reduced by the vendor to be $\ll 100 \mu\Omega$ from VR to load, a value not conventionally attainable in practical compute applications. Furthermore, there was an unusually large quantity of decoupling capacitors on the test PCB spread throughout the pin array. Given those two unique conditions, the addition of a bypass subcircuit illustrated in Fig. 5 (a) was also necessary to compensate for the AC and DC voltage gradient ignored by the grouped pins in the two-port model.

Although not emphasized in the waveform plots, the predicted 10-90% current rise- and fall-times match the applied

currents very well even without the AC compensation. The quality of the distributed capacitors embedded in the extracted PDN model dominate the di/dt behavior during the initial current transition.

V. SUMMARY

Packaged high-performance ICs are growing in footprint size and the input power pin count is increasing to accommodate high current operation. As the PCB pin-field expands, the spatial effects on PDN R/L/C cannot be ignored when utilizing NICE to predict load current. Grouping pins in two-port PDN network extraction obfuscates the distributed inter-pin current flow paths, in turn requiring network compensation.

This paper specifically concentrated on multi-pin ASIC/FPGA load devices and suitable compensation strategies for mostly symmetric BGA pin-fields. However, it should be noted that the same strategies can be used at the VR side of the PDN network if a multi-phase or a highly distributed regulator design is used.

Through experimentation with discrete example circuits, it became clear that the compensation designs will be application specific. Therefore, the exact topology of R/L/C elements required to appropriately account for the pin-field impedance distribution will likely differ depending upon PDN design features. The approach described herein provides a general roadmap for correcting errors when applying NICE using two-port networks to devices with large spatial impedance distribution.

REFERENCES

- [1] Fasig, J.L., White, C.K., Gilbert, B.K., and Haider, C.R., “Introduction to Non-Invasive Current Estimation (NICE),” DesignCon, 2018, Santa Clara, CA.
- [2] Fasig, J.L., Systems and Methods for Non-Invasive Current Estimation, International Patent Application No. PCT/US2018/041632, 2018.
- [3] C. M. Smutzer, J. R. Keuseman, C. K. White, C. R. Haider and B. K. Gilbert, "Enhancements to the Non-Invasive Current Estimation Technique Through Ground Isolation," 2021 IEEE 30th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS), 2021, pp. 1-3, doi: 10.1109/EPEPS51341.2021.9609134.
- [4] X. Hu, P. Du, J. F. Buckwalter and C. Cheng, "Modeling and Analysis of Power Distribution Networks in 3-D ICs," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 21, no. 2, pp. 354-366, Feb. 2013, doi: 10.1109/TVLSI.2012.2183904.
- [5] K. Shringarpure et al., "Formulation and Network Model Reduction for Analysis of the Power Distribution Network in a Production-Level Multilayered Printed Circuit Board," in IEEE Transactions on Electromagnetic Compatibility, vol. 58, no. 3, pp. 849-858, June 2016, doi: 10.1109/TEMC.2016.2535459.
- [6] L. D. Smith, R. E. Anderson, D. W. Forehand, T. J. Pelc and T. Roy, "Power distribution system design methodology and capacitor selection for modern CMOS technology," in IEEE Transactions on Advanced Packaging, vol. 22, no. 3, pp. 284-291, Aug. 1999, doi: 10.1109/6040.784476.
- [7] W. Xu, J. Fang, J. He and T. Kim, "Switching voltage regulator modeling and its applications in power delivery design," 2014 IEEE International Symposium on Electromagnetic Compatibility (EMC), 2014, pp. 855-860, doi: 10.1109/ISEMC.2014.6899087.
- [8] Koether, Ethan. “EDICon 2017 Transient Load Tester for Time Domain PDN Validation.” (2017).

Efficient Modeling of Random Jitter Due to Stochastic Power Supply Noise in CMOS Inverters

Ahsan Javaid*, Ramachandra Achar⁺, *Fellow IEEE*, and Jai Narayan Tripathi^Δ, *Senior Member IEEE*

* + Dept. of Electronics, Carleton Uni, Ottawa, Canada, ^Δ Dept. of EE, IIT Jodhpur, Jodhpur, India.

Email: *ahsanjavaid@cmail.carleton.ca, +achar@doe.carleton.ca, ^Δjai@iitj.ac.in

Abstract—In this paper, analytical expressions are developed for estimating random jitter (RJ) in the presence of stochastic power supply noise for CMOS inverter circuits. The proposed approach employs probability density function of the propagation delay associated with a CMOS inverter in the presence of supply variations with normal distribution. The closed-form relations are further advanced to include the effects of load. The proposed model demonstrates a reasonably accurate prediction of RJ and yields significant speed-up compared to using a circuit simulator (HSPICE) for a case study with 22nm CMOS technology.

Index Terms—CMOS inverter, probability density functions (pdfs), power integrity, random jitter (RJ), root mean square (RMS).

I. INTRODUCTION

Modern low-power consumption technologies operating at higher frequencies require quality power supply modules in order to power billions of devices contained in them. Even though the process technology continues to improve while the package design remains relatively unchanged, the desired quality and performance of these power distributing modules can be greatly affected due to reduced noise margins. Hence, designing to ensure power integrity becomes increasingly challenging [1]-[3] in modern electrical products.

Sensitivity of a circuit to the supply noise is an important contributor to the timing jitter [4] and becomes very critical to meet the timing budget of digital signals. Jitter can be classified into two groups: deterministic jitter (DJ) and random jitter (RJ). DJ is bounded and mainly correlates with the data pattern. On the other hand, RJ is unbounded and unpredictable. It is generally represented by a standard deviation, or root mean square (RMS) value of a normal distribution. RJ is often caused due to imperfection in semiconductor processing and thermal issues. Estimating the amount of jitter associated with digital signals is essential to predict the overall system performance.

For a reasonable jitter estimation, conventional approach requires simulation of large number of bits that can make the process prohibitively CPU expensive. To address the associated computational burden, several models for random jitter have been extensively researched [5]-[19]. Tail fitting algorithm based on normal distribution for random and deterministic jitter measurements was discussed in [5]. In [6], random jitter extraction techniques in high-speed signaling were investigated. Correlation between DJ and RJ using BER curve was addressed in [7]. Improved tail-fitting techniques can be found in [8]. [9] and [10] address delay-line based

techniques combined with cdfs to estimate the RMS value of the random jitter for BIST applications. Other statistical approaches using white and colored random jitter spectrums modeled in StatEye and LinkLab engines can be found in [11] and [12], respectively. In [13] and [14], the Gaussian mixture models are proposed for jitter decomposition. Frequency domain methods where RJ is modeled as an average noise of power spectral density are discussed in [15] and [16]. [17] addresses FFT based technique combined with time-lag-correlation to estimate different jitter components. Neural network approach to obtain RJ using jitter histogram is investigated in [18] and [19].

This paper presents development of closed-form expressions for predicting RMS value of the random jitter in the presence of stochastic power supply for CMOS inverter circuits. In the proposed work, probability density functions are considered for both, normally distributed power supply as well as the propagation delay [20], to derive an efficient closed-form analytical model for random jitter. The closed-form model is further advanced to include the effects of load. Results from a case study of inverters based on 22 nm CMOS PTM technology demonstrate that the proposed model predicts the RJ reasonably well while providing computational speed up.

II. DEVELOPMENT OF THE PROPOSED ANALYTICAL BASED MODEL FOR RANDOM JITTER ANALYSIS

In the recent years, computationally efficient analytical models are commonly used in signal and power integrity analysis to mimic a specific system. On the other hand, conventional methods require large number of bits for simulating low probability events, such as random jitter, that can make the process prohibitively CPU expensive. Development of the proposed analytical model that yields good speed-up compared to conventional approaches is presented in the following.

A. Proposed Method

In this section, a closed-form expression for predicting RMS value of the random jitter is developed. For this purpose, a general CMOS inverter is considered with n-channel (M_n) and p-channel (M_p) transistors connected to a load capacitor (C_L), a pulse-train input (V_{in}) and a voltage supply (v_{dd}) as can be seen in Fig. 1.

The probability distribution of the propagation time delay [20] at the output of the inverter can be obtained as

$$\tau(v) = v^2(v - v_{th})^{-(1+\alpha)} \quad (1)$$

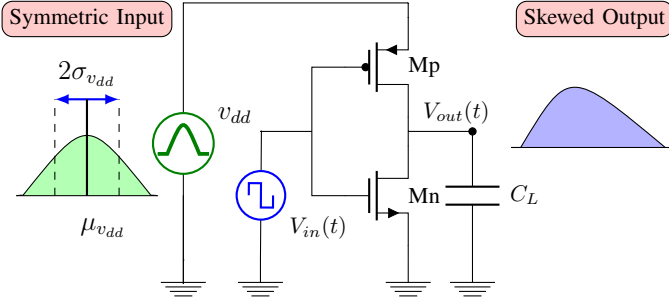


Fig. 1: CMOS inverter with a stochastic power supply noise

where v represents the random variations in the power supply, v_{th} is the threshold voltage and α represents the velocity saturation index [21]. Also, the pdfs of both, the supply voltage and the propagation time delay developed in [20] can be represented as

$$f_{v_{dd}}(v) = \frac{1}{\sqrt{2\pi}\sigma_{v_{dd}}} \exp\left[-\frac{1}{2}\left(\frac{v - \mu_{v_{dd}}}{\sigma_{v_{dd}}}\right)^2\right] \quad (2)$$

$$f_{\tau}(v) = -f_{v_{dd}}(v) \frac{dv}{d\tau} \quad (3)$$

where $\sigma_{v_{dd}}$ and $\mu_{v_{dd}}$ are the RMS (standard deviation) and nominal (mean) voltages associated with the input power supply, respectively. It is noted in (2) that the power supply is normally distributed around the nominal voltage while the time-delay distribution (3) at the output of a CMOS inverter is skewed, as illustrated in Fig. 1.

Next, using (2) and (3), the standard deviation (σ_{τ}) of the propagation delay distribution can be extracted (which also represents the RMS value of random jitter) as

$$RJ = \sigma_{\tau} = \sigma_{v_{dd}} \Gamma \left. \frac{d}{dv} \tau(v) \right|_{v=v_H} \quad (4)$$

In (4), a technology-based constant Γ [22] is introduced to properly adjust the units, given as

$$\Gamma = \frac{1}{4}(1 + \alpha)C_L(v_{ref} - v_{th})^{\alpha}(i_{ref})^{-1} \quad (5)$$

where v_{ref} and i_{ref} are the voltage and drain current when $v_{GS} = v_{DS}$, respectively.

Let the subscript v_H in (4) be defined as the voltage of interest at which the derivatives of (1) are evaluated. The voltage (v_H) represents the supply voltage associated with the propagation delay of a CMOS inverter and is dependent on the load capacitor (C_L). In the next section, new relations in the presence of load capacitor are developed to obtain the v_H (the voltage of interest), which is used while obtaining an accurate estimation of random jitter.

B. Proposed Method for Obtaining v_H for RJ Analysis including Effects of Load

In this section, proposed analytical model for obtaining v_H is developed including the effects of load. It is important to

consider an appropriate modeling method that relates CMOS inverter parameters with the time delay distribution as well as the effect of the load capacitor. For this purpose, pdf of propagation time delay (3) is used to investigate the behaviour of permutation; $n f_{\tau}$ (where $n = 1, 2$ and 3). The corresponding graphs for each n are shown in Fig. 2. The $n f_{\tau}$ response shows that as n increases, the corresponding voltage difference (v_{τ_1}, v_{τ_2} or v_{τ_3}) between the adjacent responses evaluated at the maximum of the noise distribution ($f_{v_{dd}}^{\max}$) decreases and follows the inverse power law (IPL) model [23].

Using a similar IPL model as described in [24], an expression for v_H can be developed as

$$v_H = \frac{a}{(1+x)^b} \quad (6)$$

where a and b are the unknown parameters. Also, parameter x is defined as $x = c - 1$ where c is a ratio of load to reference capacitance (C_L/C_{ref}). In the subsequent discussions, C_{ref} is set to equal to 1 and with the same units as C_L such that C_L/C_{ref} is unitless.

Next, closed-form expressions for parameters a and b are derived using voltage differences (v_{τ_1} and v_{τ_2}) between two adjacent responses of $n f_{\tau}$ (when $n = 1$ and 2) (Fig. 2). Also, Taylor series method [25] is used to expand (6) and to obtain derivative terms associated with (4). Hence, comparing both the series along with v_{τ_1} and v_{τ_2} , coefficients a and b can be obtained as

$$a = \frac{v_{th}^{-\alpha}(v_{\tau_2} - v_{\tau_1})}{(1-\alpha)(v_{th} + 5/7)^{-\alpha}} ; \quad b = -\frac{v_{\tau_1}(1+\alpha)}{2v_{th}} \quad (7)$$

In the experiments conducted, for reasonable accurate matching of the output jitter, low order Taylor series expansion is found to be sufficient (such as order 2).

Proposed analytical model for v_H (6) directly relates the input power supply with the load capacitor and is used while obtaining the corresponding derivatives of the propagation delay of a CMOS inverter which are required while estimating the random jitter. Using (4) along with (6), eliminates the

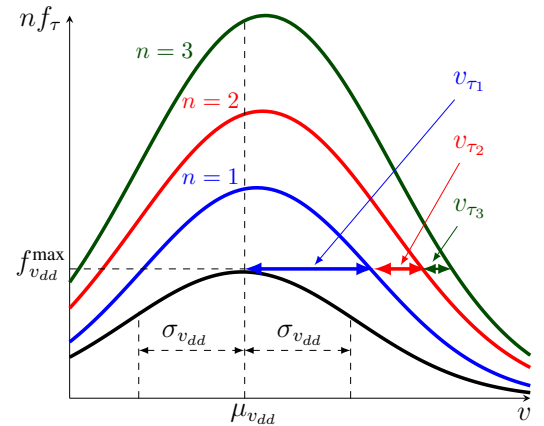


Fig. 2: PDFs of $n f_{\tau}$ evaluated at the maximum of $f_{v_{dd}}$ for $n = 1, 2$ and 3 (with $\alpha = 1.5$ and $v_{th} = 286$ mV).

need for direct time-domain or frequency-domain simulations. It only requires one DC simulation in order to obtain the technology-based coefficient (5) which leads to significant speed-up.

III. RESULTS AND DISCUSSION

In this section, the proposed methodology to evaluate RJ is validated. For this purpose, CMOS circuits operating at 125 MHz (with a rising edge of 0.4 ns) and Level-54 device parameters for 22 nm PTM technology [26] are used to simulate n-channel and p-channel devices. Also, RJ is evaluated at the midpoint of the rising edge for 1 million bits.

In this example, a normally distributed power supply (v_{dd}) of a CMOS inverter having 10 fF load capacitor with a nominal voltage ($\mu_{v_{dd}}$) of 1.2 V is considered. The input RMS voltage is varied from 0 to 100 mV with a uniform step size of 10 mV and corresponding results for random jitter are compared against the conventional approach using a circuit simulator (HSPICE).

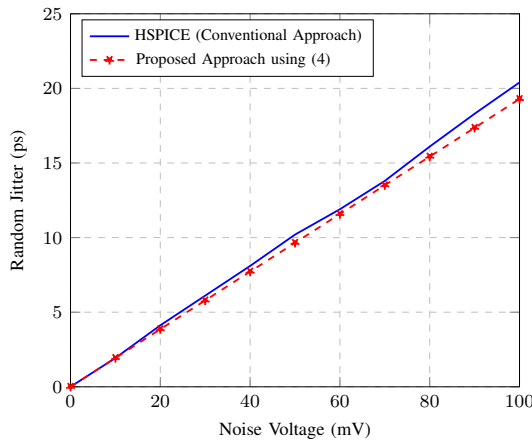


Fig. 3: Comparison of RJ response using proposed model versus the conventional approach (using HSPICE)

Corresponding plots of random jitter for each of the approaches are shown in Fig. 3. Also, CPU times for evaluation of RJ outputs using the proposed (4) and the conventional approaches at 10 equally spaced points from 0 V to 0.1 V are measured to be 0.005167 sec and 260 sec, respectively.

It can be seen from the above results that the proposed approach matches reasonably well with the conventional approach while achieving a speed up of 50319 compared to the conventional approach. The discrepancy between HSPICE and the proposed model can be further reduced by updating the parameters in (6) using high order Taylor series terms along with increasing the number of permutations of $n_{f\tau}$.

IV. CONCLUSION

In this paper, an efficient closed-form model is developed to estimate the RMS value of the random jitter for a CMOS inverter. Proposed approach relates stochastic supply noise with propagation time delay via probability density functions to provide closed-form analytical relations for RJ analysis.

The RJ model is also advanced to include the effect of load. Validating example demonstrates the reasonable accuracy and efficiency achieved using the proposed method.

REFERENCES

- [1] H. M. Tong, Y. S. Lai and C. P. Wong, "Advanced Flip Chip Packaging", Springer, 2013.
- [2] K. J. Wolter, "System Integration by Advanced Electronics Packaging", Chapter 2 in Bio and Nano Packaging Techniques, Springer 2012.
- [3] M. P. Li, "Noise and Signal Integrity at High-Speed", Englewood Cliffs, NJ, USA: Prentice-Hall, 2007.
- [4] A. V. Mezhiba and E. G. Friedman, "PDN in High Speed Integrated Circuits", Kluwer Academics, 2004.
- [5] M. P. Li, J. Wilstrup, R. Jessen, and D. Petrich, "A new method for jitter decomposition through its distribution tail fitting," in *Proc. IEEE Int. Test Conf.*, 1999, pp. 788–794.
- [6] C. K. Ong, D. Hong, K. T. Cheng, Li-C Wang, "Random Jitter Extraction Technique in a Multi-gigahertz Signal", *Design Automation Conference*, 2004, pp. 286–291.
- [7] V. K. Sharma, J. N. Tripath, R. Nagpal, S. Deb, and R. Malik, "A comparative analysis of jitter estimation techniques," in *Proc. IEEE Int. Conf. Electron., Commun. Comput. Eng.*, 2014, pp. 125–130.
- [8] G. Soliman, "Improved jitter distribution tail-fitting algorithm for decomposition of random and deterministic jitter," *IEEE Trans. Electromagn. Compat.*, vol. 62, no. 5, pp. 1852–1858, Oct. 2020.
- [9] J.-L. Huang, "A random jitter extraction technique in the presence of sinusoidal jitter," in *Test Symposium, 2006. ATS '06. 15th Asian*, Nov. 2006, pp. 318–326.
- [10] J. W. Lee, J. H. Chun and J. A. Abraham, "A Random Jitter RMS Estimation Technique for BIST Applications", in *Test Symposium, 2009.*
- [11] D. Oh, F. Lambrecht, S. Chang, Q. Lin, J. Ren, C. Yuan, J. Zerbe, and V. Stojanovic, "Accurate system voltage and timing margin simulation in high-speed I/O system designs," *IEEE Transactions on Advanced Packaging*, vol. 31, no. 4, pp. 722–730, Nov. 2008.
- [12] A. Sanders, M. Resso, and J. D' Ambrosia, "Channel compliance testing utilizing novel statistical eye methodology," presented at the IEC Design-Con, Santa Clara, CA, 2004.
- [13] F. Nan, Y. Wang, F. Li, W. Yang, and X. Ma, "A better method than tail fitting algorithm for jitter separation based on Gaussian mixture model," *J. Electron. Test.*, vol. 25, no. 6, pp. 337–342, Dec. 2009.
- [14] D. Mistry, S. Joshi, and N. Agrawal, "A novel jitter separation method based on Gaussian mixture model," in *Proc. Int. Conf. Pervasive Comput.*, 2015, pp. 1–4.
- [15] J. Kho and T. Y. Ling, "Fast and accurate technique to decompose jitter for very long pattern length waveform," in *Proc. IEEE Elect. Des. Adv. Packag. Syst. Symp.*, 2014, pp. 93–96.
- [16] S. Tabatabaei, M. Lee, and F. Ben-Zeev, "Jitter generation and measurement for test of multi-Gbps serial IO," in *Proc. IEEE Int. Conf. Test*, 2004, pp. 1313–1321.
- [17] H. Pang, J. Zhu, and W. Huang, "Jitter decomposition by fast fourier transform and time lag correlation," in *Proc. Int. Conf. Commun., Circuits Syst.*, 2009, pp. 365–368.
- [18] C. K. Ku, C. H. Goay, N. S. Ahmad, and P. Goh, "Jitter decomposition of high-speed data signals from jitter histograms with a pole-residue representation using multilayer perceptron neural networks," *IEEE Trans. Electromagn. Compat.*, vol. 62, no. 5, pp. 2227–2237, Oct. 2020.
- [19] N. Ren, Z. Fu, D. Zhou, H. Liu, Z. Wu, and S. Tian, "Jitter decomposition by Convolutional Neural Networks," *IEEE Trans. Electromagn. Compat.*, vol. 62, no. 5, pp. 2227–2237, Oct. 2021.
- [20] D. Dietz, Stochastic Propagation Delay Through a CMOS Inverter as a Consequence of Stochastic Power Supply Voltage—Part II: Modeling Examples, *IEEE Trans. Electromagn. Compat.*, vol. 61, no. 1, Feb. 2019.
- [21] M. Al-Mosawy, "Current and Delay Estimation in Deep Sub-micrometer CMOS Logic Circuits," M.Sc. thesis, Dept. of Elec. Carleton Univ., Ottawa, ON, Canada, 2007.
- [22] X. Gao et al., "Modeling timing variations in digital logic circuits due to electrical fast transients," in *Proc. IEEE Int. Symp. Electromagn. Compat.*, Aug. 2013, pp. 484–488.
- [23] ReliaWiki by HBM Prencscia, I.P.L Relationship, Feb. 8, 2017
- [24] G. Massobrio and P. Antognetti, "Semiconductor Device Modeling with SPICE," McGraw-Hill, 1st ed., Dec. 1998
- [25] WolframAlpha, Series Calculator, 2022 Wolfram Alpha LLC.
- [26] PTM, Nanoscale Integration and Modeling (NIMO) Group, 2011.

Advanced Measurement and Simulation Approach for DDR5 On-chip SI/PI with the Probing Package

WonSuk Choi, SangKeun Kwak, Jaeseok Park, Jiyoung Do, Byeongseon Yun, Yoo-jeong Kwon, Dongyeop Kim, Kyudong Lee, Tae young Kim, Wonyoung Kim, Kyoungsun Kim, Sung Joo Park, Jeonghyeon Cho and Hoyoung Song
Samsung Electronics, Hwaseong, South Korea
 E-mail: wonsuk10.choi@samsung.com

Abstract—As the operation of server system is accelerated, the importance of signal integrity (SI) and power integrity (PI) measurement methodology and modeling of dual in line memory module (DIMM) products is increasing. In this paper, we introduce the advanced methodology for on-chip SI/PI measurement and DRAM signal recovery with the DDR5 probing package development. Comparing with the conventional interposer, a new method of probing package has proved to be advantageous for high-speed signal measurement and signal recovery, and it can also be used as a useful tool for DRAM on-chip SI measurement and signal prediction in post DDR5 speed (beyond 6.4 Gbps).

Keywords—DDR5, SI, PI, Interposer, Probing package, Measurement

I. INTRODUCTION

Data centers and server systems are applying DDR5 as the main memory for high-speed and large-capacity data processing, including autonomous driving and meta-verse following 5G communication and artificial intelligence. For insuring a stable operation performance of server system, DRAM operation characteristics need to be predicted based on the accurate modeling from the beginning of DRAM development, and SI/PI measurement of DRAM on-chip is important as a base technology for this purpose.

In this paper, we present the advanced method for recovering and predicting on-chip signals of DDR5 by correlating SI/PI measurement and simulation using the on-chip measurement technology through the top side of a commercial DRAM package.

II. PROPOAL OF THE DDR5 PROBING PACKAGE

In general, SI of lower speed DRAM could be measured using the interposer PCB, however, a new method of minimizing the SI measurement loss has been required for higher speed over 3.2 Gbps (shortest distance from the DRAM pad). Fig. 1 is the concept of probing package, which forms probing pads in the empty space outside a silicon chip in a commercial DRAM and penetrates epoxy mold compound (EMC) of the package with laser drill applying the through mold via (TMV) technique to measure signals through the holes from the package top [1-3]. This idea is creative in that TMV which used for connecting packages vertically at first, however, used for observing signals on the package.

Fig. 2 shows the probing package structure and specification applied to DDR5. A total of 15 probing pads were placed on the package edge without changing the existing commercial package size. It is a structure capable of measuring DQ, CA, and differential signals DQS, CK, and VDD/VDDQ/VPP. The diameter of the probing hole is 0.6 mm, the width between the holes is 0.83 to 1.0 mm, and the depth of the hole is 0.37 mm.

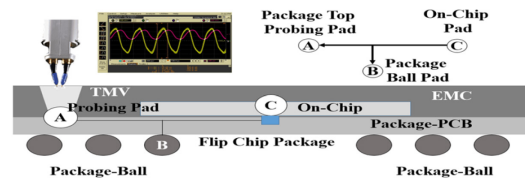


Fig. 1. Concept of the probing package

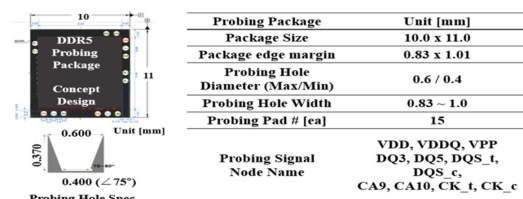


Fig. 2. Structure and specification of DDR5 probing package

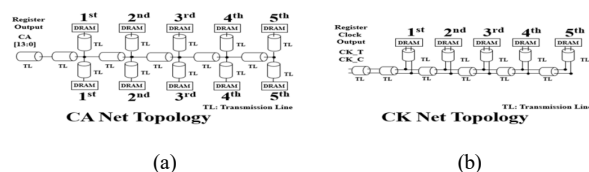


Fig. 3. DDR5 2Rank x4 RDIMM topologies (a) CA (b) CK

When attaching the probing package to a registered dual in-line memory module (RDIMM) we tested, DQ and power have no restrictions in probing point of view, however, CA and CK are more vulnerable to operation error due to signal distortion caused by overlapping of additional traces embedded in the probing package with multiple implementation on the DIMM because they have daisy chain net topologies defined by JEDEC [4] as shown in Fig 3.

To verify the effect of SI degradation of CA and CK, we compared three cases by the number of the probing packages on DIMM with simulation. Fig. 4 is the result of transient simulations of CA and CK pair probed at the last (5th) DRAM pad according to how many probing packages are attached in the daisy chain topology. First, when a single probing package attached to the last (5th) DRAM only, CA eye height (EH) and eye width (EW) are reduced by 12% and 3% respectively compared with the normal package (without the probing package). On the contrary, they are reduced 48% and 15% respectively by replacing all normal packages with probing packages.

In particular, CK signals have attenuated more severely, making them difficult to work properly. To see the SI reduction of CK quantitatively, we performed AC gain simulation for the same three cases. As shown in Fig 5, AC gain of CK_t has fallen by 13% for one probing package, and 81% for all probing packages compared with normal package

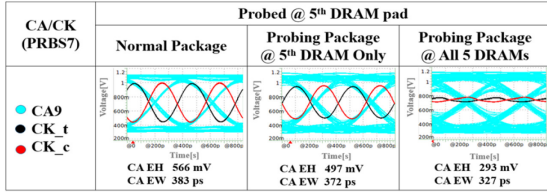


Fig. 4. SI simulation of CA and CK pair by the number of probing packages in the daisy chain topology

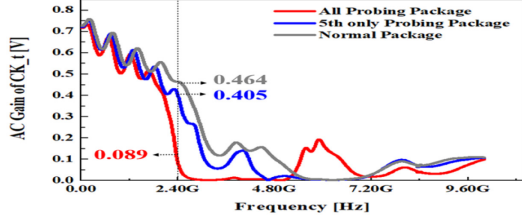


Fig. 5. AC gain of CK_t of three comparative cases

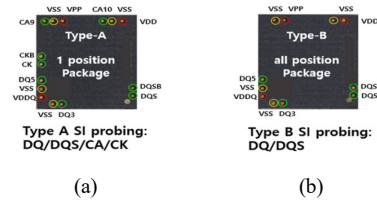


Fig. 6. DDR5 Probing package type (a) Type-A (b) Type-B

at 2.4 GHz which is the fundamental frequency of 4.8 Gbps CK operation.

To overcome the CA and CK SI limitation by the DIMM topologies, the probing package was developed with two types: Type-A and B. Type-A is targeting to measure at the single specific location for all probing signals; DQ, DQS pair, CA, CK pair and Type-B is specialized to measure DQ and DQS pair without restriction of location. Power measurement is common feature for both types as shown in Fig. 6.

III. MEASUREMENT AND SIMULATION

For measurement of DQ SI using the probing package, the automated test equipment (ATE) environment was equipped as following: ATE, test board, 2Rank x4 RDIMM with probing package, oscilloscope and probe. Fig. 7 is the picture of probing through the DDR5 probing package in 2Rx4 RDIMM on the ATE (T5511) test board.

DRAM was applied with I/O input vector of pseudo random bit sequence (PRBS) which is 127-bit (2^7-1) long for write operation at speed of 4.8 Gbps and 5.6 Gbps. Current registering clock driver (RCD) implemented on RDIMM supports 5.6 Gbps in maximum. That's why we captured real measurement eye diagram up to 5.6 Gbps.

Fig. 8(a) is a channel structure, measurement, and simulation waveforms for 1-byte DQ Write operation of the DDR5 2Rank x4 RDIMM in an ATE environment. The simulation models conducted with Hspice are the scattering parameters (S-parameters) extracted from the design files for each block and the operating speed is 4.8, 5.6 Gbps. As a result of comparing EH and EW based on DQ measurement waveforms, the measurement and simulation consistency of probing package is 90% EH and 102-115% EW in Fig. 8(b).

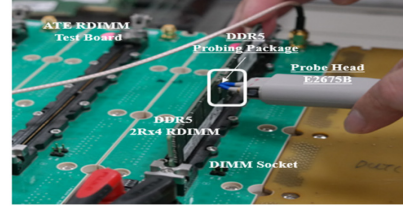


Fig. 7. Measurement through DDR5 probing package in ATE system (T5511)

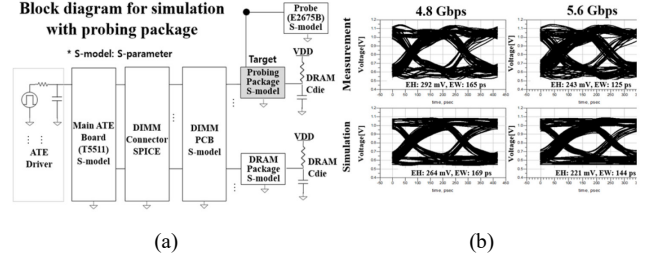


Fig. 8. Measurement and simulation of DQ SI with the probing package (Type-B) (a) Simulation block diagram (b) Comparison of measurement and simulation

IV. DRAM SI RECOVERY FROM PROBING SIGNAL USING THREE PORT DE-EMBEDDING

We describe a conventional method of utilizing the 3 port de-embedding process with interposer [5]. By restoring the original DRAM pad signals, it is shown to be effective to use a probing package rather than the interposer in terms of the accuracy of data recovery. In Fig. 9, the transfer function between the input and the interposer measurement signal is defined as the transfer function A (a), and the relationship between the input without the interposer and the output to be restored is defined as the transfer function B (b). Two transfer functions can be combined to restore DRAM internal signals from the measurement signal V_{probe} in equation (1) to (3).

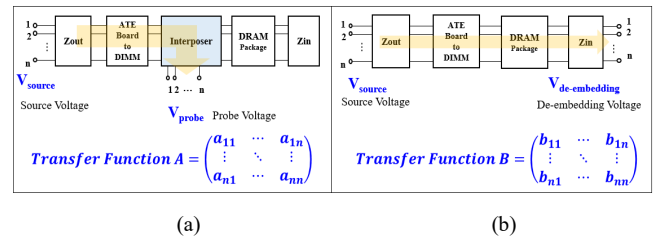


Fig. 9. Definition of transfer function (a) With interposer PCB (b) Without interposer PCB

$$V_{probe} = AV_{source} \quad (1)$$

$$V_{source} = A^{-1}V_{probe} \quad (2)$$

$$V_{de-embedding} = BV_{source} = BA^{-1}V_{probe} \quad (3)$$

Since the probing package is a tool developed for on-chip SI/PI measurement accuracy, we verified the relative improvement of measurement effectiveness by comparing it with the existing method of using an interposer PCB. Fig. 10 shows the distance between the measurement point (A), the DIMM connection point (B), and the DRAM pad (C) compared to the interposer PCB. The measurement and

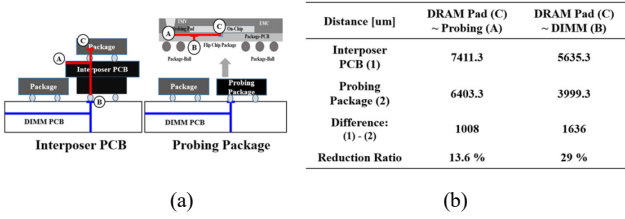


Fig. 10. Comparison of measurement distance (a) Structure (b) Physical distance of interposer PCB and probing package

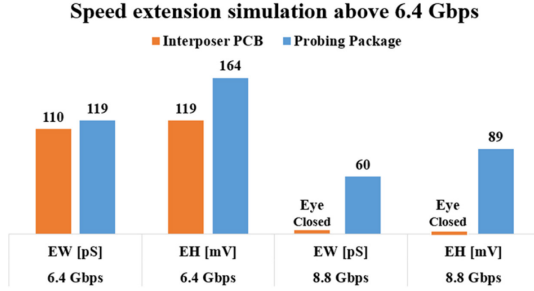


Fig. 11. Comparison of probing SI degradation due to speed-up

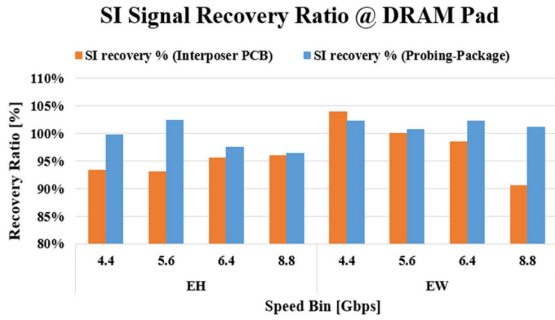


Fig. 12. Comparison of on-chip SI recovery ratio by probing methods

connection distance of the probing package was designed to be 14% and 29% shorter than those of the interposer PCB.

The SI degradation at the measurement point due to the measurement and connection distance is evident as the operating speed increases. In Fig 11, it was confirmed by simulation that eye measurement was not possible with an interposer at 8.8 Gbps, while DQ eye measurement was still possible with the shorter length of the probing package.

Fig. 12 shows simulation results of the signal recovery rate for each measurement tool up to 8.8 Gbps by applying the 3-port de-embedding method. The signal recovery error rate of the interposer is around 10%, while that of the probing package is less than 5%. That is, the probing package is a measurement method capable of more accurate DRAM signal recovery than the existing method of using interposer through the shorter on-chip measurement distance.

V. SI/PI ANALYSIS USING PROBING PACKAGE

In DDR5 DRAM, usage of on die termination (ODT) scheme has extended to CA and CK. Therefore, CA and CK termination resistors on DIMM edge, which were conventional probing positions until DDR4, have not been remained any more in DDR5. Using probing package, on-chip SI measurement of CA and CK is possible. Fig. 13 shows the

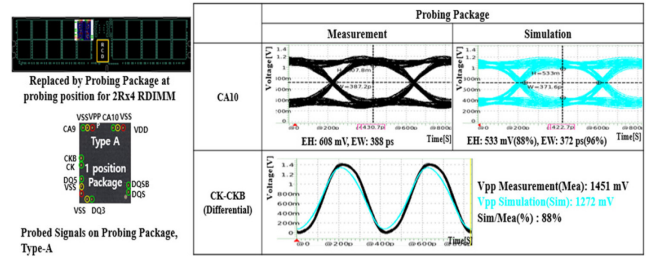


Fig. 13. Measurement and simulation of CA and CK SI (Type-A)

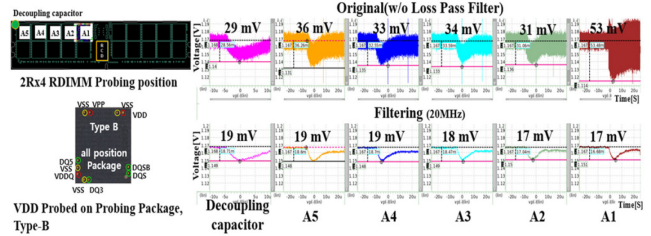


Fig. 14. Measurement of VDD power (Type-B)

measurement and simulation waveforms of CA10 and CK pair using Type A, showing consistency of about 90%.

Fig. 14 shows the result of measuring IDD5 PI noise by DRAM location. When comparing VDD noise of probing packages with decoupling capacitor, we see there is a difference in slope and size of VDD drop, which means that on-chip power can be more accurately measured with probing package.

VI. CONCLUSION

In this paper, we developed the probing package as DDR5 SI/PI signal measurement solution and introduced cases of using it for product analysis. With shorter measurement distance compared to the existing method of using interposer PCB, the DRAM signal recovery error was reduced to less than 5% in the operation of 4.8 Gbps or higher, and DQ/CA/CK/PI measurement and simulation consistency has also been verified. Using the probing package, post DDR5 speed can be tested with less measurement error than the conventional probing method. Another benefit of probing package is that double sided DIMM product can be probed. Before development of probing package, only single sided memory module products can be probed through backside of DIMM PCB without implementation of DRAM or using interposer was the only possible way for probing DRAM signals. Probing package is the advanced SI/PI probing solution applied to high speed beyond DDR5 (above 6.4 Gbps).

REFERENCES

- [1] W. Choi, S. Kwak *et al.*, "The SI/PI Modeling and Measurement of Memory System by probing on top of DRAM Package," *DesignCon*, 2020.
- [2] Jinseong Kim, Kiwook Lee *et al.*, "Application of through Mold via (TMV) as PoP base package," *ECTC*, 2008.
- [3] Yoshida, Akito *et al.*, "A Study on Package Stacking Process for Package-on-Package (PoP)," *ECTC*, 2006.
- [4] JEDEC, "DDR5 RDIMM Standard Annex A, JESD305-R8-RCA," 2022.
- [5] K. Technologies, "De-embedding Techniques in Advanced Design System," Training Materials. Available from <https://www.keysight.com/us/en/assets/9018-02221/training-materials/9018-02221.pdf>

Codimensional Optimization of Differential Via Padstacks

Jiwoong Jeon, Shivani Joshi,
Daniel de Araujo
Electronic Board Systems
Siemens EDA
jiwoong_jeon, shivani_joshi,
daniel_dearaujo@mentor.com

Bhyrav Mutnury
Infrastructure Solutions Group
Dell EMC
Round Rock, TX
bhyrav.mutnury@dell.com

Abstract—In printed circuit boards, vias are needed when signals make layer transitions. Typically, via transitions are defined through padstacks. These padstacks provide geometrical manufacturing characteristics of the transitions such as via drill size, pad size, antipad size and thermal tolerance. Differential via design consists of padstack definition and relative positioning to other padstacks such as signal and reference as well as entry and exit trace parameters. At high speeds, vias cause reflections if they are not well designed. When stack-ups have a lot of signal layers, optimizing via padstack for each signal transition is going to be cumbersome. Using a padstack optimized for one signal transition on other transitions may degrade signal performance. In this work, a simultaneous codimensional optimization to differential via padstack designs is proposed to optimize a single padstack definition so that it can handle multiple entry/exit layer transitions.

Keywords—*differential, via, printed circuit board, optimization, codimensional, padstack*

INTRODUCTION

As the signal speeds continue to increase, any small imperfection in the channel can have adverse impact on the signal quality. All high-speed printed circuit boards (PCBs) have via transitions in them. Vias are needed to help signals make transition from one layer to another. A well optimized via for impedance and crosstalk will result in good signal integrity (SI). Most midplanes, backplanes and channel cards can have many signal layers in them as shown in Figure 1a. Optimizing a signal via for each layer transition will be time consuming and cumbersome. Vias are typically defined through padstacks. These padstacks provide geometrical manufacturing characteristics of the transitions such as via drill size, pad size, antipad size and thermal tolerance. Differential via design consists of padstack definition and relative positioning to other padstacks such as signal and reference as well as entry and exit trace parameters. Optimal via dimensions for one signal transition might result in sub-optimal results for another signal transition.

Typically, a differential via can be optimized by tuning parameters such as via drill size, antipad size, reference via spacing and via pitch as shown in Figure 1b. The optimization is typically done for one signal transition at a time. In this paper, the codimensionality aspect of vias is taken into account to come up with one padstack that is optimized for multiple signal layer transitions.

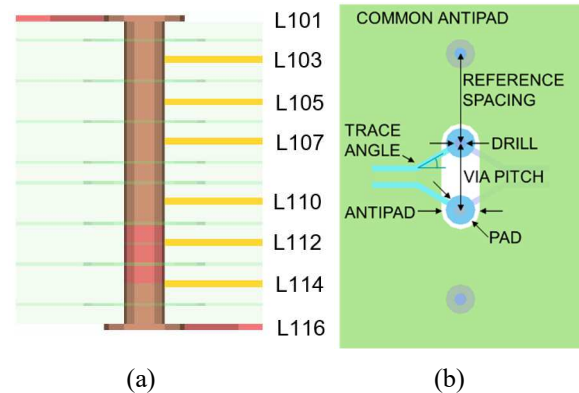


Figure 1. 16 Layer via PCB & Parameters

Previous work in via optimization include [1] using a domain decomposition approach, [2] focusing on return via location and mode conversion reduction, [3] investigating via impedance matching, and [4] and [5] where a large number of combinations using a full-factorial search of the via parameters. While the optimization is successful in these studies, the limitation is that the optimizations are done for one signal transition at a time. This work focuses on optimizing the padstack which might be used in a plurality of via layer transitions.

This paper is organized as below: Section II discusses single, multi, and codimensional optimization. Section III shows the results from individual and codimensional optimization and Section IV concludes the paper.

OPTIMIZATION

Optimization Objective: Single, Multi, Codimensional

In order to optimize a problem, a figure of merit (FOM) needs to be defined so that outcomes can be compared and improvement quantified. Examples of FOM include single objective (e.g. differential return loss), multi objective (e.g. differential insertion loss, differential return loss and differential crosstalk). With a codimensional objective the FOM is evaluated across multiple dimensions (e.g. differential return loss for multiple layer transition configurations).

Each FOM characterizes specific aspects of the design. Differential Return loss is a good quantifier for matching (as is

Differential TDR). Differential insertion loss is well suited to quantify the channel losses, whichever mechanisms they may be. Mode conversion is a good metric for quantifying channel imbalance/asymmetry and crosstalk is typically used to measure isolation.

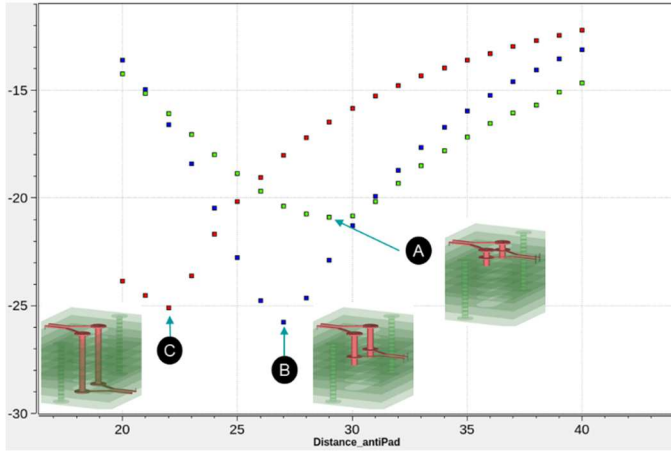


Figure 2. Max Differential Return Loss vs. Antipad size for L103(A), L107(B), and L116(C) trace exit configurations

To illustrate the challenge, a single parameter, antipad diameter was varied from 20 mils to 40 mils and a plot of the differential return loss for configurations is shown in Figure 2. The entry layer was L101, and the exit layers were L103(A), L107 (B), and L116(C).

TABLE 1. DIFFERENTIAL RETURN LOSS SPECIFIC LAYER OPTIMIZATION

Differential Return Loss Results				
	Optimization Configuration	Antipad (mils)	Optimized (dB)	Worst (dB)
A	Stripline, L103	30	-20.8	-15.8
B	Stripline, L107	27	-25.7	-18.0
C	Microstrip, L116	22	-25.1	-16.1
	All Cases	26		-19.1

When padstack optimization focuses on a specific layer transition such as cases (A) and (C), the differences between the optimized and the worst transition performance was 5.0 and 9.0 dB, respectively. The best value, 26 mils, is closer to case (B) optimal of 27 in this simple example, but this may not always be the case for other stack-ups and parameter values (Table 1).

Parameter & handling (local vs. global)

To optimize a differential via padstack which may be used for different layer transitions on a PCB, the parameters are split into two sets: Global vs. Local codimensional parameters. The local parameters define the layer transitions and are fixed for each layer transition case. The global parameters affect all simulations. Each simulation for a specific transition will generate a result for which the worst case for the global parameter set will be selected to be used for the optimization FOM as shown in Figure 3.

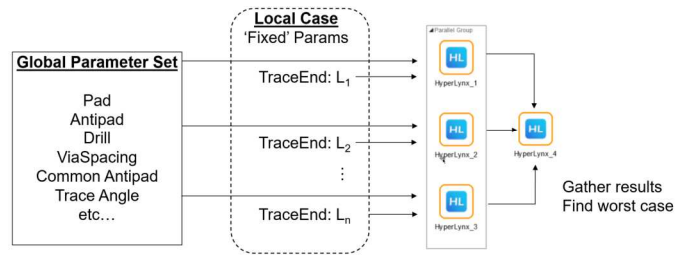


Figure 3. Codimensional parameter flow example

Layer Name	Layer Type	Cond. Material	Diell. Material	Thickness (mil)
1 ___s00	DIELECTRIC	COPPER	SM	0.5
2 L101	CONDUCTOR	COPPER	SM	1.5
3 ___d01	DIELECTRIC	COPPER	FR4	4.3
4 L102	PLANE	COPPER	FR4	0.7
5 ___d02	DIELECTRIC	COPPER	FR4	4.3
6 L103	CONDUCTOR	COPPER	FR4	0.7
7 ___d03	DIELECTRIC	COPPER	FR4	4.3
8 L104	PLANE	COPPER	FR4	0.7
9 ___d04	DIELECTRIC	COPPER	FR4	4.3
10 L105	CONDUCTOR	COPPER	FR4	0.7
11 ___d05	DIELECTRIC	COPPER	FR4	4.3
12 L106	PLANE	COPPER	FR4	0.7
13 ___d06	DIELECTRIC	COPPER	FR4	4.3
14 L107	CONDUCTOR	COPPER	FR4	0.7
15 ___d07	DIELECTRIC	COPPER	FR4	4.3
16 L108	PLANE	COPPER	FR4	0.7
17 ___core08	DIELECTRIC	COPPER	FR4	4.3
18 L109	PLANE	COPPER	FR4	0.7
19 ___d09	DIELECTRIC	COPPER	FR4	4.3
20 L110	CONDUCTOR	COPPER	FR4	0.7
21 ___d10	DIELECTRIC	COPPER	FR4	4.3
22 L111	PLANE	COPPER	FR4	0.7
23 ___d11	DIELECTRIC	COPPER	FR4	4.3
24 L112	CONDUCTOR	COPPER	FR4	0.7
25 ___d12	DIELECTRIC	COPPER	FR4	4.3
26 L113	PLANE	COPPER	FR4	0.7
27 ___d13	DIELECTRIC	COPPER	FR4	4.3
28 L114	CONDUCTOR	COPPER	FR4	0.7
29 ___d14	DIELECTRIC	COPPER	FR4	4.3
30 L115	PLANE	COPPER	FR4	0.7
31 ___d15	DIELECTRIC	COPPER	FR4	4.3
32 L116	CONDUCTOR	COPPER	SM	1.5
33 ___s16	DIELECTRIC	COPPER	SM	0.5

Figure 4. 16 Layer PCB stack-up

The stack-up used for this experiment is shown in Figure 4 with the dielectric named FR4 (Dk=3.3, TanD=0.02 @1GHz) and SM (Dk=3.7, TanD=0.03@1GHz) and copper conductivity at 5.8e7 S/m). Via Backdrilling as assumed for all the stripline transitions with via stub of 10 mils for 60 iterations and no stub for the 140 iterations runs.

The parameters used in this optimization are summarized in Table 2.

TABLE 2. OPTIMIZATION PARAMETER LIST

Parameter	Optimization Variables		
	Min	Baseline	Max
Common Antipad (True/False)	False	True	True
Drill diameter (mils) (8, 10, 12,14)	8	8	14
Pad diameter (mils)	15	20	30
Antipad diameter (mils)	16	30	60
Via pitch (mils)	35	45	75
Trace Angle (degrees)	30	30	90

RESULTS

Two sets of optimizations using SHERPA (Simultaneous Hybrid Exploration that is Robust, Progressive and Adaptive) algorithm found in [6] were run with a FOM of impedance deviation (Z_{0Dev}) for $Z_0 = 100\text{ohms}$ defined in (1):

$$Z_{0Dev} = \text{Max}(\text{Abs}(Z_{TDR} - Z_0)) \quad (1)$$

Individual Optimization

With individual optimizations, one optimization was run for each of the 7 transition layers specifically to minimize that layer transition’s impedance deviation and is listed in Table 3. While the optimization attempts to reduce the transition’s FOM for the given layer transition, the worst-case impedance deviation for the other transitions is listed under ‘worst-case’ in Table 3. It can be estimated from Table 3 that the average optimized case Z_{0Dev} is 5.3 ohms, the average worst case impedance deviation is 17.6 ohms.

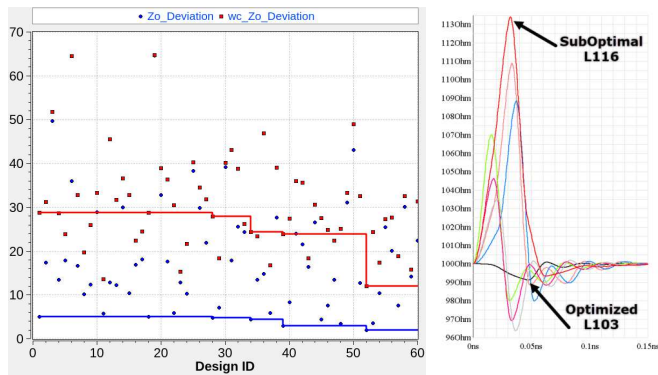


Figure 5. Differential via optimization for Top microstrip (L101) to top stripline (L103) transition (left) and TDR results for all transitions for that parameter set (right)

In Figure 5 we can see the convergence of the L101→L103 transition in blue with the worst-case impedance deviation effects in red. The TDR plot shows the optimized differential TDR of 0.9 ohms while that padstack used on other layers yielded a worst case of 13.4 ohms for the L101→L116 transition.

Codimensional Optimization

While the individual optimization in the previous section focused on a specific transition ignoring the effects on the other transitions, the codimensional optimization solves all transitions in parallel and uses the worst-case result to guide the optimization to ensure the performance across all desired configurations.

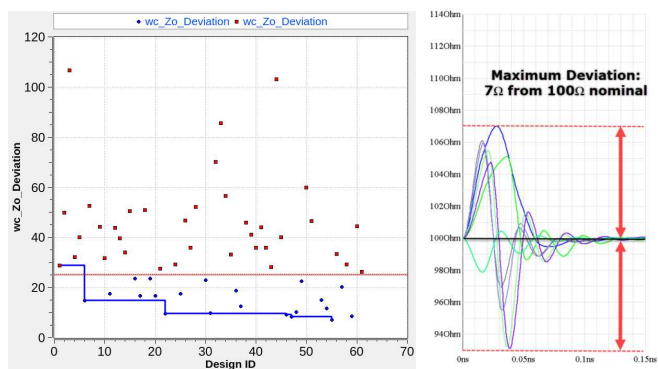


Figure 6. Design ID vs. worst case Impedance deviation (left) and Differential TDR for all transitions (right)

TABLE 3. OPTIMIZATION RESULTS

	Impedance Deviation Results				
	Optimization Configuration	Optim 60	Worst 60	Optim 140	Worst 140
i1	Stripline, L103	0.9	13.4	0.9	10.4
i2	Stripline, L105	5.1	9.7	3.5	16.6
i3	Stripline, L107	7.8	29.0	2.7	7.0
i4	Stripline, L110	9.9	10.6	2.7	3.1
i5	Stripline, L112	7.9	22.4	3.9	6.1
i6	Stripline, L114	4.6	12.4	3.9	9.2
i7	Microstrip, L116	1.2	25.0	1.5	21.4
ia	Average of Individual	5.3	17.5	2.7	10.5
c1	Codimensional, All layers		7.0		3.9

With a complex, non-linear search space, it is not guaranteed that the global minimum will always be found given the simulation budget of 60 simulations for each optimization as we can see in the individual optimizations for layers L107, L110, and L112 since a better padstack configuration was found in the codimensional optimization. When the simulation budget was increased to 140 simulations per optimization, all the individual optimized values were at or lower than the codimensional worst case.

With the codimensional optimization, a worst-case impedance deviation of 7.0 ohms across the 7 configurations was achieved for 60 iterations and 3.9 ohms for 140 iterations.

CONCLUSION

In this work, we applied codimensional optimization to differential via padstack design to optimize a single padstack definition to multiple entry/exit layer combinations and achieved 7 ohms impedance deviation across 7 different layer transition configurations. With this methodology a single padstack design can be used for multiple transitions which facilitates the padstack definitions and layout design and verification. This becomes more important as the number of layers in the PCB increases as seen with larger server boards and telco backplanes.

REFERENCES

1. Carmona-Cruz, K. Scharff, J. Cedeno-Chaves, H.-D. Bruns, R. Rimolo-Donadio, and C. Schuster, “Via transition optimization using a domain decomposition approach,” 2019 IEEE 23rd Workshop on Signal and Power Integrity (SPI), 2019.
2. M. Cracraft, S. Connor, and B. Archambeault, “Optimizing return via locations to reduce mode conversion in connector pin fields,” 2013 IEEE International Symposium on Electromagnetic Compatibility, 2013.
3. Vardapetyan and C.-J. Ong, “Via design optimization for high speed differential interconnects on circuit boards,” 2020 IEEE 29th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS), 2020.
4. 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), 2016.
5. Hardock, R. Rimolo-Donadio, S. Muller, Y. H.Kwark and C. Schuster, “Efficient, physics-based via modeling : return path, impedance and stub effect control,” *IEEE EMC Magazine*, vol 3., pp. 76-84, 2014
6. <https://www.plm.automation.siemens.com/global/en/products/simcenter/simcenter-heeds.html>

Distributed PDN Modeling Approach for Accurate Jitter Estimation in High-Speed NAND Flash Memory

Sayed Mobin
Western Digital Corporation
Milpitas, CA, USA

Pranav Balachander
Western Digital Corporation
Milpitas, CA, USA

Asha Sharma
Western Digital Corporation
Bengaluru, India

Venkatesh Ramachandra
Western Digital Corporation
Milpitas, CA

Abstract— Due to aggressive storage capacity demands, multiple NAND Flash die are often stacked in a highly integrated, complex package system. As data-rate increases, bit time (UI) is shrinking, and accurate measurement of the data valid window and jitter become very important. Power distribution network (PDN) noise affects the overall system timing. The conventional way of PDN modeling approach in NAND Flash memory System level analysis, cannot accurately predict the system level jitter and deviates from the actual product level performance.

In this paper, an accurate method for PDN-induced jitter analysis in NAND Flash system-level operation is described. Simulated PDN-induced jitter results are validated through characterization system and product level measurement jitter.

Keywords— Distributed PDN Modeling, NAND Flash SI Simulation, PDN impact on I/O jitter

I. INTRODUCTION

In the last decade, NAND Flash Memory use has evolved from simply the component in typical Flash storage devices (USB Flash drives), to the primary data storage system in mobile devices (embedded applications); in solid state drives (SSDs); and in traditional computing storage. The operating speed requirements for NAND Flash Memory are increasing exponentially, accompanied by demand for significant reductions in power requirements [1]. With increasing speed requirements, the overall system-timing margin is becoming steadily tighter. PDN induced jitter is a key parameter that must be controlled and maintained at a minimal level to enable any high-speed NAND Flash system operation.

II. CONVENTIONAL PDN MODELING APPROACH

Conventional ways of NAND I/O model creation in the industry includes either IBIS modeling or composite transistor level model of the output buffer. Generally, the internal power bus connections between I/O pads are not included in the NAND I/O model. In addition, the system PDN is extracted as a lumped network. All the VCCQ pins are shorted together and extracted as one lumped port. Because of this lumped modeling approach, all the I/O pads of the NAND Flash receive the same power. The lumped system PDN (VCCQ rail) modeling approach is shown in Figure 1. At lower speed operation, the lumped approach was sufficient, the bit time was large enough to account for the timing delta due to the extra amount jitter not being present due to the lumped approach.

I/O Model, without power aware feature of NAND die, cannot accurately estimate the NAND I/O timing penalties (t_{QHS}

& t_{DQSQ}), Clock duty cycle jitter (t_{QSH} & t_{QSL}) and per-pin data valid window (t_{DVVP}).

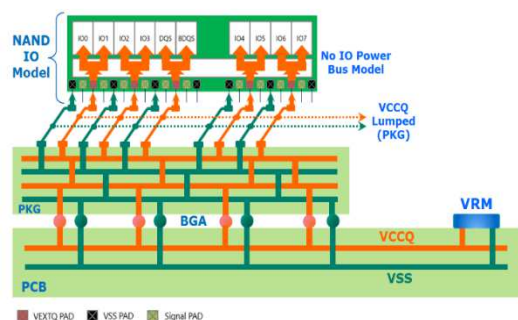


Fig. 1. NAND system PDN modeling representation (lumped approach)

With higher operating-speed, the timing window shrinks, making it extremely important to accurately model all the jitter components affecting the overall timing budget calculation.

III. DISTRIBUTED PDN MODELING APPROACH

At higher speeds, a power-aware NAND I/O model is necessary to simulate I/O jitter properly. To improve the NAND I/O modeling approach, the internal power bus (VCCQ rail) connectivity between the I/O pads is included in the NAND I/O netlist. System-level PDN modeling is also updated from the lumped approach to a per-pin distributed modeling approach. As shown in Figure 2, each VCCQ pin is isolated in the system PDN model creation, providing isolated power to individual I/O pad. I/O jitter due to the per-pin VCCQ PDN difference is thus modeled effectively [2].

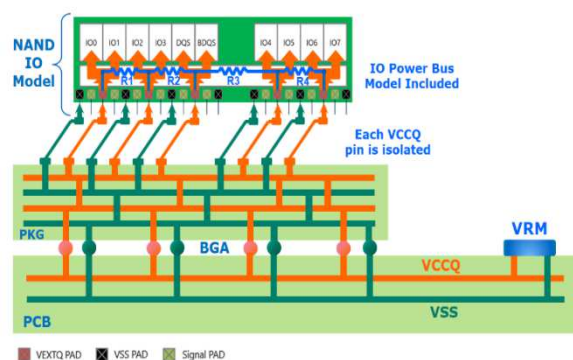


Fig. 2. Distributed system PDN modeling approach

Inclusion of the NAND internal power-bus model adds additional complexity in the NAND I/O model creation flow. In addition, the transient simulation time increases by approximately 20-25%.

IV. PDN INDUCED JITTER ANALYSIS

To study the PDN induced jitter impact on the NAND Flash Memory system, SI analysis was performed on a typical NAND Flash characterization environment. NAND die is in packaged form, and characterization is performed using semiconductor automated test equipment, or ATE.

Impact of the system PDN is studied on the following NAND AC timing parameters

- a) Clock Duty Cycle Jitter (t_{QSH} & t_{QSL})
- b) Per-pin I/O Jitter (t_{QHS} & t_{DQSQ})
- c) Per-pin data valid window (t_{DVWP})

SI simulation was performed on a WDC NAND Flash device at 800 Mbps, and worst-case operating conditions.

A. Impact of PDN on Clock Duty Cycle (t_{QSH} & t_{QSL}) Jitter

Clock duty cycle jitter is measured as the minimum DQS/BDQS high pulse and low pulse widths. The t_{QSH} & t_{QSL} duty cycle is measured at the cross point of the DQS and BDQS signals. The simulated t_{QSH} & t_{QSL} result is shown in Figure 3. The results show that without the system PDN, there is no jitter in the t_{QSH} & t_{QSL} parameters, indicating that the system PDN contributes a significant amount of jitter in the clock duty cycle [3].

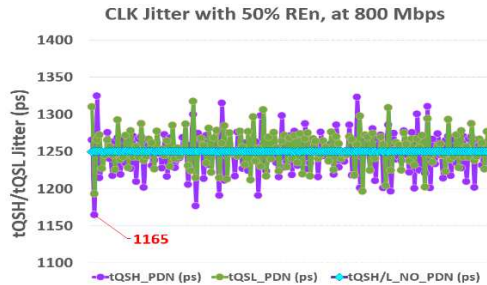


Fig. 3. Simulated t_{QSH} & t_{QSL} data at the worst-case corner

From SI simulation, the worst-case peak-to-peak jitter is 160ps (6.4% of t_{RC}) and absolute jitter is 85ps (3.4% of t_{RC}).

B. Impact of PDN on Per-Pin I/O Jitter (t_{QHS} & t_{DQSQ})

To demonstrate the impact of the distributed modeling approach with power-aware NAND I/O model, SI simulations are performed in the following three scenarios:

- a) Single I/O, No PDN, No power-aware NAND I/O
- b) Lumped PDN approach, No power-aware NAND I/O
- c) Distributed per-pin PDN, with power-aware NAND

The I/O with minimum eye window opening (worst-case) defines the maximum operating speed of the interface, in per-pin eye window training, and was used for simulation and

measurement data analysis. Simulated, the worst-case I/O jitter at 800 Mbps is shown in Figure 4.

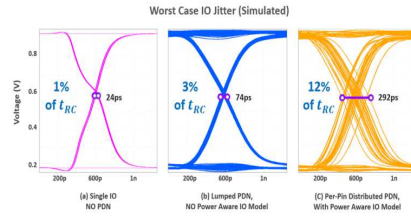


Fig. 4. Simulated worst-case I/O jitter at 800 Mbps (distributed PDN model)

Based on the simulation, memory-only I/O jitter without PDN is 24ps (1% of t_{RC}). With lumped PDN and a no-power-aware I/O model, worst-case I/O jitter is 74ps (3% of t_{RC}). With distributed PDN and a power-aware I/O model, worst-case, jitter increases to 292ps (12% of t_{RC}). The result clearly shows the importance of distributed per-pin PDN and a power-aware I/O Model.

C. Impact of PDN on Per-Pin Data Valid Window (t_{DVWP})

Because the system PDN has a significant impact on clock duty cycle and I/O jitter, the PDN would eventually influence the per-pin data valid window (t_{DVWP}). Simulated worst-case, per-pin data valid window for the three scenarios are shown in Figure 5. Based on the simulation, the per-pin data valid window of bare memory is 1220ps (49% of t_{RC}). Per-pin data valid window shrinks to 1150ps (46% of t_{RC}) using the lumped PDN approach. Finally, worst-case, per-pin data valid window is reduced to 934ps (37% of t_{RC}) with a distributed and power-aware NAND I/O model.

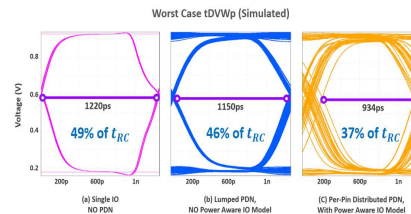


Fig. 5. Simulated worst-case per-pin data valid window jitter at 800 Mbps

V. CHARACTERIZATION SYSTEM MEASURED RESULT

To understand the impact of PDN on NAND SI system-level performance, critical AC timing parameters were measured in the characterization system environment. Measurements were performed like the simulation environment, to establish the correlation between measurement and simulation data.

A. Measured Clock Duty Cycle (t_{QSH} & t_{QSL}) Jitter

The characterized value of clock duty cycle jitter is shown in Figure 6. As depicted in the measured data, clock duty cycle peak-to-peak jitter is around 200ps (8% of t_{RC}). Absolute jitter is around 100ps (4% of t_{RC}). Measurement data shows that the PDN-induced jitter in the clock duty cycle is 4%. The average value is the duty cycle of the input RE signal (50% of t_{RC})

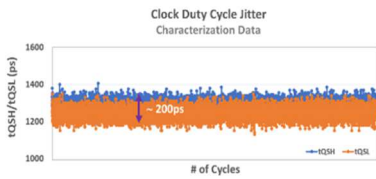


Fig. 6. Clock duty cycle jitter measurement result

B. Measured Per-Pin I/O Jitter (t_{QHS} & t_{DQSQ})

Worst-case, per-pin I/O jitter characterized data is shown in Figure 7. Characterization data shows that the worst-case, per-pin I/O jitter measured at the VREF level is 305ps (12.2% of t_{RC}).

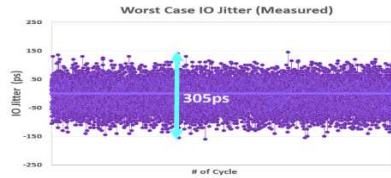


Fig. 7. Per-pin I/O jitter measurement result

VI. MEASUREMENT AND SIMULATION CORRELATION

To establish the validity and accuracy of the distributed PDN modeling approach, simulated results are correlated with the actual measurement.

A. Clock Duty Cycle Jitter Correlation

The clock duty cycle jitter correlation data is shown in Figure 8. The difference between the simulated and measured peak-to-peak clock duty cycle jitter is 1.6% of t_{RC} – 3% of t_{RC} . The absolute clock duty cycle jitter delta between the simulation and measurement is approximately 0.6% of t_{RC} – 1.3% of t_{RC} .

The delta between the simulation and the measurement data is due to the absence of NAND internal logic stages in the NAND I/O model.

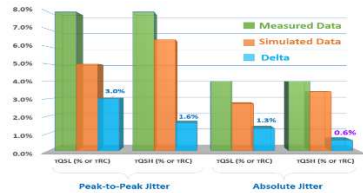


Fig. 8. Clock duty cycle jitter simulation and measurement correlation

Acceptable correlation is established between the simulated and the measured clock duty cycle jitter. The maximum absolute clock duty cycle jitter delta is only 0.6% of t_{RC} .

B. Per-Pin I/O Jitter Correlation

Per-pin I/O jitter correlation data is shown in Figure 9. The lumped PDN approach, predicts reduced jitter compared to distributed approach. The simulation and measurement worst-case I/O jitter delta is within 0.5% of t_{RC} , indicating the validity of the distributed PDN modeling approach. A very good correlation between the simulated and measured per-pin I/O jitter is established.

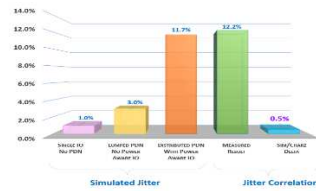


Fig. 9. Worst-case, Per-pin I/O simulation and measurement correlation

C. Per-Pin Data Valid Window Correlation

At the system level, the maximum operating speed of the NAND interface is determined by the minimum per-pin data valid window opening. The worst-case, per-pin data valid window correlation data is shown in Figure 10. The worst-case data valid window mismatch between the simulation and measurement is 1.2% of t_{RC} .

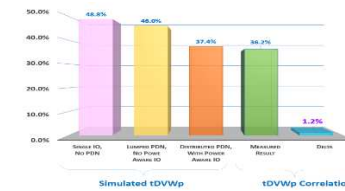


Fig. 10. Worst-case, Per-pin data valid window correlation

VII. PRODUCT LEVEL PERFORMANCE CORRELATION

To demonstrate the validity of the proposed simulation methodology, actual product level system performance was correlated on a WDC embedded multi-die product at nominal operating condition. Product level correlation data is shown in Figure 11. Measurement and simulated delta is 1.4% of t_{RC} .

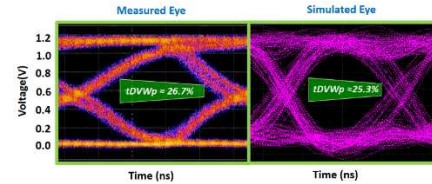


Fig. 11. Product level performance correlation

VIII. CONCLUSION

This paper demonstrates that the system PDN contributes a significant amount of jitter in critical AC timing parameters, which directly affects NAND SI analysis. Lumped approach of PDN modeling cannot predict per-pin I/O jitter accurately. Distributed per-pin System PDN and integrated NAND internal power-bus model I/O netlist is required for the accurate per-pin I/O jitter analysis. Finally, simulation and measurement correlation on critical AC timing parameters is established with reasonable consistency.

REFERENCES

- [1] S. Mobin, B. Raghunathan, A. Katz, "Impact of Read Enable (RE) Signal Duty Cycle Distortion (DCD) in NAND Flash SI Simulation", EPEPS, Oct. 2017
- [2] S. Mobin, V. Ramachandra, P. Balachander, J. Lee, C. Nguyen, A. Sharma, "PDN-Induced Jitter Analysis in High-Speed NAND Flash Memory Interfaces", DesignCon, Jan, 2020
- [3] S. Mobin and P. Balachander, "Understanding NAND AC Timing Parameters and How to Accurately Implement them in SI Simulations," 23rd IEEE SPI, Jun. 2019

Scalable Transformer Network-based Reinforcement Learning Method for PSIJ Optimization in HBM

Hyunwook Park¹⁾, Taemin Shin¹⁾, Seongguk Kim¹⁾, Daehwan Lho¹⁾, Boogyo Sim¹⁾, Jinwook Song²⁾, Kyubong Kong³⁾ and Joungho Kim¹⁾

¹⁾School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea

²⁾Solution Development Team (Memory Business), Samsung Electronics, Hwaseong, South Korea

³⁾SK Hynix, Icheon, South Korea

E-mail: hyunwookpark@kaist.ac.kr

Abstract— In this paper, we first propose a scalable transformer network-based reinforcement learning (RL) method for power supply induced jitter (PSIJ) optimization in high bandwidth memory (HBM). The proposed method can provide an optimal power distribution network (PDN) decoupling capacitor (decap) design to satisfy the target PSIJ with the minimum number of NMOS decaps. For the given number of decaps, the network is trained to maximize the impedance reduction from 10 MHz to 20 GHz compared to the initial PDN. Also, the network has scalability on the number of decap assignments. Therefore, for given any number of decaps, the scalable network can provide minimized PDN impedance profiles by one inference without re-training. Then, by increasing the decap assignments, the network can find out the minimum number to meet the given target PSIJ. For verification, the proposed network is applied to the HBM2 I/O interface. The network successfully provides the optimized decap designs to satisfy the given target PSIJ values.

Keywords—Decoupling capacitor, Power Supply Induced Jitter, Scalability, Transformer network, Reinforcement Learning

I. INTRODUCTION

As the timing budget becomes tighter due to the higher speed requirements in digital systems, PSIJ becomes more challenging in both the serial and parallel I/O interfaces [1]. Especially, the HBM module has 1024 parallel IOs between the graphic processing unit (GPU) and HBM to provide TB/s scale bandwidth as shown in Fig. 1 [2]. Moreover, the data rate has increased from 1 Gbps at gen 1 to 6.4 Gbps at gen 3 [3]. The numerous I/O buffers in the physical layer (PHY) draw the current at the same time which is called simultaneous switching current (SSC). When the broadband SSC spectrum meets the high-peak anti-resonance of the PDN impedance, a large simultaneous switching noise (SSN) occurs [2]. The induced voltage fluctuation causes jitter which severely deteriorates the timing margin at the data output. Therefore, it is important to design robust PDN to lower the impedance in the broadband frequency band.

Decap design is one of the most important processes for PDN optimization. However, it is a combinatorial optimization problem that has high computational complexity and requires high computing cost. Recently, to tackle this problem, deep RL (DRL)-based decap optimization methods have been actively investigated [4]–[6]. The RL is an algorithm to solve the problem defined by Markov decision process (MDP) parameters – state s , action a , reward r , and policy π [7]. In the DRL, the probability distribution function (PDF) π of a given s is approximated by the policy neural network (NN). Thus, DRL-based methods train the policy network to find optimal a for given s in order to maximize r .

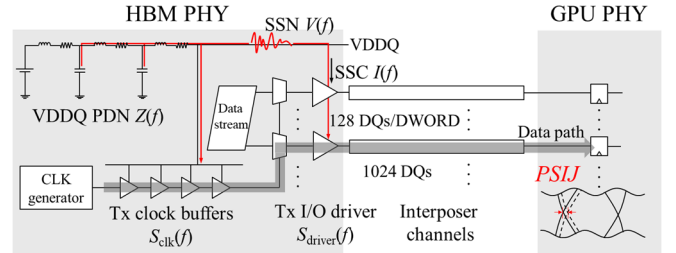


Fig. 1. Schematic of HBM I/O Interface configured of VDDQ PDN, Tx clock buffers, Tx I/O drivers and interposer channels. SSN generated by 1024 switching I/Os occurs PSIJ which deteriorates timing margin.

All the previous works [4]–[6] proposed decap optimization NNs to meet the target impedance with the minimum number of decaps. However, those networks are trained by using one PDN data to satisfy a specific target impedance. In other words, those are only useful to solve the one PDN to satisfy the one target impedance. Whenever the PDN or target specification is varied, they needed to be re-trained. Also, those do not consider the PSIJ for the optimization.

When developing the DRL-based methods for high-speed digital systems, the generality of the policy network is important to reduce the computing cost [8], [9]. From that point of view, [9] proposed a scalable transformer network for PDN optimization. The scalability refers to the property of whether the trained policy network can respond to the scale of the problem without re-training. In this paper, as the extension of our previous work [9], a scalable transformer network-based RL method for PSIJ optimization in HBM is proposed.

II. SCALABLE TRANSFORMER NETWORK-BASED RL METHOD FOR PSIJ OPTIMIZATION IN HBM

The transformer network for PDN decap optimization proposed in our previous work is used for this work [9]. The network is trained to solve the decap n/m problem in the HBM I/O interface – deriving the assignment sequence of m number of unit NMOS decaps ($\mathbf{a}=(x_{a1}, x_{a2}, \dots, x_{am})$) for given n number of positions for decap assignment candidates ($\mathbf{X}=(x_1, \dots, x_n)$) to maximize the reduction of 10 self- and transfer impedances ($Z_{11}, Z_{22}, Z_{33}, Z_{44}, Z_{12}, Z_{13}, Z_{14}, Z_{23}, Z_{24}, Z_{34}$) seen at 4 probing ports ($\mathbf{P}=(x_{p1}, x_{p2}, x_{p3}, x_{p4})$). \mathbf{P} is from the HBM PHY region and each port represents 256 IOs of 2 DWORDS. \mathbf{X} is from on-chip and on-interposer PDNs. The unit NMOS decap has 1.055 nF and 0.7 m Ω ESR. x is the feature vector of a port that contains the coordinates and the information on whether it is a probing port [9, eq. (3)]. Therefore, for a given input state $s=\{\mathbf{P}, \mathbf{X}\}$, the network is trained to derive \mathbf{a} to maximize the expected reward r by the RL. r is defined as the sum of the weighted mean of the self- and transfer impedance

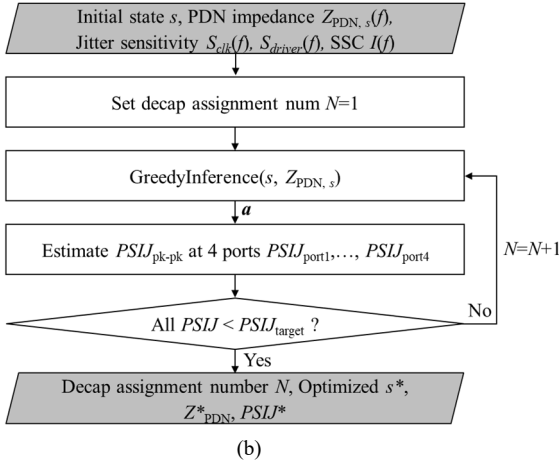
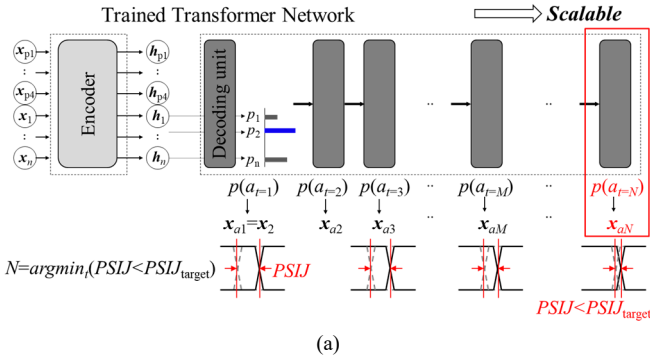


Fig. 2. (a) Overall concept of the proposed scalable transformer network-based PSIJ optimization method. (b) Flowchart of the proposed method.

reduction at \mathbf{P} [9, eq. (5)]. Details on the defined MDP parameters, the network and the training algorithm are explained in [9].

Fig. 2(a) shows the overall concept of the proposed method. The transformer network is trained to have scalability both in the encoder and decoder [9]. Therefore, the network trained in the smaller-scale decap n/m problem can solve the larger-scale problems either on n and m without re-training. The scalability is due to the shared weight properties and context embedding process in the transformer network. Details are described in [9]. Therefore, for given any decap assignment number m , the network can provide minimized PDN impedance profile. The PSIJ is expressed as the multiplication of PDN impedance, current spectrum and jitter sensitivity. Then, by increasing the decap assignments (decoding unit), the network can find the minimum number to meet the given target PSIJ.

Fig. 2(b) shows the detailed flowchart of the proposed method. Inputs are initial state s , PDN Z -matrix $Z_{\text{PDN},s}$, jitter sensitivity of clock buffer S_{clk} and Tx driver S_{driver} , and SSC spectrum I drawn by 256 IOs. For every loop, the network performs the greedy inference to output \mathbf{a} whose length is equal to the given assignment number N . The inference includes both computations in the encoder and the decoder as shown in Fig. 2(a). The encoder embeds raw feature vectors s into high-dimensional node embeddings \mathbf{h} by the attention computation [9, eqs. (7)–(12)]. In the decoder, the decoding unit sequentially outputs the next position a_t of a unit decap every time-step t depending on the PDF $p(a_t|s, a_{1:t-1})$. The context node \mathbf{h}_c configured of the previous assignment $\mathbf{h}_{a_{t-1}}$ and 4 probing ports $\mathbf{h}_{p1}, \mathbf{h}_{p2}, \mathbf{h}_{p3}$ and \mathbf{h}_{p4} becomes a query \mathbf{q} [9,

TABLE I. PSIJ OPTIMIZATION RESULTS BY THE PROPOSED METHOD

	PSIJ _{pk-pk} [ps]				Decap # (N)
	Port 1	Port 2	Port 3	Port 4	
Initial	83.94	54.29	83.99	54.26	0 EA
Case #1 (Target=25 ps)	24.52	21.68	24.96	21.93	72 EA
Case #2 (Target=22.05 ps)	21.86	20.69	22.03	20.6	143 EA

eq. (13), (14)]. Then, it computes $p(a_t)$ by the attention with the keys \mathbf{k} of the decap candidate nodes $\mathbf{h}_1, \dots, \mathbf{h}_n$ [9, eqs. (14)–(16)]. In the greedy inference, for every time-step, a_t is selected where the probability is maximized. In other words, it is chosen where the reduction of the PDN impedances is maximized. With the \mathbf{a} solved by the network, PSIJ_{pk-pk}'s are estimated at the 4 probing ports \mathbf{P} . If all the PSIJ_{pk-pk}'s do not satisfy the target PSIJ, then increase the N and repeat the loop until all the PSIJ's meet the target.

As depicted in Fig. 1(a), the total PSIJ is the sum of the PSIJ's at the clock buffer and IO driver [2]. Assuming the same SSC $I(f)$ drawn by 256 IOs at each port, PSIJ at the probing port 1 can be expressed as:

$$\begin{aligned} \text{PSIJ}_{\text{port1}}(f) &= S_{\text{clk}}(f) \times V_{\text{port1}}(f) + S_{\text{driver}}(f) \times V_{\text{port1}}(f) \\ &= S_{\text{clk}}(f) \times Z_{\text{total}}(f) \times I(f) + S_{\text{driver}}(f) \times Z_{\text{total}}(f) \times I(f). \end{aligned} \quad (1)$$

where V_{port1} is SSN at port 1. Z_{total} is the sum of the self-impedance Z_{11} and transfer impedances Z_{12}, Z_{13} and Z_{14} . Z_{11} indicates self-switching noise and $Z_{12}+Z_{13}+Z_{14}$ indicates coupled noises from ports 2, 3 and 4 to port 1. PSIJ(f) at the probing ports 2, 3 and 4 also can be represented in the same way. Then, PSIJ(t) and PSIJ_{pk-pk} can be derived as followings:

$$\text{PSIJ}(t) = \text{IFFT}(\text{PSIJ}(f)). \quad (2)$$

$$\text{PSIJ}_{\text{pk-pk}} = \max(\text{PSIJ}(t)) - \min(\text{PSIJ}(t)). \quad (3)$$

A global clock network is implemented on the PHY of the HBM logic die and Tx clock buffers are designed as 3-5 stages depending on the positions of DWORDS (probing ports) [2, Fig. 4]. The IO driver is designed based on the TSMC 65 nm process referring to HBM JEDEC [7]. S_{clk} and S_{driver} are modeled using the propagation delay-based methods [1, eq. (16)]. Since the HBM PHY is divided into 4 regions as shown in Fig. 3(a), each region contains 256 IOs of 2 DWORDS. Considering the data inversion bus (DBI), the worst-case SSC spectrum of 128 switching IOs is used for $I(f)$. The hierarchical VDDQ PDN includes an on-chip grid P/G plane, on-interposer meshed P/G plane, μ -bump array, multi-array TSVs and PKG PDN [9].

III. APPLICATION TO HBM I/O INTERFACE

The proposed method is applied to optimize the HBM2 I/O interface configured of 404 decap assignment candidates. The transformer network trained in decap 300/150 is used. The data rate of the switching IOs is 2 Gbps. The rise/fall time is 60 ps (0.12 UI). The peak transient current of a Tx driver is 9 mA. The max/min propagation delays $T_{p,\text{max}}$ and $T_{p,\text{min}}$ of Tx driver are 51.97 ps and 42.58 ps respectively. Those of the clock buffer for 3 and 5 stages are 405.4 ps/ 348.9 ps and 239.8 ps/ 208.1 ps respectively.

Table I shows the optimized peak-to-peak PSIJ results by the proposed method. Initial PSIJ's at the probing port 1, 2, 3

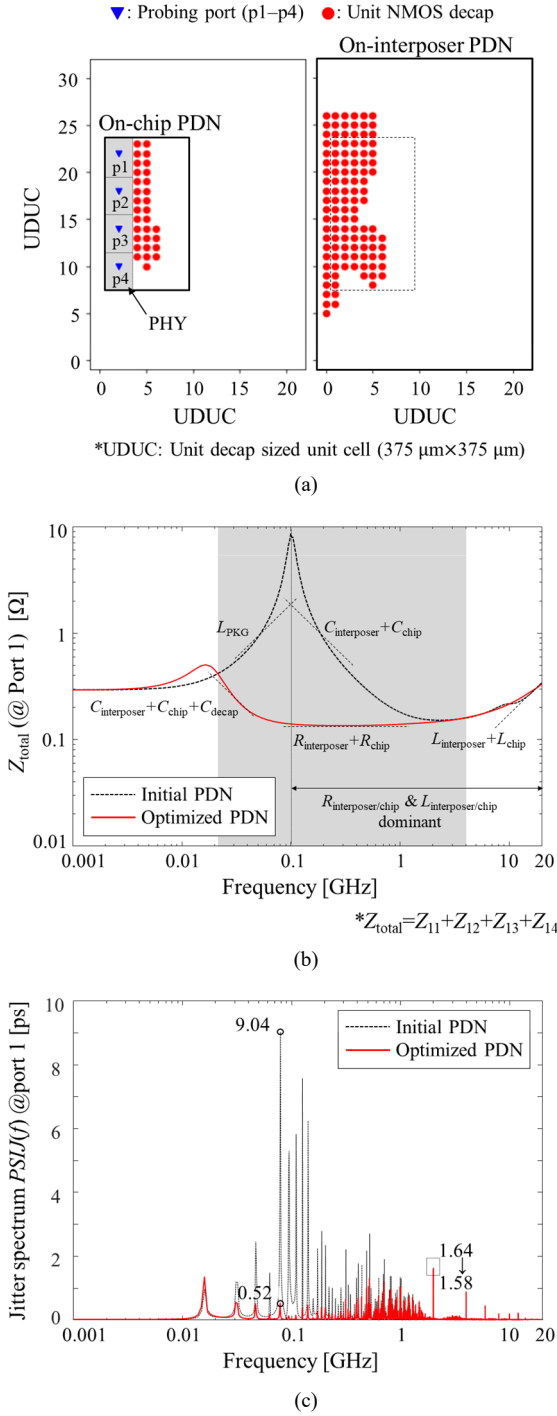


Fig. 3. Results of the Case #2. (a) Decap assignment result. (b) Z_{total} of the initial and optimized PDN seen at the probing port 1. (c) PSIJ spectrum of the initial and optimized PDN seen at the probing port 1.

and 4 are 83.94 ps, 54.29 ps, 83.99 ps and 54.26 ps respectively. The reason why the PSIJ's at ports 1 and 3 are larger than those at ports 2 and 4 is the difference in the number of clock buffer stages. Two target PSIJ's 25 ps and 22.05 ps are given. For both Case #1 and #2, all the $PSIJ_{pk}$'s at the probing ports satisfy the target. 72 EA and 143 EA of unit NMOS decaps are assigned respectively.

Detailed results of the Case #2 are plotted in Fig. 3. Fig. 3(a) depicts the decap assignment results. Fig. 3(b) shows $Z_{total}(f)$ of the initial and optimized PDN seen at the probing port 1 respectively. The SSC $I(f)$ frequency components are distributed from 10 MHz to 10 GHz, especially high peaks at

2 GHz and its harmonics. Hence, it is important to reduce $Z_{total}(f)$ in the broadband frequency range from 10 MHz to 10 GHz. As shown in Fig. 3(a), to minimize the $Z_{total}(f)$, the transformer network assigns decaps near and between the 4 probing ports both in the on-chip and on-interposer PDNs. For the anti-resonance between L_{PKG} and $C_{interposer} + C_{chip}$ around 100 MHz, total capacitance of assigned decaps (C_{decap}) is a dominant factor. Over the 100 MHz range, the positions of the decaps are dominant. In other words, effective resistance and inductance of the on-chip and on-interposer PDN $R_{interposer} + R_{chip}$ and $L_{interposer} + L_{chip}$ dominate the impedance profile. The assigned decaps suppressed the Z_{total} well from 20 MHz to 4 GHz including the anti-resonance. However, over 4 GHz, the reduction is limited by the self-inductance of the P/G plane itself. Fig. 3(c) illustrates the corresponding $PSIJ(f)$ of the initial and optimized PDN seen at the probing port 1 respectively.

IV. CONCLUSION

A scalable transformer network-based RL method for PSIJ optimization in HBM is proposed. The proposed method provides optimal decap design in on-chip and on-interposer PDNs to satisfy the target PSIJ with the minimum decaps. The decap optimization NN trained to have the generality is used. Thanks to its scalability, the network can provide the minimized PDN impedance profile for every decap assignment by one fast inference. Thus, the network can easily find the minimum number to meet the target PSIJ by increasing the decaps. The proposed method is applied to the HBM2 I/O interface and successfully provides the solutions.

ACKNOWLEDGMENT

This work was supported by Samsung Electronics Co., Ltd (IO201207-07813-01). We would like to acknowledge the technical support from ANSYS Korea. This research was supported by National R&D Program through the National Research Foundation of Korea(NRF) funded by Ministry of Science and ICT (NRF-2022M3I7A4072293).

REFERENCES

- [1] X. J. Wang and T. Kwasniewski, "Propagation Delay-Based Expression of Power Supply-Induced Jitter Sensitivity for CMOS Buffer Chain," in *IEEE Transactions on Electromagnetic Compatibility*, vol. 58, no. 2, pp. 627-630, April 2016.
- [2] T. Shin *et al.*, "Modeling and Analysis of System-Level Power Supply Noise Induced Jitter (PSIJ) for 4 Gbps High Bandwidth Memory (HBM) I/O Interface," *2021 IEEE Electrical Design of Advanced Packaging and Systems (EDAPS)*, 2021.
- [3] *High Bandwidth Memory DRAM (HBM3)*, Standard JESD238, 2022.
- [4] H. Park *et al.*, "Policy Gradient Reinforcement Learning-based Optimal Decoupling Capacitor Design Method for 2.5-D/3-D ICs using Transformer Network," *2020 IEEE Electrical Design of Advanced Packaging and Systems (EDAPS)*, 2020, pp. 1-3.
- [5] L. Zhang, W. Huang, J. Juang, H. Lin, B. -C. Tseng and C. Hwang, "An Enhanced Deep Reinforcement Learning Algorithm for Decoupling Capacitor Selection in Power Distribution Network Design," *2020 IEEE International Symposium on Electromagnetic Compatibility & Signal/Power Integrity (EMCSI)*, 2020, pp. 245-250.
- [6] S. Han, O. W. Bhatti and M. Swaminathan, "Reinforcement Learning for the Optimization of Decoupling Capacitors in Power Delivery Networks," *2021 IEEE International Joint EMC/SI/PI and EMC Europe Symposium*, 2021, pp. 544-548.
- [7] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.
- [8] A. Mirhoseini *et al.*, "A graph placement methodology for fast chip design," *Nature*, vol. 594, no. 7862, pp. 207-212, 2021.
- [9] H. Park *et al.*, "Transformer Network-based Reinforcement Learning Method for Power Distribution Network (PDN) Optimization of High Bandwidth Memory (HBM)," 2022, *arXiv:2203.15722*.

Crosstalk Analysis for PCIe 6.0 (PAM4) Under Different Transmitter Conditions

Fabio A. Ruiz-Molina
Data Center and AI Group
Intel Corporation
Zapopan, Mexico
fabio.a.ruiz.molina@intel.com

Jingbo Li
Data Center and AI Group
Intel Corporation
Hillsboro, USA
jingbo.li@intel.com

Kai Xiao
Data Center and AI Group
Intel Corporation
Hillsboro, USA
kai.xiao@intel.com

Abstract— Previous studies have shown PAM4 signaling is ~ 3x more sensitive to noise interference comparing to NRZ mode. Furthermore, because of the backward compatibility requirement, PCIe needs to support both PAM4 and NRZ signaling mode. As a result, the crosstalk scenario can be much more complicated than a sole signal modulation mode. In this paper, we present a thorough analysis for the crosstalk impact on PCIe 6.0 link (PAM4) with backward compatibility of previous generations, such as PCIe 5.0 (NRZ). Several key factors, including voltage swing, equalization, modulation type, number of aggressors and port partition or bifurcations, have been studied to show their impact on the crosstalk behavior and the related effect on the full link performance. This methodology will serve as a guideline to thoroughly study the crosstalk impact for platforms with PCIe 6.0 links. It is demonstrated that, after including all the considered variables about the crosstalk, the total degradation could add up to ~22% for EH and ~19% for EW.

Keywords— PCIe 6.0, PAM4, crosstalk, Tx equalization, high-speed serial links, signal integrity

I. INTRODUCTION

To continue succeeding in the industry, electronic designers need to enable the next PCIe generation, PCIe 6.0, at 64 Gbps with PAM4 signaling. It brings several challenges [1]. From signal integrity point of view, one of the most important topics is the crosstalk impact. It's well known that crosstalk will degrade the receiver capability to interpret the information being transmitted, and that it also gets worse with the transfer rate increment. The higher the frequency is, the higher the undesired coupling could be.

However, beyond the operation frequency, there are other elements contributing and shaping the crosstalk behavior. Identify the key elements driving the crosstalk impact is critical to take the risk out. This paper puts on the microscope the following variables: Tx voltage swing, Tx equalization, data modulation and number of aggressors.

The first variable is defined in the PCIe Base Spec [2] as differential peak-peak Tx voltage swing ($V_{TX-DIFF-PP}$). The spec allows a range for this value. The minimum value is 0.8 V, but it could go up to 1 V (PCIe 6.0 Spec) or 1.2 V (PCIe 5.0 spec). Taking as reference the 1.2 V, there could exist an increment of ~ 50% in the energy being coupled onto the neighbors. Given that this is a substantial increment, it becomes critical to include this parameter in the analysis. Additionally, in a crosstalk scenario where the aggressors and victims come from different link partners, such as a near-end crosstalk (NEXT) from transmitters to receivers, the voltage swings of the PHYs are independent.

Also, a well-known technique to compensate the signal degradation introduced by the channel is by implementing

TXLE (transmitter linear equalizer), which de-emphasizes the low frequency components [3]. Therefore, an aggressor with TX Equalization (TXEQ) will transmit less energy comparing to without TXEQ case. Since that, it is expected that less energy is coupled to the victims when the equalization is on.

In addition, before PCIe 6.0, which uses PAM4 coding, PCIe links based their operation on NRZ coding. A requirement in PCIe Base Spec is the backward compatibility [2], meaning that PCIe 6.0 links must be capable to work at lower transfer rates like PCIe 5.0 or PCIe 4.0. Besides, it will be possible on a platform that the PCIe ports neighboring to a PCIe 6.0 port on the same CPU are running at PCIe 5.0 speeds. In other words, different ports of a given CPU could operate at PCIe 6.0 and PCIe 5.0 simultaneously. This could be derived from having different ports or a bifurcated port (which adds complexity to the analysis). As result, it is necessary to evaluate some scenarios where PCIe 6.0 signals coexist with PCIe 5.0 (or other lower transfer rates) signals. In that condition, it can be found PAM4 victims surrounded by aggressors operating in NRZ mode.

Moreover, as the demand of I/O bandwidth goes up, it is needed to incorporate a higher density of signal pins for the IOs. Derived from this, the number of pins coupling energy into other pins continue increasing. It is necessary to perform an in-depth study to select the number of aggressors that must be considered as part of any signal integrity analysis.

The following section describes the methodology used to assess the impact of the 4 parameters mentioned before. Later, section III presents the results of the performed experiments. Finally, section IV shows the conclusion of the investigation.

II. METHODOLOGY OF ANALYSIS

A. Topology

As an investigation vehicle, it was selected a typical 1-connector PCIe topology (Fig. 1), where the total end-to-end insertion loss (IL) > -32 dB at 16 GHz. The package is contributing with -8 dB as is defined by the spec [2], the add-in-card (AIC) with -8.5 dB and the rest of IL is coming from main board. As is usual in many platforms, there is a socket between package and main board. Also, there are vias and transmission lines making up the main board channel. Finally, there is a connector compliant with PCIe 5.0 CEM spec [4] (CEM 6 model is under definition).

B. Choosing the number of aggressors

Fig. 2 shows a typical example of high-speed IO pin pattern. In the north we have Tx pins and, in the south, we have Rx pins. These pins will not necessarily be working at the same transfer rate, they can operate from PCIe 1.0 to PCIe 6.0 or belong to a totally different protocol. The number of aggressors will change from case to case. A good ground

isolation and favorable pin assignment will always help to reduce the number of significant aggressors. After picking the victim pair, a good strategy to select the adequate number of aggressors it is by looking at crosstalk responses in frequency domain. Assuming the configuration depicted in Fig. 2, it was decided to consider 4 far-end crosstalks (FEXT), i.e., crosstalk between transmitters, and 3 NEXTs. The directly adjacent differential pairs to the victim were selected as aggressors (first order aggression from the geometric point of view).

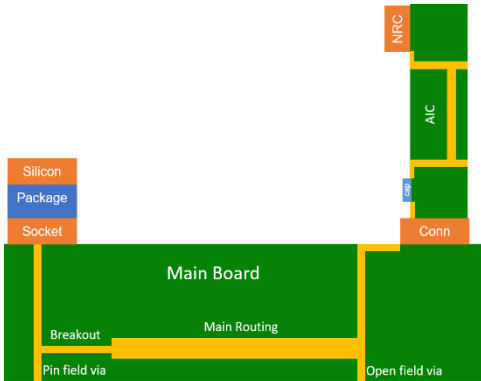


Fig. 1. 1-connector PCIe 6.0 topology.

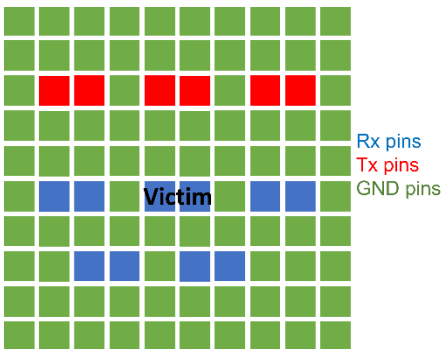


Fig. 2. Generic high-speed IO pin map.

C. Data coding

PCIe 6.0 uses PAM4 coding while the other transfer rates use NRZ. When a random bit sequence is transmitted, the total energy in a PAM4 coded waveform is 5/9 of NRZ coded waveform for the same sequence. Therefore, the energy coupled into the victim tends to be higher when NRZ is used [5]. This makes modulation an important factor in crosstalk analysis.

The amplitude of a transition between two adjacent levels in PAM4 is 1/3 of the signal at NRZ. This has two important implications, NRZ signals produce stronger instigations and PAM4 signal is more susceptible to noise.

Assuming the pin arrangement in Fig. 2, a realistic configuration could consider that 3 out of 4 FEXTs use NRZ, leaving only 1 FEXT operating with PAM4. At the same time, all the NEXTs could be working with NRZ coding. This configuration can be seen as a worst case. Section III will discuss this and other configurations.

D. TX voltage swing

The amplitude of coupled noise is directly proportional to the strength of the buffer. That is why $V_{TX-DIFF-PP}$ must be accounted as a key variable in the crosstalk analysis.

Latest PCIe spec [2] introduced a change for the maximum $V_{TX-DIFF-PP}$. The new maximum is 1 V, while in the previous revision it was allowed 1.2 V. Therefore, to support legacy PCIe devices, a maximum of 1.2 V needs to be assumed. As consequence, this paper considers Tx pairs ranging between 0.8 and 1.2 V (PCIe 5.0 compliant), while Rx pairs range between 0.8 and 1 V (PCIe 6.0 compliant).

E. TX Equalization

Like its predecessor, PCIe 6.0 will use TXLE. Typically, full link simulations assumed that aggressors would use the same TXLE configuration as the victim. This is true only when the aggressors are within the same link. Aggression could come from a different PCIe port or a different interface. Besides that, when bifurcated, links within the same PCIe port can communicate to different devices with different topologies. This means that equalization might be different even for adjacent lanes.

As it was stated in section I, when TXLE is off, instigation signal will remain at its full strength, allowing to couple more energy onto the victim. Then, the extreme condition to be tested is when it is chosen the preset Q0 (no equalization) for PCIe 6.0 or P4 for the legacy speeds [2]. The opposite case is when equalization is on. For the purpose of this paper, it is assumed that aggressors within the same link will have the same equalization setting as the victim.

III. RESULTS

All components in the channel (as shown in Figure 1) have been created for full link simulations. Most models are created by 3D EM simulator, except the transmission lines, created from a 2D simulator. We then simulated the link performance based on PCIe 6.0 standard requirements to evaluate the impact under different crosstalk conditions, as shown in Fig. 3. Each bar represents a different setup for the aggressors. The details of each setup are depicted in the Table 1. For readability, it was bolded the parameter that changes from case to case. Notice that the effects have been added progressively.

Crosstalk amplitude is linearly proportional to aggressors' voltage swing. From 0.8 to 1.2 V the crosstalk increases by 50%. From PAM4 to NRZ coding, the average transmitted energy in aggressor increases as well. Collectively, the two effects degraded EH by ~11% and EW by ~10% in simulation results (Steps 2-4 in Figure 3 and Table 1). After removing the TX equalization, another ~4% drop in EH and ~3% drop in EW are observed (step 4-5 in Fig. 3 and Table 1). Steps 6 & 7 shows the impact of voltage swing, modulation and TXEQ impacts on NEXT aggressors. Similar phenomena have been observed as FEXT in steps 2-5. In addition, the number of aggressors also have an impact on the link margin. However, that is specific to a pin pattern and component level performance, which will not be discussed in detail here.

IV. CONCLUSIONS

As analyzed, the backward compatibility of PCIe 6.0 (PAM4 signaling) and legacy speeds (PCIe 5.0 and early, NRZ signaling) can result in multiple crosstalk scenarios, which can result in significantly different impact on PCIe 6.0 link performance. In this paper, we analyzed the crosstalk

impact as a function of four parameters: number of aggressors, Tx voltage swing, Tx equalization and data codification. The results shown that each of these variables has a noticeable contribution.

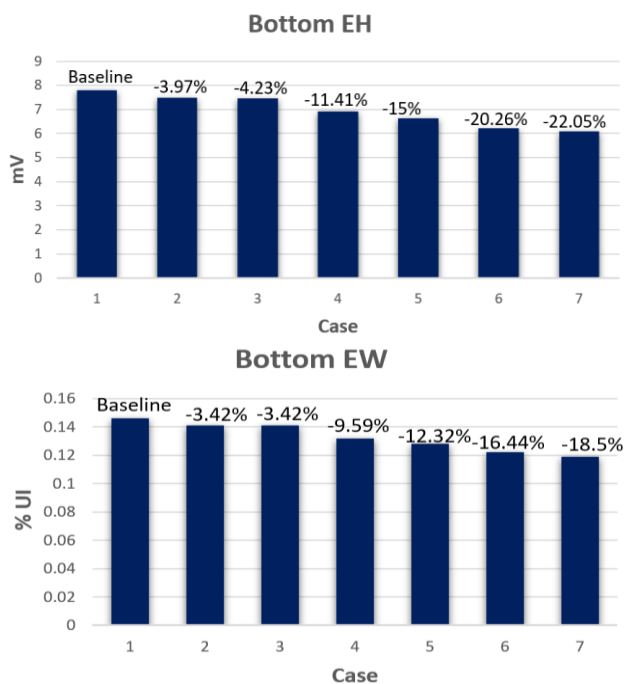


Fig. 3. Margin impact due to different crosstalk assumptions for 1-connector PCIe 6.0 topology. Bottom eye is choosing since it is the worst case among the 3 PAM4 eyes.

In terms of percentage, both EH and EW degrade similarly, even though EH is suffering more. Specific eye degradations and limiter will depend on topologies. The one connector topology in this paper is characterized by having more difficulties to pass EH spec than EW. However, topologies with more transitions, e.g., two or three connector topologies, tend to be more constrained in EW.

Based on the studies in this work, we provide the methodology to ensure a correct prediction of the crosstalk behavior as following.

1. The number of aggressors must be a function of the pin map, instead of having a predetermined fixed number. As first step, the selection can be done based on frequency domain analysis and proximity. In case of doubt, a single case based on eye margins can be conducted.
2. Once it is known the interfaces (speed and configuration) surrounding the lanes under study, it is necessary to run analysis assuming that those lanes are being driving with the maximum voltage swing allowed by the corresponding interface. Figure out the right swing used by those lanes would deliver a more realistic result.
3. Also, as extreme case, the TXLE of the aggressors should be assumed to be zero. If the TXLE configuration is known, then it would be better to configure the right TXLE setting to get rid of pessimism.
4. Finally, PAM4 links should be placed away from NRZ links. If not possible, it is needed to include the right codification in the simulation setup.

REFERENCES

- [1] N. Dikhaminija et al., "PAM4 signaling considerations for high-speed serial links," 2016 IEEE International Symposium on Electromagnetic Compatibility (EMC), 2016, pp. 906-910.
- [2] PCI Express Base Specification Revision 6.0, Version 1.0, December 2021.
- [3] ALTERA, "Understanding the Pre-Emphasis and Linear Equalization Features in Stratix IV GX Devices," Application Note, November 2010.
- [4] PCIe Express Card Electromechanical Specification Revision 5.0, Version 1.0, June 2021.
- [5] Intel, "PAM4 Signaling Fundamentals," Application Note, March, 2019.

TABLE I. DEGRADATION DUE TO DIFFERENT CROSSTALK ASSUMPTIONS FOR 1-CONNECTOR PCIe 6.0 TOPOLOGY.

Case	NEXTs	FEXTs	EQ (TXLE)	Voltage swing [V]	Data Coding	Bottom EH [mV]	Bottom EW [UI]
1	3	4	NEXT: as victim FEXT: as victim	NEXT: 0.8 FEXT: 0.8	NEXT: PAM4 FEXT: PAM4	7.8	0.146
2	3	2	NEXT: as victim FEXT: as victim	NEXT: 0.8 FEXT: 0.8	NEXT: NRZ FEXT: PAM4	7.49	0.141
3	3	4	NEXT: as victim FEXT: as victim	NEXT: 0.8 FEXT: 0.8	NEXT: NRZ FEXT: PAM4	7.47	0.141
4	3	4	NEXT: as victim FEXT1 to FEXT3: as victim FEXT4: as victim	NEXT: 0.8 FEXT1 to FEXT3: 1.2 FEXT4: 0.8	NEXT: NRZ FEXT1 to FEXT3: NRZ FEXT4: PAM4	6.91	0.132
5	3	4	NEXT: as victim FEXT1 to FEXT3: Preset Q0 FEXT4: as victim	NEXT: 0.8 FEXT1 to FEXT3: 1.2 FEXT4: 0.8	NEXT: NRZ FEXT1 to FEXT3: NRZ FEXT4: PAM4	6.63	0.128
6	3	4	NEXT: as victim FEXT1 to FEXT3: Preset Q0 FEXT4: as victim	NEXT: 1 FEXT1 to FEXT3: 1.2 FEXT4: 0.8	NEXT: NRZ FEXT1 to FEXT3: NRZ FEXT4: PAM4	6.22	0.122
7	3	4	NEXT: Preset Q0 FEXT1 to FEXT3: Preset Q0 FEXT4: as victim	NEXT: 1 FEXT1 to FEXT3: 1.2 FEXT4: 0.8	NEXT: NRZ FEXT1 to FEXT3: NRZ FEXT4: PAM4	6.08	0.119

Impedance and Cost based PDN Decoupling Optimization using Reinforcement Learning

Allan Sánchez-Masis¹ and Sameer Shekhar²

¹Client Computing Group, Intel Corporation, Heredia, Belén, Costa Rica

²Client Computing Group, Intel Corporation, Hillsboro, Oregon, USA.
allan.sanchez.masis@intel.com, sameer.shekhar@intel.com

Abstract—PDN optimization involves selection of capacitors to meet the target impedance. This paper uses reinforcement learning to solve decoupling stuffing problem based on impedance-based reward and then with both impedance & cost-based reward. It is shown how the agent can be biased when trained only on impedance-based reward. Key results including attainment of target impedance and overall achieving cost and impedance optimized solution are reported.

Keywords—Reinforcement learning, power delivery network, capacitor selection, power integrity.

I. INTRODUCTION

Microprocessor power supply requires low power delivery network (PDN) impedance peaks. This is obtained via careful capacitor (Cap) stuffing in the power integrity design. This process of evaluating different decoupling capacitor option relies on engineering knowledge, is time consuming and typically lacks cost consideration for practical design.

This paper addresses PDN optimization problem via utilization of deep reinforcement learning (RL). This work contributes to literature like prior references [1], [2] which too address capacitor stuffing using RL in a frequency band. Reference [1] features silicon interposer based 2.5D/3D integrated circuits; however, it doesn't account for cost. In [2], only the total number of capacitors are used in the reward function. This work provides an approach of attaining target impedance in a frequency band at the same time ensuring cost-based stuffing. Paper also features different reward functions, states, and possible actions along with techniques to calibrate impedance and cost rewards for joint utilization. Training time saving techniques like reutilization of previously run impedance simulations are also discussed.

Paper is organized as follows. Section II covers RL overview and explains the design problem. Section III provides RL results from impedance-based reward and Section IV provides results from of cost and impedance-based reward. Conclusions are summarized in Section V.

II. RL AND PDN OPTIMIZATION PROBLEM

This section provides a brief overview of reinforcement learning and its application to a common power integrity problem with specific details.

A. Reinforcement Learning Overview

RL is a field of machine learning which is gaining significant popularity [3]. RL does not use label, unlike supervised learning [4], [5]. Instead, it trains an agent (e.g., Q-table, DQN, etc.) via interacting with the environment. RL training involves many episodes with each episode consisting of many iterations or steps. As shown in Fig. 1, during each training step the agent takes an action in the environment for which it receives a reward. Data including states, action, reward, and next state is stored in memory buffer for utilization during training. Agent's goal is to maximize reward during training by striking a balance between exploration and exploitation. Finally, after all episodes, an agent is learnt which can be used for testing or deployment.



Fig. 1. Reinforcement learning framework.

B. Problem Description

This paper addresses a conventional power integrity problem of PDN capacitor selection after a touchstone file is extracted capturing PCB and/or package parasitics. The process of trying different decoupling option is cumbersome, time consuming and lacks cost consideration for optimal design. Paper trains RL agent to obtain optimal capacitor stuffing.

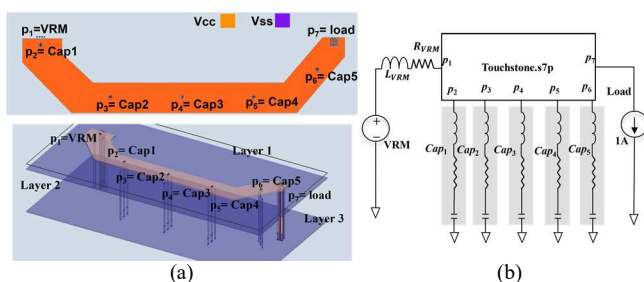


Fig. 2. PDN optimization problem: (a) 2D & 3D illustration of extracted 7-port PDN model (b) circuit representation.

Fig. 2(a) illustrates 2D and 3D view of the PCB under consideration. A touchstone file (.s7p) consisting of a voltage regulator module (VRM), capacitors (5 count) and SoC load port is generated. PDN impedance profile is defined as the AC response obtained via application of 1A AC current at the SoC load location. VRM is modeled as a RL element to represent its

bandwidth. The circuit details are shown in Fig 2(b). The capacitors use a higher order model containing parasitic inductance and resistance.

Fig. 3 shows the optimization target of achieving $< 25 \text{ m}\Omega$ in the frequency band of interest [6].

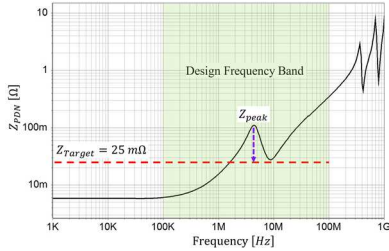


Fig. 3. PDN optimization goal to meet Z_{target} in the design frequency band.

III. TARGET IMPEDANCE BASED RL LEARNING

This section covers RL model details, including the results from impedance peak minimization and meeting of target impedance.

A. States, Action, and Agent

The work begins by sorting available commercial capacitors in an increasing order of capacitance to create a vector. Then we use the index of this vector in RL modeling for any further utilization. This is for better interpretability of results via power integrity domain knowledge.

Fig. 4(a) shows the action and state vector details. Action has one hot encoding, i.e., at one time one capacitor is changed. States are current capacitor index (or effectively stuffing) along with Z_{peak} during the iteration. The used DQN model is shown in Fig. 4(b).

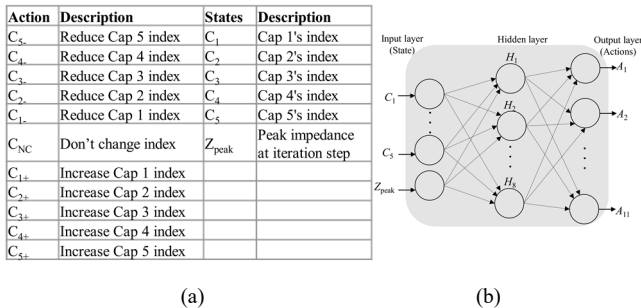


Fig. 4. Model details (a) states, actions and (b) employed DQN agent.

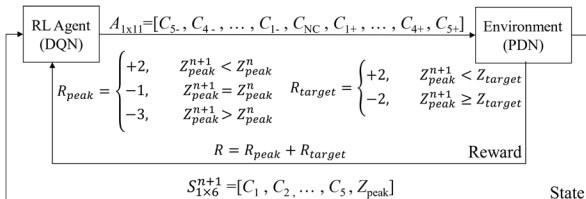


Fig. 5. RL model for impedance reduction factoring peak and target.

B. Applied RL Approach and Algorithm

Fig. 5 shows the RL modeling details. We can see the reward (R) details. Iteration reward is sum of two rewards: (a) R_{peak}

which factors reduction in impedance peak and (b) R_{target} which reflects attainment of $25 \text{ m}\Omega$ impedance.

Algorithm 1 illustrates training process, which is done for 140 episodes each with 20 iterations. A key delta to conventional approach is utilization of previously computed Z_{peak} in order to reduce training runtime. Other steps in algorithm are more or less consistent with a typical RL model [3].

Algorithm 1. DQN Training.

```

Memory Initialization:  $D$  of size 5000
Random weights initialization for DQN agent
for episode = 1, 140 do:
  for t = 1, 20 do:
    If t=1 do: Initialize state randomly
    Select action:
       $\mathbf{a}_t$  randomly
      Otherwise based on DQN policy
    Execute  $\mathbf{a}_t$  in PDN environment to get states
      if  $Z_{peak}$  obtained previously, re-use
    Compute  $\mathbf{R}$ 
    Save  $\{S^t, A, R, S^{t+1}\}$  to buffer
    Update DQN weights based on MAE
  end for
end for

```

C. Episode Reward and Test Results

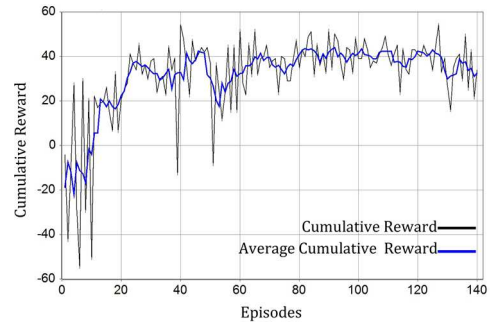


Fig. 6. Episode reward during training without considering cost in the reward.

Fig. 6 shows each episode's cumulative reward alongside a 5-episodes moving average. We see the expected trend of increasing reward over episodes. Note the maximum possible reward is 80 which the agent could converge to given appropriate episodes and perhaps by additional hidden layers in DQN network. For intent of this work this training was deemed sufficient as will be discussed via agent's test run results in Table I.

IV. REWARD SHAPING FACTORING IMPEDANCE AND COST

The approach of minimization of impedance peak as shown in Fig. 3 is limiting as it does not factor cost of capacitors for practical reasons. Therefore, reward reshaping is needed to account for cost of capacitors.

A. Addressing Cost of Capacitors

In this approach, capacitor index is proportional to capacitance (as in Section III) and cost, however, note that capacitor cost depends on many aspects like supply chain, form-factor and is not only limited by capacitance value. A linear fit via least squares approximation is obtained as shown in Fig. 7.

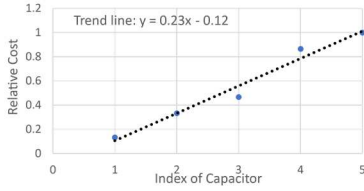


Fig. 7. Relative cost vs capacitor index.

B. Reward Reshaping

Fig. 8 shows the modification in the reward. Reward now considers R_{peak} and R_{target} along with a new reward for capacitor cost, R_{cost} . Constants, m & b are 0.23 and -0.12 respectively (refer Fig. 7). For ensuring the numerical range of impedance and cost rewards are similar a ‘scaling factor’ β is introduced. Finally, a ‘priority co-efficient’ α ($[0, 1]$) is introduced to decide the importance of impedance vs cost during training. An α of 0.3 was used in the simulations.

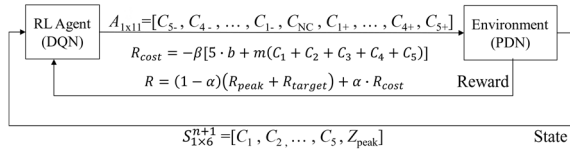


Fig. 8. RL model for impedance reduction factoring peak, target impedance and cost.

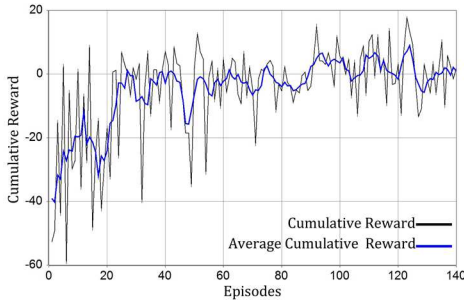


Fig. 9. Episode reward during training considering cost in the reward.

C. Episode Reward and Test Results

Fig. 9 shows the episode reward during training with the expected trend of increasing reward. The test results for 10 test runs using agent from Section III and Section IV are tabulated in Table I. Initial state has the random starting point capacitor index along with corresponding peak impedance. After the runs the results from agent in Section III and Section IV are tabulated to its right. Overall, both the agents meet the target impedance of 25 mΩ. In some cases, agent from Section III does not meet the target impedance. It is observed that this agent learns to choose capacitor 5 with highest capacitance value nearest to the SoC load depicting a bias originating due to a local minimum. This can perhaps be resolved via higher episode count and/or adding layers to the DQN network. Agent from Section IV however always meets the target impedance as it is forced to explore other minima of the overall solution due to additional cost reward. This agent also has a trend of generally choosing cheaper capacitors as expected. This can be seen by observing the sum of capacitor indices. In some cases, at slightly higher cost the agent from Section IV provides a significantly lower

impedance peak. Also, at times Section IV achieves an equal or lower impedance at a higher cost. This can be improved by increase the value of α in the model.

TABLE I. TEST RUNS ON AGENTS FROM SECTION III & IV.

#	Initial state		Agent Section III		Agent Section IV	
	Initial Capacitor Index	Initial Z_{peak} [mΩ]	Predicted Capacitor Index	Test Z_{peak} [mΩ]	Predicted Capacitor Index	Test Z_{peak} [mΩ]
1	[1,1,1,1,3]	52.5	[2,4,1,1,5]	62.63	[1,1,5,1,3]	16.44
2	[4,4,1,1,1]	219.23	[4,4,1,1,5]	62.6	[4,4,5,2,5]	10.35
3	[1,1,2,5,4]	11.43	[1,4,2,5,5]	9.89	[1,1,5,5,4]	9.95
4	[1,3,4,1,5]	11.62	[1,5,4,1,5]	10.78	[1,3,5,1,5]	10.67
5	[2,1,4,2,5]	11.67	[2,4,4,2,5]	11.29	[2,1,5,2,5]	10.74
6	[3,2,5,3,2]	45.16	[3,4,5,3,5]	10.21	[3,2,5,3,4]	12.03
7	[5,1,1,2,1]	112.46	[5,1,1,2,5]	27.85	[5,1,5,2,3]	16.24
8	[5,1,4,4,2]	43.37	[5,1,4,4,5]	10.59	[5,1,5,4,4]	10.91
9	[1,3,2,3,3]	27.6	[1,4,2,3,5]	11.78	[1,3,5,3,4]	11.88
10	[4,4,3,1,1]	176.95	[4,4,3,1,5]	11.82	[4,4,5,2,5]	10.35

V. CONCLUSIONS

This paper addresses a common design optimization problem of capacitor stuffing in a PDN via use of reinforcement learning. Analysis presented shows how reinforcement learning can be used to perform impedance peak minimization to meet target impedance. The contribution also demonstrates accounting for cost of capacitors in reward function to reduce both cost and impedance peak to achieve a target impedance.

This work has practical significance as trying out different decoupling combinations is a cumbersome approach. Furthermore all PDN network simulations ($O \sim n^n$) are not practical in design process. Therefore, the proposed method can be used for determining decoupling stuffing of PDN. From a RL model perspective paper proposes a unique approach of saving previous Z_{peak} to speed up training. Finally, it is also shown how optimization of impedance and cost can be incorporated via reward function. Future work can include to study trained agent on degraded PDNs, using more episodes and more powerful neural net for training and studies involving variation of priority co-efficient α .

REFERENCES

- [1] H. Park *et al.*, "Deep Reinforcement Learning-Based Optimal Decoupling Capacitor Design Method for Silicon Interposer-Based 2.5-D/3-D ICs," in *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 10, no. 3, pp. 467-478, March 2020.
- [2] L. Zhang *et al.*, "Decoupling capacitor selection algorithm for PDN based on deep reinforcement learning," in *Proc. IEEE Int. Symp. Electromagn. Compat., Signal Power Integr. (EMC+SIPI)*, Jul. 2019, pp. 616-620.
- [3] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529-533, 2015.
- [4] A. Sánchez-Masis, S. Shekhar, C. Chaves, and M. Aguilar, "Parameter estimation of silicon metal grid using supervised learning," *IEEE EMC+SIPI*, 2022.
- [5] A. Sánchez-Masis *et al.*, "ANN Hyperparameter Optimization by Genetic Algorithms for Via Interconnect Classification," *2021 IEEE 25th Workshop on Signal and Power Integrity (SPI)*, 2021, pp. 1-4.
- [6] S. Shekhar, A. K. Jain, and A. Gullapalli, "Analytical decomposition of power delivery network response for ramp loads," in *2016 IEEE International Symposium on Electromagnetic Compatibility (EMC)*, 2016, pp. 657-662.

Memory Speed Enhancement via SI/PI Optimization in Constrained Tablet Designs

Simon Chun Kit See
Client Computing Group
Intel
Penang, Malaysia
simon.chun.kit.see@intel.com

Asmah Truky
Client Computing Group
Intel
Penang, Malaysia
asmah.truky@intel.com

Gaurav Hada
Client Computing Group
Intel
Bangalore, India
gaurav.hada@intel.com

Sameer Shekhar
Client Computing Group
Intel
Hillsboro, Oregon, USA
sameer.shekhar@intel.com

Chandru Raman
Customer Enabling Debug Group
Intel
Penang, Malaysia
chandru.raman@intel.com

Abstract— Achieving high speed memory in small form-factor and low PCB layer count is critical for tablets designs. This paper reports several practical SI/PI design improvements such as PTH patterning suited for common memory ball-maps, etc. Simulation data showing voltage noises and eye diagrams are presented. Finally, measurement data with speeds of 3200 MT/s are provided for successful demonstration.

Keywords—memory, LPDDR4x, power integrity, signal integrity

I. INTRODUCTION

Modern tablets have high memory capacity and performance requirements. Tablet designs are always constrained due to form-factor (FF) and at times cost constrained depending upon end device's business considerations. Therefore, it is imperative to fully utilize available engineering design resources for platform electrical solutions.

Memory interface is seeing an increasing trend in transfer rate mainly driven by the growing demands for higher bandwidth applications such as video streaming, storage, and advanced biomedical sciences[1,2,3]. To obtain higher memory transfer rate compared with reference design requires improvement in both power integrity (PI) and signal integrity (SI) solutions. This is always a challenge in tablet designs due to desired cost savings and to support sustainability initiatives. This challenge can be addressed by better SI/PI design and take key inputs into consideration such as modification of printed circuit board (PCB) to better incorporate the memory ball-maps, improvements by accounting for JEDEC specs[4], and employing different stackup approach.

This paper presents SI/PI design results in a type-3 six layers PCB configuration to achieve LPDDR4x[5] memory speeds up to 3200 MT/s, a design improvement over reference design, i.e. type-3 10 layers. The paper is organized as follows. Section II addresses the problem statement with system targets. Section III describes the proposed design improvements. Section IV provides electrical simulation and memory validation data. Conclusions are summarized in Section V.

II. PROBLEM STATEMENT

A. Tablet Design Overview

Tablet designs are constrained due to smaller PCB FF and cost considerations. This typically makes meeting electrical design target harder gen over gen due to reduction in PCB FF and constrained electrical routing due to larger battery size and reduction in layer count. Also gen over gen the speed targets increase for all intellectual properties (IPs) making the overall design challenging.

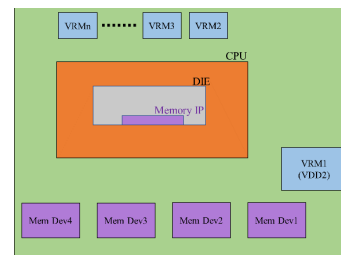


Fig. 1. PCB overview depicting CPU, VRM and memory devices.

B. System Targets and JEDEC specifications

Tablet system target was to achieve speed of 2933 MT/s which was based upon a 10 layers reference design. However, the layer count requirement was achieving the solution in a 6 layers type-3 PCB configuration.

According to JEDEC specification[4], three voltage supplies namely the VDD2, VDDQ (also common supply to the central processing unit (CPU) Memory IP) and VDD1 are needed to power the memory devices. VDDQ and VDD1 supplying to memory devices are low current and are easy to design to meet specification. Furthermore, the VDDQ is also shared with the CPU Memory IP and the noise coupling between CPU and memory devices is mitigated with capacitors placed near the CPU package (not shown in figure). The focus of this paper is on VDD2 which is higher in current and hence requires wider spaces to route the power delivery network (PDN) to meet

JEDEC spec V_{min} of 1.06V and of V_{p2p} of 45mV for frequencies ≤ 20 MHz at memory device ball grid arrays.

III. MEMORY DOWN CHANNEL DESIGN

We start by working on the biggest delta of reducing 4 board layers for the design. Fig. 2 shows the reference design stackup with 10 layers and the optimized stackup with 6 layers. The figure also shows the Cu layer assignment for power (both VDD2 and VDDQ) and signals. As VDD1 is low current, it can be routed as narrow plane and where space is available.

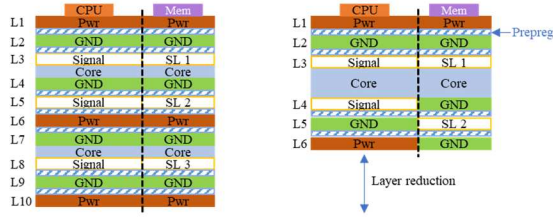


Fig. 2. PCB stackup: 10 layers reference design and 6 layers optimized design with dual stackup strategy.

It was observed from the reference design the plated through hole (PTH) patterns can be further improved for better PDN quality for the type-3 board – this is depicted in Fig. 3. The signal design was improved with clean linear routing in order to avoid transition and length asymmetry (see Fig. 4). After changes a better SI/PI optimized PTH pattern and surface layer routing was engineered.

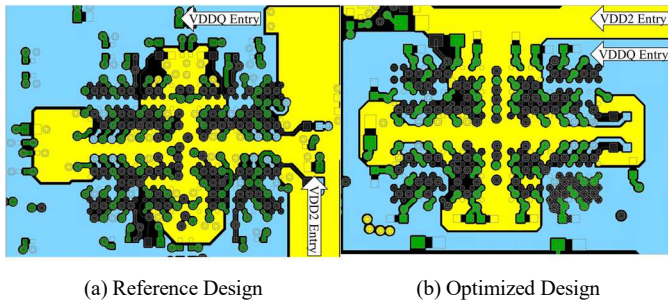


Fig. 3. Comparison one memory device routing and the improved type-3 PTH patterning for better SI/PI (VDD2 is output of VRM1, see Fig. 1).

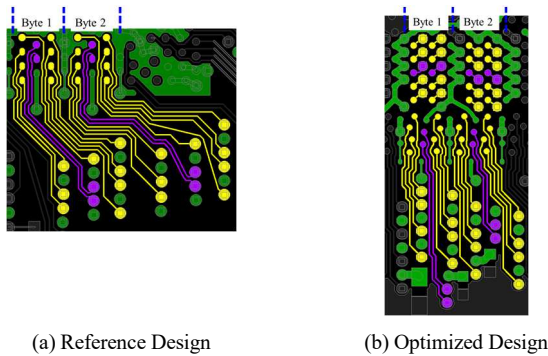


Fig. 4. PCB surface optimized signal routing.

Dual stackup studies, i.e. regions of PCB to have different routing for different interfaces, were performed. In particular,

stackup assignment under CPU and memory devices were changed to enhance the memory signalling. Finally, prepreg thickness reduction investigations were made to study trade-off between crosstalk and impedance.

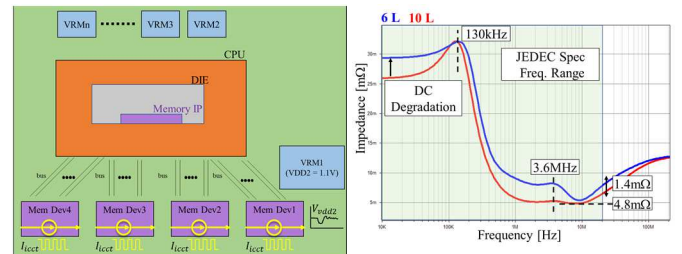
IV. SIMULATION AND VALIDATION RESULT

This section features electrical results from the layout changes described in Section III including the post-silicon validation data. Electrical studies gave sufficient confidence to adopt the design changes.

A. Memory Device Power Integrity

Simulations were performed to study the power integrity differences between 10 layers and 6 layers PCB with dual stackup, as shown in Fig. 2, and the PTH optimization, as shown in Fig. 3.

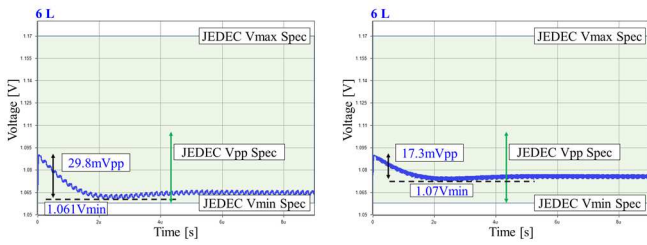
Fig. 5(a) shows the simulation setup for AC and transient analyses using the analog circuit simulator HSPICE. The PDN is modeled as a sub-circuit s-parameter (for AC) and SPICE netlist (for transient) extracted from the physical layout using the Cadence Sigriety PowerSI. We excite the PDN by drawing current at the four memory devices. Fig. 5(b) shows the impedance plots comparing the 6 layers with 10 layers PDN designs. We can immediately see the degradation in the low frequency or DC-region of the plot. This is due to loss of Cu layers. This however does not manifest in memory speed degradation due to three reasons: (a) the modeling of $I_{cc}(t)$ (current load) uses a burst-idle-burst (BIB) pattern at the highest PDN peak[6] which does not factor DC, (b) DC degradation manifest in lower nominal DC voltage due to delta from the voltage regulator module (VRM) sense to the load, this is typically low for memory PDN due to lower current draw compared to other high current power supplies like cores and graphics and (c) JEDEC requirements are from DC to 20MHz range[4] where decoupling can be adjusted to minimize the impact.



(a) AC and transient simulation setups. (b) Impedance plots comparing 6 layers with 10 layers designs.

Fig. 5. Power integrity simulation setup and impedance data.

Fig. 6(a) & (b) show the verifications using transient analyses with BIB patterns at 3.6MHz (worst-case) and 20MHz frequencies. Here current is applied for each of the four memory devices simultaneously. The transients showed the worst-case scenario is still within the JEDEC specifications.

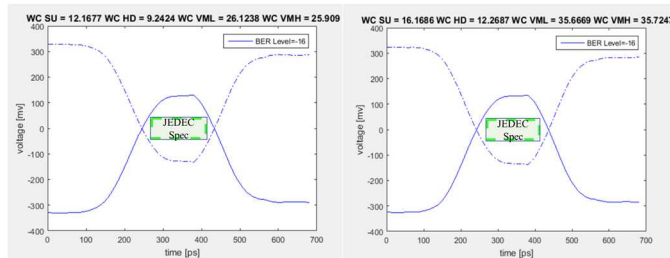


(a) Transient plot with 3.6MHz BIB pattern. (b) Transient plot with 20MHz BIB pattern.

Fig. 6. Power integrity transient data.

B. Memory Device Signal Integrity

For the tablet design it was obvious to use optimized strategy with PTH patterning and signal routing. With such optimization in a 6 layers configuration and the adoption of the dual PCB stackup, investigations showed the signal eye diagram is meeting JEDEC specification as shown in Fig. 7(a). Fig. 7(b) shows the improved eye-height and eye-width margins with reduction in di-electric thickness, which is an advantage to thinner tablet design. This finding is consistent with memory channel expectation of being cross talk limited.



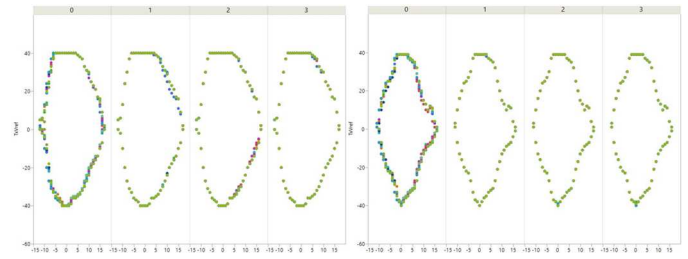
(a) 6 layers with optimized PTH, signal routing and dual stackup strategy. (b) Further reduction in di-electric thickness.

Fig. 7. Signal integrity simulation data.

C. Validation Data

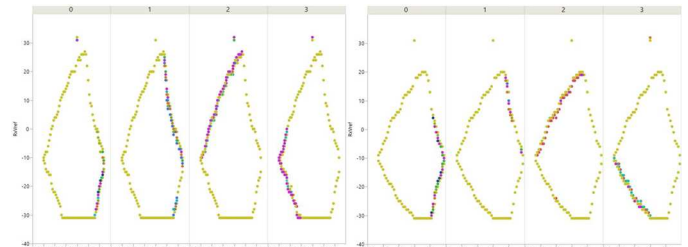
Validation data at 2933MT/s with all memory stress tests performed showed good eye openings for both DQ WRITE and READ modes with lower margins seen at READ and therefore being the limiter in the design. READ mode is largely dependent on the DRAM devices used and therefore the eye performance varies from one device to another. Additional validation data shows the speed can reach the highest speed with stability at 3200MT/s.

The validation coverage was limited and did not cover all the process, voltage and temperature (PVT) corners, but all memory stress tests were performed. The available data showed good passing of eye margins up to 3200MT/s for all four memory devices.



(a) 2933 MT/s (b) 3200 MT/s

Fig. 8. Tx eye diagram for 4 memory devices.



(a) 2933 MT/s (b) 3200 MT/s

Fig. 9. Rx eye diagram for 4 memory devices.

V. CONCLUSION

This paper presents memory electrical design details for a tablet FF system. Design modification, simulation results and validation results are reported with type-3 six layers PCB. Memory interface was successfully demonstrated to achieve speeds of 3200 MT/s.

REFERENCES

- [1] G. Zhu et al., "Parallel simulation of electromagnetic and thermal characteristic in RF component," 2017 IEEE Electrical Design of Advanced Packaging and Systems Symposium (EDAPS), 2017, pp. 1-3, doi: 10.1109/EDAPS.2017.8277057.
- [2] X. Lecoq, A. Goulahsen, P. Derouet and D. Rousseau, "EMI/EMC and co-existence analyses of digital and RF wireless interfaces on consumer and mobile products," 2016 6th Electronic System-Integration Technology Conference (ESTC), 2016, pp. 1-8, doi: 10.1109/ESTC.2016.7764743
- [3] F. Ali, "Direct conversion receiver design for mobile phone systems - challenges, status and trends," 2002 IEEE Radio Frequency Integrated Circuits (RFIC) Symposium. Digest of Papers (Cat. No.02CH37280), 2002, pp. 21-22, doi: 10.1109/RFIC.2002.1011501.
- [4] JEDEC Specification, Committee Item: 1824.42D, Low Power Double Data Rate 4 (LPDDR4), JESD209-4D, Jun 2021.
- [5] Wikipedia, 'LPDDR', <https://en.wikipedia.org/wiki/LPDDR>.
- [6] S. Shekhar, A. K. Jain, and A. Gullapalli, "Analytical decomposition of power delivery network response for ramp loads," in 2016 IEEE International Symposium on Electromagnetic Compatibility (EMC), 2016, pp. 657-662.

Intra-pair Skew Impact Analysis of High-speed Cables for HDMI Interface

Boogyo Sim¹, Keunwoo Kim¹, Taein Shin¹, Hyunwook Park¹, Seongguk Kim¹, Daehwan Lho¹, Keeyoung Son¹, Kyubong Kong², Seungtaek Jeong³, Seonguk Choi¹, Jihun Kim¹ and Joungho Kim¹

¹ School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea

²SK Hynix, Icheon, South Korea,

³Missouri University of Science and Technology, Rolla, USA

boogyo@kaist.ac.kr

Abstract— In this paper, the coaxial and shielded-twisted pair (STP) cables with intra-pair skew was analyzed by frequency domain analysis method by the equations in close-to-real intra-pair skew setup for high definition multimedia interface (HDMI) interface. For the high-speed interface, the intra-pair skew has become critical factor in terms of signal integrity. Heretofore, the intra-pair skew has not been commonly investigated in frequency domain for the high-speed data rate. To verify the intra-pair skew impact on the high-speed interface, the intra-pair skew location was distributed in the coaxial and STP cable assemblies for the real cable setup. Next, the skew impact was verified using frequency domain analysis and the eye-diagram. As the result, by the frequency-domain analysis, the differential loss and the mode conversion of coaxial cable assembly deteriorated due to the intra-pair skew, but STP did not deteriorate. Furthermore, the eye-diagram of the coaxial cable assembly with skew had smaller opening than without any skew, but of STP had little change.

Keywords— Eye diagram, Mode conversion, High definition multimedia interface (HDMI), High-speed cables, High-speed interface, Intra-pair skew, Signal integrity,

I. INTRODUCTION

Recently, the display market such as 4K/8K/Smart TV and high-end portable laptop has increased dramatically for the demand of high-quality video by the consumers. In market, high definition multimedia interface (HDMI) cables are commonly used for the TVs and the laptops. For 12 Gbps HDMI 2.1 interface, signal integrity (SI) performance has to be guaranteed to transfer data without bunch of distortion. A few of general SI performance criterion are insertion loss, crosstalk, electromagnetic interference and so on [1], [2]. Nowadays, the intra-pair skew has become one of the critical issues since the data rate has increased. Therefore, it is invaluable to analyze the intra-pair skew impact on high-speed cables for HDMI interface and the developing next-generation interfaces.

The definition of intra-pair skew is time gap between P and N channel. The intra-pair skew is occurred in unbalanced signal pair caused by inhomogeneous dielectric material, length difference, weave effect and so on [3], [4]. Since the intra-pair skew causes fatal mode conversion, the electrical performance has been distorted as the data rate has increased. To verify the intra-pair skew effect on high-speed serial link interface, various research has been proposed and investigated [3-7]. The previous works verified severe effect of the skew impact on

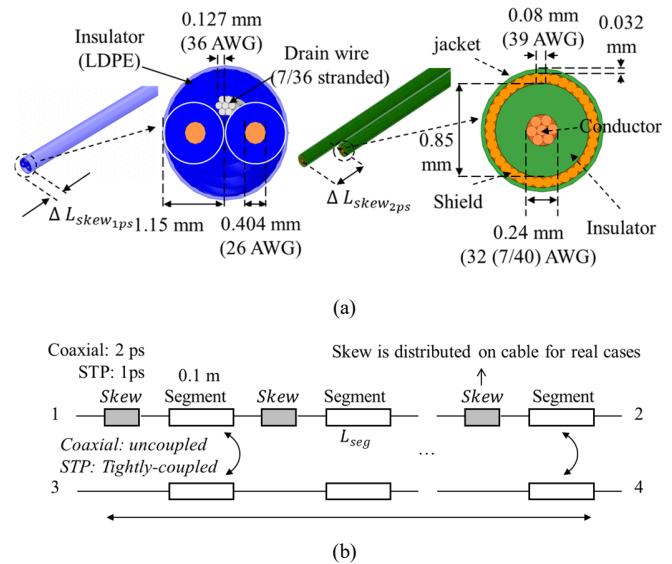


Fig. 1. (a) The model of the shielded-twisted cable (STP) and the coaxial cable for the intra-pair skew analysis in the frequency domain and the eye-diagram. (b) The intra-pair skew was distributed in the STP and coaxial cables for close-to-real setup. In real case, the intra-pair skew was not located either end of the cables as a lumped skew.

high-speed serial link in frequency and time-domain. However, most of studies did setup the skew as a lumped skew which was located either end of the cables. Furthermore, the tendency of the skew in frequency domain was not clarified on the previous papers. And the characteristics of coaxial and STP cables for the interfaces were not compared in the other papers. Therefore, it is necessary to scrutinize the electrical characteristic of the coaxial and STP cables and the intra-pair skew impact on high-speed serial interface such as HDMI.

In this paper, the electrical characteristic and eye-diagram of the coaxial and STP cables were verified with the intra-pair skew in frequency and time domain. The coaxial and STP cables were designed using HFSS 3D simulation tool shown in Fig. 1(a) for the accurate close-to real skew generation. In addition, in Fig. 1(b), the skew was distributed on the cables for the real unbalanced cables. Therefore, the intra-pair skew impact on the both type of HDMI cables was compared and verified by using the accurate setup and method.

II. FREQUENCY-DOMAIN INTRA-PAIR SKEW ANALYSIS AND EYE-DIAGRAM OF COAXIAL, STP CABLES AND ASSEMBLIES ON HDMI INTERFACE

As depicted in Fig. 1(a), the model of the coaxial and STP cables are designed by the parameters with 50Ω matching. The coaxial cable had smaller diameter (39 AWG) than the STP cable (36 AWG) for coaxial cable's flexible and thin characteristics. Therefore, for the precise comparison of each cables, the coaxial cable and the STP cable were designed as 2 m and 4 m respectively to have similar insertion loss.

The skew value was designated as about 40 ps which is 0.5 UI. 1 UI is 83.3 ps for 12 Gbps of HDMI 2.1. Fig. 1(b) shows that the intra-pair skew was located between the segment of each cables. Since the skew was distributed and segment length of cables was 0.1m, each skew level was 2 ps and 1 ps for the coaxial and STP cables respectively.

For the frequency-domain analysis, the equations for the skew calculation is needed. As port is depicted in Fig. 1(b), S2d1 means that relationship between incident differential wave at port 1,3 and received single wave at port2. S4d1 has same meaning instead of single wave port4. S2d1 and S4d1 is called the modified mixed-mode insertion loss. The equations are derived from the modified mixed-mode S parameter [5-7].

$$S2d1 = \frac{1}{\sqrt{2}}(S21 - S23) \quad (1)$$

$$S4d1 = \frac{1}{\sqrt{2}}(S43 - S41) \quad (2)$$

S2d1 and S4d1 contain the coupling information between P and N channels. Therefore, to calculate the intra-pair skew between P and N channels, the skew is calculated from the phase differences between S2d1 and S4d1.

$$\begin{aligned} t_1(f) &= \frac{-\text{unwrap}(\text{phaserad}(S2d1))}{2\pi f} \\ t_2(f) &= \frac{-\text{unwrap}(\text{phaserad}(S4d1))}{2\pi f} \\ \text{Skew}(f) &= t_2(f) - t_1(f) \end{aligned} \quad (3)$$

For the differential loss of 4 port transmission line, the loss can be written as (4) using equations (1), (2).

$$SDD21 = \frac{1}{\sqrt{2}}(|S2d1| + |S4d1|) \cdot \cos(\pi \cdot f \cdot \text{skew}(f)) \quad (4)$$

$$SDD21 = 0 \text{ if } f = \frac{1}{2 \cdot \text{skew}(f)} \quad (5)$$

The resonant frequency of the dip on SDD21 can be estimated using the equation (5) when the intra-pair skew is happened.

Fig.2 shows the intra-pair skew impact on the coaxial cable and STP cable. Fig. 2(a) demonstrates the coaxial cable with 40 ps skew had constant 40 ps skew in frequency-domain. In addition, the cosine equation (4) had zero value at 12.78 GHz. At this frequency, SDD21 of the coaxial cables with the skew had a resonant dip which magnitude was even -65 dB.

In contrast, in Fig. 2(b), the intra-pair skew in frequency-domain of the STP cable represented 40 ps value only at DC frequency and approached to zero value with fluctuation. Due to the tightly-coupled characteristic between P and N channels

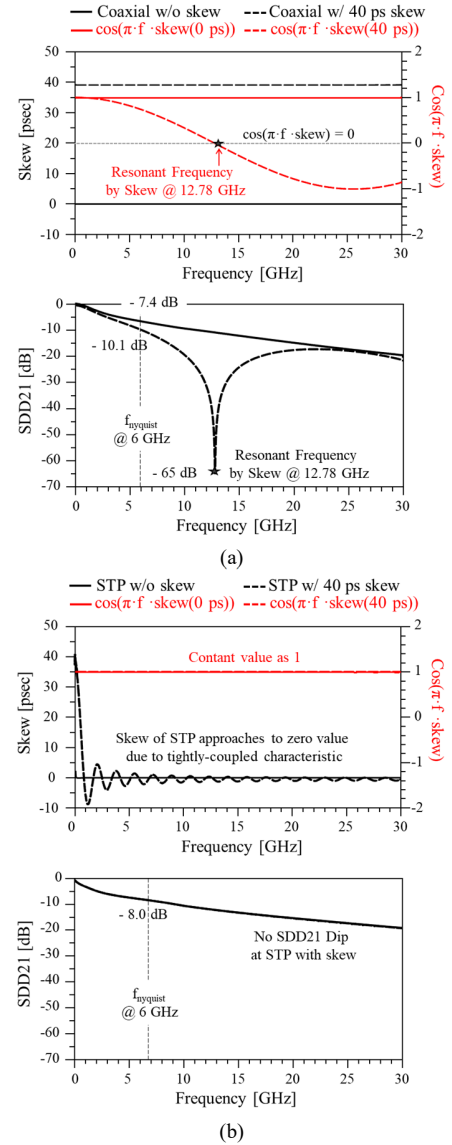


Fig. 2. Frequency-domain analysis of 40 ps intra-pair skew on (a) coaxial and (b) STP cables only.

in STP cable, the intra-pair skew was coupled to the other single-ended channel and the phase difference of modified mixed-mode insertion loss became zero. Furthermore, the value of cosine had constant 1 value in all frequency range. Therefore, SDD21 without and with the intra-pair skew were same.

Fig. 3 shows the differential loss and the mode conversion of the coaxial and STP cable assemblies with connectors when 40 ps skew is implemented or not. As shown in Fig. 3(a), SDD21 of the coaxial cable with the skew had large dip at 12.78 GHz which was calculated using the equations and verified in Fig. 2. Besides, since the cables were made in unbalanced cable, the mode conversion level increased by the intra-pair skew. On the other hand, SDD21 of the STP cable assembly did not have any dip and was not distorted by the intra-pair skew. In addition, even the STP cable was being unbalanced cable by the adding skew, the mode conversion had little change compared to the mode conversion without the skew.

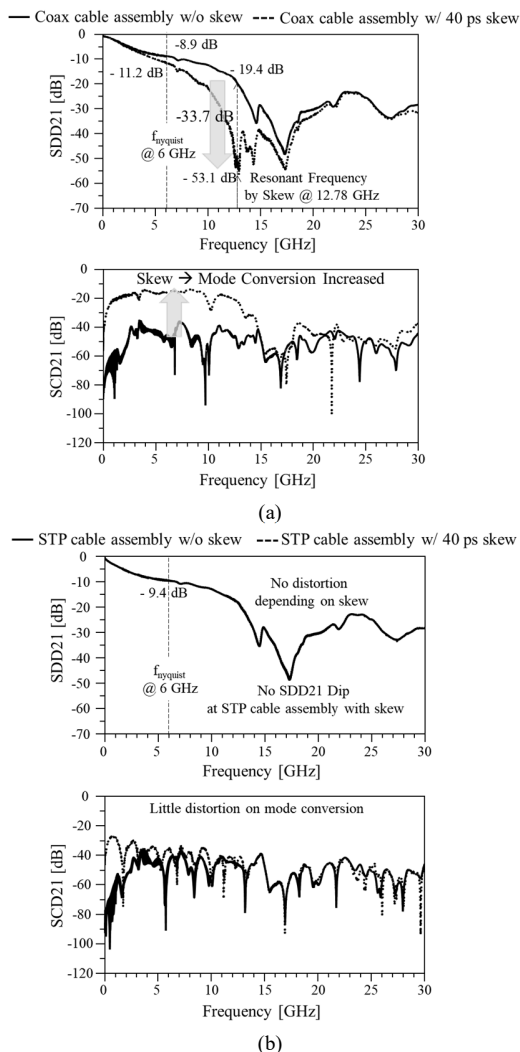


Fig. 3. Frequency-domain analysis of 40 ps intra-pair skew on (a) coaxial and (b) STP cables assemblies.

For eye-diagram simulation of HDMI interface, ADS circuit tool with 1-tap DFE adaptation and Continuous Time Linear Equalization (CTLE) function was used. As shown in Fig.4, in the coaxial cable assembly case, the eye with the skew had smaller eye opening than the eye without the skew due to the decrease of SDD21. The eye width and height were decrease by 7 ps and 77 mV respectively. In contrast, the eye opening of the STP cable assembly did not have change depending on the skew.

III. CONCLUSION

In this paper, the coaxial and STP cable model were proposed and used for the accurate intra-pair skew generation and unbalanced cable design. The intra-pair skew was located in distributed position between the each segment of the cables for the real cable skew cases. For the coaxial cable, due to uncoupled characteristic of the cable, the intra-pair skew represented in frequency-domain and demonstrated that the differential loss, mode conversion and eye-diagram was degraded. For the STP cables, even the intra-pair skew was implemented to P channel, it seemed there were not any skew in

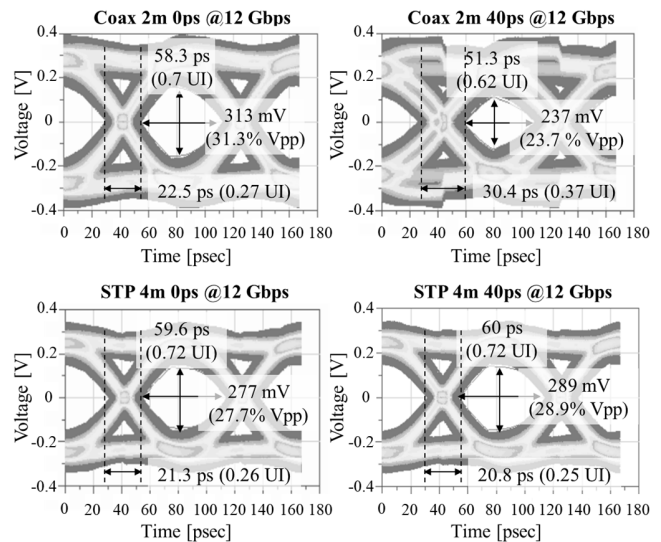


Fig. 4. The eye diagram of the coaxial and STP cable assembly with intra-pair skew and without intra-pair skew.

cables. It was verified that the intra-pair skew fluctuated in frequency-domain and approached to zero value. Furthermore, the eye-diagram was not distorted by the skew.

ACKNOWLEDGMENT

We would like to acknowledge the technical support from ANSYS. This work was supported by the Technology Innovation Program (20015559) funded by the Ministry of Trade, Industry & Energy(MOTIE, Korea). This research was supported by National R&D Program through the National Research Foundation of Korea(NRF) funded by Ministry of Science and ICT (NRF-2022M317A4072293). This work was supported by Samsung Electronics Co., Ltd (IO201207-07813-01).

REFERENCES

- [1] H. Park et al., "Signal Integrity Design and Analysis of a HDMI 2.1 Connector for Improved Electrical Characteristics," 2021 IEEE Electrical Design of Advanced Packaging and Systems (EDAPS), 2021, pp. 1-3, doi: 10.1109/EDAPS53774.2021.9657005.
- [2] D. Lho et al., "Design and Analysis of HDMI 2.1 Connector for Crosstalk Reduction using Tabs and Inverse Tabs," 2021 IEEE 30th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS), 2021, pp. 1-3, doi: 10.1109/EPEPS51341.2021.9609231.
- [3] Amendra Koul et al. "Fiber weave effect: Modeling, measurements, challenges and its impact on differential insertion loss for weak and strong-coupled differential transmission lines," DesignCon 2018
- [4] Christopher White et al. "Skew impact estimation on High Speed Serial Channels using mathematical analysis and accurate lab measurements," DesignCon 2010
- [5] Hansel Dsilva et al. "Mathematically De-mystifying Skew Impacts on 50G SERDES Links," DesignCon 2017
- [6] S. -J. Moon et al., "Intra-Pair Skew Metric, EIPS (Effective Intra-Pair Skew)," 2021 IEEE 25th Workshop on Signal and Power Integrity (SPI), 2021, pp. 1-4, doi: 10.1109/SPI52361.2021.9505188.
- [7] S. Baek, E. Lee and B. Sung, "Computation of Intra-pair Skew for Imbalance Differential Line using Modified Mixed-mode S-parameter," 2007 IEEE Electrical Performance of Electronic Packaging, 2007, pp. 179-182, doi: 10.1109/EPEP.2007.4387154.

NEXT Effect in Pin-area Routing at Receiver End from Via to Trace Coupling in a 32 Gb/s Channel

Pavel Roy Paladhi*, Yanyan Zhang, Xianbo Yang, Nam Pham, Megan Nguyen, Mahesh Bohra, Junyan Tang, Sungjun Chun, Joshua Myers, Wiren Becker, and Daniel Dreps

IBM Corporation, Austin, TX 78758

* Pavel.Roy.Paladhi@ibm.com

Abstract— With increasing bandwidth and higher transmission data rates in each generation, routing density in motherboards especially under module area have also increased proportionally. Maintaining signal integrity of high-speed channels under such dense routing conditions is becoming more challenging in each new product generation. This paper shows how via to trace coupling in under LGA area can give rise to increased NEXT values thereby causing channel margin loss and failure at high data rates.

I. INTRODUCTION

In every generation of server computers, high speed buses evolve drastically with increasing speeds of operation and larger data bandwidths. With higher operating frequencies in each generation, as the structural dimensions become proportional to the wavelengths, many propagation effects through PCB laminates, factors which could have been ignored as higher order effects only one generation prior become serious constraints, e.g. see [1]. In a high-speed channel's physical design, maintaining signal integrity involves controlling these major aspects like loss, impedance match, noise immunity and crosstalk. Dense wiring areas especially through via fields need to be carefully planned as crosstalk becomes more likely and can cause channel failure without upfront mitigation through design rules and this has been an active field of research [2],[3]. Coupled noise increases with frequency and becomes a fundamental barrier towards maintaining the desired data rates at the acceptable bit error rates and jitter tolerances. This paper focuses on a specific type of near end cross talk effect (NEXT). The authors show how coupling effect from a field of aggressor differential via pairs to a differential stripline victim trace could easily stifle channel performance. While the effect is expected, the denser layouts needed to handle larger number of high-speed lanes in successive generations in addition to higher speeds of operation make such crosstalk coupling scenarios almost inevitable and some investigations are being seen in recent literature [4]. Studies like in this paper provides some baseline to make decisions and rules of thumb on how much coupling may be allowed in the layout while still maintaining signaling performance goals. This can be useful towards cost control during the layout phase. The effects shown here would depend very much on the topology and each scenario in successive generations would need detailed analysis to avoid potential channel failures in future high-speed channels. As such, studies like this will become an important part of the PCB design process in successive generations.

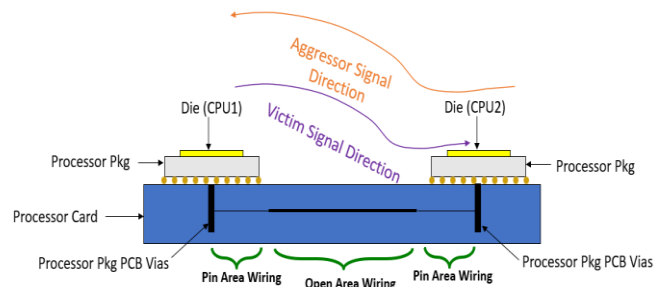


Figure 1. Topology of a 32 Gb/s CPU to CPU high speed channel.

II. CROSSTALK FROM VIA TO TRACE COUPLING

In this paper the authors show the effect of crosstalk coupling from a via-field of aggressors to a victim differential trace. We consider here a generic topology where signaling occurs between two processors on a motherboard. A cartoon representation of the topology is shown in Fig. 1. The victim net carries the signal from CPU1 (TX side) to the CPU2 (RX side). The path includes package wiring, HLGA transition into PCB and under package area wiring on both TX and RX side CPUs and PCB open area wiring in between. The aggressor nets start at the opposite side CPU2 i.e., receiving end of the victim signal and end at CPU1. The coupling and crosstalk which is the focus of this study occurs at the victim net's receiver side pin area wiring under the package where the victim is made to pass through a via field of aggressor differential pairs inducing a near end crosstalk coupling to the victim trace near its receiving end.

This channel was modeled and simulated to see the crosstalk effects. Further, a test setup representative of the above topology was also used to generate experimental data and perform some qualitative correlations. The modelling and measurement results are presented in the following sections.

III. MODELING & SIMULATION STUDIES

The modeled topology included the following: 30mm of package wiring at CPU1 side; the PCB wiring was done on Megtron-6 like material with total length of 15.7". Of that, 0.4" of wiring was in the under-package pin area at CPU1 and 1.7" in the under-package pin area under CPU2. The PCB had PTH vias with maximum stub length of 15 mil. The crosstalk effect studied occurs in the latter pin area of the victim's receiver side. Finally, there is about 40 mm of package wiring at CPU2 side.

Full channel 3D modeling was done using ANSYS HFSS. Each subsection mentioned above were modeled separately and then s-parameters were cascaded to generate the full channel model. In this topology, the operational data rate was 32 Gb/s, hence the fundamental frequency is 16 GHz. The s-parameters were generated from 20 MHz to 50 GHz, in steps of 20 MHz. Thus, the model captures the signaling effect up to the third harmonic. The full channel had one victim differential pair, eight differential pairs as far end crosstalk (FEXT) aggressors and another nine differential pairs as near end crosstalk (NEXT) aggressors. All the differential pairs were designed with the target onboard differential impedance of 85Ω . The differential insertion and return losses (DIL and DRL respectively) of the victim channel have been plotted in Fig. 2. At the operating frequency, the DIL of the channel is 24.6 dB.

The victim net is made to pass through a via field of nine differentially paired NEXT aggressors as shown in the Fig. 3 below. This occurs at the under-package pin area at the receiving end (CPU2). The aggressor vias have a breakout layer lower than the victim trace layer in the PCB. Ideally such via fields need to be avoided completely. However, in a tightly packed escape region, in practical scenarios complete bypassing may be unrealizable due to the combined global constraints which are applicable while designing the PCB layout. It would be interesting to know how many such via pairs could be encountered by the victim channel before its performance breaks.

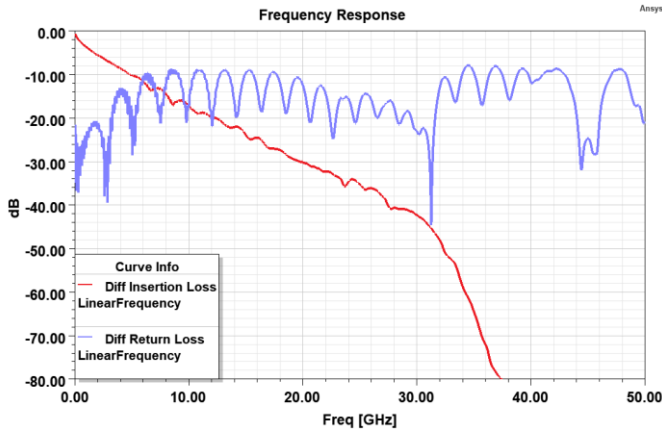


Figure 2. Frequency response of the victim channel

The structure in Fig. 3 was modelled in HFSS to calculate the crosstalk coupling from the via pairs to the victim in this section. This effect is shown in Fig. 4 where the power sum of near end crosstalk (PS_NEXT) from this section is plotted as the aggressor via pairs are turned on sequentially. As expected, the PS_NEXT from the via field can be seen to proportionately increase as the via pairs are sequentially turned on. This increase in PS_NEXT would have negative effect on the channel performance. To capture the extent of this degradation on the actual performance, a time domain transient simulation was performed on the full channel to observe the effect on the eye opening.

As mentioned earlier, the s-parameters for the full channel were generated through 3D modeling. These were then used in IBM's proprietary tool HSSCDR to perform time domain simulations and obtain eye opening values. To maximize the eye opening, the transmitters have FFE equalization (with 2 taps including 1 precursor) while the receiver has CTLE and DFE capabilities. The simulation was performed in statistical mode and the eye openings were recorded at a bit error rate (BER) of 10^{-15} . There were 32 CDR steps per unit interval.

The NEXT aggressors were turned on sequentially. Eye simulations showed a proportional decrease in the eye opening as the aggressors were sequentially turned on. With a single aggressor turned on, the horizontal eye (Heye) opening was 33.2%UI. The eye margin reduced by about a percentage each time a new aggressor was turned on. With all the NEXT aggressors turned on simultaneously the horizontal eye opening reduced all the way to 25.2% UI. The crosstalk from the HFSS model thus shows direct correlation with the full channel time domain eye simulation values. The simulation results, i.e. the eye margin values are tabulated in Table-1 below.

Generally, this eye reduction value would depend on the relative strength of the victim and aggressor signals. In long and lossy channels, the NEXT aggressors would be stronger, and it would be expected that the %UI reduction per via pair would be larger. The converse would also hold true. More importantly, based on a specific topology, such a scenario can be modeled and simulated to estimate the eye-opening reduction per via pair and wiring guidelines could be created on how many via to trace couplings could be allowed in a given topology during the PCB physical design phase.

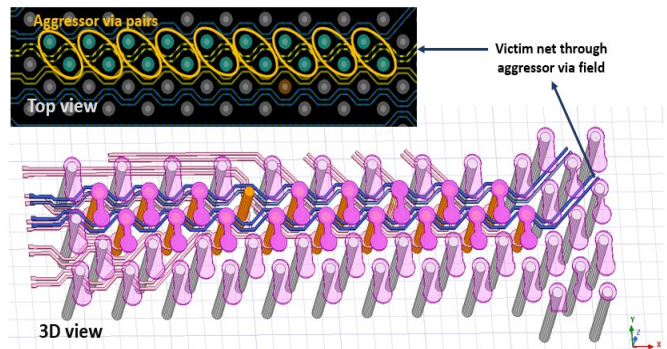


Figure 3. Victim channel (differential trace) passing through a via field of aggressors at the receiver side pin area wiring under CPU package. The victim passes in between the 'p' and 'n' vias of the differential aggressor nets

IV. MEASUREMENTS

The experimental setup used to perform a qualitative evaluation of NEXT coupling from via pairs to the trace is described in this section. To generate the measurement data, a setup similar to that shown in Fig. 1 was used. At the receiver end (CPU2), bathtub curves were collected and compared for two scenarios: a) once with the NEXT aggressors turned off and then b) with all NEXT aggressors turned on. The FEXT aggressors remain turned on in both cases. The eye opening is measured in terms of 'open ticks'. In this setup, 64 ticks

represent a 100%UI Heye opening, thus having 63 steps in a unit division and the available resolution (of Heye opening) is $\sim 1.6\%$ UI per tick or division.

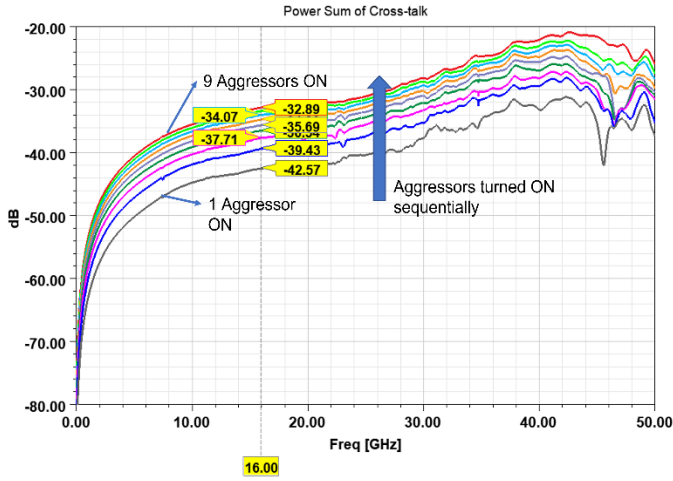


Figure 4. Power sum of crosstalk as via pairs are sequentially incorporated into the power sum.

Table 1 Eye simulation results: Heye progressively decreases as NEXT aggressors are turned on sequentially

No. of Aggressor via pairs turned on	Horizontal eye width (%UI)	Vertical height (mV)
0	33.2	51
1	32.1	49
2	31.1	46
3	30.2	44
4	29.3	43
5	28.3	42
6	27.4	40
7	26.3	38
8	25.2	35

The measured bathtub curves are plotted together in Fig. 5 below. The x-axis represents the measured time steps at the receiver CPU2 ('open ticks'). Without the aggressors turned on, the bathtub showed 12 ticks at BER-12, which equates to a 19% UI Heye opening. With all the NEXT aggressors turned on, the bathtub shows about 6 ticks which would equate to 9.5%UI Heye opening. It is worth noting that in the bathtub curve with all aggressors ON, there is no data point recorded at BER-12 which suggests that no bit error was reported at that level. Hence the nearest available datapoints to BER12 level were used to calculate the minimum horizontal eye width from the bathtub curve. The drastic eye reduction is caused by the crosstalk coupling from the via field of aggressors to the victim trace. The effect is magnified as the aggressors are 'near end', being located at the pin area near to the receiver (CPU2) of the victim channel. Thus, at the coupling area, the aggressor signal strength is stronger than the victim signal. Also, turning all the aggressors on, the total Heye reduction was about 9.5%UI meaning about 1.05% horizontal eye reduction occurred per aggressor being

turned on. This follows very closely the pattern observed in the simulation results which showed about 1%UI Heye margin reduction per aggressor.

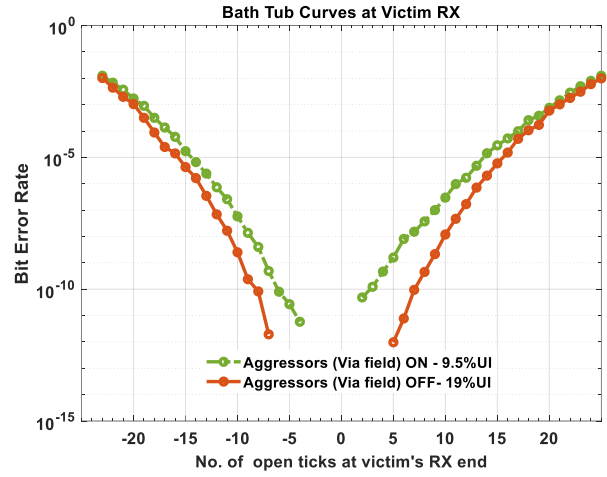


Figure 5. Experimental data: Bath-tub curves are shown at victim channel's receiver end with and without tuning on the via field of NEXT aggressors.

V. CONCLUSION

This study showed the effect of near end crosstalk from via to trace coupling on a high-speed channel. The study was performed at 32 Gb/s data rate. With increasing operational bandwidths and data rates, the crosstalk from via to trace coupling has increased impact on the eye margins especially if the coupling occurs near the receiver end. The effect on eye opening is studied when the victim trace passes through a field of aggressor differential vias in the pin area wiring under its receiver CPU package. The effect will depend on multiple factors, such as the relative signal strengths of victim and aggressor signals, escape layer arrangement between the aggressors and the victim trace, the topology of the high-speed net, data rates of operation etc. As wiring density increases with each generation, sometimes avoiding such via fields become hard to implement. Modeling and simulation studies become important to determine the extent to which such via to trace coupling can be tolerated while maintaining the target channel performance.

REFERENCES

- [1] P.R. Paladhi et al., "Effect of NEXT coupling in close proximity to receiver of 25Gb/s bus", 2017 IEEE 26th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS), 2017, pp. 1-3.
- [2] Y. Shim and D. Oh, "Improved PCB via pattern to reduce crosstalk at package BGA region for high speed serial interface," 2014 IEEE 64th Electronic Components and Technology Conference (ECTC), 2014, pp.1896-1901.
- [3] H. Johnson, "BGA Crosstalk", http://www.sigcon.com/vault/pdf/8_03.pdf
- [4] G. Pitner et al., "'BGA Routing Impact on High-Speed Signals," 2021 IEEE 30th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS), 2021, pp. 1-3.

Realistic Stripline Corner Modeling Using Surrogate Model and Topographic Fitting

¹Andrew Page, ¹Matteo Cocchini, ¹Zhaoqing Chen, ²Xu Chen

¹IBM Systems, Poughkeepsie, NY 12601, USA

²Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61820, USA
ajpage2@illinois.edu, mcocchi@us.ibm.com, zhaoqing@us.ibm.com, xuchen1@illinois.edu

Abstract—This paper demonstrates a method to extract impedance-attenuation corners of a stripline with user-prescribed confidence levels. This is done using a sparse-grid-based surrogate model to quickly generate vast Monte Carlo datasets from which the impedance-attenuation distribution is calculated. Ellipses are fit to this distribution as equi-density contours to enclose a proportion of the solution data. Appropriate corners can be read off these ellipses and applied to broadband simulation. The results are compared against three measured test coupons, showing capability to analyze a PCIe Gen. 5 link. Realistic modeling of geometries and material variations is emphasized.

Index Terms—corner model, sparse grid, Monte Carlo

I. INTRODUCTION

Electrical interconnect performance is becoming increasingly difficult to characterize and control as the demand for fast data rates grows. Imperfect manufacturing processes lead manifested designs to not possess their intended physical properties and provide uncontrollable variation in performance characteristics. Interconnect characterization based on the nominal design is thus unrealistic. The worst-case behavior can be studied but represents a performance outlier. Corner modeling is a way to capture realistic variation from the nominal behavior. Current corner modeling algorithms rely on fast but unreliable boundary scan methods or expensive Monte Carlo (MC) procedures [1], [2].

The method proposed by this paper combines the best of both approaches by creating a surrogate model that accurately and efficiently maps cross-section and material parameters to corresponding attenuation and impedance (α, Z_0) values. This model is used to generate vast amounts of MC data, which is then binned into a 2D density. Ellipses centered on the nominal solution are fit as effective equi-density contours using an iterative algorithm. The high/low impedance/attenuation (HZLA, LZLA, HZHA, LZHA) corners can be read from the contour, and the corresponding parameter configurations can be found easily. They are then inputted to Ansys 2D Extractor for broadband S-parameter extraction and eye diagram simulation.

Section II introduces the EM simulation methodology, the surrogate model development, the contour fitting and the corner identification algorithm. Section III discusses the results

This material is based upon work supported by the National Science Foundation under Grant No. CNS 16-24811 - Center for Advanced Electronics through Machine Learning (CAEML) and its industry members.

978-1-6654-5075-1/22/\$31.00 ©2022 IEEE

of the procedure and the eye characteristics of each calculated corner, with concluding remarks offered in Section IV.

II. THE CORNER MODELING PROCESS

A. EM Simulation & Parameter Variation

The corner location procedure will be done using both a single-ended and a differential stripline, defined in Table I and Fig. 1. These models are solved with Ansys 2D Extractor at a single frequency to extract the RLGC parameters. The single-ended line's attenuation α and characteristic impedance Z_0 can be calculated at solution frequency $f = 20\text{GHz}$ via

$$\alpha = \text{Re}[(R + j\omega L)(G + j\omega C)], \quad (1)$$

$$Z_0 = \text{Re}[(R + j\omega L)/(G + j\omega C)], \quad (2)$$

where $\omega = 2\pi f$. The differential impedance for the two-line structure is found by extracting RLGC for the left half of Fig. 1(b) and doubling the calculated impedance from (2). The attenuation is found directly from (1). Placing a conducting boundary as shown rejects all common-mode fields.

The 2D models are built to reflect consequences of realistic manufacturing procedures, including a trapezoidal conductor with angle ϕ to model chemical etching effects and holding pitch $s + w$ constant due to tight photolithography tolerances. The dielectric constant ϵ_r and tangent delta $\tan \delta$ were measured at 20GHz using short pulse propagation (SPP) [3].

A parameter ξ with a symmetric tolerance $\xi = \mu \pm 3\sigma$ is modeled as a Gaussian random variable with mean μ and standard deviation σ , i.e. a parameter's tolerance represents its

TABLE I
PARAMETERS & TOLERANCES

Name	Symbol	Nominal	Tolerance	Unit
Perm.	ϵ_r	3.25	+0.25/-0.15	-
Loss Tan.	$\tan \delta$	0.00325	+0.0024/-0.0020	-
Thickness	t	0.6	+/-0.1	mil
Width	w	4.7	+/-0.6	mil
Prepreg size	h_1	4.2	+/-1.0	mil
Core size	h_2	4.0	+/-0.7	mil
Pitch	$s + w$	9	-*	mil
Gnd. width	w_{gnd}	200	-	mil
Gnd. thick.	t_{gnd}	1.2	-	mil
Edge angle	ϕ	78.23	-	deg
Cu resist.	ρ	1.9	-	$\mu\Omega\text{-cm}$

*A dashed tolerance represents a fixed parameter.

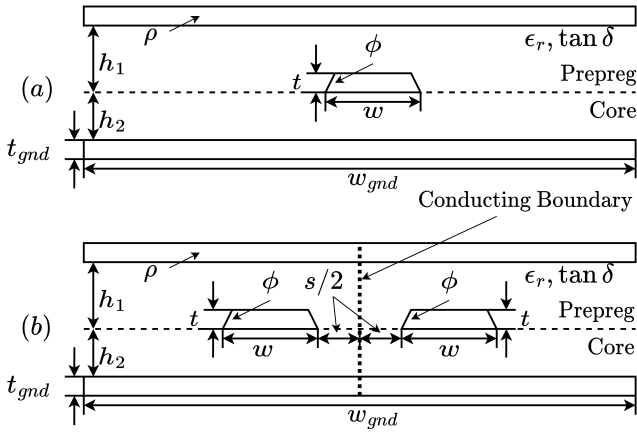


Fig. 1. Cross section of single-ended (a) and differential-mode (b) striplines. Conducting boundary can be used for efficient differential-mode impedance extraction.

3σ value. Asymmetric parameters, e.g. $\xi = \mu + 3\sigma_2 / -3\sigma_1$, are modeled with a modified Gaussian distribution, characterized by probability density function with a spread of σ_1 below and σ_2 above the center value μ :

$$f(u) = \frac{2}{(\sigma_1 + \sigma_2)\sqrt{2\pi}} \begin{cases} \exp\left(-\frac{(u-\mu)^2}{2\sigma_1^2}\right), & u \leq \mu, \\ \exp\left(-\frac{(u-\mu)^2}{2\sigma_2^2}\right), & u > \mu. \end{cases} \quad (3)$$

B. Sparse Grid Surrogate Model

Surrogate modeling provides a way around the computational expense of MC simulation. The Tasmanian software package allows creation of such a model based on sparse grid collocation [4]. The sparse grid is a set of points in the design space that serves as a high-dimensional interpolation grid. Simulations can be run for each point in the grid, and the results can be interpolated. This interpolation can be evaluated in a split-second, and yields a good approximation to the 2D simulation of the same input. Details can be found in [5].

Interpolations for α and Z_0 at 20GHz are formed by simulating RLGC for each point on the sparse grid with 2D Extractor and interpolating the calculated (α, Z_0) over the sparse grid. Several surrogate models were tested against 4,000 MC datapoints, which each model computed in less than 110ms with mean errors less than 1% and 0.1% in α and Z_0 , respectively, demonstrating the sparse grid's reliability. The selected model used a precision 3 'level' grid based on the Clenshaw-Curtis rule, taking 389 simulations to form, which is a one-time cost similar to a boundary scan [1].

C. Fitting the Equi-Density Contour

Locating the corners begins by first finding the distribution of (α, Z_0) based on parameter values from Table I. This is done with a massive MC batch to be run with the surrogate model. This paper uses two million samples, a job that would take 2D Extractor days to calculate but the surrogate model can perform in seconds. The distribution can be approximated

by binning the (α, Z_0) scatter data into $n_x \times n_y = 200 \times 200$ cells and counting the number of solutions lying within each cell. The resulting density histogram resembles a hill peaking near the nominal solution. The shape and orientation of the distribution is due to the near-Gaussian input parameters and the slight nonlinear dependence of (α, Z_0) on the input parameters. This histogram can be characterized by equi-density contours, as in a topographical map. The histogram's shape suggests the ellipse as a good equi-density contour.

Three coordinate systems are defined to help identify the ellipse: (α, Z_0) is the main working system, (x, y) results from centering and normalizing (α, Z_0) , and (η, ν) is a version of (x, y) rotated by θ_f so the η axis aligns with the semimajor axis of the ellipse, as summarized in (4) and (5). Attenuation and impedance ranges are denoted $\Delta\alpha$ and ΔZ_0 .

$$R = \begin{pmatrix} \cos \theta_f & -\sin \theta_f \\ \sin \theta_f & \cos \theta_f \end{pmatrix}, \quad M = \begin{pmatrix} \Delta\alpha/n_x & 0 \\ 0 & \Delta Z_0/n_y \end{pmatrix}, \quad (4)$$

$$\begin{pmatrix} \alpha \\ Z_0 \end{pmatrix} = \begin{pmatrix} \alpha_{nom} \\ Z_{0,nom} \end{pmatrix} + M \begin{pmatrix} x \\ y \end{pmatrix}, \quad \begin{pmatrix} x \\ y \end{pmatrix} = R \begin{pmatrix} \eta \\ \nu \end{pmatrix}. \quad (5)$$

The ellipse will lie in (α, Z_0) space centered on the nominal solution $(\alpha_{nom}, Z_{0,nom})$. Its orientation can be found by first specifying a threshold percentage of the peak density, like an equipotential contour on a topographical map. A sweep of the angle from the x -axis, θ , tracking the distance r to the prescribed threshold in (x, y) -space is fit into (6). The tilt angle θ_f and semimajor/minor axes a and b thus can be recovered.

$$r^2(\theta) = \left(\frac{\cos^2(\theta - \theta_f)}{a^2} + \frac{\sin^2(\theta - \theta_f)}{b^2} \right)^{-1}. \quad (6)$$

This procedure is iterated by lowering the threshold until the ellipse encloses a certain proportion of the solution data, i.e. a given inclusion rate. A solution's inclusion within the contour can be tested by evaluating the face equation of (7) after appropriate transformation to (η, ν) using (4) and (5):

$$\text{contour: } \frac{\eta^2}{a^2} + \frac{\nu^2}{b^2} = 1; \quad \text{face: } \frac{\eta^2}{a^2} + \frac{\nu^2}{b^2} \leq 1. \quad (7)$$

D. Corner Identification

The corners can be read off this ellipse through any desired means. This paper employs a procedure based on [1]. The ellipse can be bound by a rectangle sharing its extreme α and Z_0 values, i.e. $\alpha = \alpha_{nom} \pm \delta\alpha$, $Z_0 = Z_{0,nom} \pm \delta Z_0$. The impedance is then scaled by $C = \delta\alpha / \delta Z_0$ after shifting to the origin, which stretches the rectangle into a square of length $2\delta\alpha$ and scales the ellipse impedance extrema to $\pm\delta\alpha$. The corners are read as each intersection of the ellipse and the diagonals of the square, reverting to (α, Z_0) using the scaling.

III. NUMERICAL EXAMPLES

The procedure outlined in Section II was applied to single-ended and differential models with 1σ , 2σ and 3σ inclusions (68%, 95% and 99%). The corners were located using the scaling method, as shown in Fig. 2 for the 3σ differential case. The three differential-line ellipses are shown with the

MC scatter data in Fig. 3. Table II lists the parameters for each 3σ corner, each behaving as expected; low-Z corners share high widths, low-A corners share low loss tangents.

Three test coupons were measured for model validation. The (α, Z_0) of each coupon was found using VNA measurements and simulated TDR, and are plotted in Fig. 3. Each fall within the 2σ contour, demonstrating the inclusion as a confidence in the performance of a manufactured board. The four 3σ corners were run in a broadband simulation to compute their S-parameters as 7cm lines. The dielectric was modeled with a multipole Debye fit based on measured data [6]. The insertion losses of each model, summarized in Table III, are compared at the fundamental frequencies of PCIe Gen. 4 and 5 (8 and 16GHz, respectively), and the material characterization frequency of 20GHz. The corners form a bound on the insertion loss, bounded below by HZLA and above by LZHA.

The eye was simulated using a PRBS-23 sequence with a rate of 32Gbps and a 7ps rise/fall time, to emulate the line as a PCIe Gen. 5 link. The results are summarized in Table IV. The high-A corners showed the smallest eye heights. The eye width sensitivity is accentuated by the line length, showing low sensitivity in general. A similar bounding on the eye characteristics based on the inclusion level is expected.

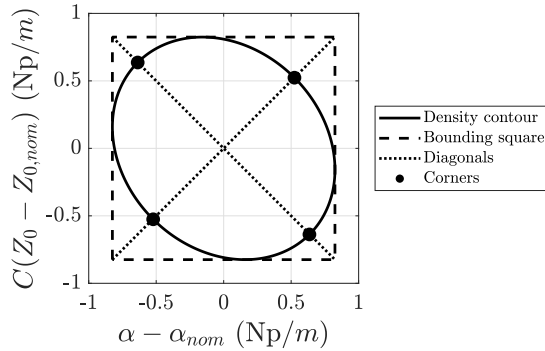


Fig. 2. Differential-line 3σ -inclusion ellipse and bounding square in scaled/centered attenuation-impedance space.

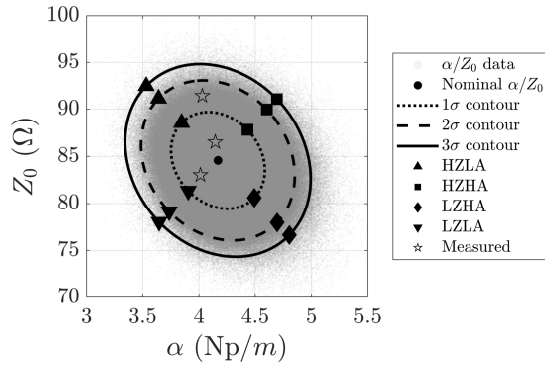


Fig. 3. Samples of differential-line MC data with derived equi-density contours and corners for 1σ , 2σ and 3σ inclusion.

IV. CONCLUSION

The proposed corner modeling scheme is capable of characterizing PCIe Gen. 5 interconnect performance. A well-trained surrogate model makes the process as expensive as a boundary scan. The scheme offers flexibility in the precise control of the inclusion rate. The simplicity of the contour choice allows for easy implementation while maintaining accurate results. Further work may include a more robust density-finding algorithm based on kernel density estimation.

REFERENCES

- [1] Z. Chen, "Transmission line attenuation-impedance realistic corner modeling by scaled-down tolerance boundary scan," in *2007 IEEE International Symposium on Electromagnetic Compatibility*, 2007, pp. 1–6.
- [2] N. Lu, "Statistical and corner modeling of interconnect resistance and capacitance," in *IEEE Custom Integrated Circuits Conference 2006*, 2006, pp. 853–856.
- [3] A. Deutsch, R. Krabbenhoft, K. Melde, C. Surovic, G. Katopis, G. Kopsay, Z. Zhou, Z. Chen, Y. Kwark, T.-M. Winkel, X. Gu, and T. Standaert, "Application of the short-pulse propagation technique for broadband characterization of pcb and other interconnect technologies," *Electromagnetic Compatibility, IEEE Transactions on*, vol. 52, pp. 266 – 287, 06 2010.
- [4] M. Stoyanov, D. Lebrun-Grandie, J. Burkardt, and D. Munster, "Tasmanian," 9 2013. [Online]. Available: <https://github.com/ORNLTasmanian>
- [5] M. Stoyanov, "User manual: Tasmanian sparse grids," Oak Ridge National Laboratory, One Bethel Valley Road, Oak Ridge, TN, Tech. Rep. ORNL/TM-2015/596, 2015.
- [6] A. E. Engin, I. Ndip, K.-D. Lang, and J. Aguirre, "Closed-form multipole debye model for time-domain modeling of lossy dielectrics," *IEEE Transactions on Electromagnetic Compatibility*, vol. 61, no. 3, pp. 966–968, 2019.

TABLE II
3 σ CORNER PARAMETER CONFIGURATIONS

Param.	ϵ_r	$\tan \delta$	t	w	h_1	h_2
Unit	-	-	mil	mil	mil	mil
Differential (fixed 9mil pitch)						
HZLA	3.25	0.00170	0.607	4.47	5.25	4.15
LZLA	3.23	0.00159	0.689	4.98	3.79	4.13
HZHA	3.22	0.00503	0.591	4.42	4.45	4.24
LZHA	3.26	0.00449	0.605	4.82	3.29	4.07
Single-Ended						
HZLA	3.12	0.00190	0.633	4.43	4.66	4.20
LZLA	3.34	0.00155	0.576	5.11	3.80	4.20
HZHA	3.43	0.00472	0.650	4.43	4.78	4.56
LZHA	3.23	0.00461	0.646	4.80	3.36	4.10

TABLE III
INSERTION LOSS COMPARISON AGAINST MEASUREMENT

f (GHz)	Insertion loss (dB)						
	3 σ Corners				Measured		
	HZLA	LZLA	HZHA	LZHA	#1	#2	#3
8	1.15	1.42	1.36	1.67	1.31	1.30	1.26
16	1.82	2.20	2.21	2.69	2.13	2.08	2.06
20	2.14	2.42	2.65	3.02	2.52	2.44	2.45

TABLE IV
EYE CHARACTERISTICS OF 7CM & 20CM LINE

	7cm line		20cm line	
	Height (mV)	Width (UI)	Height (mV)	Width (UI)
Nominal	379.34	0.9202	247.58	0.7545
HZLA	385.15	0.9261	260.61	0.7725
LZLA	368.73	0.9202	242.02	0.7345
HZHA	347.85	0.8603	213.38	0.7345
LZHA	355.33	0.9022	180.10	0.6826

Fast LDO Simulations via Parameter-Varying Linearized Macromodels

Tommaso Bradde, Stefano Grivet-Talocia

Dept. Electronics and Telecommunications, Politecnico di Torino, Italy
tommaso.bradde@polito.it, stefano.grivet@polito.it

Abstract—An approach for generating time-varying linearized macromodels of analog circuit blocks is presented. These models can be used to perform fast small-signal analyses characterized by nonstationary operating conditions, thanks to their certified stability. We validate the proposed approach by performing post-layout simulations of a Low DropOut (LDO) voltage regulator, in view of power integrity assessment applications.

I. INTRODUCTION

Analog Circuit Blocks (CBs) are fundamental components in virtually all electronic systems. When considered in advanced design stages, the behavior of these components is properly described in terms of large equivalent netlists that must take into account both the semiconductor models and the electrical characterization of the parasitics due to the circuit layout and packaging. As the number of CBs involved in modern Systems-on-Chip (SoC) and Systems-in-Package (SiP) is usually large, straight exploitation of such accurate yet complex descriptions within system-level simulations is often unfeasible due to an excessive computational cost. Thus, the availability of behavioral models for this kind of components is highly desirable [1], [2].

This contribution focuses on the generation of macromodels for CBs operating under small-signal conditions, characterized by a nonstationary working point. The latter can be determined, e.g., by changes of the system operation mode performed for the sake of energy management. The approach relies on the generation of a Linear-Parameter-Varying (LPV) reduced order model that approximates the local dynamics of the original system around its dynamic working point, as it evolves within prescribed limits [3]. The model is generated starting from samples of the circuit small-signal transfer function, retrieved in correspondence of a finite number of admissible bias configurations. Model parameterization is performed at runtime, by extracting the low frequency components of the electrical quantities at the circuit interface ports, which determine the instantaneous bias condition. Suitable numerical constraints are embedded in the model generation to guarantee the stability of the resulting LPV system for arbitrary working point trajectories. The method is applied to perform a fast post-layout simulation of a LDO circuit design, in view of possible applications for advanced power integrity optimization and assessment. Experimental evidence shows that the proposed models are accurate and guarantee up to 50× speedup in transient simulations.

II. PROBLEM SETTING

We consider a mildly nonlinear analog circuit block accessible from P electrical interface ports, whose behavior is described by the nonlinear differential equations

$$\begin{aligned}\dot{\xi}(t) &= F(\xi(t), u(t)), \\ \eta(t) &= G(\xi(t), u(t)),\end{aligned}\tag{1}$$

being $u(t), \eta(t) \in \mathbb{R}^P$ the system input and output signals, and $\xi(t) \in \mathbb{R}^N$ is the system state vector, with N large. F, G are nonlinear differentiable maps not known in closed form, but encrypted in an available SPICE netlist.

We want to obtain a reduced order behavioral model of (1) for small-signal analyses characterized by nonstationary working conditions, compatible with the input decomposition

$$u(t) = U_0(t) + \tilde{u}(t),\tag{2}$$

satisfying the following assumptions

- 1) $\tilde{u}(t)$ is a small-signal component with $\tilde{u}(0) = 0$.
- 2) $U_0(t)$, henceforth denoted as *bias component*, attains values within a (not necessarily small) hyper-rectangle

$$U_0(t) \in \mathcal{U}_0 = [a_1, b_1] \times \cdots \times [a_P, b_P], \forall t \geq 0.\tag{3}$$

Additionally, at each $t^* \geq 0$, there exists a small constant $\delta_\xi \geq 0$ such that

$$\|\xi(t^*) - \Xi_0(t^*)\|_2 \leq \delta_\xi,\tag{4}$$

being $\Xi_0(t^*)$ the unique asymptotically stable equilibrium point satisfying

$$\begin{aligned}0 &= F(\Xi_0(t^*), U_0(t^*)), \\ Y_0(t^*) &= G(\Xi_0(t^*), U_0(t^*)).\end{aligned}\tag{5}$$

Under the above assumptions, since (5) admits a unique solution, at each time instant system (1) operates in the neighborhood of the operating point determined solely by the instantaneous value of $U_0(t)$. This condition is practically verified when the bias component varies slowly with respect to the dynamics of the circuit of interest. Also, when (4) and (5) hold, the output of (1) can be decomposed as

$$\eta(t) = Y_0(t) + \tilde{\eta}(t)\tag{6}$$

where $Y_0(t)$ is the solution of (5) and $\tilde{\eta}(t)$ is the deviation from the corresponding equilibrium. Our macromodeling approach is based on the construction of a LPV model that approximates the local dynamics of (1) around the trajectory

$(\Xi(t), U_0(t), Y_0(t))$, induced by the instantaneous value of the bias component $U_0(t)$. Exploiting the model linearity, we proceed as follows

- 1) We build a LPV reduced order model for the map $\tilde{u}(t) \rightarrow \tilde{\eta}(t)$ (Sec. III-A).
- 2) We modify the above to guarantee that it also recovers the mapping $U_0(t) \rightarrow Y_0(t)$ (Sec. III-B).
- 3) We address a self-parameterizing approach to update the model working point at runtime (Sec. III-C).

III. MODELING FRAMEWORK

A. Reduced Order LPV Small-Signal Model

The analytical expression for the linearization of (1) around the trajectory $(\Xi(t), U_0(t), Y_0(t))$ reads

$$\begin{aligned} \dot{\tilde{\xi}}(t) &\approx \tilde{A}(U_0(t)) \cdot \tilde{\xi}(t) + \tilde{B}(U_0(t)) \cdot \tilde{u}(t), & \tilde{\xi}(0) &= 0 \\ \tilde{\eta}(t) &\approx \tilde{C}(U_0(t)) \cdot \tilde{\xi}(t) + \tilde{D}(U_0(t)) \cdot \tilde{u}(t). \end{aligned} \quad (7)$$

where we dropped the dependencies of the Jacobians linearizations on the state $\Xi(t)$ since we assumed that (5) has a unique solution. For the above, we desire the following reduced order representation of order $n \ll N$

$$\begin{aligned} \dot{\tilde{x}} &= A(U_0(t))\tilde{x} + B(U_0(t))\tilde{u}, & \tilde{x}(0) &= 0 \\ \tilde{y} &= C(U_0(t))\tilde{x} + D(U_0(t))\tilde{u}, & \tilde{y}(t) &\approx \tilde{\eta}(t). \end{aligned} \quad (8)$$

Since the maps F, G are unknown in closed form, we build (8) starting from data. To do this, we observe that for frozen time instants, (7) is associated with a transfer function parameterized by $U_0 \in \mathcal{U}_0$, that reads

$$\tilde{H}(s, U_0) = \tilde{D}(U_0) + \tilde{C}(U_0)(sI_N - \tilde{A}(U_0))^{-1}\tilde{B}(U_0). \quad (9)$$

We can thus build an approximation $H(s, U_0)$ of order n for (9) and then cast this approximation into a state space parameterized by the instantaneous value of U_0 to obtain (8). We start by retrieving samples of (9) via AC sweeps, performed for a finite number of frequency values and static bias configurations

$$\tilde{H}_{k,m} = \tilde{H}(j\omega_k, U_{0m}), \quad k = 1, \dots, K \quad m = 1, \dots, M. \quad (10)$$

The reduced order transfer function is obtained by enforcing

$$H(j\omega_k, U_{0m}) \approx \tilde{H}_{k,m}, \quad k = 1, \dots, K \quad m = 1, \dots, M. \quad (11)$$

via PSK iteration [4], based on model structure

$$H(s, U_0) = \frac{N(s, U_0)}{D(s, U_0)} = \frac{\sum_{i=0}^n \sum_{\ell \in \mathcal{I}_\ell} R_{i,\ell} \cdot b_\ell^\ell(U_0) \varphi_i(s)}{\sum_{i=0}^n \sum_{\ell \in \mathcal{I}_\ell} r_{i,\ell} \cdot b_\ell^\ell(U_0) \varphi_i(s)}. \quad (12)$$

In this model, $R_{i,\ell} \in \mathbb{R}^{P \times P}$, $r_{i,\ell} \in \mathbb{R}$ are unknowns, and $\varphi_i(s) = (s - q_i)^{-1}$ are partial fractions with $\Re\{q_i\} < 0$. The functions $b_\ell^\ell(U_0)$ are multivariate Bernstein polynomials with multidegree $\ell = (\ell_1, \dots, \ell_P)$, while \mathcal{I}_ℓ denotes a set of admissible indices. Model structure (12) admits a representation in terms of state space (8). Technical details about the employed realization procedure are available in [3].

During model generation, we find the involved unknowns by guaranteeing that the final realization (8) remains stable for every possible trajectory of $U_0(t)$, a property known as

quadratic stability [5]. This is possible thanks to the following Theorem, proved in [3].

Theorem 1 (Sufficient conditions for quadratic stability): let A_1 and B_1 be known constant matrices and

$$C_{1,\ell} = [r_{1,\ell}, r_{2,\ell}, \dots, r_{n,\ell}], \quad d_{1,\ell} = r_{0,\ell}. \quad (13)$$

Then LPV system (8) is quadratically stable if there exists $Q_1^* \in \mathbb{R}^{n \times n}$ such that $Q_1^* = Q_1^{*\top} \succ 0$ and

$$\begin{bmatrix} A_1^\top Q_1^* + Q_1^* A_1 & Q_1^* B_1 - C_{1,\ell}^\top \\ B_1^\top Q_1^* - C_{1,\ell} & -2d_{1,\ell} \end{bmatrix} \prec 0 \quad \forall \ell \in \mathcal{I}_\ell \quad (14)$$

Condition (14) represents a Linear Matrix Inequality in the model denominator coefficients. This constraint can be incorporated in the model training phase, so that the resulting constrained fitting problem becomes equivalent to a standard convex optimization problem, which is solved through standard optimization libraries.

B. Reconstructing the Bias Component

At each time instant, the circuit output component $Y_0(t)$ is instantaneously determined by the corresponding value of $U_0(t)$, via the equilibrium mapping (5). Input-output samples of this mapping can be obtained by performing a DC sweep of the circuit netlist for different constant values of U_0 . This procedure returns samples

$$Y_{0j} = Y_0(U_{0,j}), \quad j = 1, \dots, J, \quad U_{0,j} \in \mathcal{U}_0. \quad (15)$$

When the small-signal system (8) is subject to static input $U_{0,j}$, the corresponding DC output reads

$$H(0, U_{0,j})U_{0,j} \neq Y_{0j} \quad (16)$$

and is not expected to match the observations Y_{0j} because the bias component is not necessarily small. Therefore, we add a parameterized output correction term $Y_C(U_0)$ to the output equation of (8) in order to restore the desired equilibrium output for all bias conditions. This correction is generated by requiring that

$$Y_C(U_{0,j}) \approx Y_{0j} - H(0, U_{0,j})U_{0,j}, \quad j = 1, \dots, J. \quad (17)$$

The above is a standard multivariate function approximation problem that can be tackled via any standard approach (e.g. least-squares regression). Adding the correction in the model output leads to the final model structure

$$\dot{x} = A(U_0(t))x + B(U_0(t))u \quad (18)$$

$$y = C(U_0(t))x + D(U_0(t))u + Y_C(U_0(t)), \quad (19)$$

which is fed with the total input $u(t)$ and returns the total output approximation $y(t) \approx \eta(t)$.

C. Real Time Parameterization

Model (18) is thought to be parameterized in real time with the value of the bias component $U_0(t)$ determining the current working point. However, during system operation, this input term is not directly observable, as mixed together with the small-signal at the circuit interface ports. Thus, the model is practically usable only in view of an automated procedure

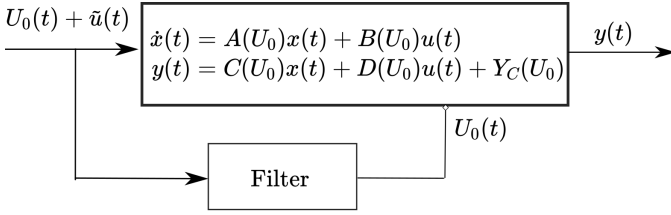


Fig. 1. Block diagram of the proposed self-parameterized LPV macromodel structure.

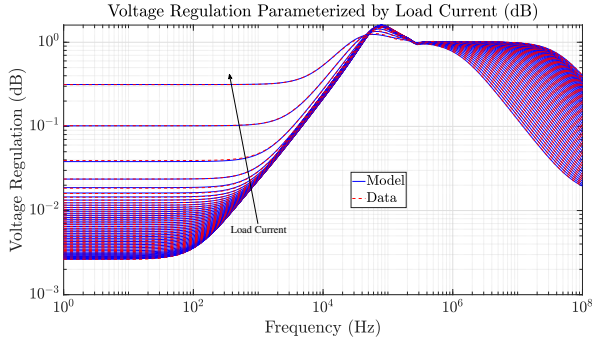


Fig. 2. Fitting of the LDO voltage regulation transfer function for $M = 50$ load current configurations.

aimed at isolating the bias and the small signal components. As assumptions (4), (5), require $U_0(t)$ to vary slowly with respect to the circuit dynamics, we perform real time parameterization as follows

- 1) During online operation we perform a low pass filtering operation over the the total input $u(t)$. This operation is performed based on a second order Butterworth filter.
- 2) We use the output of the filter to istantaneously parameterize model (18), as in Fig. 1.

In order to guarantee sufficiently slow variations of $U_0(t)$, we set the cut-off frequency of the filter to $\omega_c = 0.1\omega_p$, being ω_p the angular frequency of the slowest pole of model (12).

IV. FAST POST-LAYOUT LDO SIMULATION

To test the proposed modeling approach, we instantiated in Cadence environment the Only-MOS low-power regulator design proposed in [6]. The circuit was designed including the layout, using a 40 nm CMOS process, resulting in a 30 MB equivalent netlist.

We performed fast circuit simulation under nonstationary loading conditions, characterized by admissible bias components $U_0^1 \equiv V_{DD} = 0.9$ V, and $U_0^2 \equiv I_L \in [0, 10]$ mA, in agreement with the design specifications. To this aim we built the small-signal model (12) with $n = 9$, in Hybrid representation, considering as input the unregulated voltage V_{DD} at port 1 and the load current I_L at port 2. The model was generated in 8.6 s starting from AC data retrieved for $M = 50$ load current configurations and enforcing the required stability constraints (14). The voltage regulation transfer function of the model is compared with the reference data in Fig. 2. Once (12) is generated, the required DC correction term $Y_C(U_0)$ is computed by enforcing (17) via linear regression, using $J = 50$

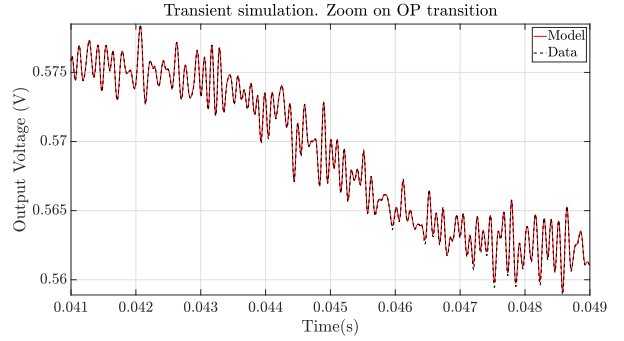


Fig. 3. Time domain validation of the proposed modeling approach.

data samples of the reference function (15). Finally, a low-pass filter with cut-off frequency $\omega_c = 2\pi 500$ rad/s was designed and an equivalent netlist for model structure of Fig. 1 was instantiated in LTSpice environment. In this environment, we performed a 0.2 seconds long transient analysis by considering a load transition from $I_L = 5$ mA to $I_L = 8$ mA, taking place in $\Delta t = 6$ ms. A small-signal of amplitude 0.2 mA and flat power spectrum in the band 1 – 10 kHz was added to the bias component I_L . The results of the simulation were compared with those obtained by performing the same analysis using the reference post-layout netlist. Fig. 3 shows the results of the comparison before ($t \in [0.41, 0.42]$ s), during ($t \in [0.42, 0.48]$ s), and after ($t \in [0.48, 0.49]$ s) the load current transition. In all these three situations, the model returns very accurate predictions of the circuit behavior. Using a common laptop, the model is simulated in 16 s, while the original netlist in 13 minutes, with a speedup of about $50\times$.

V. CONCLUSIONS

We presented an approach for generating macromodels of analog circuit blocks under small-signal operation with a nonstationary operating point. The resulting macromodels prove to be accurate at reproducing the circuit behavior, while at the same time guaranteeing significant runtime reduction.

REFERENCES

- [1] Z. Naghibi, S. A. Sadrossadat, and S. Safari, "Time-domain modeling of nonlinear circuits using deep recurrent neural network technique," *AEU-International Journal of Electronics and Communications*, vol. 100, pp. 66–74, 2019.
- [2] J. C. Pedro and S. A. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE transactions on microwave theory and techniques*, vol. 53, no. 4, pp. 1150–1163, 2005.
- [3] T. Bradde, S. Grivet-Talocia, P. Toledo, A. V. Proskurnikov, A. Zanco, G. C. Calafiore, and P. Croveti, "Fast simulation of analog circuit blocks under nonstationary operating conditions," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 11, no. 9, pp. 1355–1368, 2021.
- [4] D. Deschrijver, T. Dhaene, and D. De Zutter, "Robust parametric macromodeling using multivariate orthonormal vector fitting," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 7, pp. 1661–1667, 2008.
- [5] C. Briat, *Linear parameter-varying and time-delay systems*. Springer, 2014.
- [6] T. Y. Man, P. K. Mok, and M. Chan, "A high slew-rate push-pull output amplifier for low-quiescent current low-dropout regulators with transient-response improvement," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 54, no. 9, pp. 755–759, 2007.

Signal Integrity and Power Leakage Optimization for 3D X-Point Memory Operation using Reinforcement Learning

Kyungjune Son, Keunwoo Kim, Gapyeol Park, Daehwan Lho, Hyunwook Park, Boogyo Sim, Taein Shin, Joonsang Park, Haeyeon Kim and Joungho Kim
 School of Electrical Engineering, KAIST
 Daejeon, South Korea
 kyungjuneson@kaist.ac.kr

Kyubong Gong
 Mixed Design Team
 SK Hynix
 Icheon, South Korea
 kyubong.kong@sk.com

Abstract—As the signal integrity (SI) issues become critical with high bandwidth and density applications, the SI analysis and optimization are necessary. The SI optimization loop including design, modeling, simulation, analysis and revision is repetitive and confined to specific applications. To overcome the recurrent issues, we proposed reinforcement learning (RL) model for SI and power leakage optimization in 3D X-Point memory operation. We defined the MDP components to reflect the optimization problem and the RL model shows learning convergence. The optimal design shows 6.2 % of crosstalk, 17.7 % of IR drop and 25.3 % of power leakage improvement than original design.

Keywords—3D X-Point memory, power leakage, reinforcement learning, signal integrity (SI)

I. INTRODUCTION

With finer process technology the signal integrity (SI) issues such as crosstalk, IR drop and etc. are critical and SI optimization becomes more important in memory design [1]. To optimize signal integrity performance, the engineers need to iterate the SI optimization loop. As shown in Fig. 1(a), the sequence of design, modeling, simulation, analysis and revision construct the SI optimization loop. For SI analysis, we need to consider diverse SI issues such as crosstalk, IR drop, power leakage and etc. In conventional memory SI/PI design, a loop of repetitive processes derives the optimal memory design. However, the conventional method is time-consuming with modeling, simulation and iterations for revision and one-time method with limited design rule. To overcome these issues in conventional method, applying machine learning is promising solution [2].

Many factors including SI issues are linked and complex to optimize. Also, high time-cost for 3D electromagnetic (EM) simulations and insufficient training sets are critical issues to use machine learning for optimization. The reinforcement learning (RL) methods shows powerful performance rather than other machine learning methods in reusability [3], [4]. As shown in Fig. 1(b), the memory design loop can be linked to the optimization loop by RL. Unlike other ground-up optimization methods, the RL model can derive the optimal solution with one inference using reusable policy net.

In previous work [4], the optimal interconnection design shows better performance in same area of conventional design

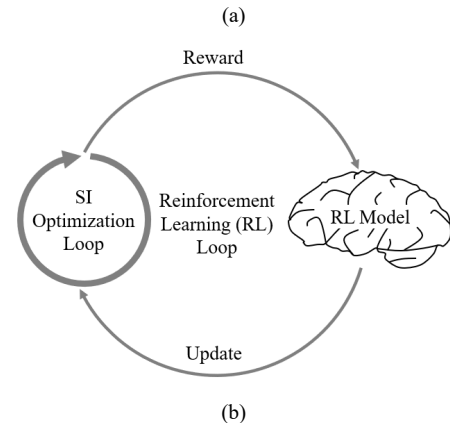
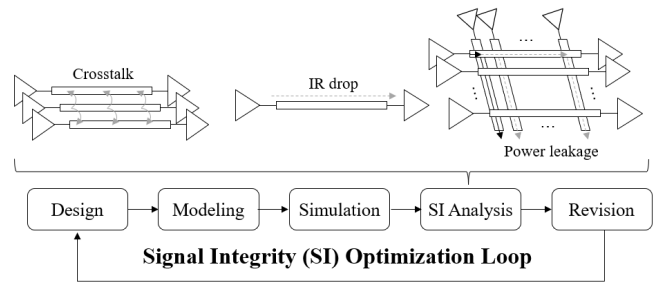


Fig. 1. (a) Signal integrity optimization loop considering crosstalk, IR drop and power leakage issues. The iteration of SI optimization loop is time-consuming and confined to specific applications. (b) RL loop for SI optimization loop. Based on the reward from SI optimization loop, the RL model updates and consists RL loop.

with irregularity. However, with irregularity size of interconnection lines, the practical conditions such as the inconsistent characteristics issues of unit memory cell, unfixed memory size and manufacturing issues need to be considered.

In this paper, we proposed RL model for 3D X-Point memory operation considering signal integrity and power leakage issues. We defined the 4 cases of interconnection width and space and size of MOSFETs in sense amplifier circuit to Markov decision process (MDP) components. The proposed model converged well and optimized the SI performance and power leakage compares to original design and random search design.

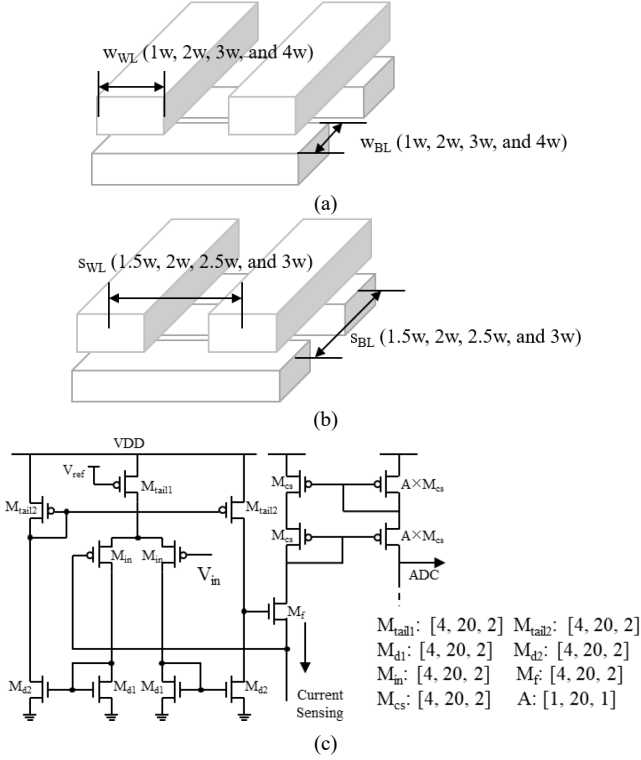


Fig. 2 MDP components in proposed RL model (a) 4 cases of width size for word line and bit line (b) 4 cases of space size for word line and bit line (c) Parameters of CMOS in sense amplifier circuit in 3D X-Point memory.

II. PROPOSAL AND VERIFICATION OF RL MODEL FOR 3D X-POINT MEMORY OPERATION

A. MDP Components of Proposed RL Model

To construct RL model, the MDP components need to be defined. As shown in Fig. 2(a) and (b), the width of word lines (WLs) and bit lines (BLs) are fixed to 1 to 4 times of unit width. The space of word lines and bit lines are fixed to 1.5 to 3 times of unit width. Also, we included the size of MOSFET in sense amplifier circuits. The sense amplifier is designed to 2-stage amplifier and current sensing circuit with feedback. We set the size of MOSFET for differential pairs as same and detailed parameter variations are shown in Fig. 2(c).

In reward factor, we defined the memory density as total number of bits / total area and considered crosstalk and IR drop issues that are critical SI issues in 3D X-Point memory. During read/write operation in 3D X-Point memory, the half-voltage is applied to WLs and BLs except selected lines for minimum operating power consumption. Along the selected lines, the half-selected memory cells are existing and the selector controls on/off status of the cells. However, both memory cell and selector are resistive components so, the sneak current is inevitable issue. Considering high density characteristics in 3D X-Point memory, the power leakage by sneak current is still critical issue.

In reward function, each hyperparameters is important. As the proposed RL model learns based on which factor to give more weight to, the hyperparameters can affect the direction of learning. We considered memory density, crosstalk, IR drop, and power leakage by sneak current to reward function. In the

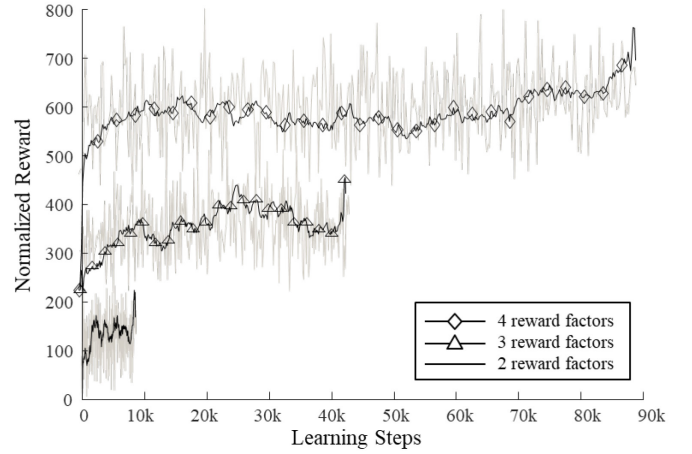


Fig. 3 Verification of RL model by ablation study with reward factor variations. The proposed RL model converged under 10 k, 45 k and 90 k learning steps with 2, 3, and 4 reward factors respectively.

case of memory density term, it is the most important factor in memory design. Therefore, the memory density term is set to the largest value. Next, the SI issues' term has a trade-off relationship with each other, so they were set equally in consideration. Lastly, for the power leakage issue due to sneak current, the same scale as the SI issue.

When reflecting optimization constraints in real world to MDP components, the learning variance increases and the convergence is not guaranteed in RL. It is important to secure learning convergence in RL. We experimentally proved that convergence by applying stability methods in policy update and policy optimization. For policy update, we used long short-term memory (LSTM) and temporal difference (TD) regularized actor-critic network. The LSTM stabilize the learning by memorizing the previous actions and features. With TD regularized actor-critic network, the learning became more stable by decreasing the divergence of the gradient estimation. For stably training policy, proximal policy optimization (PPO) updates the policy proximally by clipping methods [5]. As shown in Fig. 3. with 4 reward factors case, the proposed RL model well learned and converged.

B. Ablation Study of Proposed RL Model with Reward Factor Variations

The ablation study is one of the figure of merit in optimization method using machine learning. The ablation study is experiment to figure out that which components in model affect overall performance. We can understand the contribution of the component to the overall system with causality. For verification of the proposed RL model, we conducted ablation study with reward factor variations. As shown in Fig. 3, the proposed RL model easily converged under 10 k and 45 k learning steps with 2 and 3 reward factors. Considering 4 reward factors, the proposed RL model converged around 90 k learning steps. With the results of ablation study, we also verified the expandability of proposed RL model. By adding more reward factors or revising the reward factors, we can easily realize the optimal design with a desirable size and feasible performance capabilities and specifications.

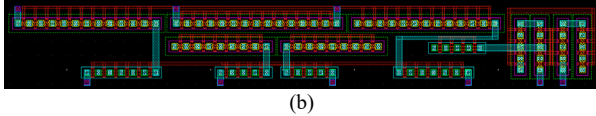
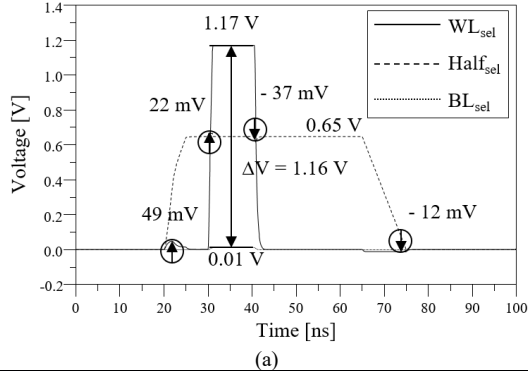


Fig. 1 Optimal solution by proposed RL model ($w_{WL}, s_{WL}, w_{BL}, s_{BL}, M_{tail1}, M_{tail2}, M_{d1}, M_{d2}, M_{in}, M_f, M_{cs}, A$) = (2, 3, 2, 3, 14, 12, 4, 6, 8, 4, 8, 1) (a) Time-domain simulation results of optimal solution by proposed RL model (b) Layout of optimal sense amplifier circuit design for 3D X-Point memory.

III. ANALYSIS OF THE OPTIMAL SOLUTION BY RL MODEL

We conducted time-domain simulation for SI analysis of the optimal solution by proposed RL model. We used copper as metal, low-k material for dielectric material and 20 nm process technology for interconnection lines. The unit interconnection designs are cascaded to $2k \times 2k$ for practical comparison.

As shown in Fig. 4(a), the crosstalk issues are mainly occurred in rising and falling edge of voltage pulse because of capacitance between interconnection lines. The average percentage of coupled voltage is calculated to reward function. The IR drop issue is unavoidable in high density memory. We calculated the percentage of degraded voltage margin to reward function. As previously mentioned, the power leakage by sneak current also critical issue in 3D X-Point memory. Due to IR drop through interconnection line, the large amount of sneak current increases power leakage of non-ideal half-selected memory cells. Based on these reward factors, the optimal solution by RL model improves SI performance: 6.2 % of crosstalk, 17.7 % of IR drop and 25.3 % of power leakage.

To secure the CMOS driving performance, the size of PMOS is 2 times larger than the size of NMOS considering carrier mobility. Also, the size of current driving MOSFET pair should be similar. As shown in Fig. 4(b), the proposed RL model designed the MOSFETs in sense amplifier circuit : The width of $M_{tail1}-M_{in}-M_{d1}$ is $14L-8L-4L$, the width of $M_{tail2}-M_{d2}$ is $12L-6L$ and the width of $M_{cs}-M_f$ is $8L-4L$. The width of $M_{d1}-M_{d2}$ is $4L-6L$ which is current driving MOSFET pair. To minimize the layout area of sense amplifier, the RL model designs sense amplifier with minimum number of the scale factor 'A'. In [6], the size of sense amplifier to unit memory array tile size is 0.45 %. However, in the optimal design, the size of sense amplifier to unit memory array tile size is 0.32 % considering process technology node size.

The performance evaluation including memory density, crosstalk, IR drop and power leakage is summarized in Table II. We compared the optimal solution using proposed RL model with the case of original design and random search design. For

Table I Performance Evaluation with Optimization Method Variations

Learning Algorithm	Memory Density	Crosstalk	IR Drop	Power Leakage
No Optimization	1 / 1	10 %	28.5 %	40.9 %
Random Search	1 / 9	5.1 %	16.2 %	22.5 %
Ours	1 / 4	3.8 %	10.8 %	15.6 %

the random search, we conducted 10,000 iterations. The optimal design by RL model shows lower memory density than original design. However, the optimal design by RL model outperforms in terms of crosstalk, IR drop and power leakage than original design and optimal design by random search. With the reusable policy net, the proposed RL model solves new problem with one inference and extremely reduces the computational cost than other ground-up optimization methods.

IV. CONCLUSION

In this paper, we proposed RL model for SI and power leakage optimization in 3D X-Point memory operation. The MDP components considering practical conditions such as interconnection manufacturing, sense amplifier circuit, SI issues and power leakage by sneak current are defined. The optimal design by RL model surpass the SI performance of original design and optimal design by random search. The proposed RL model can apply in future memory design applications with expandability and reusability.

ACKNOWLEDGMENT

We would like to acknowledge the technical support from ANSYS. This research was supported by National R&D Program through the National Research Foundation of Korea (NRF) funded by Ministry of Science and ICT (NRF-2020M3F3A2A01081587, NRF-2022M3I7A4072293) and the Technology Innovation Program (20015559, Development of 1000V Class High Voltage Relay Technology Based on High Durability and Arc Resistant Materials) funded by the Ministry of Trade, Industry & Energy.

REFERENCES

- [1] K. Son, *et al.*, "Signal Integrity Design and Analysis of 3-D X-Point Memory Considering Crosstalk and IR Drop for Higher Performance Computing," in *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 10, no. 5, pp. 858-869, May 2020
- [2] M. Swaminathan, *et al.*, "Demystifying machine learning for signal and power integrity problems in packaging," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 10, no. 8, pp. 1276-1295, Aug. 2020
- [3] K. Kim *et al.*, "Deep Reinforcement Learning-based Through Silicon Via (TSV) Array Design Optimization Method considering Crosstalk," *2020 IEEE Electrical Design of Advanced Packaging and Systems (EDAPS)*, 2020, pp. 1-3.
- [4] K. Son, *et al.*, "Reinforcement-Learning-Based Signal Integrity Optimization and Analysis of a Scalable 3-D X-Point Array Structure," in *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 12, no. 1, pp. 100-110, Jan. 2022.
- [5] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, arXiv:1707.06347.
- [6] J. Kwon *et al.*, "A Low-Power TDC-Configured Logarithmic Resistance Sensor for MLC PCM Readout," in *IEEE Sensors Journal*, vol. 16, no. 14, pp. 5524-5535, July 15, 2016.

Interconnect Modeling using a Surface Admittance Operator Derived with the Fokas Method

Dries Bosman¹, Martijn Huynen¹, Daniël De Zutter¹, Xiao Sun²,
Nicolas Pantano², Geert Van der Plas², Eric Beyne², Dries Vande Ginste¹

¹*Quest, IDLab, Department of Information Technology, Ghent University/imec, Ghent, Belgium*

²*imec, Leuven, Belgium*

dries.bosman@ugent.be

Abstract—In this contribution, we propose a novel approach to rigorously model interconnect structures with an arbitrary convex polygonal cross-section and general, piecewise homogeneous, material parameters. A full-wave boundary integral equation formulation is combined with a differential surface admittance approach, invoking an extended form of the numerically fast Fokas method to construct the pertinent operator. Several examples validate our method and demonstrate its applicability to per-unit-of-length resistance and inductance characterization.

Index Terms—differential surface admittance, Fokas method, interconnect modeling

I. INTRODUCTION

In our modern society, where information technology is omnipresent, the development of sophisticated devices at ever higher operating frequencies poses serious challenges, e.g., in terms of electromagnetic compatibility and signal integrity. Combined with the continuing miniaturization, this evolution renders a proper analysis of the occurring electromagnetic fields and their wave nature indispensable. More specifically, in high-frequency interconnects, phenomena such as the skin and proximity effect should be taken into account in a rigorous fashion. For electromagnetic solvers employing a volumetric mesh, such as the versatile finite elements method (FEM), the exponential nature of the current crowding enforces an intractably fine discretization. The boundary integral equation (BIE) method and many other surface-based techniques, on the other hand, require particular attention to deal with the numerical integration of the Green's function in highly conductive media [1].

A popular procedure to circumvent this strenuous situation replaces the conductive material by its surrounding medium, while introducing additional boundary conditions. For instance, in the class of approximate techniques, (local) surface impedances are invoked [2]. Alternatively, the differential surface admittance (DSA) operator [3] captures the substituted material's properties in an exact, global way. Its implementation requires the eigenfunctions of the considered cross-sections, imposing a *de facto* limitation to circular and rectangular shapes. An extension to triangles, not relying on the Dirichlet eigenfunctions, was presented in [4]. However, this approach involves special measures to eliminate a prominent Gibbs effect degrading the initial solution. Moreover, a combination of multiple triangular components is necessary

for the analysis of arbitrary polygonal cross-sections. In yet other formulations [6], numerical issues may arise, in particular in the case of high material contrasts [1].

Here, on the other hand, we invoke and extend the Fokas method [5] to construct the DSA operator, automatically expanding its applicability to arbitrary convex shapes, while combined magnetic and dielectric contrast is allowed. As such, our method can, e.g., account for etching effects during the manufacturing of integrated circuits, resulting in trapezoidal structures. Coupled with the discretized electric field integral equation (EFIE), a formalism to accurately characterize interconnect structures is obtained, even for high material contrast and a strongly developed skin effect.

II. FORMULATION OF THE METHOD

Consider the two-dimensional (2-D) transverse magnetic (TM) polarized electromagnetic regime with a $e^{j\omega t}$ time dependence. We study a polygonal cylinder (typically a conductor) with M corner points (x_m, y_m) in the xy -plane, denoted as complex numbers $\zeta_m = x_m + jy_m$. The cylinder is characterized by its permittivity ϵ_i , permeability μ_i , conductivity σ_i and wavenumber k_i , and is situated in a homogeneous background medium with material properties ϵ_e , μ_e , σ_e and wavenumber k_e , as depicted in Fig. 1(a). Its longitudinal dimension is aligned with the z -axis. By applying the single source equivalence theorem, introducing an equivalent surface current density $\mathbf{j}_s = j_{s,z}\hat{\mathbf{z}}$ on the boundary \mathcal{C} , we can replace the cylinder's material by its surrounding medium, preserving the outside fields, while the inside fields $(\mathbf{e}_i, \mathbf{h}_i)$ are modified to the fictitious quantities $(\mathbf{e}'_i, \mathbf{h}'_i)$, as in Fig. 1(b). This equivalent surface current density is given by

$$\mathbf{j}_s = \hat{\mathbf{n}} \times (\mathbf{h}_i - \mathbf{h}'_i). \quad (1)$$

At the boundary of the structure, and only there, we find that $\mathbf{e}_i = \mathbf{e}'_i \triangleq \mathbf{e} = e_z\hat{\mathbf{z}}$, which is mapped to its normal derivative in the original and the equivalent situation via Dirichlet-to-Neumann (DtN) operators \mathcal{X} and \mathcal{X}' , resp.:

$$-j\omega\mu_i(\hat{\mathbf{n}} \times \mathbf{h}_i) = \mathcal{X}\mathbf{e}, \quad (2)$$

$$-j\omega\mu_e(\hat{\mathbf{n}} \times \mathbf{h}'_i) = \mathcal{X}'\mathbf{e}. \quad (3)$$

By combining (1), (2) and (3), we obtain

$$\mathbf{j}_s = \left(\frac{\mathcal{X}'}{j\omega\mu_e} - \frac{\mathcal{X}}{j\omega\mu_i} \right) \mathbf{e} \triangleq \mathcal{Y}\mathbf{e}, \quad (4)$$

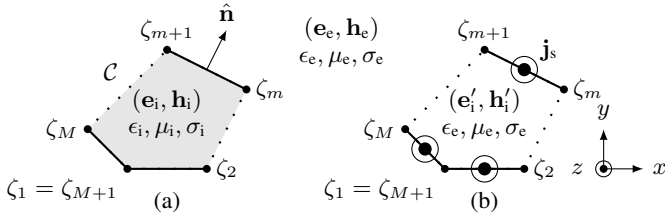


Fig. 1: Geometry of the problem, illustrating the equivalence theorem, with (a) the original and (b) the equivalent situation.

where \mathcal{Y} is the desired DSA operator [3].

The tangential electric field e_z satisfies the Helmholtz equation with eigenvalues k_i and k_e :

$$\nabla^2 e_z + k_{\{i,e\}} e_z = 0. \quad (5)$$

To solve the boundary value problems posed by (2), (3) and (5) we propose a Fokas-like method [5]. The following Fourier transform, the so-called global relation, is at its core:

$$F(\lambda) = \int_C \exp \left[-\frac{j k}{2} \left(\frac{\tilde{\zeta}}{\lambda} + \lambda \zeta \right) \right] \times \left[\frac{k \phi}{2} \left(\lambda d\zeta - \frac{d\tilde{\zeta}}{\lambda} \right) + \frac{\partial \phi}{\partial n} d\zeta \right] = 0, \quad \forall \lambda \in \mathbb{C}, \quad (6)$$

where $\zeta_m = x_m + jy_m$, $\tilde{\cdot}$ indicates the complex conjugate and \mathcal{C} denotes the polygonal boundary. Furthermore, $\phi = e_z$ and $\frac{\partial \phi}{\partial n} = -j\omega\mu_{\{i,e\}} h_{\tan}^{\{i,e\}}$ in our case. Equation (6) is cast onto an appropriate basis of P orthogonal Legendre polynomials on each polygon side and evaluated at Λ well-chosen spectral collocation points $\lambda \in \mathbb{C}$:

$$\lambda = -\frac{l/k + \sqrt{(l/k)^2 - |h_m|^2}}{h_m}, \quad (7)$$

for $l \in \{0, 1, 2, \dots, \Lambda - 1\}$ and $m \in \{1, 2, \dots, M\}$, where $h_m = (\zeta_{m+1} - \zeta_m)/2$ and k is the wavenumber. This way, one ends up with an overdetermined, but very quickly solved, linear system with a solution that finally yields a discrete approximation of the pertinent DtN operators.

To incorporate this result in a BIE framework, a transformation to local, pulse-shaped basis functions is performed. By collecting the corresponding expansion coefficients of \mathbf{j}_s and \mathbf{e} into vectors \mathbf{J} and \mathbf{E} , we obtain the discretized version of (4):

$$\overline{\overline{\mathbf{G}}}\mathbf{J} = \left(\frac{\overline{\overline{\mathbf{X}}'}}{j\omega\mu_e} - \frac{\overline{\overline{\mathbf{X}}}}{j\omega\mu_i} \right) \mathbf{E} \triangleq \overline{\overline{\mathbf{Y}}}\mathbf{E}, \quad (8)$$

with $\overline{\overline{\mathbf{G}}}$ the Gram matrix of the local basis functions.

To find the per-unit-of-length (p.u.l.) resistance and inductance matrices $\overline{\overline{\mathbf{R}}}$ and $\overline{\overline{\mathbf{L}}}$ for a configuration with N conductors, we invoke the procedure outlined in [3], yielding

$$\overline{\overline{\mathbf{R}}} + j\omega\overline{\overline{\mathbf{L}}} = \left(\overline{\overline{\mathbf{T}}}^T \left(\overline{\overline{\mathbf{G}}}\overline{\overline{\mathbf{Y}}}^{-1}\overline{\overline{\mathbf{G}}} + j\omega\overline{\overline{\mathbf{A}}} \right)^{-1} \overline{\overline{\mathbf{T}}} \right)^{-1}, \quad (9)$$

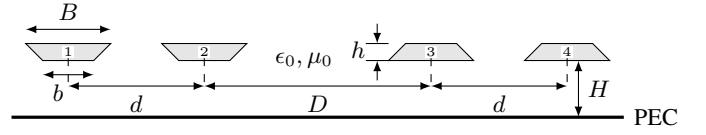
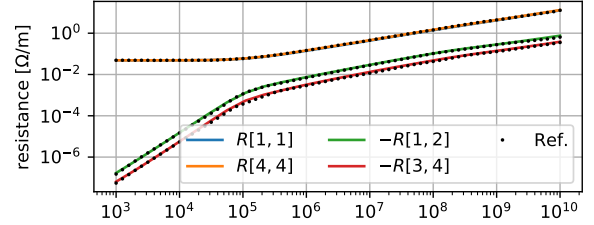
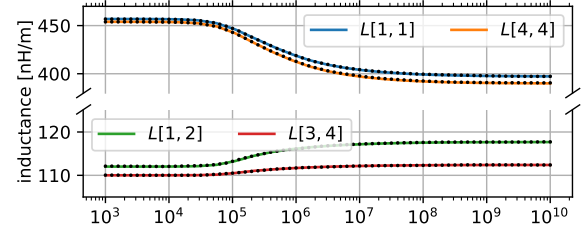


Fig. 2: Configuration with four trapezoidal conductors ($\sigma = 5.72 \times 10^7$ S/m) with dimensions in mm: $B = 1.5$, $b = 0.9$, $h = 0.3$, $D = 4$, $d = 2.4$ and $H = 1.5$, situated above an infinite PEC ground plane.



(a) P.u.l. resistance in Ω/m .



(b) P.u.l. inductance in nH/m.

Fig. 3: Relevant elements of the resistance and inductance matrices $\overline{\overline{\mathbf{R}}}$ and $\overline{\overline{\mathbf{L}}}$, for the configuration of Fig. 2.

where the elements of the matrix $\overline{\overline{\mathbf{A}}}$ are obtained through

$$\left(\overline{\overline{\mathbf{A}}} \right)_{ij} = -\mu_e \int_C \int_{C'} G(\mathbf{r}, \mathbf{r}') b_i(c) b_j(c') dc' dc. \quad (10)$$

with $G(\mathbf{r}, \mathbf{r}') = \ln |\mathbf{r} - \mathbf{r}'| / (2\pi)$, the 2-D static Green's function. The matrix $\overline{\overline{\mathbf{T}}}$ is defined as

$$\left(\overline{\overline{\mathbf{T}}} \right)_{in} = \begin{cases} \ell_i, & \text{if segment } i \in \text{conductor } n \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

with ℓ_i the length of segment i in the mesh.

III. NUMERICAL EXAMPLES

Consider the configuration with two oppositely oriented trapezoidal line pairs and conductivity $\sigma = 5.72 \times 10^7$ S/m above an infinite ground plane, shown with annotated dimensions in Fig. 2. Relevant elements of the corresponding resistance matrix $\overline{\overline{\mathbf{R}}}$ and inductance matrix $\overline{\overline{\mathbf{L}}}$, determined by means of the procedure outlined above, are compared to the reference solution provided by [4] in Fig. 3. The pertinent system matrix is constructed invoking $P = 20$ Legendre polynomials per side of the trapezoids, and is evaluated in $\Lambda = 40$ collocation points λ per side as well. These values for the parameters (P, Λ) will also be utilized in the remaining examples. An excellent agreement between our proposed method and the result found in literature is observed.

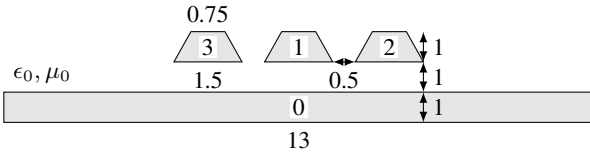
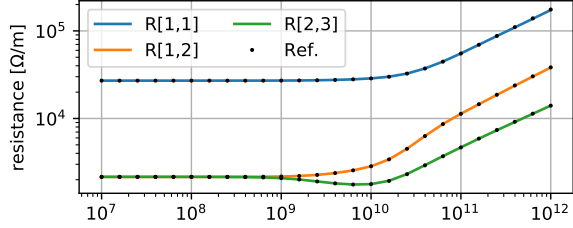
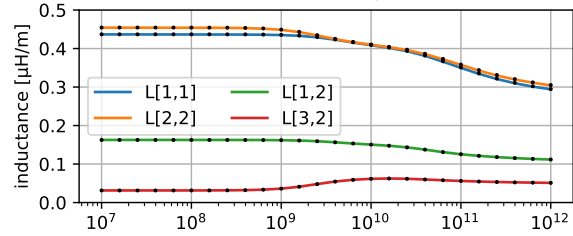


Fig. 4: Multiconductor transmission line ($\sigma = 3.57 \times 10^7 \text{ S/m}$) with three trapezoidal signal lines and a finite rectangular reference conductor. All dimensions are in μm .



(a) Pu.l. resistance in Ω/m .



(b) Pu.l. inductance in $\mu\text{H}/\text{m}$.

Fig. 5: Relevant elements of the resistance and inductance matrices $\overline{\overline{R}}$ and $\overline{\overline{L}}$, for the configuration of Fig. 4.

Next, we study the multiconductor transmission line depicted in Fig. 4, with reference conductor 0. The dimensions annotated on the figure are all given in μm . We obtain the curves plotted in Fig. 5, validated by means of the results in [6]. Once again, both sets of results match excellently.

Finally we investigate the influence of the conductor ($\sigma = 1 \times 10^7 \text{ S/m}$, $\mu_r = 5$) shape in the configuration of Fig. 6, evolving from triangular (solid lines), over asymmetric trapezoidal (dashed) to rectangular (dotted). This example includes a conductive, magnetic medium and therefore demonstrates the capability of our method to model this novel class of materials, present in state-of-the-art interconnect applications [7]. The elements of the matrices $\overline{\overline{R}}$ and $\overline{\overline{L}}$ are given in Fig. 7, for these three shapes. Note that $R_{11} = R_{22}$ and $L_{11} = L_{22}$ owing to the symmetry of our problem.

For all of the above examples, the calculation of the DSA matrix by means of a Python code on a system with a 1.9GHz CPU and 16GB of RAM required less than 0.5s per frequency point, a value comparable to the times reported in [6], confirming the efficiency of our method.

IV. CONCLUSIONS

We presented a novel interconnect modeling technique, combining a boundary integral equation framework with a differential surface admittance operator, constructed through application of the numerically fast Fokas method. Our approach

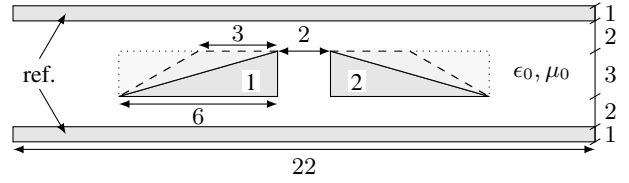
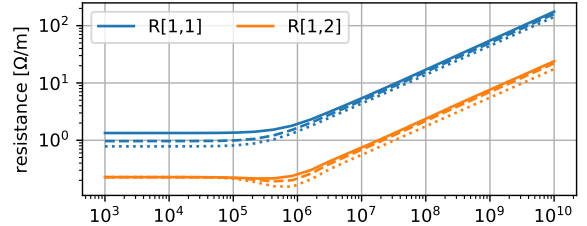
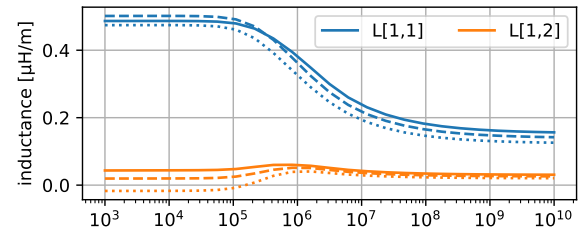


Fig. 6: Multiconductor transmission line ($\sigma = 1 \times 10^7 \text{ S/m}$, $\mu_r = 5$) with two triangular/trapezoidal/rectangular signal lines and two finite rectangular reference conductors. All dimensions are in 0.1 mm.



(a) Pu.l. resistance in Ω/m .



(b) Pu.l. inductance in $\mu\text{H}/\text{m}$.

Fig. 7: Relevant elements of the resistance and inductance matrices $\overline{\overline{R}}$ and $\overline{\overline{L}}$, for the configuration of Fig. 6. The solid, dashed and dotted lines correspond to the triangular, trapezoidal and rectangular conductor shapes, resp.

supports multiconductor configurations with arbitrary, piecewise homogeneous material properties and convex polygonal shapes. By means of per-unit-of-length resistance and inductance characterization of these structures, we demonstrated our method's accuracy, efficiency and broadband applicability.

REFERENCES

- [1] J. Peeters, I. Bogaert, and D. De Zutter, "Calculation of MoM interaction integrals in highly conductive media," *IEEE Trans. Antennas Propag.*, vol. 60, no. 2, pp. 930–940, Feb 2012.
- [2] T. B. A. Senior, and J. L. Volakis, "Approximate boundary conditions in electromagnetics," Institution of Electrical Engineers, 1995.
- [3] D. De Zutter, and L. Knockaert, "Skin effect modeling based on a differential surface admittance operator," *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 8, pp. 2526–2538, Aug 2005.
- [4] T. Demeester, and D. De Zutter, "Construction of the Dirichlet to Neumann boundary operator for triangles and applications in the analysis of polygonal conductors," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 1, pp. 116–127, Jan 2010.
- [5] A. S. Fokas, N. Flyer, S. A. Smitheman, and E. A. Spence, "A semi-analytical numerical method for solving evolution and elliptic partial differential equations," *J. Comput. Appl. Math.*, vol. 227, no. 1, pp. 59–74, May 2009.
- [6] U. R. Patel and P. Triverio, "Skin effect modeling in conductors of arbitrary shape through a surface admittance operator and the contour integral method," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 9, pp. 2708–2717, Sept 2016.
- [7] N. A. Lanzillo *et al.*, "Exploring the limits of cobalt liner thickness in advanced copper interconnects," *IEEE Electron Device Lett.*, vol. 40, no. 11, pp. 1804–1807, Nov 2019.

CISPR 25 Radiated Emission Simulation and Measurement Correlation of an Automotive Reinforced Isolated Switch Driver

Jie Chen^[1], Rajen Murugan^[1], Sooping Saw^[1], Francisco Lauzurique^[1], John Broze^[1], Craig Greenberg^[1], Alex Triano^[1], Bibhu Nayak^[2], Harikiran Muniganti^[2], Joe Sivaswamy^[2], and Dipanjan Gope^[2]

^[1] Texas Instruments Incorporated, Dallas, TX, USA; ^[2] Simyog Technology, Pvt., Ltd.

r-murugan@ti.com

Abstract — Applications of power electronics that integrate high-switching isolated gate drivers in switch mode power converters create excessive transient di/dt and dv/dt loops that exacerbate electromagnetic emissions. In this work, we developed a robust system-level coupled circuit-to-electromagnetic modeling and analysis methodology to predict the CISPR 25 radiated emission performance of a reinforced isolated switch driver during product development. The coupled method accurately captures the electromagnetic interactions between the nonlinear time-variant switchers and the system. Preliminary silicon validation measurements on an automotive high-switching isolated switch driver with an integrated power supply are presented to validate the integrity of the predictive modeling methodology. In an EMC pre-compliance lab, good correlations between modeling and measurements are achieved (i.e., within +/- 3dB for resonant peaks within the frequency band of 30MHz - 1GHz). The predictive EMC modeling methodology can be implemented to assess the performance of the initial silicon design during early IC development.

I. INTRODUCTION

A gate/switch driver is a buffer circuit that amplifies a low-power input from a microcontroller or any other source to drive semiconductor power switches efficiently. Gate driver designs typically fall into two broad categories - non-isolated or isolated. One key industry trend is integrating the gate driver with an isolator (the device that performs the isolation function) known as the isolated gate driver. A basic isolation integrated circuit (IC) allows data and power to transfer between high-voltage and low-voltage domains while preventing any hazardous DC or uncontrolled transient current from flowing across the domains [1]. Isolation is critical in high voltage automotive and industrial applications to maintain functionality and protect against electric shocks. In addition, many applications control power rails of 800V or higher, and, as such, additional reinforced isolation is required. For these applications, the additional protection is provided through reinforced isolation, the equivalent of two basic isolation layers enabled through a single isolation barrier [2]. A particular implementation of reinforced isolation is enabled through a single laminate

air-core transformer. Inductive coupling provides power and control signaling transmission for the driver stage.

Data transitions through the reinforced isolation transformer typically have sharp edge rates that potentially can cause conducted and radiated emissions due to the generation of high di/dt and dv/dt transient loops in the system (viz. package and PCB). For radiated emission, the focus of this work, two primary mechanisms have been identified as significant contributors to noise. These include PCB edge and input-to-output dipole emissions [3]. The main component of emissions is the common-mode current injected across the isolation barrier. Since isolators drive common-mode current across gaps in ground planes, the disruption in the return current path across the voltage domains creates an equivalent dipole antenna. The ability to predict radiated emission through simulation is highly desirable to achieve first-pass design success for stringent automotive EMC regulatory standards (e.g., CISPR 25).

In this work, we develop a predictive EMC modeling methodology to assess CISPR 25 radiated emission for an automotive reinforced isolated gate driver IC. Section II provides some key features and functionalities of the automotive device. The description of the device and system under test (i.e., package and PCB) is detailed in Section III. The modeling flow is discussed in detail in Section IV. Finally, a comparative analysis between simulation and EMC-certified laboratory measurements is presented in Section IV.

II. REINFORCED ISOLATED SWITCH DRIVER DETAILS

The device under test is a fully integrated, reinforced isolated power switch driver (see Fig. 1). It features a 10-V gate drive with a 1.5/3-A peak source and sink current with a reinforced isolation rating of 5-kV_{RMS}. It generates its own secondary supply from the power received from its primary side, so no isolated secondary bias supply is required. The primary side includes a transmitter that drives an alternating current into the primary winding of an integrated transformer at a rate determined by the setting of the PXFR (power control) pin and the logic state

of the EN (enabled) pin. The transmitter operates at a high frequency to optimally drive the transformer to its peak efficiency. In addition, the transmitter uses spread-spectrum techniques to improve EMI performance significantly, allowing many applications to achieve CISPR 25 automotive EMC regulatory standards. During transmission, signal information transfers to the secondary side alongside the power. On the secondary side, the voltage induced on the secondary winding of the transformer is rectified, and the shunt regulator regulates the output voltage level of VDDH (primary side power). Lastly, the demodulator decodes the received data information and drives VDRV (secondary side power) based on the logic state of the EN (enable) pin.

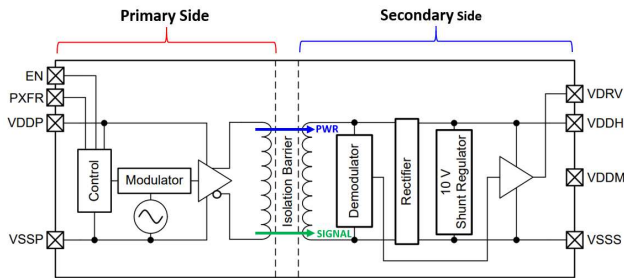


Fig. 1. Functional block diagram of an automotive reinforced isolated switch driver with integrated gate supply.

III. SYSTEM & MEASUREMENTS DETAILS

The device is packaged in a 7.5x5.85 mm² 8-pin SOIC (small outline integrated circuit) package. The primary die, secondary die, and transformer are all integrated into the package. The transformer is designed in a laminate-based substrate. The packaged device was mounted on an evaluation module (i.e., EVM) PCB for EMC characterization (see Fig. 2). It is a 2-layer PCB of size 55.13mm x 60.12mm. The EVM was designed and optimized for CISPR 25 EMC requirements.

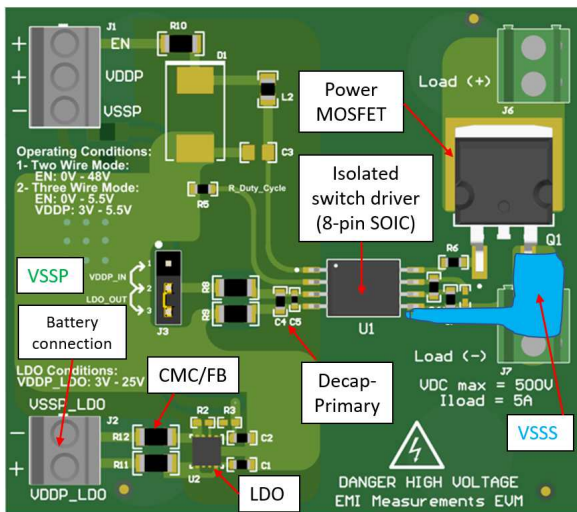


Fig. 2. CISPR 25 EVM PCB for characterization.

Appropriate PCB layout techniques to minimize signal and power integrity issues were employed. Provisioning for additional filter components (e.g., ferrite bead, common-mode chokes, among others) was designed accordingly. Additionally, the PCB design implements a large ratio of primary to secondary ground planes to minimize common-mode noise injection through the CISPR 25 cable. The CISPR 25 radiation emission measurement was set up, in a pre-compliance EMC chamber, following CISPR 25 standard requirements [4].

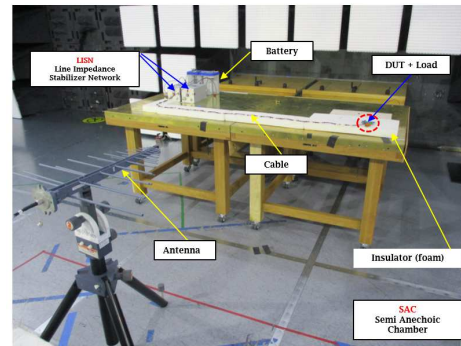


Fig. 3. CISPR 25 radiated emission measurements setup.

The power input of the system is a 12-V car battery (Fig. 3). A 12V-to-5V LDO was used to provide a lower voltage to supply the device. Biconical antenna was used for the 30MHz-300MHz and log-periodic antenna for 200MHz-1GHz measurements as per the specification. An EMI receiver was used to cover all the measurement frequency bands outlined in the CISPR 25 standard.

IV. EMC MODELING METHODOLOGY

The modeling methodology developed and implemented here is formulated through a coupled circuit-to-electromagnetic scheme (see Fig. 4). A SPICE-based circuit simulator coupled with a 3D full-wave low-rank method-of-moment (MoM) solver solution captures the electromagnetic interaction between the circuit and the system. For full details on methodology see Ref. [5]. The mathematical formulation is fully detailed in [6] and is not covered here. In the first Step, an industry-standard de-facto system-level circuit simulator is employed to set up the complete system-level transient analysis. The top-level isolated switch drive circuit includes appropriate circuit block models for the complete device under test. Physics-based device level reduced-order models were developed to optimize simulation runtime while not impacting performance. From laboratory measurements and characterization, fully validated off-chip components (viz. package, EVM, surface-mount components) are appropriately connected to the top-level circuit from the project model library. Transient analysis is then performed to extract voltage/current versus time waveforms at any circuit node of interest. For the analysis conducted here, the voltage between primary GND and

secondary GND was extracted in step 2. Once the voltage/current versus time waveforms are extracted at the appropriate nodes in the transient system setup, the data is parsed to Step 3 for the complete system-level electromagnetic analysis.

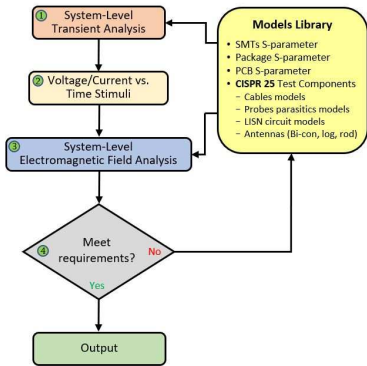


Fig. 4. Steps of the coupled circuit-to-electromagnetic modeling flow.

Using the voltage/current vs. time waveforms as stimuli/excitations, the field analysis is set up and performed in Step 3. Fig. 5 below shows the setup in the EMC virtual modeling and simulation platform. Finally, the simulated emission profile is then compared to CISPR 25 radiated emission measurements envelope.

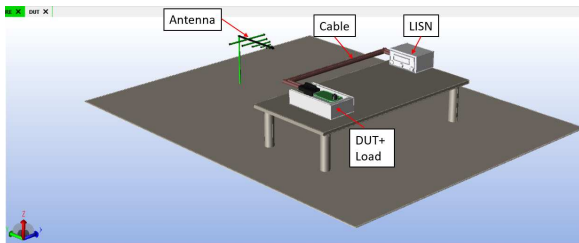


Fig. 5. Modeling and simulation of CISPR 25 in EMC virtual platform.

V. SIMULATION VS MEASUREMENT CORRELATION

Fig. 6 shows the radiated emission simulation versus measurement results comparison. The results correlate well for frequency band 30-300MHz (within +/-3dB magnitude). Every resonant peak that corresponds to switcher fundamental and harmonic frequencies is recovered. For the 450MHz and 540MHz, over a 10dB difference was observed between simulation and measurement.

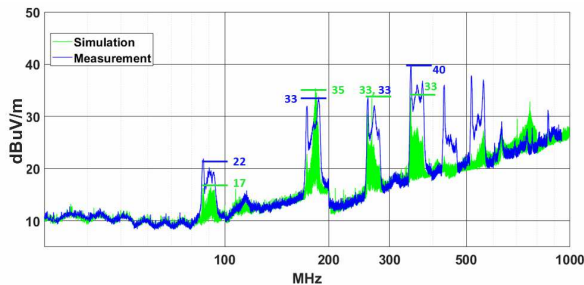


Fig. 6. Simulation vs measurement for CISPR 25 radiated emission.

The observed discrepancies between simulation and measurements at higher frequencies can be partially attributed to published findings on inter-labs comparative EMC CISPR 25 testings [7-9]. Reflection and scattering due to chamber effects, variations in the test setup (EMI vs. spectrum analyzer), and cable orientations likely contribute to the observed discrepancies. The correlation achieved with the predictive modeling methodology on our preliminary silicon has enabled optimization of the system (viz. silicon, package, and PCB) to meet CISPR 25 radiated emission regulatory standard successfully.

CONCLUSIONS

A robust system-level EMC coupled circuit-to-electromagnetic modeling and analysis methodology for predicting CISPR 25 radiated emission has been demonstrated in this work. The coupled method accurately captures the electromagnetic interactions between the nonlinear time-variant power switchers and the system. A good correlation was observed between simulation and measurements, considering the known inter-labs variations in CISPR 25 measurements. The predictive modeling methodology developed here is being implemented to optimize the design performance before the final regulatory compliance testing at certified EMC laboratories.

REFERENCES

- [1] N. Sridhar, "Impact of an isolated gate driver", Texas Instruments, Inc., Application Note SLYY140a, February 2019.
- [2] T. Bonifield, "Enabling high voltage signal isolation quality and reliability", Texas Instruments, Inc., Application Note SSZY028, March 2017.
- [3] B. Kennedy & M. Cantrell, "Recommendations for Control of Radiated Emissions with iCoupler Devices", Analog Devices, Application Note, AN-1109, 2011.
- [4] Vehicles, boats and internal combustion engines – Radio disturbance characteristics – Limits and methods of measurement for the protection of on-board receivers." CISPR 25:2016, fourth edition (or EN 55025:2017), October 2017.
- [5] R. Murugan et al., "Multiscale EMC Modeling, Simulation, and Validation of a Synchronous Step-Down DCDC Converter", Special Section on Industry Applications of Computational Electromagnetics, IEEE Journal on Multiscale and Multiphysics Computational Techniques, submitted 22nd August 2022.
- [6] G. Chatterjee, A. Das, S.V. Reddy and D. Gope, "Mesh Interpolated Krylov Recycling Method to expedite 3D Full-Wave MoM Solution for Design Variants", IEEE Transactions on Microwave Theory and Techniques, vol. 65, No. 9, Mar. 2017.
- [7] F. Lafon, J. Davalan and R. Dupendant, "Inter-laboratory comparison between CISPR25 chambers, identification of influent parameters and analysis by 3D simulation," 2015 Asia-Pacific Symposium on Electromagnetic Compatibility (APEMC), 2015.
- [8] F. Lafon et al., "Investigation on dispersions between CISPR25 chambers for radiated emissions below 100 MHz", EMC Europe 2014.
- [9] S. Baisakhiya, A. Albin, B. Subbarao, "Interlaboratory comparison of radiated emission measurements. 2008 10th International Conference on Electromagnetic Interference & Compatibility", 283-285, 2008.

Deterministic Policy Gradient-based Reinforcement Learning for DDR5 Memory Signaling Architecture Optimization considering Signal Integrity

Daehwan Lho¹⁾, Hyunwook Park¹⁾, Keunwoo Kim¹⁾, Seongguk Kim¹⁾, Boogyo Sim¹⁾, Kyungjune Son¹⁾, Keeyoung Son¹⁾, Jihun Kim¹⁾, Seonguk Choi¹⁾, Joonsang Park¹⁾, Haeyeon Kim¹⁾, Kyubong Kong²⁾ and Joungho Kim¹⁾

¹⁾School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea

²⁾SK Hynix, Icheon, South Korea

E-mail: daehwan.lho@kaist.ac.kr

Abstract— In this paper, we propose the deterministic policy gradient-based reinforcement learning for DDR5 memory signaling architecture optimization considering signal integrity. We convert the complex DDR5 memory signaling architecture optimization to the Markov decision process (MDP). The key limitation factor was found through the analysis of the hierarchical channel, and MDP was configured to solve it. The deterministic policy is essential for optimizing high-dimensional problems that have many continuous design parameters. For verification, we compare the proposed method with conventional methods such as random search (RS) and Bayesian optimization (BO) and other reinforcement learning algorithms such as the advantage actor-critic (A2C) and proximal policy optimization (PPO). RS and BO could not be properly optimized even after 10000 iterations of 1000 times, respectively, and A2C and PPO failed to optimize. As a result of comparison, the proposed method has the highest optimality, low computing time, and reusability.

Keywords— *Deterministic policy gradient, optimization, reinforcement learning, signal integrity, DDR5.*

I. INTRODUCTION

Recently, the performance of processors is significantly increasing due to artificial intelligence and the 4th industrial revolution, and accordingly, the speed of dynamic random-access memory (DRAM) applications is also increasing. In the case of double data rate (DDR), it is used in various fields such as PCs, data centers, and servers, and is increasing up to 6.4Gbps per line in DDR5. Due to this increase in data rate, various signal integrity (SI) problems such as crosstalk, inter symbol interference (ISI), jitter, and electromagnetic interference (EMI) occur in the interconnection between DDR5 and the processor [1].

Unlike other DRAM applications such as LPDDR, HBM, and GDDR, DDR architecture has a memory module to increase versatility, and a socket is used to connect the module and board, which further degrades SI. In other words, DDR architecture is the hierarchical signaling architecture that has many SI degradation components. Many components between the processor and memory attenuate the signal, as well as the values of on-die termination (ODT) and decision feedback equalizer (DFE) greatly affect the signal. When designing hierarchical signaling architecture like DDR signaling architecture considering SI, we should design it by optimizing the necessary components after composing the entire architecture, rather than simply optimizing each component.

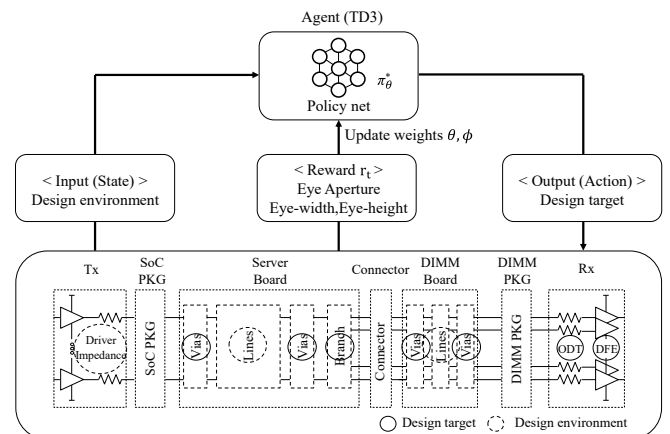


Fig. 1. Conceptual view of the proposed deterministic policy gradient-based reinforcement learning for DDR5 memory signaling architecture optimization considering signal integrity.

Many previous studies have been performed to optimize the high-speed channel considering SI. Bayesian optimization (BO) which has tremendous power in black box function optimization is applied to find the optimal design parameters of the high-speed channels [2], [3]. However, BO has the limitation that it is a non-reusable method that must be optimized whenever the environment parameters change. In addition, the BO does not optimize well when many design parameters need to be optimized. Therefore, it is difficult to use BO in the hierarchical signaling structure because many design parameters need to be optimized.

In this paper, we propose the deterministic policy gradient-based reinforcement learning for DDR5 memory signaling architecture optimization considering SI. We convert the complex DDR5 memory signaling architecture optimization to the Markov decision process (MDP). The conceptual view of the proposed method is shown in Fig. 1. We defined MDP including state, action, and reward. The environment delivers the state and reward to the agent, and the agent sends an action to the environment that increases the reward by updated policy net. We use the twin delayed deep deterministic policy gradient (TD3) algorithm, which is one of the deterministic policy gradient algorithms. The deterministic policy helps to optimize parameters with a continuous action space. The proposed method optimizes the target parameters with only one iteration even if the value of the environment parameters changes. We compare the proposed method with conventional methods such

TABLE I
DESIGN PARAMETER RANGE OF ENVIRONMENT DESIGN (STATE)

Component	Parameter	Min	Max
Tx	Impedance	14 Ohm	56 Ohm
	Width	0.0813 mm	0.3251 mm
Server board trace	Space	0.1726 mm	0.6902 mm
	Thickness	0.0132 mm	0.0528 mm
	Height1	0.0406 mm	0.1626 mm
	Height2	0.2002 mm	0.8006 mm
	Losstan	0.0245	0.0455
	Permittivity	3.14	5.85
	DIMM board trace	Width	0.0580 mm
Space		0.6000 mm	2.4000 mm
thickness		0.0060 mm	0.0240 mm
Height1		0.0240 mm	0.0960 mm
Height2		0.0820 mm	0.3280 mm
Losstan		0.009	0.0169
Permittivity		2.87	5.33

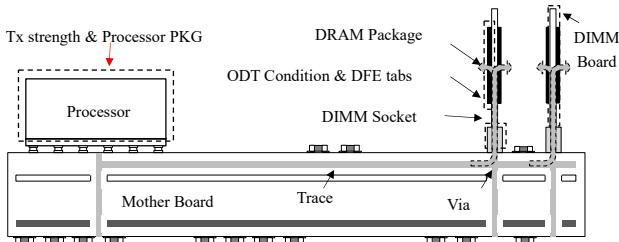


Fig. 2. DDR5 signaling architecture including signal integrity degradation components.

as random search (RS) and BO and other reinforcement learning algorithms such as the advantage actor-critic (A2C) and proximal policy optimization (PPO). As a result of the comparison, it has much higher optimality than the conventional algorithms, the optimization is faster, and the reusability is verified by obtaining optimal results for various tasks.

II. PROPOSED DETERMINISTIC POLICY GRADIENT-BASED REINFORCEMENT LEARNING FOR DDR5 MEMORY SIGNALING ARCHITECTURE OPTIMIZATION

A. Design and Analysis of DDR5 Memory Signaling Architecture

The DDR5 hierarchical signaling architecture is shown in Fig 2. The DDR5 hierarchical signaling structure was selected as the most used 2 DIMM per channel (DPC) and dual rank. By changing each component into an analytic model, the entire hierarchical model was constructed. When configuring the model, one signal line is branched into two signal lines on the server board and the DIMM board to configure 2 DPC and dual-rank. Finally, it is one signal line in Tx, but it branches into 4 signal lines in Rx. Finally, these signals use on-die termination (ODT) to change the termination value of the line that actually receives the signal and the line that does not receive the signal, so that the signal is transmitted normally. In addition, it uses a Decision feedback equalizer (DFE) to improve the ISI, so that the final signal is transmitted well. When ODT and DFE are added to the hierarchical model in this way, the model is completed. For server board and DIMM board, Olympus Intel XSP Motherboard and SK Hynix DDR5 DIMM board, which are released for free, are used. Based on the DDR5 JEDEC, the eye diagram simulation setup was set [4].

As a result of analysis based on eye height and eye width, three factors have the greatest impact on the DDR5 hierarchical

TABLE II
DESIGN PARAMETER RANGE OF TARGET DESIGN (ACTION)

Component	Parameter	Min	Max
Server board [CPU side via]	Via radius	0.2311 mm	0.4293 mm
	Antipad radius	0.4267 mm	0.7925 mm
	Stub length	0.0 mm	1.5 mm
Server board [DIMM side via]	Via radius	0.0890 mm	0.1651 mm
	Antipad radius	0.2489 mm	0.4623 mm
	Stub length	0.0 mm	1.5 mm
DIMM board Via	Via radius	0.07 mm	0.13 mm
	Antipad radius	0.21 mm	0.39 mm
Branch-DIMM0	Length	0 mm	14 mm
ODT on	Impedance	96 Ohm	384 Ohm
ODT off	Impedance	32 Ohm	128 Ohm
DFE	Tab 1	-200 mV	50 mV
	Tab 2	-75 mV	75 mV
	Tab 3	-60 mV	60 mV
	Tab 4	-45 mV	45 mV

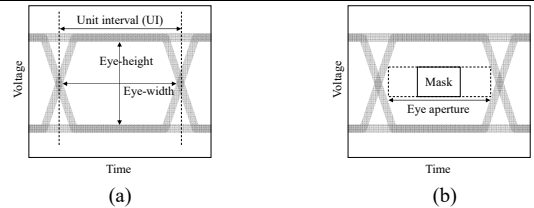


Fig. 3. The definition of (a) eye height and eye width, unit interval, and (b) eye aperture.

signaling architecture. The most influential factor is the side DIMM effect, followed by the via stub effect, and finally via impedance mismatch. Since these three things actually have a lot of influence, an off-chip solution and optimization method is essential to reduce the key limitation factors.

B. Markov Decision Process(MDP)

To optimize using reinforcement learning, we need to define Markov decision process (MDP). MDP consists of state, action, and reward. The purpose of MDP is to find the optimal value of the target variable by removing the key limitation factors. The agent updates the policy net to achieve the purpose so that the action that maximizes the reward can be obtained. The state is a design variable in the DDR5 signaling architecture. As shown in Table I, driver impedance, server board traces, and DIMM board traces are set to states to estimate optimal design goals. Additionally, since the drive impedance can be changed during design, the driver impedance is also set as a state and a total of 15 states are set. The length of the server board and DIMM board trace is set to 130 mm, 14 mm each.

Action is a design target and is a design parameter to improve the side DIMM effect, via stub, and via impedance discontinuity, which are key limitations of DDR memory signaling architecture. First, backdrill is done to remove the stubs of both vias of the server board, and the length of drilling is set as the length of the drill. Also, to consider the side DIMM effect, the length from branch to DIMM0 is set as an action. Additionally, by setting the radius of via and anti pad as an action, impedance mismatch of vias can also be resolved. Finally, not only the dimension-related design parameters were set as actions, but the on ODT resistance, off ODT resistance, and 4 DFE tap values that greatly affect the eye diagram were also set as actions. As shown in Table II, a total of 15 actions were set. The min-max value of the state and action is assumed to be an

TABLE IV
PERFORMANCE EVALUATION OF THE PROPOSED METHOD COMPARING WITH RS, BO, AND OTHER RL METHODS

Method	DIMM	Eye aperture [U]										Time (1 test)
		Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10	
RS {10000}	1	0.2016	0.1736	0.2128	0.3528	0.2856	0.2352	0.2352	0.336	0.2184	0	3798 s
	0	0.2016	0.112	0.2968	0.3528	0.2856	0.2352	0.2352	0.336	0.2184	0.1792	
BO {1000}	1	0.2765	0.616	0.308	0.3416	0.3136	0.3696	0.2912	0.336	0	0	2113 s
	0	0.2765	0.1176	0.308	0.3416	0.3136	0.3696	0.2912	0.336	0	0	
A2C {1}	1	0	0	0	0	0	0	0	0	0	0	0.3 s
	0	0	0	0	0	0	0	0	0	0	0	
PPO {1}	1	0	0	0	0	0	0	0	0	0	0	0.3 s
	0	0	0	0	0	0	0	0	0	0	0	
Proposed {1}	1	0.2912	0.3136	0.3304	0.392	0.3528	0.3752	0.3136	0.3752	0.2688	0.1624	0.3 s
	0	0.2912	0.3136	0.3304	0.392	0.3528	0.3752	0.3136	0.3752	0.2688	0.1624	

TABLE III
HYPER-PARAMETERS OF PROPOSED TD3 POLICY NET

Network	Layer	Activation function	Node
Actor (actor, target)	Input layer	-	15
	Hidden 1	ReLU	400
	Hidden 2	ReLU	300
	Output layer	Tanh	15
Critic * 2 (critic, target)	Input layer	-	30
	Hidden 1	ReLU	400
	Hidden 2	ReLU	300
	Output layer	Linear	1

arbitrary ratio based on the value in Olympus Intel XSP Motherboard and SK Hynix DDR5 DIMM board.

To set the reward, all eye height (EH), eye width (EW), and eye aperture (EA) are used as shown in Fig. 3(a) and (b). EA is the most important because it includes both EH and EW information and is used as an evaluation index for DDR system. However, when using only the EA as a reward, the training did not converge well due to zero reward unless the eye is larger than the eye mask. Therefore, EH and EW are also added. Instead of adding them as it is, since the EH is the most important parameter, α is added to adjust the ratio of the three values and it is set the value to 0.9. In 2 DPC, performance should be improved in both cases where the signal goes to DIMM1 and DIMM0, so the reward is set as the sum of the two results. The reward is expressed as follows:

$$Reward = \alpha * EA_{DIMM0} + (1-\alpha) * (EW_{DIMM0} + EH_{DIMM0}) + \alpha * EA_{DIMM1} + (1-\alpha) * (EW_{DIMM1} + EH_{DIMM1}) \quad (1)$$

As the agent, a TD3 algorithm called twin delayed deep deterministic policy gradient was used. The TD3 algorithm is one of the deterministic policy gradient algorithms because, unlike the existing stochastic policy gradient algorithms, it can be trained stably in the continuous action space [5].

III. VERIFICATION OF THE PROPOSED METHOD

For verification, 10 states of random values were set. The data rate was set to 5.6 Gbps. The hyperparameters of the proposed TD3 policy net is shown in Table III. Additionally, learning rate, batch size, action noise, and gradient descent algorithm are set to 0.001, 100, 0.01, and Adam-optimizer each. The proposed method trains the policy net for 150,000 epochs using the set hyperparameters. Table IV shows the performance evaluation of the proposed method compared with random search (RS), Bayesian optimization (BO), advantage actor-critic (A2C), and proximal policy optimization (PPO). A2C and PPO also train the policy net for 150,000 epochs using the set

hyperparameters. The results of the signals passed to DIMM1 and DIMM0 are shown in the table respectively. In the case of RS and BO, optimization was performed separately for each task. When repeated more than 1000 times, BO greatly increases the execution time, making it difficult to use. It can be seen that the proposed method, which obtains the result with only one inference, has the highest optimality than RS with 10000 times and BO with 1000 times. In the case of A2C and PPO, which are other reinforcement learning methods, they are stochastic policy gradient methods, but they did not optimize properly in multidimensional continuous action space. Through the above results, the proposed method has the best performance for high optimality, reusability, and low computing time as well.

IV. CONCLUSION

In this paper, we proposed the deterministic policy gradient-based reinforcement learning for DDR5 memory signaling architecture optimization considering SI. We analyzed the effect of key limitations and we defined the MDP as improving the key limitation factors. Since the problem with continuous 15 states and 15 actions was a very multidimensional problem, it had not been optimized with the conventional optimization methods such as RS and BO, as well as other reinforcement learning methods such as A2C and PPO. The proposed method achieved the optimization in 0.3 seconds, proving high optimality, reusability, and computing time.

ACKNOWLEDGMENT

We would like to acknowledge the technical support from ANSYS Korea. This research was supported by National R&D Program through the National Research Foundation of Korea (NRF) funded by Ministry of Science and ICT (NRF-2022M317A4072293, NRF-2020M3F3A2A01081587).

REFERENCES

- [1] E. Bogatin, "Essential Principles of Signal Integrity," in IEEE Microwave Magazine, vol. 12, no. 5, pp. 34-41, Aug. 2011.
- [2] D. Lho et al., "Bayesian Optimization of High-Speed Channel for Signal Integrity Analysis," 2019 IEEE 28th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS), 2019, pp. 1-3, doi: 10.1109/EPEPS47316.2019.193211.
- [3] J. Kim et al., "PAM-4 based PCIe 6.0 Channel Design Optimization Method using Bayesian Optimization," 2021 IEEE 30th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS), 2021, pp. 1-3, doi: 10.1109/EPEPS51341.2021.9609213.
- [4] DDR5 SDRAM Standard, Standard JESD79-5A, 2021.
- [5] Scott Fujimoto, Herke van Hoof, David Meger, "Addressing Function Approximation Error in Actor-Critic Methods," International conference on machine learning (ICML). PMLR, 2018

An Efficient Methodology to Parse and Mesh Large Interconnect Layouts for Electromagnetic Analysis

Qinghao Zhang[†], Ruoyi Xie[†], Fei Guo[‡], Shashwat Sharma[†], Damian Marek[†], Piero Triverio[†]

[†]*Edward S. Rogers Sr. Department of Electrical & Computer Engineering, University of Toronto, Toronto, Canada*

[‡]*Advanced Micro Devices, Markham, Canada*

{qinghao.zhang, ruoyi.xie, shash.sharma, damian.marek}@mail.utoronto.ca, fei.guo@amd.com, piero.triverio@utoronto.ca

Abstract—We present a complete methodology to import and mesh the layout of complex interconnect networks for electromagnetic analysis. At first glance, these tasks may seem straightforward. In reality, they require complex geometrical operations, which are not trivial to perform efficiently and robustly for realistic layouts. The methodology is based on publicly-available libraries and generates a conformal surface mesh suitable for the boundary element method. The method is tested on an entire IC package from the Packaging Benchmark Suite.

Index Terms—electromagnetic analysis, layout processing, meshing.

I. INTRODUCTION AND MOTIVATION

Electromagnetic (EM) analysis is essential for the correct design of interconnect networks at the chip, package and board level. A critical part of EM analysis is importing the interconnect layout and generating a quality mesh suitable for the numerical solution of Maxwell equations. These operations may seem relatively straightforward to perform combining open-source libraries to read standard layout files such as GDSII [1], manipulate 3D objects [2], and mesh them [3]. In reality, importing and meshing a realistic interconnect layout is a very challenging task to perform efficiently and robustly. Limited literature is available on this problem, and few software libraries exist for parsing, meshing, and in some cases, analyzing interconnect layouts.

Some open-source libraries exist for reading, writing and visualizing GDSII files, such as `gdstk` [1], `gds3xtrude` [4], and `GDS3D` [5]. For finite-difference methods, the `gds2Para` library was recently released to support finite difference time domain analysis of ICs, packages, and boards [6]. Alternatively, `libGDSII` [7] was designed to be used in both finite difference and integral equation methods. A limitation of both `GDS3D` [5] and `libGDSII` [7] is that they rely on `Gmsh` [3] for meshing. While `Gmsh` is an excellent meshing library, our numerical results will show how `Gmsh`, by working in 3D, cannot scale to realistic layouts, and a custom solution that exploits their 2.5D nature is mandatory.

Parsing and meshing a real interconnect layout is challenging for many reasons. In most industrial formats (e.g.

This work was partially supported by Advanced Micro Devices, by the Natural Sciences and Engineering Research Council of Canada, and by the Canada Research Chairs Program.

`GDSII`), only the base (footprint) of each conductor is stored, for performance reasons. In a realistic layout, there can be thousands of polygonal footprints. Each footprint can be a very complex polygon in itself, with hundreds of vertexes and holes, as in the case of a ground plane traversed by many signal vias. After each footprint is extruded vertically into the actual 3D object, one must detect all portions of its surface that are in contact with other conductors in adjacent layers. This is necessary to properly enforce boundary conditions during electromagnetic analysis and to, during meshing, enforce mesh conformity. Conformity is required by several, but not all, computational EM methods, such as the boundary element method (BEM) [8]–[10], which is taken as end point in this work.

Overall, the complexity of parsing and meshing interconnect layouts, and the limitations of the state of the art, are an obstacle to scientific research, industrial developments, and to the adoption of quality benchmarks, such as the Packaging Benchmark Suite [11]. We describe a complete methodology to read a complex layout from industry-standard formats, generate the actual 3D structure, and mesh it in a conformal manner. The methodology is based on publicly-available libraries and is shown to scale to an entire IC package.

II. METHODOLOGY

Interconnect layouts are typically stored in two files:

- a layout file which contains the 2D base of each object in metal layers, here referred to as *footprint*;
- a technology file that maps metal layers to the dielectric layers of the substrate. This file will also provide the height and material properties of both metal and dielectric layers.

Our end goal is to reconstruct the 3D structure of the interconnect, identify where metal objects in different layers touch each other, and generate a triangular mesh for the surface of all metallic objects. The mesh on the corresponding faces of touching objects must be identical (conformal), as required by several state-of-the-art BEM formulations [9], [12].

A. Reading and Simplifying Footprints in Each Layer

For reading the layout file from GDSII format, we chose `gdstk` [1]. In GDSII, footprints with voids are represented as self-reentrant contours. The `gdstk` and `Clipper` [13] library translate this representation to a more convenient one where

the outer contour of the footprint, and the inner contour of each void, are represented as separate polygonal paths. The latter representation is simpler and preferable for later manipulations.

In layout files, one can find multiple footprints that overlap or are in flush contact. These footprints must be merged with a union operation, as they represent a single conductive entity with homogeneous material properties. Among available libraries to manipulate 2D polygons, we found Clipper [13] to be the most reliable and efficient. Furthermore, to make each polygon strictly simple, we apply contour simplification operations to all objects.

B. Detecting Contacts Between Objects in Adjacent Layers

After footprints have been prepared, the next crucial step is extruding them along the vertical direction to generate the actual 3D objects, detect contacts, and generate a conformal mesh. Two routes exist at this point.

The simplest approach is to first extrude vertically each footprint using Gmsh [3]. Next, Gmsh’s boolean operations (called boolean fragments) can be used to identify, for a given object, all portions of its surface that are in contact with other objects in adjacent layers. Once identified, Gmsh will identically mesh these surfaces on touching objects, ensuring mesh conformity. However, we found that a direct application of these built-in functions does not scale to even moderately complex layouts, due to the cost of performing these operations in three dimensions.

We propose an alternative solution that relies only on 2D polygonal operations, and later extrudes footprints to their 3D shape while ensuring mesh conformity. First, each footprint is duplicated into what will be the lower and upper base of each conductive object. Then, using Clipper [13], we check if the upper base of an object in a layer intersects with the lower base of any object in the layer above. If so, we use the difference and intersection operations in Clipper [13] to partition the base into contact and non-contact portions. Note that each partition may consist of disjoint polygons, even if the original footprint was a single polygon. This is the case, for example, of a ground plane underneath a via fence. The upper base of the ground plane will be partitioned into many circular regions, representing the contact surfaces with the vias above, and the remainder, which will be a large polygon with many circular holes.

C. 3D Extrusion

Next, we describe how each 3D object is created given its lower and upper bases, which are possibly partitioned. While the lower and upper base still have the overall same shape, they may have been partitioned into remarkably different geometrical entities, making extrusion a non-trivial problem. Fig. 1 shows a simple scenario of a square ring which touches an object in the layer above (e.g. a stripline). While the lower base is still a single polygon with a hole, the upper base now consists of two polygons without any holes. Furthermore, partitioning introduced additional vertexes and edges in the

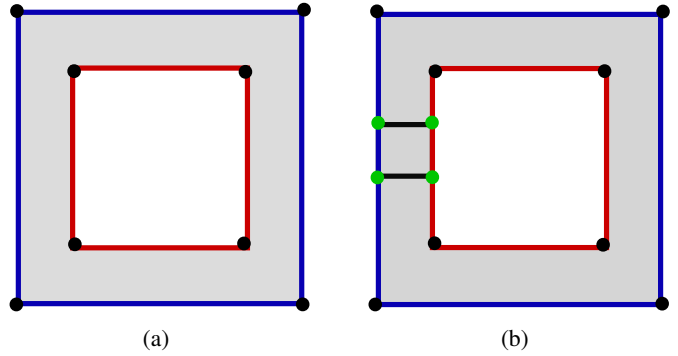


Fig. 1: Example of a square, hollow footprint: (a) lower base. (b) upper base. Red lines belong to the internal contour of the object, while blue lines belong to the external contour. Black lines do not belong to neither the external nor the internal contour.

upper base that do not exist in the lower base. These changes make 3D extrusion quite challenging for complex layouts. To address this challenge, we devised the following algorithm:

- 1) the lower and upper bases are decomposed into vertexes, lines, and surfaces. This representation will also facilitate mesh generation with Gmsh [3];
- 2) all vertexes and lines are re-indexed looping through the upper and lower bases. Lines that constitute the outer or inner contours of the bases are identified during this step (shown in red and blue in Fig. 1). However, at this point it is only known whether or not a line belongs to one of the contours;
- 3) a graph representation is created to relate vertexes on the internal and external contours to their corresponding lines. Then, a depth-first search is applied to find disjoint subgraphs, which correspond to closed contours of line segments. The graph representation is used to identify matching contours on the upper and lower bases and facilitate traversing around each of the closed contours;
- 4) the vertexes in the lower and upper bases that are vertically aligned are detected based on their (x, y) coordinates. Vertical lines are drawn between such vertexes to form the side faces of the extruded object.

D. Meshing and Visualization

At this point, the actual layout has been reconstructed in 3D, and is ready for meshing. Since the contact surface between each pair of touching objects has been identified, it can be meshed once ensuring conformity on all interfaces. Due to this pre-processing, any meshing library can be used at this point, providing flexibility on the type of mesh to be generated. In our case, we used Gmsh [3]. The 3D layout and mesh can be visualized with Gmsh’s viewer or, for better scalability, with Paraview [14].

III. RESULTS

The proposed method was implemented in C++ and tested on an IC package from the Packaging Benchmark Suite [11].

Test Case	Method	# of Layers	# of 2D Footprints	# of Gmsh Surfaces	# of Triangles	CPU Time
IC Package (portion)	Previous	15	335	1,291	56,742	25.9 s
	Proposed	15	335	1,371	56,066	2.14 s
IC Package (full)	Proposed	15	51k	326k	10M	27.4 min

TABLE I: Layout, mesh and CPU time statistics for the power integrity test package.

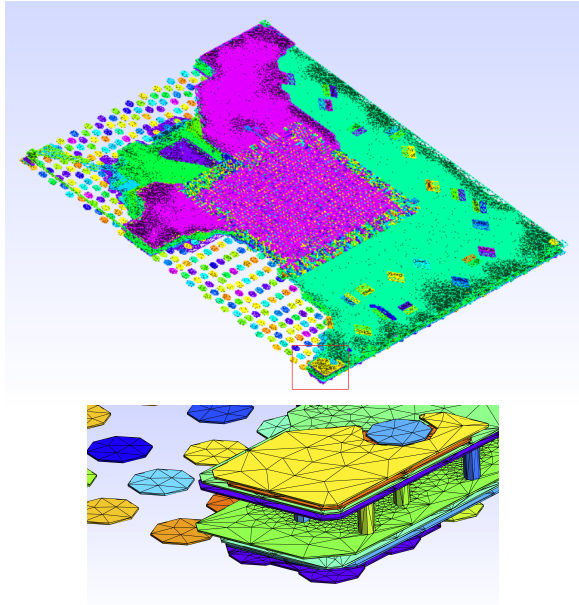


Fig. 2: Visualization of the mesh generated by the proposed method for the IC package.

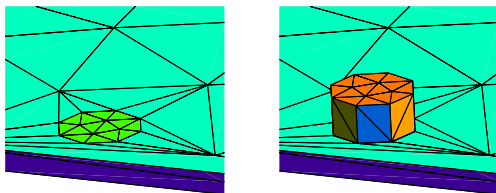


Fig. 3: A via connected to a ground plane, showing the mesh used for the lower base of the via (left panel) and for the upper base (right panel).

This is a medium-size package with vias, plated through holes, 15 layers and about 2,900 bumps. All tests were run on an Intel Core i7-2600 CPU running at 3.4 GHz.

We consider first a small portion of the IC package, consisting of 335 footprints out of the 51,000 of the whole package. The goal is to compare the proposed method against the simpler solution of using Gmsh and its boolean operations to extrude footprints and ensure mesh conformity. While both approaches result in a good, conformal mesh, their computational cost is remarkably different. As shown in Table I, the Gmsh approach takes almost 26 s, since contact detection and mesh conformity are performed in full 3D mode. The proposed approach is about 12X faster, taking only 2.1 s. Most

of the speed-up arises from the fact that contact detection is performed in 2D rather than in 3D.

Next, we consider the whole IC package. The 3D approach based on Gmsh’s boolean operations does not scale to this complexity, and can only handle, in several hours, selected pairs of layers of the structure. The proposed method parses and meshes the whole structure in 27.4 minutes on a single thread. With multithreading and code optimization, this time can likely be reduced further. The generated 3D structure consists of about 326,000 surfaces. The mesh contains about 10 million triangles and is shown in Fig. 2. The figure shows the whole structure and a local area with several power/ground planes, vias, and bumps. Fig. 3 provides a detailed view of a via connecting to a ground plane underneath. The figure shows how the contact surface between the via and the plane has been correctly identified and meshed to ensure conformity.

REFERENCES

- [1] *Gdstk Documentation*, 2020. [Online]. Available: <https://heitzmann.github.io/gdstk/>
- [2] *Open CASCADE Technology: Be free with your 3D modeling kernel*, 2000. [Online]. Available: <https://www.opencascade.com/open-cascade-technology/>
- [3] C. Geuzaine and J.-F. Remacle, “Gmsh: A 3-d finite element mesh generator with built-in pre- and post-processing facilities,” *International Journal for Numerical Methods in Engineering*, vol. 79, no. 11, pp. 1309–1331, 2009.
- [4] *Simple 3D viewer for GDS2 files. Based on OpenSCAD and KLayout.*, 2018. [Online]. Available: <https://codeberg.org/tok/gds3xtrude>
- [5] *GDS3D - An application used for rendering IC (chip) layouts in 3D.*, 2017. [Online]. Available: <https://github.com/trilomix/GDS3D>
- [6] *GDSII File Parsing, IC Layout Analysis, and Parameter Extraction*, 2018. [Online]. Available: <https://github.com/purdue-onchip/gds2Para>
- [7] *libGDSII - A C++ library and command-line utility for reading GDSII geometry files.*, 2019. [Online]. Available: <https://github.com/HomerReid/libGDSII>
- [8] W. S. Hall, *Boundary Element Method*. Dordrecht: Springer Netherlands, 1994, pp. 61–83.
- [9] Z. G. Qian, W. C. Chew, and R. Suaya, “Generalized impedance boundary condition for conductor modeling in surface integral equation,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, no. 11, pp. 2354–2364, 2007.
- [10] D. Marek, S. Sharma, and P. Triverio, “A parallel boundary element method for the electromagnetic analysis of large structures with lossy conductors,” *IEEE Transactions on Antennas and Propagation*, 2022, (in press).
- [11] *IEEE EPS Technical Committee on Electrical Design, Modeling and Simulation, ‘Packaging Benchmark Suite’*, 2021. [Online]. Available: <https://packaging-benchmarks.org/>
- [12] S. Sharma and P. Triverio, “SLIM: A well-conditioned single-source boundary element method for modeling lossy conductors in layered media,” *IEEE Antennas and Wireless Propagation Letters*, vol. 19, no. 12, pp. 2072–2076, 2020.
- [13] *Clipper - an open source freeware library for clipping and offsetting lines and polygons.*, 2010. [Online]. Available: <http://www.angusj.com/delphi/clipper.php>
- [14] U. Ayachit, *The ParaView Guide: A Parallel Visualization Application*. Kitware, 2015.

Novel Closed-Form 2-D Green's Function of Shielded Layered Media And Its Use in Transmission Lines Inductance Extraction

Shucheng Zheng

Dept. of Electrical and Computer Engr.
University of Manitoba
Winnipeg, Canada
umzheng6@myumanitoba.ca

Vladimir I. Okhmatovski

Dept. of Electrical and Computer Engr.
University of Manitoba
Winnipeg, Canada
vladimir.okhmatovski@umanitoba.ca

Abstract—New closed-form expression for quasi-static 2-D Green's function of fully-shielded layered medium is proposed. Along the vertical direction of the medium stratification the spectrum of the Green's function can be obtained through the 1-D ordinary differential equation and expressed as pole-residual form. This rational function representation of the spectrum allows to evaluate space domain Green's function in layered medium shielded from top and bottom with horizontal ground planes in a closed-form. This closed-form expression allows to analytically add up contributions from infinite number of source images with respect to the vertical side walls of the rectangular enclosure, hence, producing the new closed-form expression for the 2-D Green's function of rectangular enclosure vertically filled with multilayered medium. Availability of such Green's functions enable construction of integral equation based magneto-quasi-static and electro-quasi-static 2-D extractors. Such 2-D magneto-quasi-static extractor based on solution of Surface-Volume-Surface Electric Field Integral Equation (SVS-EFIE) for multi-conductor transmission lines (MTLs) situated in shielded layered media is demonstrated.

Index Terms—Green's function, inductance, integral equations, method of moments, resistance, transmission lines

I. INTRODUCTION

EFFICIENT computation of periodic Green's functions is essential for expedient analysis of shielded transmission lines [1]. In this work we propose a new approach to evaluation of double-periodic 2-D Green's functions featuring planar layered medium along one of the dimensions, which leads to fully close-form representation of such Green's functions. The infinite spectral integrals are evaluated in closed-form through casting of the layered media Green's function spectra into the pole-residual form. The latter is enabled through discretization of the 1-D ordinary differential equation governing the layered media Green's function spectrum followed by the eigenvalue decomposition in the resultant matrix equation. The method is commonly called Spectral Differential Equation Approximation Method in our prior work [2]–[4]. After closed-form evaluation of the spectral integrals the remaining infinite series ends up in the form of geometric series which can be summed analytically. To demonstrate application of the

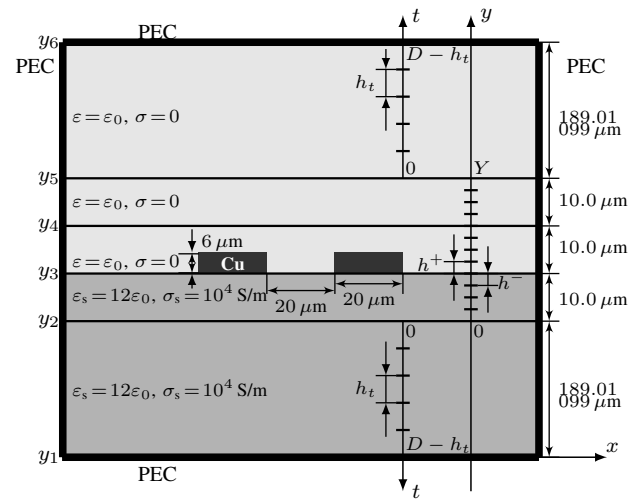


Fig. 1. The shielded two-conductor MTLs located over lossy substrate. The PEC planes situated at $y_1 = -189.01099\mu\text{m}$ and $y_6 = 219.01099\mu\text{m}$ enable termination of the FD grid for spectral domain 1-D ODE along the y coordinate. The region of interest is $y \in [y_2, y_5]$, where $y_2 = 0$ and $y_5 = Y$. In order to model the case of top PEC far away from the bottom PEC easily, the buffer regions $y \in (y_1, y_2)$ and $y \in (y_5, y_6)$ are mapped to regions of parametric variable $t \in (0, D - h_t)$ by using coordinate transformation [4].

new representation of double periodic 2-D Green's function featuring layered media, we use it to construct the 2-D Green's function of rectangular enclosure filled with layered medium. This 2-D Green's functions of the shielded layered medium is subsequently used in 2-D magneto-quasi-static formulation of the SVS-EFIE [4] to enable analysis of MTLs with complex cross-sections shielded by the rectangular enclosures. This paper is an abbreviated version of the complete descriptions provided in the chapter 3 of the thesis [5].

II. LAYERED MEDIA GREEN'S FUNCTION COMPUTATION

A. Computation of Primary Green's Function

The primary Green's function $G_{\epsilon_0}(\rho, \rho')$ is computed by considering only the top PEC plane, the bottom PEC plane,

and the multilayered substrate sandwiched in between. The layered medium Green's function satisfies the 1-D Helmholtz equation in spectrum domain [4]. The numerical solution of it is obtained in pole-residual form by using FD technique as discussed in details in [4] and is not repeated here. This pole-residual form enables the closed form evaluation of the inverse Fourier transformation analytically. The closed form space domain primary Green's function $G_{\epsilon 0}$ at the discrete samples along y coordinate is expressed as

$$G_{\epsilon 0}(|x-x'|, y_n, y'_m) = \sum_{j=0}^{N-1} [T]_{nj} [T^{-1}]_{jm} [d]_m \frac{e^{-\sqrt{S_j}|x-x'|}}{2\sqrt{S_j}}, \quad (1)$$

where y_n is the observation point elevations, y'_m is source point elevations, $[T]$ is the matrix storing the eigenvectors and S_j is the j th eigenvalue.

B. Accounting for Sidewalls Using Image Theory

The layered media Green's function has contributions from the left PEC plane and right PEC plane through image theory. The images of the point source located at x' created by the left PEC plane situated at x_1 and the right PEC plane situated at x_2 can be classified into four categories according to the distance between the position of the images to the PEC planes. Source has infinite number of images beyond the left hand side of left PEC plane and the right hand side of right PEC plane. The distance between left PEC plane and the right PEC plane is d . The point source and all the images of the point sources will have the contributions to the magnetic vector potential Green's function at observation location x .

From the image theory, the total Green's function can be obtained from the summation of the primary contribution due to the point source and the contributions due to all the images of the source point with respect to the sidewall, which after analytic evaluation of geometric series over infinite number of images yields

$$G_{\epsilon}(x, x', y_n, y'_m) = \sum_{j=0}^{N-1} [T]_{nj} [T^{-1}]_{jm} \frac{[d]_m}{2\sqrt{S_j}} \left(e^{-\sqrt{S_j}|x-x'|} + \frac{1}{1 - e^{-\sqrt{S_j}2d}} \left[-e^{-\sqrt{S_j}(2x_2-x'-x)} - e^{-\sqrt{S_j}(-2x_1+x'+x)} + e^{-\sqrt{S_j}(x_2-x_1+x'-x+d)} + e^{-\sqrt{S_j}(-x_1+x_2-x'+x+d)} \right] \right). \quad (2)$$

The above formula (2) constitutes the new closed-form expression for the 2-D magneto-quasi-static Green's function of planar layered medium inside the rectangular enclosure. In the following Section we briefly revisit the 2-D magneto-quasi-static SVS-EFIE and the use of new proposed Green's function in its formulation.

III. SVS-EFIE FORMULATION IN MULTILAYERED MEDIA

Magneto-quasi-static SVS-EFIE is formulated for auxiliary surface current density J_z residing conductor boundaries. Substitution of single-source field representation into 2-D Volume EFIE [4] and enforcement of the latter on the boundary of the conductor ∂S while approaching the boundary ∂S from inside its volume S yields [4]

$$-\iota\omega\mu_0 \oint_{\partial S} G_{\sigma}(\boldsymbol{\rho}, \boldsymbol{\rho}') J_z(\boldsymbol{\rho}') d\boldsymbol{\rho}' + \sigma\omega^2\mu_0 \oint_{\partial S} \left[\iint_S G_{\epsilon}(\boldsymbol{\rho}, \boldsymbol{\rho}') G_{\sigma}(\boldsymbol{\rho}', \boldsymbol{\rho}'') ds' \right] J_z(\boldsymbol{\rho}'') d\boldsymbol{\rho}'' = V_{p.u.l.}, \quad \boldsymbol{\rho} \in \partial S. \quad (3)$$

In (3), $G_{\sigma}(\boldsymbol{\rho}, \boldsymbol{\rho}')$ is the Green's function of the homogeneous conducting space having wavenumber k_{σ} , conductor bulk conductivity σ . ι is imaginary unity $\sqrt{-1}$, $\omega = 2\pi f$, and f is frequency. For conductor or group of conductors forming a MTLs, the Green's function $G_{\epsilon}(\boldsymbol{\rho}, \boldsymbol{\rho}')$ in (3) is given by (2).

IV. NUMERICAL RESULTS

In order to validate the proposed computation of the shielded multilayered media Green's function, the layered medium SVS-EFIE formulation is applied to model 2-D shielded microstrip transmission lines embedded in the multilayered media (Fig. 1). The numerical results obtained from proposed approach are compared to those obtained using COMSOL FEM commercial solver.

In the experiments, two rectangular cross-section conductor transmission lines made of copper with $\sigma = 5.8 \cdot 10^7$ S/m are situated in the air layer above lossy substrate with $\sigma_s = 1 \cdot 10^4$ S/m, $\epsilon_s = 12\epsilon_0$ and enclosed in a perfect electric conductor (PEC) box as depicted in Fig. 1. Each conductor has width $w = 20 \mu\text{m}$, thickness $t = 6 \mu\text{m}$. The distance between the two conductors is $w = 20 \mu\text{m}$. The height of the PEC box is fixed at $408.02198 \mu\text{m}$, the distance between the left conductor to the left PEC wall is the same as the distance between the right conductor to the right PEC wall. The width of the PEC box d varies in the experiments to test the impact of the PEC sidewalls on the fields and extracted network parameters of such shielded MTLs. In these experiments the y coordinate is discretized by the FD grid as shown in Fig. 1. The intervals of interest $y \in [0\mu\text{m}, 10\mu\text{m}]$, $y \in [10\mu\text{m}, 20\mu\text{m}]$, and $y \in [20\mu\text{m}, 30\mu\text{m}]$ are discretized with the uniform samples. The conductors are contained in the interval $y \in [10\mu\text{m}, 20\mu\text{m}]$ and the sampling step size in each interval may be different. Then the boundary locations of the top PEC and the bottom PEC can be determined by the extension of the upper buffer regions $y \in (30\mu\text{m}, 219.01099\mu\text{m})$ and the lower buffer regions $y \in (0\mu\text{m}, -189.01099\mu\text{m})$ through mapping the physical y coordinate to the parametric variable t . This transformation enables modeling the top PEC and the bottom PEC far away from each other easily without applying dense uniform FD samples. Distribution of the volumetric electric current density in the two-conductor MTLs situated above the lossy substrate at 5GHz computed using SVS-EFIE and

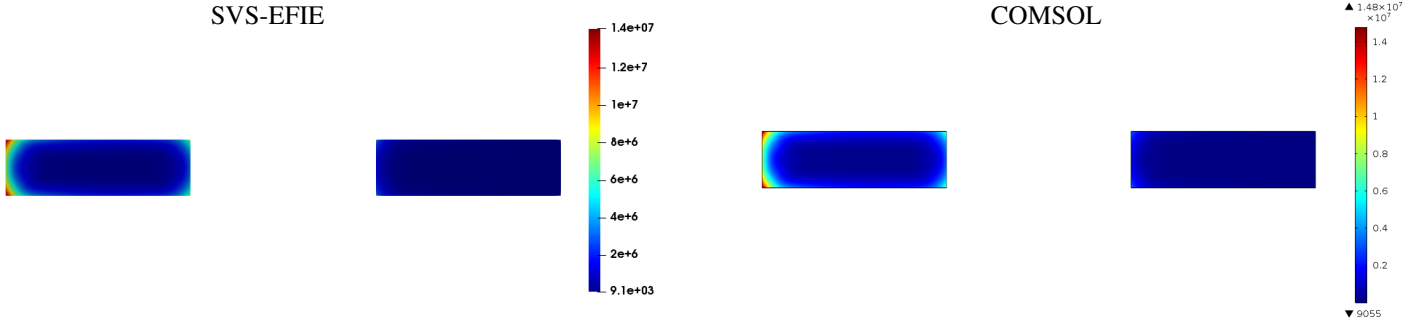


Fig. 2. Volumetric current density in the two-conductor MTLs situated above the lossy substrate at 5GHz. The box width is $70.0 \mu\text{m}$. The conductors are driven with 1V in the left conductor and 0V in the right conductor. The current distribution computed using SVS-EFIE is shown in the left. The validated result by COMSOL FEM is shown in the right.

COMSOL FEM solver with the enclosure width $d = 70.0 \mu\text{m}$ is shown in Fig. 2. The conductors are driven with voltage 1V in the left conductor and 0V in the right conductor. The self-resistance R_{11} and self-inductance L_{11} (mutual-resistance R_{12} and mutual-inductance L_{12} were also studied and omitted to save space) for various enclosure widths d at 5GHz are shown in Fig. 3 and Fig. 4, respectively. The numerical performance of SVS-EFIE based RL extraction is demonstrated in [6].

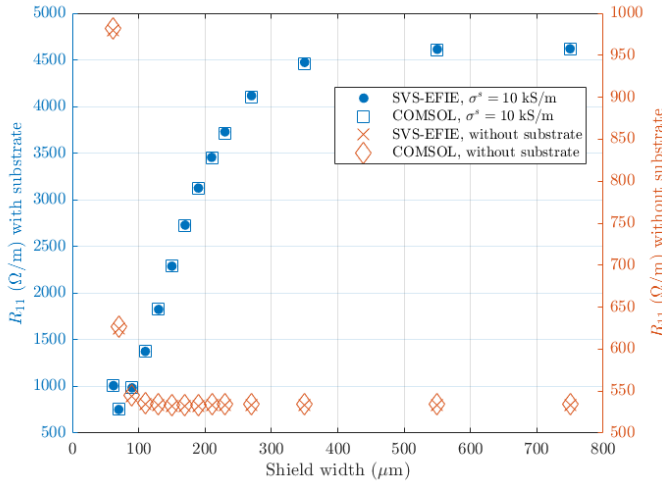


Fig. 3. Per-unit-length self-resistance R_{11} in two-conductor MTLs at 5GHz shown in Fig. 1 as functions of box width for the homogeneous box fill and box containing substrate of conductivity $\sigma_s = 10\text{kS/m}$.

V. CONCLUSIONS

Paper presents new approach to closed-form evaluation of the 2-D Green's function of the rectangular enclosure filled with planar layered medium. The method allows for analytic evaluation of the Fourier transform integrals as well as subsequent analytic addition of infinite geometric progression series over the images occurring due to vertical walls of the rectangular enclosure. The resultant new Green's function representation can be used for analysis of complex cross-section multi-conductor transmission lines in layered media.

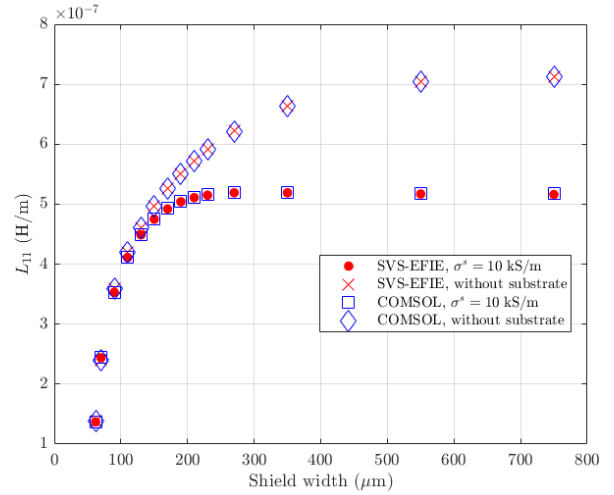


Fig. 4. Per-unit-length self-inductance L_{11} in two-conductor MTLs at 5GHz shown in Fig. 1 as functions of box width for the homogeneous box fill and the box containing substrate of conductivity $\sigma_s = 10\text{kS/m}$.

REFERENCES

- [1] Y. Wang, M. Yu, and K. Ma, "Substrate integrated suspended slot line and its application to differential coupler," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 12, pp. 5178–5189, Dec. 2020.
- [2] X. Li, I. Jeffrey, M. Al-Qedra, and V. I. Okhmatovski, "Error-controlled static layered-medium Green's function computation via hp-adaptive spectral differential equation approximation method," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 11, no. 9, pp. 1329–1342, Sep. 2021.
- [3] X. Li, S. Zheng, I. Jeffrey and V. Okhmatovski, "Closed-form evaluation of mixed potential shielded layered media Green's functions with spectral differential equation approximation method," *IEEE Trans. Microw. Theory Tech.* vol. 70, no. 5, pp. 2553–2565, Mar. 2022.
- [4] S. Zheng, A. Menshov, and V. I. Okhmatovski, "New single-source surface integral equation for magneto-quasi-static characterization of transmission lines situated in multilayered media," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 12, pp. 4341–4351, Dec. 2016.
- [5] S. Zheng, "New single-source surface integral equation for solution of electromagnetic problems in multilayered media," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Manitoba, Winnipeg, MB, Canada, 2022. [Online]. Available: <http://hdl.handle.net/1993/36682>.
- [6] M. Shafieipour, Z. Chen, A. Menshov, J. De Silva and V. Okhmatovski, "Efficiently computing the electrical parameters of cables with arbitrary cross-sections using the method-of-moments," *ELSEVIER Electr. Power Syst. Research*, vol. 162, pp. 37–49, Sep. 2018.

An Efficient Parallel Electromagnetic Solver for Extracting Scattering Parameters from Large Electrical Interconnects With Many Ports

Damian Marek, Piero Triverio

Edward S. Rogers Sr. Dept. of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada
damian.marek@mail.utoronto.ca, piero.triverio@utoronto.ca

Abstract—An efficient parallel solver is proposed for extracting port parameters from electrical interconnects with many ports. We demonstrate that block iterative methods can be used to improve convergence rate and parallel efficiency. The proposed method is up to $16\times$ faster than an existing method on structures with up to 128 ports and 3 million unknowns.

Index Terms—boundary element method, adaptive integral method, block iterative methods, skin effect modeling

I. INTRODUCTION

The boundary element method (BEM) is a popular technique for the electromagnetic analysis of large complex structures, which abound in modern electrical interconnects and electronic packages. Unlike volumetric formulations, the BEM only requires a surface mesh of conductive objects. Therefore, the number of unknowns can be considerably reduced compared to volumetric methods. However, the BEM results in a dense system of equations, which limits the size of problems that can be solved. As a result, the BEM is commonly partnered with an acceleration method, such as the adaptive integral method (AIM), and an iterative algorithm is used to compute the solution, e.g., the generalized minimum residual (GMRES) algorithm [1].

Although existing parallel BEM solvers can simulate large structures, most of them are not suitable for the electromagnetic analysis of electrical interconnects. The few BEM solvers that are applicable [2], [3], require considerable improvement before they may be used to efficiently analyze realistic structures. For example, silicon interposers used for high-bandwidth memory may contain thousands of signal lines. Extracting all scattering parameters of these structures would require repeatedly solving a system of equations thousands of times. For direct solvers, which are not easily applicable to the BEM, increasing the number of excitation vectors remains computationally efficient, since the dominant computational workload is the matrix factorization step. On the other hand, the computational cost of solving the system of equations using the GMRES algorithm [1] is proportional to the number of excitation vectors. Therefore, the computation time can grow considerably when thousands of ports need to be characterized.

This work was partially supported by: Advanced Micro Devices, Natural Sciences and Engineering Research Council of Canada, Digital Research Alliance of Canada, CMC Microsystems.

978-1-6654-5075-1/22/\$31.00 © 2022 IEEE

In this work, we propose an efficient parallelization strategy for an AIM-accelerated BEM solver that enables rapid extraction of port parameters from large electrical interconnect structures. The proposed strategy leverages block iterative methods [4] to reduce the number of iterations associated with multiple port excitations and improves the overall parallel efficiency. The solver is validated on the package microstrip benchmark [5]. Subsequently, the proposed method is tested on stripline array structures of increasing size and port count. The proposed method is found to be $16\times$ faster than an existing method on the largest structure, which is composed of 64 striplines and 3 million mesh edges.

II. FORMULATION

We consider structures comprising rough lossy conductors embedded in a layered medium. Our objective is to model the structure using Maxwell's equations and extract scattering (S) parameters. In order to avoid low-frequency breakdown issues, we use the augmented electric field integral equation (AEFIE) [6], along with an approximate surface impedance boundary condition (SIBC) [7], to formulate the system of equations. After discretization with Rao-Wilton-Glisson (RWG) [8] and pulse basis functions, the AEFIE can be written as:

$$\begin{bmatrix} jk_0\mathbf{L}_A + \eta_0^{-1}\mathbf{Z}_s & -\mathbf{D}^T\mathbf{L}_\Phi\mathbf{B} \\ \mathbf{FD} & jk_0\mathbf{I} + \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{J}_s \\ c_0\boldsymbol{\rho}_r \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{I}_s \end{bmatrix}. \quad (1)$$

In (1), \mathbf{L}_A and \mathbf{L}_Φ are the discretized vector and scalar potential parts of the \mathcal{L} operator [9], which involve the Green's function of layered media [10]. Coupling to a Thevenin-equivalent circuit excitation is provided by matrix \mathbf{C} [11]. The SIBC is enforced by \mathbf{Z}_s , which includes the Groiss correction factor for conductors with surface roughness [12]. The identity matrix is given by \mathbf{I} , while \mathbf{I}_s is the excitation current. The remaining matrices are identical to [6].

In the AIM, (1) is solved iteratively, and the required matrix-vector products (MVPs) involving \mathbf{L}_A and \mathbf{L}_Φ are accelerated with fast Fourier transforms. To improve the convergence properties of the system of equations in (1), the constraint preconditioner from [6] is used, which requires the solution of a sparse system of equations. The cost of solving (1) scales linearly with the number of port excitations N_P . When N_P is large, an improved approach is needed to reduce the computation time.

III. PROPOSED METHOD

A simple parallelization strategy for solving (1) is to treat each excitation as a separate problem assigned to a small fraction of the total available processes. Although this would lead to an embarrassingly parallel computational problem, duplicating the system matrix for each group of processes would require an excessive amount of memory. The proposed method distributes the system matrix over all processes and uses memory sparingly.

We propose to use block iterative methods [4] in order to reduce the total time needed to extract S parameters. In block iterative methods, excitation and unknown vectors are concatenated into tall skinny dense matrices. Then, the system of equations is solved for all unknown vectors simultaneously. For example, a block iterative method can be used to solve linear systems of the form

$$\mathbf{A}\mathbf{X} = \mathbf{B}, \quad (2)$$

where \mathbf{A} is the system matrix, \mathbf{X} is a dense matrix of concatenated unknown vectors, and \mathbf{B} is a dense matrix of concatenated excitation vectors.

Block iterative methods have two main advantages over conventional iterative methods when multiple excitations are present. First, they can generate larger subspaces, which can improve the convergence rate, thus, reducing the number of iterations [4]. Second, the block nature of the matrix-matrix products (MMPs) increases the arithmetic intensity and can improve computational efficiency [4]. This second point is especially valid for distributed-memory solvers, where parallel efficiency can be improved by packing more data into communication buffers. On the other hand, a block iterative method will use additional memory to store vectors that are needed for the block Krylov subspace and any intermediate computations required in system matrix multiplications. However, memory usage in the AIM is dominated by the near-region portions of \mathbf{L}_A and \mathbf{L}_Φ , so the effect of any additional vectors on the total memory consumption will be small. In any case, the number of excitation vectors may always be reduced to fit the available memory of the system with remaining excitation vectors being solved in a subsequent batch solve.

A state-of-the-art parallel solver was modified [3], [13], so that the system matrix in (1) could efficiently multiply dense matrices. This was achieved by swapping all MVPs involving submatrices of (1) with MMPs. This includes the AIM-accelerated matrices \mathbf{L}_A and \mathbf{L}_Φ . Vectors required to store intermediate results of a MVP with (1) were replaced by dense matrices. The preconditioner was enhanced to enable its efficient application to dense matrices using a similar approach. These changes allowed for the parallel solver to make use of the block GMRES algorithm implemented in [4].

IV. RESULTS

A. Package Microstrip Benchmark

In this section, the proposed method is validated on the package microstrip benchmark from [5]. This benchmark

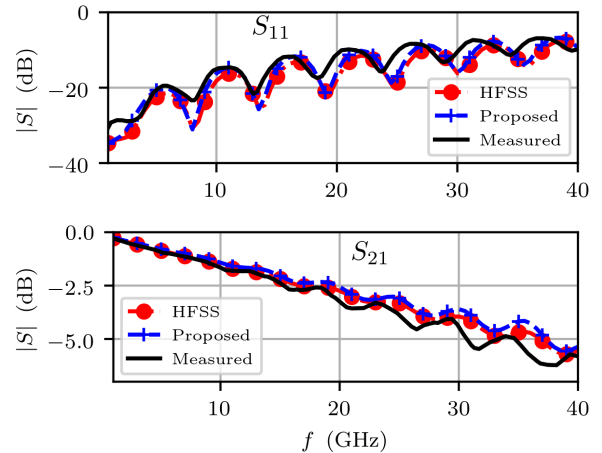


Figure 1. S parameter plots for the package microstrip benchmark [5]. Top panel: return loss $|S_{11}|$ Bottom panel: insertion loss $|S_{21}|$.

consists of a single microstrip with an excitation port at each end. A detailed description is available in [5].

The structure was simulated from 1 to 40 GHz using the proposed method and Ansys HFSS. Measured results were extracted from plots in [5] using an online digitizer and are plotted alongside the simulation results in Fig. 1. The S parameters computed by simulation are in good agreement. As expected, the simulated S parameters deviate slightly from the measured S parameters as observed in the benchmark description [5].

B. Stripline Array

Next, the performance of the proposed solver is tested by extracting S parameters from structures of increasing size. A schematic of the smallest structure is shown in Fig. 2. These structures are inspired by the package microstrip benchmark and the set of structures in [2]. Each structure consists of two ground planes, N_S striplines, and $2N_S$ other conductors, which are not connected to ports. An example structure with $N_S = 4$ is presented in Fig. 2. All conductors are assigned a conductivity of 4.5×10^7 S/m and an RMS surface roughness of $0.3 \mu\text{m}$. These structures are embedded in a layered medium composed of a single dielectric material of thickness $180 \mu\text{m}$ surrounded by air. The dielectric has an electric permittivity of 3.4 and a loss tangent of 0.018. The length of the structures is set to 28.26 mm, which is comparable to the size of the package power integrity benchmark problem [5]. An excitation frequency of 10 GHz was used for all simulations. Table I contains supplementary information for each structure, including the number of mesh edges N_E , the number of AIM grid points N_x, N_y, N_z , and the number of effective MVPs N_{MVP} required by the GMRES algorithm and the proposed method to reach a relative tolerance of 10^{-4} . The value of N_{MVP} for the block iterative method is computed by multiplying the number of MMPs by N_P .

The execution time of the proposed method was compared to the existing strategy which uses the GMRES algorithm [1].

Table I
PARAMETERS FOR THE STRIPLINE ARRAY STRUCTURES IN SECTION IV-B.

N_S	N_E	N_x	N_y	N_z	N_{MVP}	
					GMRES	Proposed
4	399,603	600	40	10	1,718	1,952
16	936,747	600	64	10	11,910	5,856
64	3,089,775	600	160	10	93,179	21,888

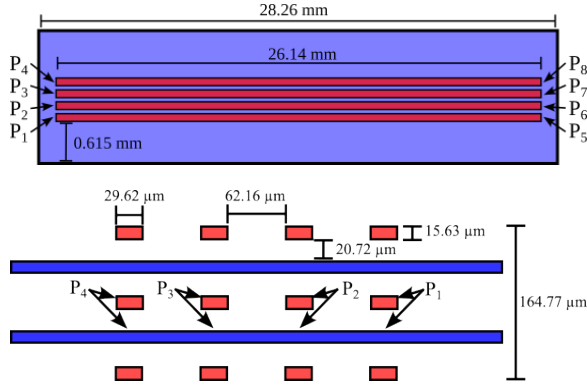


Figure 2. Schematic of the $N_S = 4$ stripline array structure. Top panel: top view. Bottom panel: cross sectional view.

To compare the computational efficiency of each method independently of their convergence rate, the accumulated time spent in solver steps was divided by N_{MVP} , which results in an average time for a single excitation vector. All simulations were run on the Scinet Niagara cluster where each node has 40 Intel Skylake cores, running at 2.4 GHz, and 202 GB of memory. The total number of CPU cores used ranged from 20 to 1,280. The results of the scalability experiments for each stripline array structure are plotted in Fig. 3.

The results indicate that the reduction in solver iterations reduces the total solve time. Furthermore, the plots averaged by N_{MVP} demonstrate that many solver steps, such as, applying the preconditioner, are more computationally efficient when they are applied to multiple vectors at once. The proposed method was $16\times$ faster at extracting the 128 S parameters of the largest structure, when 1,280 processes were used.

V. CONCLUSION

We presented a new strategy for accelerated BEM solvers that improves the efficiency of computing scattering parameters of large electrical interconnect structures with many ports. This approach leverages the superior convergence and computational efficiency properties of advanced block iterative methods. The proposed method was $16\times$ faster than existing strategies when compared on a structure with 128 ports.

REFERENCES

- [1] Y. Saad and M. H. Schultz, "GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems," *SIAM J. Sci. Stat. Comput.*, vol. 7, no. 3, pp. 856–869, Jul. 1986.
- [2] C. Liu, K. Aygün, and A. E. Yılmaz, "A parallel FFT-accelerated layered-medium integral-equation solver for electronic packages," *Int. J. Numer. Model.: Electron. Networks, Devices and Fields*, pp. 1–17, Sep. 2019.

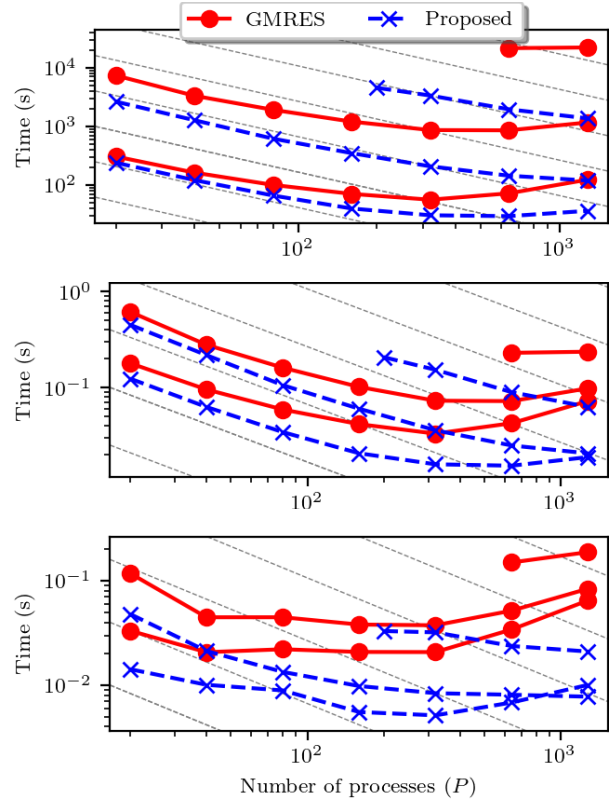


Figure 3. Execution times for the three stripline array structures described in Table I. Dashed lines represent ideal scalability slopes and are spaced by a factor of 4. Top panel: total solve time. Center panel: average time to compute one iteration. Bottom panel: average time to apply the preconditioner.

- [3] D. Marek, S. Sharma, and P. Triverio, "A parallel boundary element method for the electromagnetic analysis of large structures with lossy conductors," *IEEE Trans. Antennas Propag.*, to be published.
- [4] P. Jolivet, J. E. Roman, and S. Zampini, "KSPHPDDM and PCHPDDM: Extending PETSc with advanced Krylov methods and robust multilevel overlapping Schwarz preconditioners," *Comput. Math. Appl.*, vol. 84, pp. 277–295, 2021.
- [5] IEEE EPS Technical Committee on Electrical Design, Modeling and Simulation, "Packaging Benchmark Suite," 2021. [Online]. Available: <https://packaging-benchmarks.org/>
- [6] Z. G. Qian and W. C. Chew, "Fast Full-Wave Surface Integral Equation Solver for Multiscale Structure Modeling," *IEEE Trans. Antennas Propag.*, vol. 57, no. 11, pp. 3594–3601, 2009.
- [7] S. Yuferev and N. Ida, *Surface Impedance Boundary Conditions*. CRC Press, 2009.
- [8] S. Rao, D. Wilton, and A. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. Antennas Propag.*, vol. 30, no. 3, pp. 409–418, May. 1982.
- [9] W. C. Gibson, *The Method of Moments in Electromagnetics*. Chapman and Hall/CRC, Jul. 2014.
- [10] K. A. Michalski and D. Zheng, "Electromagnetic Scattering and Radiation by Surfaces of Arbitrary Shape in Layered Media, Part I: Theory," *IEEE Trans. Antennas Propag.*, vol. 38, no. 3, pp. 335–344, Mar 1990.
- [11] Y. Wang, D. Gope, V. Jandhyala, and C. J. R. Shi, "Generalized Kirchoff's current and voltage law formulation for coupled circuit-electromagnetic simulation with surface integral equations," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 7, pp. 1673–1682, 2004.
- [12] S. Groiss, I. Bardi, O. Biro, K. Preis, and K. R. Richter, "Parameters of lossy cavity resonators calculated by the finite element method," *IEEE Trans. Magn.*, vol. 32, no. 3, pp. 894–897, 1996.
- [13] D. Marek, S. Sharma, and P. Triverio, "An efficient strategy for distributing the mesh of parallel electromagnetic solvers based on the AIM," in *2022 16th Eur. Conf. Antennas Propag.*. IEEE, 2022, pp. 1–5.

A Transfer Learning Approach to Expedite Training of Artificial Neural Networks for Variability-Aware Signal Integrity Analysis of MWCNT Interconnects

Surila Guglani[#], Km Dimple[#], Avirup Dasgupta[#], Rohit Sharma^{*}, Brajesh Kumar Kaushik[#], and Sourajeet Roy[#]

[#] Department of Electronics and Communication Engineering, Indian Institute of Technology Roorkee, Roorkee, India

^{*} Department of Electrical Engineering, Indian Institute of Technology Ropar, Ropar, India

Email: sourajeet.roy@ece.iitr.ac.in

Abstract — In this paper, an artificial neural network (ANN) trained using a novel transfer learning approach is presented for the variability-aware signal integrity analysis of on-chip multi-walled carbon nanotube (MWCNT) interconnects. In the proposed transfer learning approach, initially a secondary ANN is trained to emulate the signal integrity quantities of interest of an approximate equivalent single conductor (ESC) model of the MWCNT interconnects. Thereafter, the values of the weights and bias terms of this secondary ANN are used to expedite the training of the primary ANN that will emulate the signal integrity quantities of the more rigorous multiconductor circuit (MCC) model of the MWCNT interconnects.

Keywords — Artificial neural networks (ANNs), high-speed interconnects, multi-walled carbon nanotubes (MWCNTs), transfer learning, variability analysis, signal integrity.

I. INTRODUCTION

At sub-22 nanometer technology nodes, multi-walled carbon nanotube (MWCNT) on-chip interconnects display significantly lower scattering, higher current carrying capacity, and higher thermal conductivity than conventional copper interconnects [1]. For right-the-first-time design of such MWCNT interconnects, electronic design automation (EDA) tools need to explore how variability in the geometrical, material, and physical parameters of the interconnect structure affect the signal integrity quantities of interest such as the signal delay, peak crosstalk, and eye diagram characteristics. Typically, a Monte Carlo framework is employed to perform the variability-aware signal integrity analysis of MWCNT interconnects [2]. However, because of the poor convergence of the Monte Carlo technique, this entails performing thousands of SPICE simulations of the highly rigorous and complex multiconductor circuit (MCC) model of the MWCNT interconnects at extremely high computational time costs [3].

To mitigate this problem, surrogate models based on machine learning (ML) regression techniques such as artificial neural networks (ANNs) have been reported in the literature [4]-[6]. These surrogate models take the form of analytic functions emulating the nonlinear dependency of the signal integrity quantities of interest on the geometrical, material, and physical parameters of the interconnect structures. In particular, ANNs have the capacity to emulate highly nonlinear

relationships that other methods such as those based on smooth polynomial basis functions cannot [7]. Once trained, the ANN surrogate models can be probed at a miniscule fraction of the computational time cost of a SPICE MCC model simulation to estimate the values of the signal integrity quantities for different input parameter values. Hence, these ANN surrogate models can replace the repeated and expensive SPICE MCC model simulations in a Monte Carlo framework. Unfortunately, the training of ANNs is a very tedious process where massive amount of training data may be required using repeated SPICE simulation of the MWCNT interconnects.

In this paper, a transfer learning (TL) approach is developed to address the above computational costs to train conventional ANNs [8]. The TL approach focuses on extracting and reusing the knowledge gained from solving a secondary or a related problem to expedite the training of the primary ANN to solve the target problem. In this work, the secondary problem is defined to be the training of an ANN surrogate model to emulate the signal integrity quantities of an MWCNT interconnect modeled by the inaccurate equivalent single conductor (ESC) model [9]. Importantly, generating the training dataset for this secondary problem is numerically very cheap given the relatively small SPICE simulation cost of the ESC model compared to the MCC model. Next, the entire ANN architecture including the optimized weights and bias values of the secondary problem is used as an informed starting guess to train the primary ANN to emulate the signal integrity quantities of the interconnect modeled by the rigorous MCC model. Given the use of the informed initial guess of the weights and bias terms, the number of SPICE MCC model simulations required for training the primary ANN will be a significantly reduced.

II. PROPOSED TRANSFER LEARNING APPROACH

A. Problem Statement

Consider a general MWCNT interconnect network consisting of M conductors, each with N_s concentric shells as shown in Fig. 1. The variability in the geometrical, physical, and material parameters of the interconnect structure is mapped to N mutually uncorrelated random variables $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_N]$ located within the multidimensional support Ω . The task is to

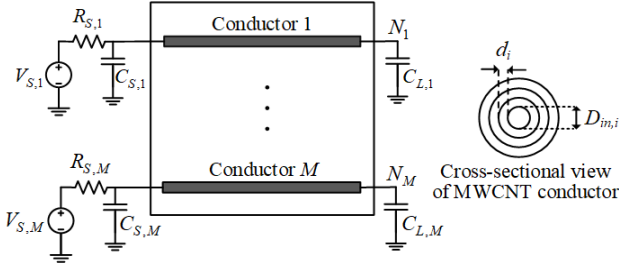


Fig. 1: Circuit schematic of the MWCNT interconnect network with the cross-sectional view of the MWCNT conductors.

further map the impact of these random variables on the signal integrity quantities of interest of the interconnects. To that end, fully connected conventional ANNs can be developed to emulate the signal integrity quantities as

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \mathbf{M} \\ y_p \end{bmatrix} = \begin{bmatrix} F_1(\boldsymbol{\lambda}) \\ \mathbf{M} \\ F_p(\boldsymbol{\lambda}) \end{bmatrix} = \mathbf{F}(\boldsymbol{\lambda}) \quad (1)$$

In (1), \mathbf{y} represents the vector of P signal integrity quantities of interest which are also the outputs of the ANN. The vector \mathbf{F} represents the vector of the analytic nonlinear mapping functions identified by the ANN. The training dataset for this ANN is represented as $\{\boldsymbol{\lambda}^{(k)}, \mathbf{y}_{MCC}(\boldsymbol{\lambda}^{(k)})\}_{k=1}^{N_t}$ where $\mathbf{y}_{MCC}(\boldsymbol{\lambda}^{(k)})$ is the set of the signal integrity quantities determined from SPICE MCC model simulation at the k -th training sample $\boldsymbol{\lambda}^{(k)} = [\lambda_1^{(k)}, \lambda_2^{(k)}, \dots, \lambda_N^{(k)}]$. Here, the key challenge is that the number of training samples, N_t , used to reliably train the ANN is usually massively large, thereby requiring prohibitively high time costs to extract the training dataset from repeated SPICE MCC model simulations. To address this problem, a transfer learning (TL) approach is developed as explained in the next few subsections.

B. Transfer Learning: Training the Secondary Problem

The first step in the proposed TL approach is to solve a related secondary problem – to train an ANN to emulate the target signal integrity quantities of (1) as functions of $\boldsymbol{\lambda}$ where the training dataset is generated using SPICE ESC model simulations instead of MCC model simulations. In this case, the training dataset will be $\{\boldsymbol{\lambda}^{(k)}, \mathbf{y}_{ESC}(\boldsymbol{\lambda}^{(k)})\}_{k=1}^{N_t}$ where $\mathbf{y}_{ESC}(\boldsymbol{\lambda}^{(k)})$ is the signal integrity quantities obtained from a SPICE ESC model simulation at $\boldsymbol{\lambda}^{(k)}$. The outputs of this ANN will be

$$\mathbf{z} = \begin{bmatrix} z_1 \\ \mathbf{M} \\ z_p \end{bmatrix} = \begin{bmatrix} H_1(\boldsymbol{\lambda}) \\ \mathbf{M} \\ H_p(\boldsymbol{\lambda}) \end{bmatrix} = \mathbf{H}(\boldsymbol{\lambda}) \quad (2)$$

where for a standard three-layer multi-perceptron architecture

$$z_i(\boldsymbol{\lambda}^{(k)}) = \sigma_{i,3} \left(b_{i,3} + \sum_{j=1}^{N_h} w_{2,3}^{j,i} \sigma_{j,2} \left(b_{j,2} + \sum_{p=1}^N w_{1,2}^{p,j} \lambda_p^{(k)} \right) \right) \quad (3)$$

In (3), $\sigma_{q,p}$ refers to the q -th nonlinear activation function used in the neurons of the p -th layer, $b_{q,p}$ is the bias value entering

TABLE I
UNIFORMLY DISTRIBUTED DEVICE PARAMETERS

S. No.	Parameter	Nominal Value	Variation
1-3	$D_{m,i}$ (Diameter of inner shell for i -th conductor)	2.28 nm	+/- 15%
4-6	d_i (Intershell distance for i -th conductor)	0.14 fF	
7-9	$C_{s,i}$ (Driver capacitance of i -th conductor)	0.049 fF	
10-12	$C_{L,i}$ (Load capacitance of i -th conductor)	50 nm	
13	w (Conductor spacing)	22 nm	
14	H (Height of conductor above ground plane)	50 nm	

the q -th neuron of the p -th layer, and $w_{q,p}^{\alpha,\beta}$ is the synaptic weight linking the α -th neuron of the q -th layer to the β -th neuron of the p -th layer. Once the weights and bias terms of (3) are optimized to minimize the mean squared error loss function

$$f_2(\mathbf{w}, \mathbf{b}) = \frac{1}{N_t} \sum_{k=1}^{N_t} \left\| \mathbf{y}_{ESC}(\boldsymbol{\lambda}^{(k)}) - \mathbf{z}(\boldsymbol{\lambda}^{(k)}) \right\|_2^2 \quad (4)$$

the secondary ANN is said to be trained. Here, \mathbf{w} and \mathbf{b} refer to the full set of weights and bias term in (2)-(3). At this point, it is emphasized that the cost of generating the training dataset for the secondary ANN is very cheap given that the ESC model is substantially smaller in size than the MCC model [7]. Moreover, given that the ESC model is an approximation of the MCC model, the secondary task of fitting the mapping functions of (2)-(3) is expected to be similar to the target task of fitting the original mapping functions of (1). Thus, the secondary ANN is ready for transfer learning.

C. Transfer Learning: Training the Primary ANN

In the second step, the secondary ANN architecture along with the optimized values of the weights and bias terms in (3) is used as an initial guess to start the optimization process of the primary ANN of Section II A. The rationale behind this choice is that given the similarity of the tasks performed by the secondary and the primary ANNs, the optimal values of the weights and bias terms in the primary ANN will be close to the values of the secondary ANN. Therefore, the number of training samples, N_t , in the training dataset for the primary ANN required to refine the weights and bias terms from their initial guesses to their optimized values will be significantly reduced compared to the conventional training process where the initial guess of the weights and bias terms are randomly selected. In other words, the number of SPICE MCC model simulations required to train the primary ANN will be reduced. This benefit is validated by a numerical example in the next section.

III. NUMERICAL RESULTS AND DISCUSSIONS

In this section, a 3-line MWCNT interconnect network as shown in Fig. 1 is considered. The variability in the geometrical

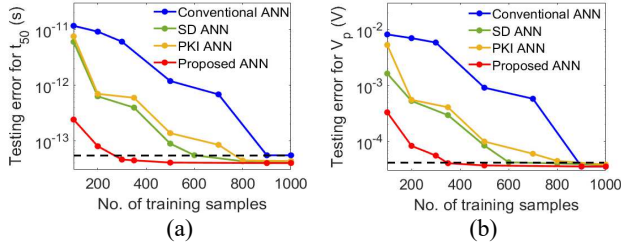


Fig. 2: Decay of the RMS testing error with the increasing number of training points for different ANNs. (a) 50% delay time for response at node N_1 , (b) peak crosstalk at node N_2 .

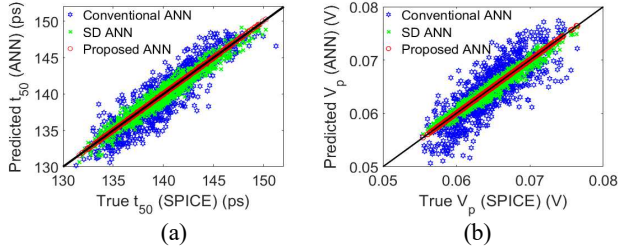


Fig. 3: Scatter plot of (a) 50% delay time for response at node N_1 and (b) peak crosstalk at node N_2 at 1000 testing points for different ANNs trained using 350 MCC samples.

and physical parameters of the interconnect is provided in Table I. In this example, Line 1 and 3 are the active lines while Line 2 is quiet. The SI quantities of interest for this example are the 50% delay time (t_{50}) for the active line and peak crosstalk (V_p) for the victim line. The statistics of these SI quantities are evaluated using a Monte Carlo based analysis with 20,000 samples. The Monte Carlo analysis is performed using the following methods – the direct method using 20,000 SPICE MCC model simulations, an ANN trained using the conventional training method of Section II A, an ANN trained using the well-known source difference (SD) [6] and prior-knowledge input (PKI) methods [10], and an ANN trained using the proposed TL approach. All ML techniques utilize the same training dataset of $N_t = \{100, 200, 300, 350, 500, 600, 700, 800, 900, 1000\}$ Latin hypercube sampling points and a common testing dataset of 1000 points. All ANN models use a single hidden layer, hyperbolic tangent activation function, and the Levenberg-Marquardt optimizer. The conventional ANN requires 11 hidden neurons and 4000 epochs for training, the SD ANN requires 8 hidden neurons and 2000 epochs for training, and the PKI ANN requires 9 hidden neurons and 2000 epochs for training. The proposed ANN trained using the TL approach requires 10 hidden neurons for both the secondary and target ANN. In Fig. 2, the decay of the testing error with the increasing number of training points for all the above ANNs is displayed. From Fig. 2, it is clear that the proposed TL approach requires the smallest number of training samples (350 MCC and 500 ESC samples) compared to not only the conventional ANN (900 MCC samples) but also the SD and PKI ANNs (600 and 800 MCC samples respectively, each with a further 500 ESC samples) to reach the error threshold of 5 fs and 4.5e-5 V for

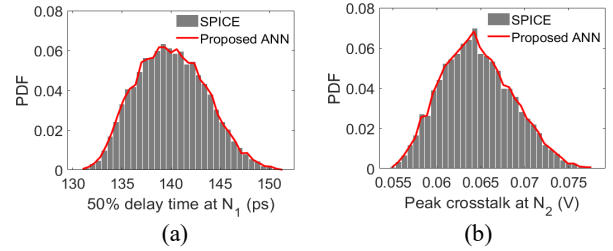


Fig. 4: PDF of (a) 50% delay time for response at node N_1 and (b) peak crosstalk at node N_2 for 20,000 Monte Carlo samples.

delay time and peak crosstalk, respectively. This corresponds to the proposed TL approach exhibiting the best speedup of roughly 2.5x over the conventional ANN during training. This is illustrated in the scatter plots of Fig. 3. Finally, the probability density function of the SI quantities of interest obtained using the proposed TL approach and the direct Monte Carlo approach are compared in Fig. 4.

IV. CONCLUSION

In this work, a novel transfer learning approach is developed to expedite the training of conventional ANNs when performing variability-aware signal integrity analysis of on-chip MWCNT interconnects. The key attribute of the proposed transfer learning approach is its ability to reuse the knowledge gained from training a related secondary ANN to intelligently accelerate the optimization of the primary ANN.

REFERENCES

- [1] A. Todri-Sanial, J. Dijon, and A. Maffucci (editors), *Carbon Nanotubes for Interconnects*. Switzerland: Springer International, 2017
- [2] A. Nieuwoudt and Y. Massoud, "On the optimal design, performance, and reliability of future carbon nanotube-based interconnect solutions," *IEEE Trans. Electron Devices*, vol. 55, no. 8, pp. 2097–2101, Aug. 2008.
- [3] H. Li, W. Y. Yin, K. Bannerjee, and J. F. Mao, "Circuit modeling and performance analysis of multi-walled carbon nanotube interconnects," *IEEE Trans. Electron Devices*, vol. 55, no. 6, pp. 1328–1337, June 2008
- [4] Q.-J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design*, Norwood, Massachusetts: Artech House 2000
- [5] A. Veluswami, M. S. Nakhla, and Q.-J. Zhang, "The application of neural networks to EM based simulation and optimization of interconnects in high-speed VLSI circuits," *IEEE Trans. Microw. Theory and Techn.*, vol. 45, no. 5, pp. 712–723, May 1997.
- [6] Y. Li, S. Bhatnagar, A. Merkley, D. Weber, and S. Roy, "A predictor-corrector algorithm for fast polynomial chaos-based uncertainty quantification of multi-walled carbon nanotube interconnects" *IEEE Trans. Comp., Packag. and Manuf. Technol.*, vol. 9, no. 10, pp. 1963–1975, Oct. 2019
- [7] R. Kumar *et al.*, "Knowledge based neural networks for fast design space exploration of hybrid copper-graphene on-chip interconnect networks," *IEEE Trans. Electromag. Comp.*, vol. 64, no. 1, pp. 182–195, Feb. 2022
- [8] T. G. Karimpanal and R. Bouffanais, "Self-organizing maps for storage and transfer of knowledge in reinforcement learning," *Adaptive Behavior*, vol. 27, no. 2, pp. 111–126, Dec. 2018
- [9] M. S. Sarto and A. Tamburrano, "Single-conductor transmission line model of multiwall carbon nanotubes," *IEEE Trans. Nanotechnol.*, vol. 9, no. 1, pp. 82–92, Jan. 2010
- [10] P. M. Watson, K. C. Gupta, and R. L. Mahajan, "Applications of knowledge-based artificial neural network modeling to microwave components," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 9, pp. 254–260, 1999

An Improved Methodology for High Frequency Socket Performance Characterization

Saikat Mondal, Dhanya Athreya, Emile Davies-Venn, Zhichao Zhang, and Kemal Aygün
Assembly Test and Technology Development, Intel Corporation
Chandler, USA

saikat.mondal@intel.com, dhanya.athreya@intel.com, emile.davies-venn@intel.com, zhichao.zhang@intel.com,
kemal.aygun@intel.com

Abstract—In this paper, we present an improved methodology to achieve good correlation between measured and modeled high frequency data for sockets. The resulting technique was applied to a land grid array (LGA) socket, designed to provide a detachable solution between a microelectronic package and a printed circuit board (PCB), while simultaneously satisfying the stringent electrical requirements for high speed signaling. Test vehicles were assembled with surface mount LGA sockets on test boards and a removable test package. 4-port and 12-port S-parameter measurements were performed on the test vehicle assembly. The socket only insertion loss and return loss performance was extracted using a de-embedding process. A good correlation was achieved between de-embedded measured and modeled differential-ended (DE) insertion loss (IL) data from DC to 16 GHz. For the first time such good correlation for de-embedded high frequency socket data has been reported as per the authors' best knowledge.

Index Terms—LGA socket, methodology, correlation, SI

I. INTRODUCTION

Within the last decade, there has been a drastic increase of data speeds in microelectronic systems, driven by applications such as big data, cloud services, and more recently by artificial intelligence, machine learning, and autonomous driving applications. To enable such data speed increases, there has been a continuous effort in developing low-loss substrates for both packages and PCBs, interconnect conductors with better surface finish, low-loss cables, connectors with low-loss and better shielding, smooth impedance transition between different channel components, and strategic routing strategies to lower the overall noise budget. Each of the component designs has its own challenges and requires individual signal integrity budgeting to successfully achieve the desired data rate. In a set of microelectronic systems, socket is an important component between the package and the PCB and one of the major contributors for impedance discontinuity and crosstalk within the channel. Hence, it is important to validate the performance of a socket on its own to effectively implement a successful channel.

LGA sockets, which are commonly used in microprocessor systems, are surface mounted (SMTed) on top of a PCB.

A certain amount of force is applied after the microprocessor package is inserted into an LGA socket to establish and maintain good electrical contact. Several factors should be considered to achieve a good high frequency electrical measurement to modeling correlation of a socket. For example, loading mechanism of the socket should be well controlled to mimic the desired loading condition. Developing a comprehensive yet focused methodology for high frequency socket characterization is important to produce high quality measurement data and build confidence in the corresponding high frequency socket models. A well correlated model will provide stronger confidence for the channel link simulation results and consequently on the overall system performance.

In literature, few earlier works exist for high frequency socket validation and correlation [1], [2]. However, majority of the previous work do not comprehend de-embedded socket measurement to modeling correlation. Post de-embedding correlation is specifically challenging, as any errors from probe to probe data can get accumulated. To reduce the error, different sources of uncertainties need to be addressed and removed, as much as possible, for better correlation. In this work, we have addressed several of these uncertainty factors, which can impact the measurement to modeling correlation, and demonstrated good post de-embedding correlation.

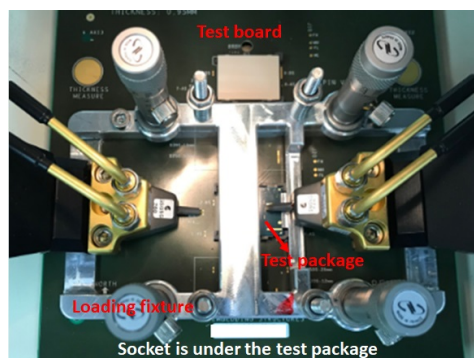


Fig. 1: Measurement setup of the SUT assembly using PNA.

II. TEST VEHICLE DESIGN

The test vehicle (TV) in this study contains three parts: a) Test package, b) socket under test (SUT), and c) test PCB. Low

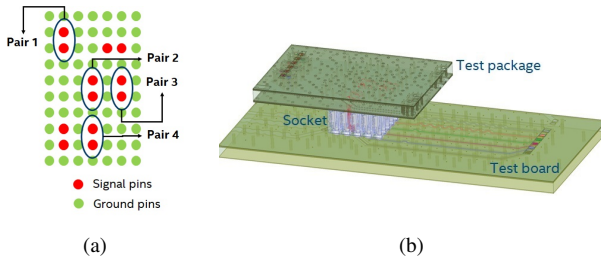


Fig. 2: (a) Test structure locations within the TV. (b) Sample model assembly of DE crosstalk test site.

loss PCB and package materials were chosen to minimize calibration and de-embedding errors. Since the primary objective of this study was to extract SUT-only characteristics, a test package and PCB with minimum number of layers were used, instead of using a typical microprocessor package and PCB stack-up with a large layer-count. The TV assembly with the loading fixture is shown in Fig. 1. S-parameter measurements were performed at different test sites within the TV to extract the key signal integrity parameters. These included: 1) DE IL, 2) DE return loss (RL), and 3) DE crosstalk. The test sites are shown in Fig. 2(a). ‘Pair 1’ is for characterizing DE IL and RL, which is measured by a 4-port VNA. ‘Pairs 2, 3, and 4’ are for characterizing DE crosstalk, which is measured by a 12-port VNA.

Once the SUT assembly is measured, the S-parameter results are de-embedded to extract socket-only parameters. Individual de-embedding structures were included within the test package and PCB with a 2X through length to remove the effect from long traces. Automated fixture removal (AFR) was used as the de-embedding methodology as AFR enables use of a small number of de-embedding structures on TV without compromising the quality [3]. Full-wave 3D simulations were performed on specific portions of the TV model using the SUT, package and PCB layer stack-up to mimic the assembly test conditions. A model assembly of the ‘Pairs 2, 3, and 4’ crosstalk site is shown in Fig. 2(b).

III. VALIDATION METHODOLOGY DEVELOPMENT

A comprehensive validation methodology was developed for the measurement to modeling correlation process, which takes into account some key uncertainties that impact the SUT measurement process and results. These are reviewed next.

A. Test PCB Impedance

The glass weave direction and fabrication process variation can impact the impedance of the manufactured traces within the test board. Too much impedance variation among the traces impacts the amount of accumulated error in the post de-embedding results. To remove this uncertainty, impedance screening of the test boards using time domain reflectometry (TDR) was performed and only boards meeting impedance target of nominal $\pm 5\%$ were used for SUT assembly and tests. Even though this manufacturing and screening process qualified approximately 30% of the total test boards, this was a crucial step to have high quality de-embedded measured data.

B. Solder Ball Shape

As the sockets are SMTed, another important parameter to consider is the resulting solder ball shape and geometry under the socket pins. High frequency performance results have been shown to be very sensitive to the solder ball shape [4]. Methodology from [4] was adopted to approximate the solder ball geometry for SUTs in this study. These were confirmed afterwards by also (destructive) cross-section measurements.

C. Socket Deflection Condition

As mentioned earlier, the high frequency response of the SUT is dependent upon the socket deflection condition. If the applied normal force on the SUT is not well controlled, then the deflection condition of the SUT will be different from the one assumed in the model, and can result in poor correlation. A controlled loading structure with high precision of deflection tuning capability ($\pm 5 \mu\text{m}$) was specifically designed for this study to provide the desired deflection condition. The two ends of the possible deflection conditions for the SUT were termed as a) the maximum deflection, and b) the minimum deflection. The numeric mean of these two cases is termed as the average deflection.

In addition to these major uncertainty factors, there can be few additional sources of error such as warpage and manufacturing tolerances, that can also induce additional errors for the correlation.

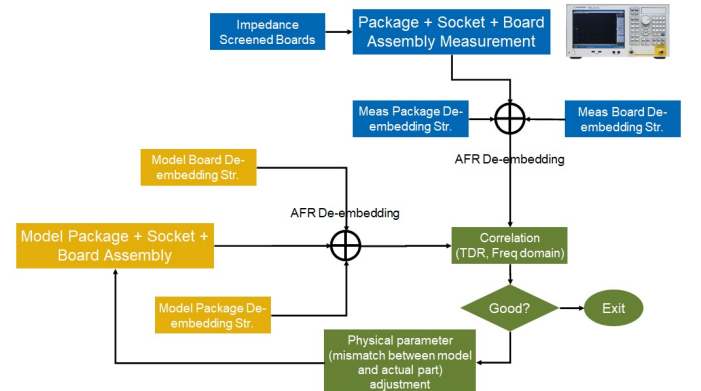
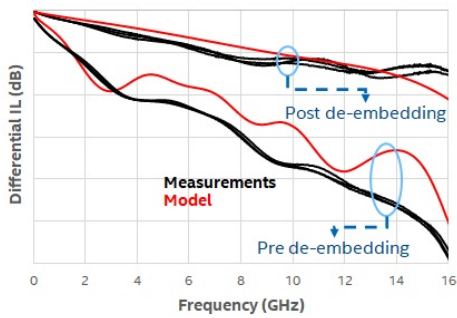


Fig. 3: Validation methodology adopted for SUT model to measurement correlation.

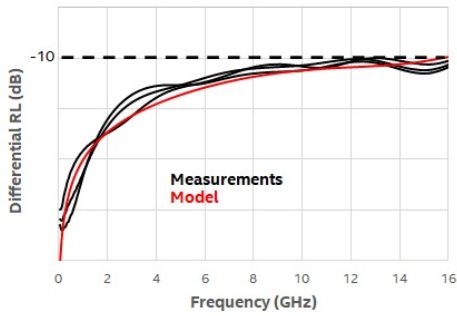
The flow chart for the resulting proposed correlation methodology is shown in Fig. 3. As a first step, the impedance screened boards are used to construct the TV assembly. AFR de-embedding was performed using the package and board de-embedding structures. On the modeling side, a similar approach was adopted with the model construction of complete TV, followed by AFR de-embedding. If the measurement to modeling correlation is not satisfactory, then further effort is spent to close the gap between model and actual TV physical geometries until good correlation quality is achieved.

IV. RESULTS

Three different test vehicles were constructed for data collection and all the measurements were performed at room



(a)



(b)

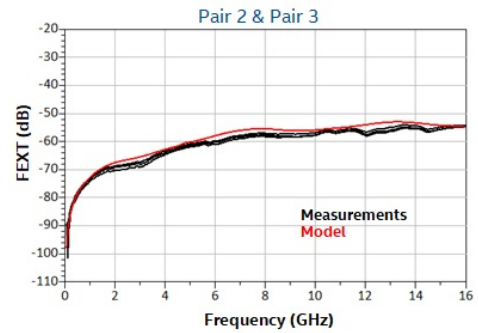
Fig. 4: Correlation at TV test site ‘Pair 1’ for: (a) Pre and post de-embedded DE IL, and (b) post de-embedded DE RL.

temperature condition. The DE IL, and RL plots at TV test site ‘Pair 1’ are shown in Fig. 4. The de-embedded IL correlation shows very good agreement between measurement and modeling up to 16 GHz. To show the correlation quality graphically, pre de-embedding and post de-embedding DE IL are plotted at the same scale and the relative values shown. As the pre-deembedded data includes effect of package and board, the process variation in those parts can impact the correlation quality. De-embedding process is specifically important in this aspect as it removes the dependency from peripheral test structures. Similarly, the post de-embedding DE RL shows a good correlation quality.

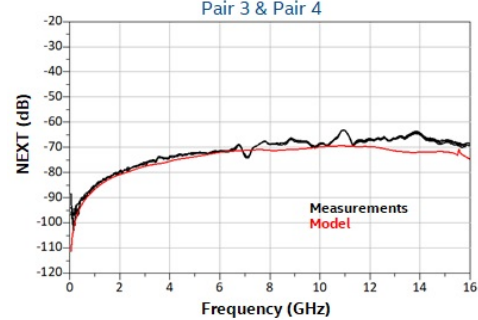
The far-end crosstalk (FEXT) and near-end crosstalk (NEXT) correlation plots among the test sites ‘Pair 2’, ‘Pair 3’, ‘Pair 4’ under average deflection condition are shown in Fig. 5. FEXT and NEXT correlation plots are without de-embedding, as the de-embedding process becomes challenging with higher order port numbers. However, the crosstalk contribution from the test package and board were kept at least 10 dB lower compared to the socket crosstalk contribution, so that the crosstalk from socket becomes dominant in pad-to-pad FEXT and NEXT measurements. In general, very good measurement-to-modeling correlation was observed for the signal patterns included in this TV, even at -70 dB crosstalk levels.

V. CONCLUSION

In this work, an improved methodology for high frequency socket validation was described to achieve good measurement



(a)



(b)

Fig. 5: Correlation for: (a) FEXT between ‘Pair 2’ and ‘Pair 3’, (b) NEXT between ‘Pair 3’ and ‘Pair 4’.

to modeling correlation for LGA sockets used in microprocessor systems. Various sources of uncertainty factors that can impact the correlation quality were investigated along with techniques to address them. Key uncertainty factors were found to be PCB impedance variations, accuracy of solder ball modeling, and the loading condition of the socket assembly. With the proposed methodology, good correlation was achieved for key signal integrity metrics such as IL, RL, and crosstalk for test LGA sockets. Use of de-embedding was another key challenge, yet an important feature of the proposed methodology to characterize and correlate the socket-only performance metrics. Future work will address the remaining uncertainty factors that have not been addressed here, to further improve the correlation quality and at higher frequencies.

REFERENCES

- [1] D.-H. Han, V. Prokofiev, W. Leigh, L. Polka, and T. Ruttan, “High frequency modeling and characterization of pin and land grid array sockets,” in *53rd Electronic Components and Technology Conference, 2003. Proceedings.* IEEE, 2003, pp. 1264–1269.
- [2] P. R. Paladhi, Y. Zhang, J. Tang, D. Rodriguez, J. Hejase, S. Chun, W. Becker, B. Beaman, and D. Dreps, “SI model to hardware correlation on a 44 Gb/s HLGA socket connector,” in *2020 IEEE 29th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*. IEEE, 2020, pp. 1–3.
- [3] S. A. Smith, Z. Zhang, and K. Aygün, “Assessment of 2x thru de-embedding accuracy for package transmission line DUTs,” in *2020 IEEE 29th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*. IEEE, 2020, pp. 1–3.
- [4] J. Sun, Z. Qian, C. S. Geyik, and K. Aygün, “Accurate BGA package solder joint modeling for high speed SerDes interfaces,” in *2020 IEEE 29th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*. IEEE, 2020, pp. 1–3.

On-Dual In-line Memory Module (DIMM) Low-Pass Filter (LPF) using Via Stubs for Enhancing Signal Integrity

Seungjin Lee*, Jonghoon J Kim, Jinseong Yun, Heejin Cho, Youngho Lee, Kyoungsun Kim, Sungjoo Park, Jeonghyeon Cho and Hoyoung Song
 DRAM Solution Team (Memory Business), Samsung Electronics, Hwaseong, Republic of South Korea
 *sjinjin.lee@samsung.com

Abstract— A novel on-dual in-line memory module (on-DIMM) low-pass filter (LPF) using via stubs is proposed to enhance signal integrity. Proposed on-DIMM LPF implemented with existing via stubs not only eliminates the via stub effects, but also enhances the signal quality due to its low-pass characteristics without additional manufacturing processes and costs. For verification, both time- and frequency-domain simulation results of proposed DQ signal line with on-DIMM LPF are shown and compared with conventional DQ signal line. Proposed on-DIMM LPF exhibits 0.67 dB improvements at 2.8 GHz (nyquist frequency of 5.6 Gbps) shows relatively flat and improved insertion loss. Simulated eye height is improved for 29 mV and 20 mV and eye width is improved for 4.2 ps (0.02 UI) and 3.6 ps (0.02 UI) respectively at 4.8 Gbps and 5.6 Gbps. Finally, measurements verification with test board of on-DIMM LPF is also conducted and shows improvements in insertion loss and eye diagram

Keywords—Dual In-line Memory Module (DIMM), Low Pass Filter (LPF), Signal Integrity, Via stub

I. INTRODUCTION

As recent server systems demand high capacity memories with higher speed, next generation double data rate 5 (DDR5) - dynamic random access memory (DRAM) is expected to achieve data rates as high as 6.4 Gbps. As the density of dual in-line memory modules (DIMMs) and data rates increase rapidly, designing DIMMs becomes more challenging than ever. To ensure signal integrity at such higher speeds, there are many non-ideal effects degrading signal integrity to be considered such as insertion loss, impedance mismatching, losses of copper and substrates, discontinuities, via stubs effect and cross-talk, etc [1]. In practical DIMMs, plated through-hole (PTH) vias are mostly used to interconnect the outer layer to inner layer because of their cheap cost and ease of manufacturing. However, PTH vias inevitably make useless portion of via such called ‘via stub’ because of their own manufacturing process. Via stub is main factor of deteriorating signal integrity because it makes signal detour [2]. To eliminate via stubs, buried or blind vias can be adopted in multilayer DIMMs. Also, back-drilling technology can remove via stubs, but these methods increase complexity of design procedure and manufacturing costs [3].

In RF systems, butter-worth filter usually used when its maximally flat insertion loss in pass-band and adequate skirt response is needed [4],[5]. Also, in high-speed digital systems, flat insertion loss without fluctuation is essential to ensure signal integrity. To obtain maximally flat insertion loss with eliminating via stub effects, we proposed a novel on-DIMM low pass filter (LPF) using via stubs in this paper. To enhance the insertion loss of conventional DQ signals, 3rd order butter-worth LPF implemented with existing via stubs and additional stub is proposed. The proposed on-DIMM LPF not only eliminates via stub effects, but also enhances the signal quality

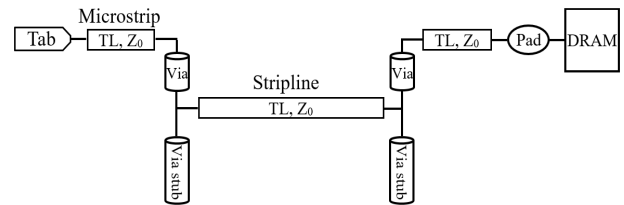


Fig. 1. Traditional DQ bus topology in DIMM.

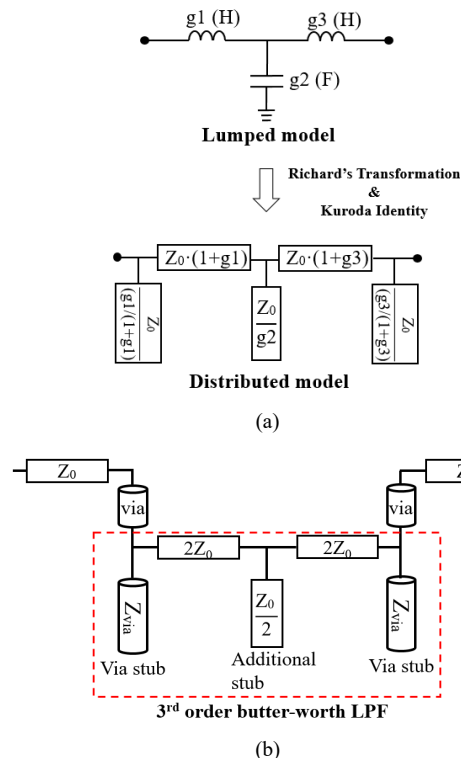


Fig. 2. Schematics of (a) ideal 3rd order LPF models and (b) proposed DQ signal line with 3rd order butter-worth LPF using via stubs in DIMMs.

due to its LPF characteristics having flat insertion loss in passband. Also, proposed on-DIMM LPF can be implemented easily without any additional manufacturing processes and costs. Finally, simulation results in time- and frequency-domain are shown for verification. Proposed On-DIMM LPF exhibits flat and improved insertion loss up to 3.91 GHz when compared with conventional DQ signal line. Simulated eye diagram of proposed DQ signal line with on-DIMM LPF also shows improvements in eye height and width.

II. ON-DIMM LOW-PASS FILTER USING VIA STUBS

A. Traditional DQ Bus Topology

Conventional DQ bus topology in DIMM is shown in Fig. 1. It shows the connection from tab (module gold finger) to DRAM including microstrip in outer layer, stripline in inner

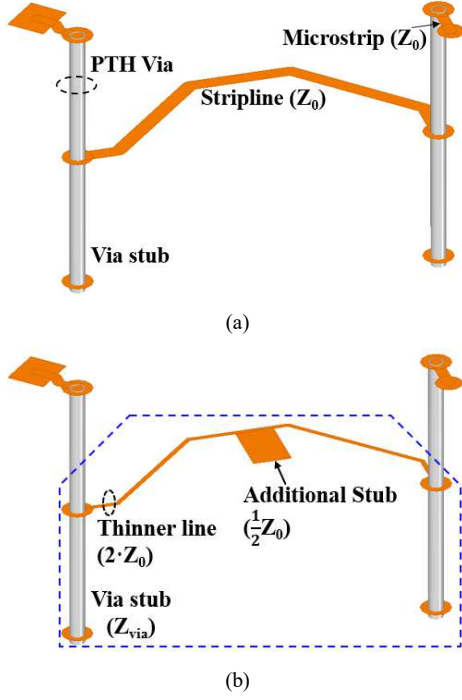


Fig. 3. Implementations in practical DIMMs. (a) Conventional DQ signal line and (b) proposed DQ signal line with 3rd order butter-worth LPF using via stubs.

layer with PTH via that connect the transmission lines between layers, back to top layer with another PTH via and pad. Two main reasons which deteriorate signal integrity are via stubs and impedance mismatches between components. Especially, via stubs are main factor which makes high-speed signaling challenging. We proposed a novel scheme to enhance signal integrity using via stubs as implementing 3rd order LPF in following chapters.

B. Low-Pass Filter Design

Microwave filters including low-pass filters can be realized with various types using open stubs, transmission lines, defected grounds structures and resonators [6]. To design low-pass filter, we can begin with LC ladder-type lumped model as shown in Fig.2 (a). Ideal 3rd order LPF can be implemented with two series inductors and one shunt capacitor for lumped model. Filters can be characterized by their own filter element values of lumped model such as g_1 , g_2 and g_3 . Because lumped element inductors and capacitors are difficult to implement in practical DIMMs at microwave frequencies, distributed models such as open or short-circuited transmission lines are used to implement filters. Distributed model in Fig. 2(a) can be transformed using Richard's transformation and Kuroda's identities from lumped model and implemented with open stub and series transmission line as shown in Fig. 2(a). In our case, we designed maximally flat (butter-worth) LPF to enhance insertion loss in fundamental frequencies. To implement 3rd order butter-worth LPF, filter element values should be specific values, $g_1, g_3 = 1$ and $g_2 = 2$. [6].

C. Proposed DQ Bus Topology with On-DIMM LPF using Via stubs

In Fig.2 (b), schematics of proposed DQ signal with 3rd order butter-worth LPF using via stubs is shown. Impedance of series and parallel transmission lines for 3rd order low-pass filter is as follows respectively:

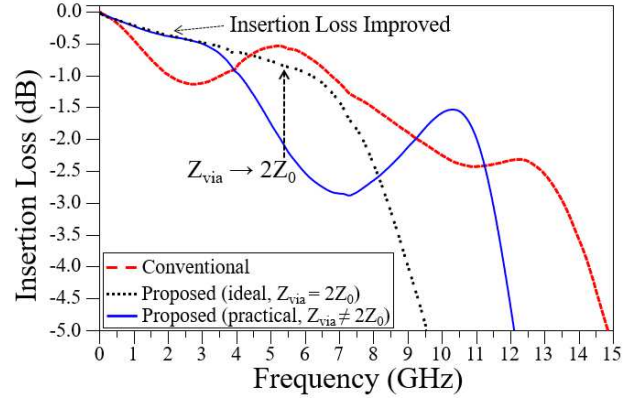


Fig. 4. Simulated insertion losses of conventional DQ signal line and proposed DQ signal lines with 3rd order butter-worth LPF using via stubs.

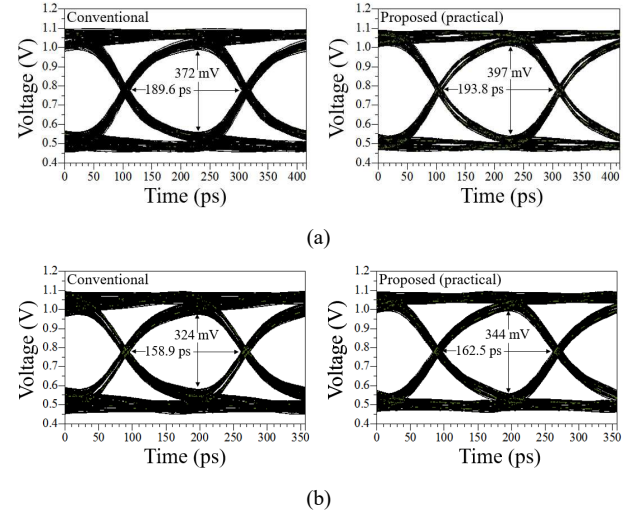


Fig. 5. Comparisons of simulated eye diagram between conventional and proposed practical DQ signal line at (a) 4.8 Gbps and (b) 5.6 Gbps.

$$Z_0 \cdot (1+g_1) = Z_0 \cdot (1+g_3) = 2 \cdot Z_0 \quad (1)$$

$$Z_0 / ((g_1)/(1+g_1)) = Z_0 / ((g_3)/(1+g_3)) = 2 \cdot Z_0 \quad (2)$$

$$Z_0 / (g_2) = Z_0 / 2 \quad (3)$$

As shown in Fig.2 (b), two existing via stubs are used for implementing 3rd order butter-worth LPF with series two transmission lines and additional open stub line. This filter eliminates open stub effect with exhibiting low-pass characteristics to enhance signal integrity.

Conventional DQ signal line in practical 1 rank registered DIMM is shown in Fig. 3 (a). Total thickness of DIMM is 1.27 mm with multiple layers. PTH vias interconnect microstrip of top layer to stripline of inner layer with two via stubs. Proposed DQ signal with on-DIMM LPF using via stubs in previous section is adopted in practical DIMMs as shown in Fig. 3 (b). Thinner inner line with $2Z_0$ and additional open stub with $0.5Z_0$ combined with two existing conventional via stubs can exhibit LPF characteristics in proposed DQ signal line, while conventional DQ signal line is composed with transmission lines with Z_0 . To implement 3rd order butter-worth filter ideally, characteristic impedance of via stubs should be $2Z_0$. But via-hole in multilayer PCBs usually exhibits relatively lower impedance due to parasitic capacitances. As Z_{via} increases approaching to $2Z_0$, proposed

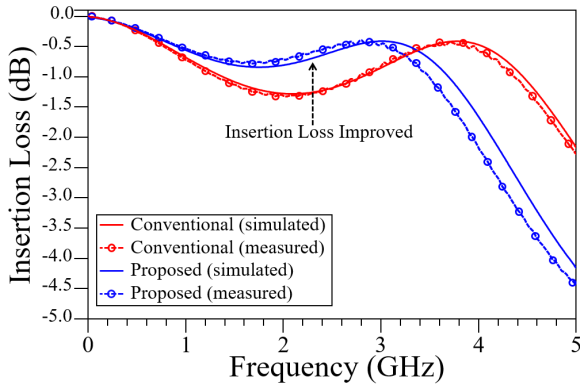


Fig. 6. Measured insertion losses of conventional and proposed test board lines with 3rd order butter-worth LPF using via stubs.

on-DIMM LPF exhibits ideal butter-worth LPF characteristics. There are some ways to increase impedance of via-hole in multilayer PCBs (e.g. increasing anti-pad size, decreasing via drill size) [7]. But in practical manufacturing process of DIMM, it is hard to tune via impedance finely. So, we proposed practical version of our proposed on-DIMM LPF without modifications on via-hole dimension to simplify design and manufacturing process. Even though implemented on-DIMM LPF is not ideal 3rd order butter-worth filter due to difference in via stubs impedance, it still can behave approximately 3rd order butter-worth filter, which exhibits flat insertion loss in its pass-bands.

III. VERIFICATIONS

For verification, both time- and frequency-domain simulations of proposed DQ signal lines with on-DIMM LPF are shown and compared with conventional DQ signal lines. In Fig. 4, insertion losses of two cases of proposed DQ signal line with LPF are shown and compared with conventional DQ signal line. The black dotted-line represents the insertion loss of ideal proposed DQ signal line with ideal 3rd order butter-worth LPF ($Z_{\text{via}}=2Z_0$). It shows flat insertion loss and relatively smaller cut-off frequency when compared with the insertion loss of conventional DQ signal line. The blue solid line exhibits proposed DQ signal line with practical 3rd order LPF ($Z_{\text{via}}\neq 2Z_0$). As Z_{via} increases approaching to $2Z_0$, insertion loss is improved and exhibits ideal LPF characteristics as shown in Fig. 4. Both practical and ideal proposed DQ signal line show similar flat insertion loss up to 3.07 GHz and show enhancements in insertion loss when compared with conventional DQ signal lines up to 3.91 GHz and 4.39 respectively. When considering data rate of 5.6 Gbps, the ideal and practical proposed DQ signals show -0.46 dB and at nyquist frequency (2.8 GHz), while conventional DQ signal line exhibits -1.13 dB. Proposed DQ signal line with practical LPF shows degradations around 4~8 GHz, but degradation in this frequency range does not affect signal integrity much up to 6 Gbps, because insertion loss of nyquist frequency is greatly enhanced. We also compared eye diagrams obtained through time-domain simulations with simplified DDR5 interface channel consist of host package, mother board, DIMM connector and DIMM. Fig. 5 shows estimated eye diagram of conventional and proposed DQ signal line with 4.8 Gbps and 5.6 Gbps data rates respectively. The eye height is improved for 29 mV and 20 mV and eye width is improved for 4.2 ps (0.02 UI) and 3.6 ps (0.02 UI) respectively at 4.8 Gbps and 5.6 Gbps.

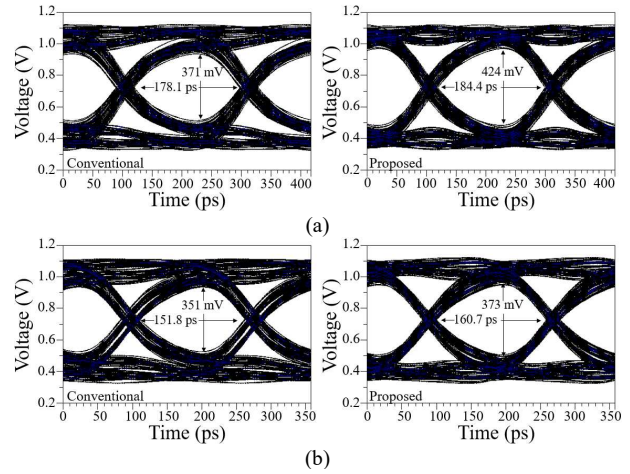


Fig. 7. Comparisons of measured eye diagrams at (a) 4.8 Gbps and (b) 5.6 Gbps.

Finally, we verified our proposed on-DIMM LPF using via stubs with measurements. We have made simplified test board of proposed on-DIMM LPF with changing design parameters including stack-up and pattern widths. As shown in Fig. 6, both simulated and measured results of test board show good agreements. Measured insertion loss of proposed on-DIMM LPF is improved for 0.689 dB and 0.55 dB at 2.4 GHz and 2.8 GHz respectively. Fig. 7 shows measured eye diagrams at 4.8 Gbps and 5.6 Gbps data rates. The eye height is improved for 53 mV and 22 mV and eye width is improved for 6.3 ps (0.03 UI) and 8.9 ps (0.05 UI) respectively at 4.8 Gbps and 5.6 Gbps.

IV. CONCLUSIONS

In this paper, a novel on-DIMM LPF using via stubs for enhancing signal integrity of DQ signals is proposed in practical DIMMs. 3rd order LPF is adopted without any additional costs and manufacturing processes. Both frequency and time-domain results are shown to verify improvements of proposed structure in signal integrity. Proposed DQ signal line with novel on-DIMM filters show great enhancements especially in eye height and can be one of solutions in enabling DDR5 DRAM to operate up to 6.4 Gbps.

REFERENCES

- [1] T.-L. Wu, F. Buesink, and F. Canavero, "Overview of signal integrity and EMC design technologies on PCB: Fundamentals and latest progress," *IEEE Trans. Electromagn. Compat.*, vol. 55, no. 4, pp. 624–638, Aug. 2013.
- [2] G.-H. Shiue, C.-L. Yeh, L.-S. Liu, H. Wei, and W.-C. Ku, "Influence and mitigation of longest differential via stubs on transmission waveform and eye diagram in a thick multilayered PCB," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 4, no. 10, pp. 1657–1670, Oct. 2014.
- [3] S. Camerlò, B. Ahmad, Y. Zou, L. Dang, M. Hu, and S. Priore, "Improving signal integrity of system packaging by back-drilling plated holes in board assembly," in *Proc. IEEE Electron. Compon. Technol. Conf.*, Jun. 2004, vol. 2, pp. 1220–1226.
- [4] R. Kaszynski and J. Piskorowski, "New concept of delay equalized lowpass Butterworth filters," in *IEEE International Symposium on Industrial Electronics*, vol. 1, Jul. 2006, pp. 171–175.
- [5] S. Nam, B. Lee, and J. Lee, "Theory for pseudo-Butterworth filter response and its application to bandwidth tuning," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 8, pp. 2847–2856, Aug. 2017.
- [6] D. M. Pozar, *Microwave Engineering*, 3rd ed. Wiley, 2005.
- [7] W.-D. Guo, W.-N. Chine, C.-L. Wang, G.-H. Shiue, and R.-B. Wu, "Design of wideband impedance matching for through-hole via transition using ellipse-shaped anti-pad," in *Proc. Electr. Perform. Electron. Packag.*, Scottsdale, AZ, Oct. 2006, pp. 245–24.

A Novel Differential Signal Routing Method for High-Speed and Large-Capacity DDR5 Dual-In-Line Memory Module

Yun-Ho Lee, Dongyeop Kim, SangKeun Kwak, Jeonghun Baek, Kyoungsun Kim, Sung Joo Park, Jeonghyeon Cho and Hoyoung Song
Samsung Electronics, Hwaseong, South Korea
 E-mail: yunhouno.lee@samsung.com

Abstract—In DDR5 dual-in-line-memory-modules (DIMMs), a registering clock driver (RCD) receives control, address, clock signals from CPU and re-drives them to dynamic random access memories (DRAMs) through transmission lines on printed circuit board (PCB). Especially, clock signals operate at double data rate (DDR) which makes it substantially difficult to guarantee signal integrity (SI), whereas control, address signals operate at single data rate (SDR). As of now, JEDEC standards prescribes that the effective characteristic impedance of clock signals be as low as 22.5 Ohm for DDR5 DIMMs. With the strict restriction on the physical dimensions of PCB set forth by JEDEC, along with the fact high permittivity materials have already reached their limit, there is a limit to achieve low characteristic impedance with current PCB fabrication processes including increasing signal width, use of high permittivity materials and close reference plane. Therefore, we propose a novel differential signal routing method so as to achieve low characteristic impedance, and consequently wider bandwidth for high-speed and large-capacity DDR5 DIMMs. With the proposed structure on DIMM, the system bandwidth can be extended by 12%, allowing 7.2Gbps operation of DDR5 Mono DRAM.

Keywords—DDR5 dual-in-line memory module (DIMM), Differential signal, Clock, Characteristic impedance, Routing method, Printed circuit board (PCB)

I. INTRODUCTION

The rise of artificial intelligent and fourth industrial revolution encourage data centers to demand high-speed and large-capacity memory modules. To meet the increasing market demand, DDR5 DIMMs are supposed to support up to 7.2Gbps and high-capacity memory modules implemented 3-dimensional die stack (3DS) chip using wire bond packaging technology. As operation speed is getting faster and the amount of loadings keeps increasing, it becomes much harder to guarantee the fastest operating clock signal's SI. DDR5 DIMMs adopted differential pair clock signal for better SI. In PCB interconnect, a clock signal pair is close to each other pair to make strong coupling between them so that fringe-out to another signal lines and crosstalk attenuate. Furthermore, it has advantages in terms of common-mode noise and non-ideal reference plane problems [1], [2].

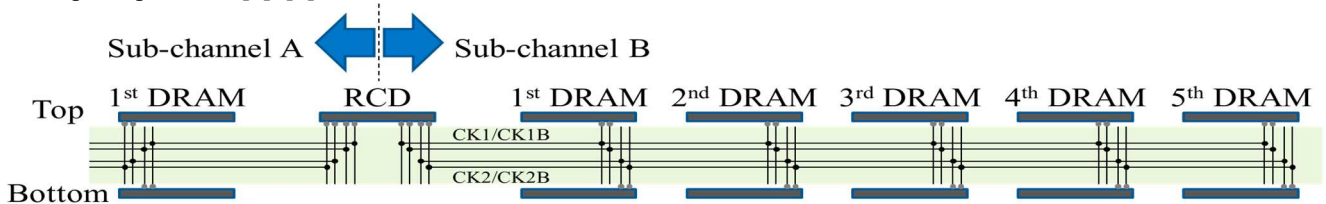


Fig 1. PCB interconnection schematic diagram of DDR5 DIMM's clock signals. Top-side DRAMs connected to CK1/CK1B and bottom-side DRAMs to CK2/CK2B

On account of DDR5 memory module's channel environment, the amount of DRAM chip loadings and operation speed, JEDEC has lowered clock signal's effective characteristic impedance than DDR4- 42.5 Ohm (differential impedance 85 Ohm) for DDR4, 22.5 Ohm (differential impedance 45 Ohm) for DDR5. Nevertheless, it is not as low as to secure enough peak-to-peak voltage for next generation of DDR5 and DDR6 DIMMs. The use of lower impedance has been studied for high-speed and large capacity DIMMs. However, PCB manufacturers struggle with conventional method to lower clock signal's effective characteristic impedance below 20 Ohm due to the limit of physical dimensions prescribed by JEDEC standard and high permittivity materials. In order to design clock signal's effective characteristic impedance lower than 20 Ohm, we propose a novel routing method for differential signals as in the following.

II. PROPOSED ROUTING METHOD

DDR5 DIMM's clock signal connection between RCD and DRAMs is shown in Fig. 1. A pair of CK1/CK1B is connected to top-side DRAMs and another pair of CK2/CK2B is connected to bottom-side DRAMs. In general, memory module company prefer to design each pair of clock signals enclosed by reference plane, edge-coupled strip-line, in different layer from one another as depicted in Fig. 2 so that no coupling occurs between pairs but within itself and with reference plane since clock switch at the highest speed and can be easily noise source to other signals. Conventionally, we design signal line wide or close to reference plane to make characteristic impedance low. However, there is no enough room for increasing signal width due to high routing density of DDR5 DIMM PCB and not easy to decrease dielectric material thickness due to manufacture difficulties and cost burden. For the aforementioned reasons, 20 Ohm is as low as characteristic impedance of clock signal line can have, given that space resource is strictly restricted by JEDEC standard and high permittivity materials have reached the limit.

Fig. 3 (a) and (b) show the proposed novel routing methods to design clock signal characteristic impedance lower than 20 Ohm. Proposed routing methods induce both

horizontal and vertical coupling by placing odd mode, 180° anti-phase, operating signals close so as to face side-by-side and up-and-down. In comparison with conventional method, proposed method has stronger odd mode coupling caused by the other clock pair so that effective total capacitance increase and effective total inductance decrease. As a result, effective characteristic impedance decrease depending on how close two clock pairs are. It is worthy to note that keep each clock signal routing length same so that differential signals maintain 180° anti-phase as far as possible. Otherwise, characteristic impedance gets higher than expected since differential pairs do not perfectly operate as odd mode due to phase delay caused by length differences. Both proposed methods have same characteristic impedance, but Fig. 3 (a) is better suited on account of RCD pin location. Therefore, we compare conventional method Fig. 2 with the proposed method Fig. 3 (a) to verify superiority of the proposed method.

In order to analyze electrical performance in terms of characteristic impedance, AC gain and peak-to-peak swing voltage, the 2Rx4 RDIMM registered on JEDEC is selected for reference design. The dielectric material used for reference design has relative permittivity $\epsilon_r=3.8$, loss $\tan \delta=0.02$ at 1 GHz and the conductivity of copper is 5.959×10^7 [S/m]. The stack-up information such as width and height is shown in Fig. 2 and Fig. 3. The trace lengths of clock signals for each section, RCD to DRAM and DRAM to DRAM, are as long as follows. Via connecting strip-lines RCD to 1st DRAM 28mm, 1st DRAM to 2nd DRAM 17mm, 2nd DRAM to 3rd DRAM 17mm, 3rd DRAM to 4th DRAM 17mm, 4th DRAM to 5th DRAM 17mm and every micro-strip lines connecting via to RCD and DRAMs 0.5mm [3].

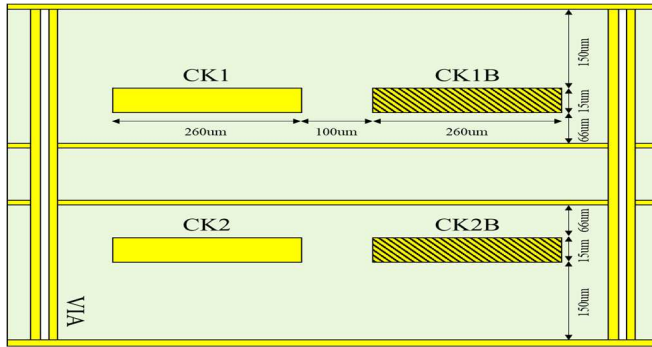


Fig. 2. Cross-section PCB view of conventional clock routing method

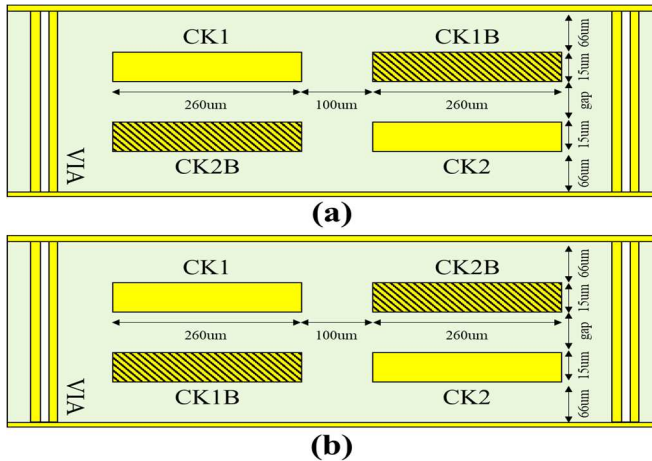


Fig. 3. Cross-section PCB view of proposed clock routing (a) method 1 (b) method 2

III. VERIFICATION OF THE PROPOSED METHOD

Table 1 shows self, mutual inductance and capacitance simulation results conducted with HSPICE field solver and effective characteristic impedance calculated by Eq. (1) and (2). In contrast to conventional method trying to lower L_{self} and raise C_{self} , the proposed method makes effective L_{mutual} and C_{mutual} higher by locating 180° anti-phase signal close - CK1 nearby CK1B and CK2B. This result shows effective characteristic impedance can be as low as 10 Ohm when the gap between CK1 and CK1B is 42 um, but 13 Ohm to keep JEDEC standard PCB stack-up since the gap is restrained by 62 um. To verify the performance of practical application, we set the gap as 62 um for proposed method in the following simulations. HSPICE is used to perform frequency and time domain simulation. Conditions and models used in simulation are depicted in Fig. 4.

TABLE I. LC AND EFFECTIVE CHARACTERISTIC IMPEDANCE (a) CONVENTIONAL METHOD (b) PROPOSED MEHOD ON GAP BETWEEN CK1-CK1B AND CK2-CK2B

Parameter	Conventional edge-coupled	Proposed		
		42um	62um	302um
Gap	-	42um	62um	302um
L_{CK1_self} [nH]	158	144	151	177
$L_{CK1-CK1B}$ [nH]	8	7	8	16
$L_{CK1-CK2B}$ [nH]	-	78	67	19
$L_{CK1-CK2}$ [nH]	-	6	7	8
C_{CK1_self} [pF]	269	417	348	244
C_{CK1_CK1B} [pF]	14	7	9	20
C_{CK1_CK2B} [pF]	-	226	153	25
C_{CK1_CK2} [pF]	-	4	5	6
Z_{CK1_eff} [Ohm]	23	10	13	23

$$Z_{Edge-coupled} = \sqrt{\left(\frac{L_{CK1_self} - L_{CK1-CK1B}}{C_{CK1_self} + C_{CK1-CK1B}} \right)} \quad (1)$$

$$Z_{Proposed} = \sqrt{\left(\frac{L_{CK1_self} - L_{CK1-CK1B} - L_{CK1-CK2B} + L_{CK1-CK2}}{C_{CK1_self} + C_{CK1-CK1B} + C_{CK1-CK2B} - C_{CK1-CK2}} \right)} \quad (2)$$

The simulated CK1 effective characteristic impedance results of the edge-coupled and proposed routing for both Mono and 3DS 2Rx4 RDIMM are shown in Fig. 5 time-domain-reflectometry(TDR) plot using 20ps rise time when CK1B and CK2B operate as differential mode and CK2 operates as common mode. Since routing dimension is symmetric, all signals have same impedance. From left to right of the plot, it shows impedance profile of the signal routing from RCD pad to 5th DRAM. As it can be seen, the impedances of routing section from RCD pad to 1st DRAM match up well with the calculated impedances shown in Table 1. The proposed routing has lower impedance so that it matches better than the edge-coupled routing counterpart due to impedance dips observed at via connecting DRAMs.

The AC characteristics of 5th DRAM, farthest from RCD, with Mono DRAM and 3DS DRAM are shown in Fig. 6 (a)

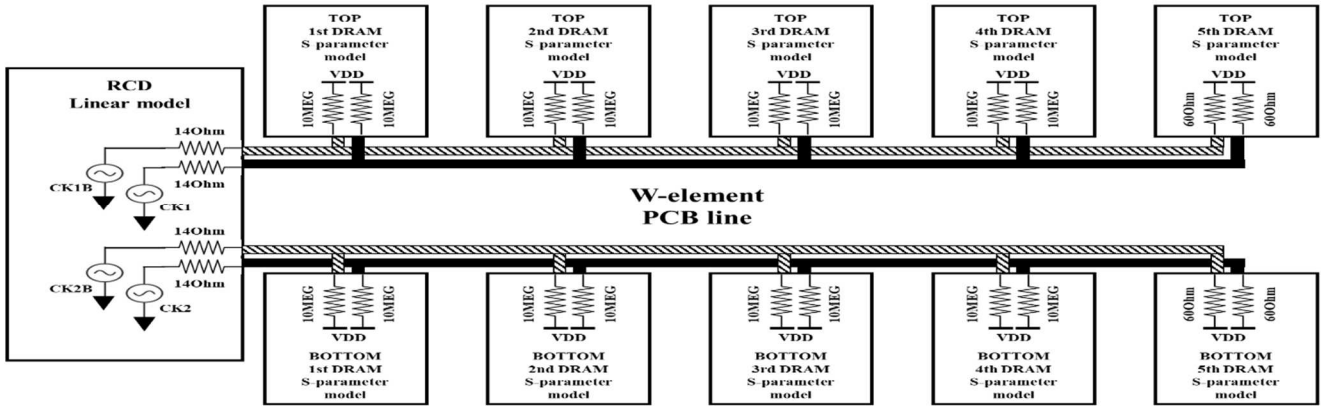


Fig. 4. Channel condition of top-side and bottom-side clock signal pairs with model descriptions used for both frequency and time domain simulations

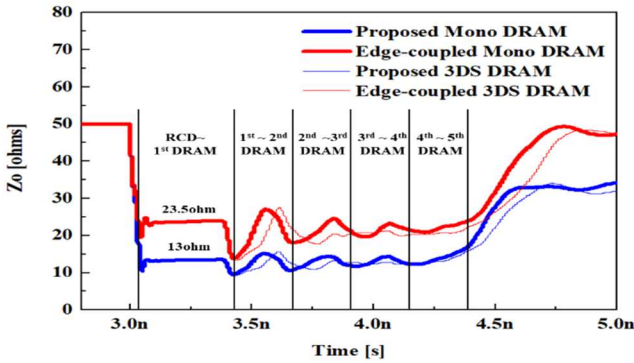


Fig. 5. CK1 signal's TDR simulation result with 50ohm source impedance

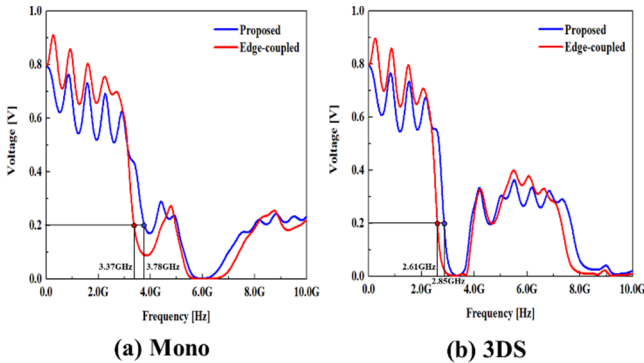


Fig. 6. AC simulation (a) Mono DRAM (b) 3DS DRAM

and (b). The farthest DRAM has the biggest loss and worst AC performance due to the longest PCB trace length and biggest loadings. As can be seen, when 3DS DRAMs are used worse AC performance than mono DRAMs are used due to bigger loadings. The proposed method exhibits substantially wider bandwidth, cut-off frequency of 200 mV, than conventional method improved from 3.37 GHz to 3.78 GHz (+ 12 %) with Mono DRAM, from 2.61 GHz to 2.85 GHz (+ 9.6 %) with 3DS DRAM. This bandwidth extension occurs as a result of proposed routing method's effective characteristic impedance which is 10ohm lower than conventional method.

For more intuitive analysis, transient simulations of 5th DRAM are shown in Fig. 7. It is notable that with the proposed method remains over 200 mV clock swing up to 7.2 Gbps for Mono DRAM and 5.2 Gbps for 3DS DRAM. This result shows the proposed routing scheme has better SI performance at high speed and with high loadings as has also been demonstrated in aforementioned AC characteristic simulation results.

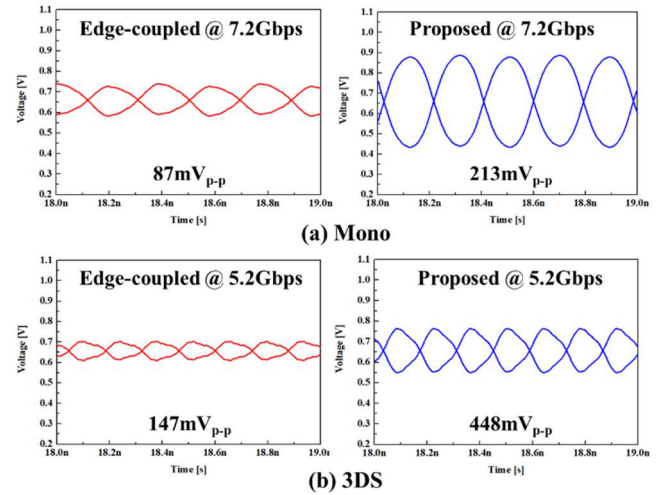


Fig. 7. Peak-to-peak voltage at 5th DRAM (a) Mono DRAM (b) 3DS DRAM

IV. CONCLUSION

In this paper, we proposed a novel PCB routing method which drives coupling between 2 pair of clock signals, consequently increases total mutual inductance and capacitance leading to decreasing effective characteristic impedance. This result is of significance to guarantee clock signals' SI for high-speed and high-capacity DDR5 memory module without any routing space overheads or even with better routing space efficiency. It is expected to adopt the proposed routing method for upcoming DDR5 memory modules supporting over 6.4 Gbps speed. We are in process of PCB artwork for the upcoming DDR5 memory modules and they will be verified in comparison with simulation results in the near future.

REFERENCES

- [1] Ivy Qin, Oranna Yauw, Gary Schulze, Aashish Shah, Bob Chylak and Nelson Wong, "Advances in wire bonding technology for 3D die stacking and fan out wafer level package," 2017 IEEE 67th Electronic Components and Technology Conference, May-Jun. 2017.
- [2] Ding-Bing Lin, Ching-Pin Huang, Hsin-Nan Ke and Wen-Sheng Liu, "Common-mode noise reduction on broadside-coupled delay line," 2015 IEEE Symposium on Electromagnetic Compatibility and Signal Integrity, Mar. 2015.
- [3] JEDEC, "DDR5 RDIMM Standard Annex A, JESD305-R8-RCA," 2022.

Improvement of Radiation Characteristics of a 300-GHz On-Chip Patch Antenna with Epoxy Mold Compound (EMC) Encapsulation

Harshpreet S. Bakshi, Rajen Murugan*, Sylvester Ankamah-Kusi
Texas Instruments Incorporated, Dallas, TX, USA
*r-murugan@ti.com

Abstract— Radiation characteristics of on-chip patch antennas are limited by the metallization and dielectric properties of the back-end of line (BEOL) silicon manufacturing processes. A 300-GHz on-chip patch antenna is designed using a radio frequency (RF) complementary metal-oxide-semiconductor (CMOS) process. The radiation efficiency, peak gain, and impedance bandwidth improve upon encapsulation of the antenna with IC packaging epoxy mold compounds (EMCs). In addition, high-frequency conduction and dielectric losses are analyzed, and their effects on antenna radiation efficiency are quantified in this paper. The overall radiation efficiency is shown to improve by 25%, peak gain by ~3 dB, and the -10-dB return loss bandwidth improves from 3 GHz to 18 GHz by encapsulating a 300-GHz on-chip patch antenna within commercially available EMCs.

Keywords—mmWave, on-chip antenna, semiconductor packaging

I. INTRODUCTION

The frequencies between microwave and infrared bands on the electromagnetic spectrum are referred to as the mmWave (30-300 GHz) and terahertz (0.3-10 THz) bands. These frequency bands are being used for numerous applications, including wireless communication, imaging, non-destructive testing, spectroscopy, and others [1-3]. Systems operating at mmWave and THz frequencies demonstrate wide fractional bandwidths, enabling high data rate applications [4]. This is made possible by implementing these systems on a single integrated circuit (IC) or using multiple chips within the same semiconductor package (system-in-package). With the increase in operational frequency, the corresponding wavelength reduces, consequently reducing the size of frequency-dependent microstrip transmission-line-based passive electronic components such as filters, baluns, couplers, and others. The size of mmWave and THz antennas also reduces when compared to similar antennas designed to operate in the microwave bands. Miniaturization of these antennas enables them to be implemented on-chip or within semiconductor packages. Moreover, losses due to interconnects are minimized by having antennas on-chip or on the package instead of routing lossy interconnects to external antennas. Implementation of antennas on-chip or within semiconductor packages also eliminates the need for mmWave routing on PCBs, thereby reducing the manufacturing cost of mmWave and THz systems

since standard materials such as FR4 can still be used for the manufacturing of PCBs housing high-frequency systems.

Prior work published in [5] and [6] has shown improvement in radiation characteristics of an on-chip patch antenna, designed in a 65-nm complementary metal-oxide-semiconductor (CMOS) process upon encapsulation using mold compounds based on silica microparticles dispersed in an epoxy matrix. The same technique is implemented to demonstrate a 25% improvement in radiation efficiency, ~3 dB improvement in the peak gain, and 6X improvement in the -10-dB impedance bandwidth (from 3 GHz to 18 GHz) of a 300-GHz on-chip patch antenna designed and simulated incorporating the design rules of a nine-metal layer RF CMOS process. Effects of placing five epoxy mold compound (EMC) encapsulation materials are analyzed, and the factors contributing to the reduction in radiation efficiency of encapsulated on-chip patch antennas are analyzed and presented in this paper.

II. ISSUES ASSOCIATED WITH ON-CHIP PATCH ANTENNAS

Patch antennas are planar structures that are comparatively easy and straightforward to implement on-chip, using the back-end of line (BEOL) metallization and dielectric stack. However, these antennas operating in the mmWave and THz frequency bands suffer from poor radiation efficiency due to ohmic (I^2R) losses from the metal layers of the BEOL and high dielectric loss tangents of the EMCs at these frequencies. Moreover, there is loss of radiated power to generated higher-order surface wave modes, which further limits the antenna radiation efficiency. The antenna aperture determines the gain, and the -10-dB return loss bandwidth of patch antennas depends on the separation of the patch from its ground plane and the properties of the dielectric material within, which is also limited in a typical silicon process. Furthermore, antennas designed on-chip, and within IC packages must conform to the process fabrication constraints. These antennas also suffer from process variation and fabrication tolerances due to their small form factor. Furthermore, proximity to off-chip components cause reflection and absorption of electromagnetic energy, further degrading the directionality and overall radiation performance of these antennas.

In references [5] and [6], it was demonstrated that typical packaging materials used in low-cost quad-flat no-lead (QFN) packages could be utilized to improve on-chip patch antenna

performance. In addition, it was shown that as the permittivity of the mold compound encapsulating the antenna increases, the antenna becomes a more efficient radiator because of the increase in the energy of fringing fields responsible for radiation. The same technique is used in this work to compare and analyze the performance of five EMC materials whose dielectric properties are listed in Table I.

TABLE I. DIELECTRIC PROPERTIES AND CHARACTERIZATION FREQUENCY OF ENCAPSULATION MATERIALS.

	Relative permittivity (ϵ_r)	Loss Tangent ($\tan \delta$)	Frequency (GHz)
Material 1 [5]	2.36	0.013	500
Material 2	3.4	0.0025	77
Material 3	3.5	0.013	70
Material 4 [5]	3.55	0.0095	220
Material 5 [5]	3.86	0.011	220

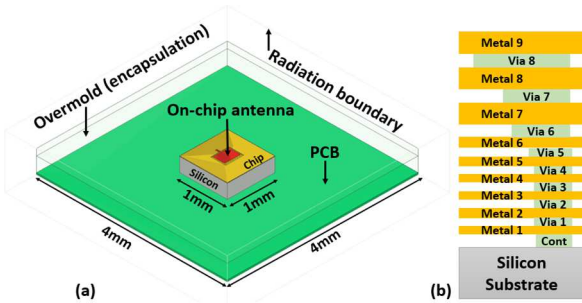


Fig. 1. (a) Simulation model of a 300 GHz on-chip patch antenna designed in an RF CMOS process. (b) Metal-via stack of the RF CMOS process used (not drawn to scale).

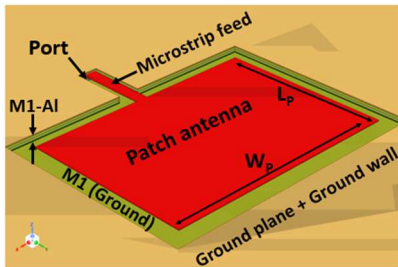


Fig. 2. Zoomed-in figure of the rectangular patch antenna.

III. SIMULATION SETUP

A full-wave 3D electromagnetic solver is used to model the antenna and conduct simulation studies. The rectangular patch antenna is designed in a nine-metal layer RF CMOS process. Metal 9 is the thick bond-pad layer of thickness $\sim 1\mu\text{m}$. Metal 1 is used as the ground plane and is $\sim 8\mu\text{m}$ below the patch. This antenna is modeled on a $1 \times 1 \text{ mm}^2$ chip which is placed within an EMC encapsulation on a printed circuit board (PCB) having a $4 \times 4 \text{ mm}^2$ footprint, as shown in Fig. 1. The patch has a length (L_p) of $240\mu\text{m}$, corresponding to $\sim \lambda/2$ at 300 GHz. The patch width (W_p) is $329.5\mu\text{m}$. Similar to that implemented in [5], metals and vias between metal 9 and metal 1 are shunted together to form a ground plane and ground walls, $\sim 25\mu\text{m}$

around the patch, as shown in Fig. 2. This is done to suppress surface wave generation and improve antenna directivity. A 50Ω microstrip feed line having a width of $17\mu\text{m}$ is designed and excited using a generic port in the 3D solver. Dielectric properties and metal conductivity values are incorporated, corresponding to the RF CMOS process (not disclosed in this paper). Properties of materials 1, 4, and 5 are obtained from [5] and [6] and properties of materials 2 and 3 are obtained from commercially available EMCs. Accuracy of the simulation studies performed in [5] and [6] were validated with experimental corroboration. Furthermore, measurement of on-chip patch antenna characteristics is challenging due to a lack of reliable technique to de-embed the effects of the high-frequency probe from the measurements. Due to these reasons, simulated results are utilized in this work.

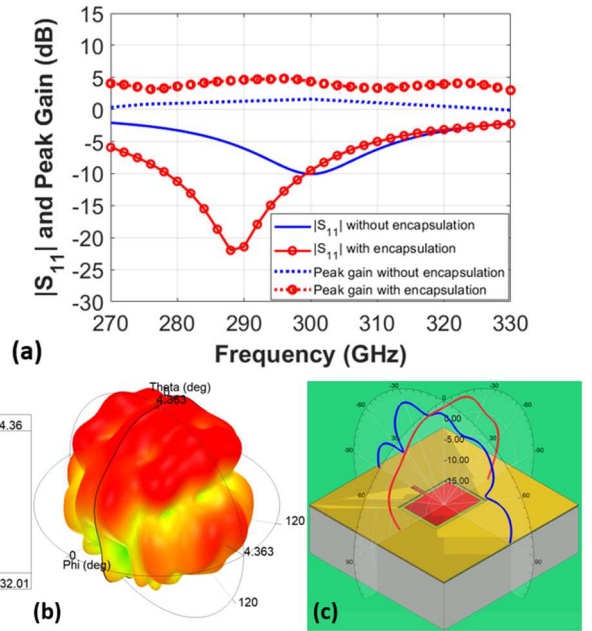


Fig. 3. Simulated (a) $|S_{11}|$ and peak gain versus frequency without and with (Material 3) $140\mu\text{m}$ thick encapsulation. ϕ is not fixed at 0° . (b) 3-D polar plot of gain at 300 GHz showing peak of 4.36 dB. (c) Radiation (gain) patterns in planes $\phi = 0^\circ$ (red) and 90° (blue) overlaid on the antenna model.

IV. SIMULATION RESULTS AND DISCUSSION

The frequency response of gain and return loss of the antenna, without and with a $140\mu\text{m}$ thick encapsulation made of material 3, is shown in Fig. 3(a). Similar to the results in [5], an increase in peak gain and impedance bandwidth is observed in the 300-GHz on-chip antenna designed using the RF CMOS process in this work. A 6X impedance bandwidth ($|S_{11}| < -10 \text{ dB}$) improvement from $\sim 3 \text{ GHz}$ to 18 GHz , an increase in peak antenna gain from 1.6 dB to 4.36 dB at 300 GHz, and an $\sim 3\text{dB}$ improvement in peak gain is observed throughout the frequency band (see Fig. 3(a)). The encapsulation thickness of $140\mu\text{m}$ corresponds to the efficiency maxima for material 3 as seen in Fig. 6 [5].

A 20% improvement in radiation efficiency is also demonstrated when the thickness of encapsulation, above the patch is $\sim \lambda/4$ in the EMC material. Among two EMC materials having the same loss tangent values, the effect of permittivity on

the fringing fields is evident (see curves for materials 1 and 3 in Fig. 4). Encapsulating the antenna with an EMC made of material 3 which has a higher permittivity (of 3.5), improves the peak radiation efficiency by ~5% more than the improvement from material 1 which has a lower permittivity (of 2.36). Material 2 which has the least value of loss tangent (0.0025) shows a 25% improvement in antenna radiation efficiency when encapsulated. Materials 1, 4 and 5, show similar trends as previously reported for the on-chip antenna designed in a 65-nm CMOS process [5,6].

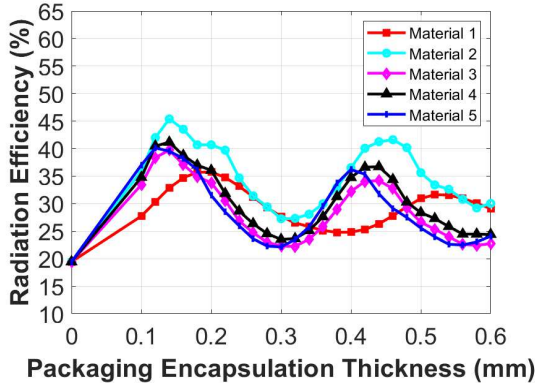
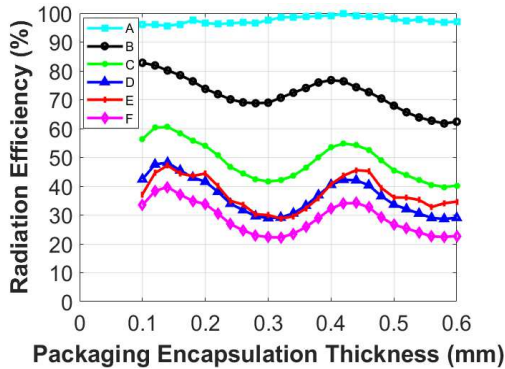


Fig. 4. Efficiency at 300 GHz versus package thickness (μm) for different materials. ($4 \times 4 \text{ mm}^2$ footprint).

The effects of conduction and dielectric losses on radiation efficiency were simulated, and results were analyzed as per [6]. The dielectric loss tangent of material 3 is set to zero, and the conduction (I^2R) losses are eliminated by making the metals perfect conductors. For the completely lossless case A (no



Case	Conduction losses		Dielectric losses of encapsulation
	Patch	Ground plane	
A	x	x	x
B	x	x	✓
C	✓	x	✓
D	x	✓	✓
E	✓	✓	x
F	x	x	x

Fig. 5. Efficiency at 300 GHz versus package thickness (μm) incorporating conduction and dielectric losses part-by-part, for an on-chip patch antenna encapsulated using material 3.

dielectric and no conduction losses), the radiation efficiency is ~100% (see Fig. 5). Case B demonstrates that ~20-30% reduction in radiation efficiency is due to the loss of the EMC material. On the other hand, case E shows that ~55-70% loss in radiation efficiency is due to the lossy metals. Case D shows a similar trend where the reduction in efficiency due to the lossy ground plane and the lossy dielectric is ~55-70%. Case C shows the contribution of a lossy patch and lossy dielectric, while the ground plane is lossless. Collectively, the patch and EMC contribute to ~40-60% reduction in radiation efficiency. Lastly, case F shows an overall decrease of ~60-80% in the antenna efficiency at 300 GHz, when the on-chip patch (designed using the RF CMOS process) is encapsulated within a package made of EMC having properties of material 3.

V. CONCLUSION

This paper extends the work published in [5] and [6] with five EMC materials that are used to show improvement in the radiation characteristics of a 300-GHz on-chip patch antenna designed in an RF CMOS process. Upon encapsulation of the antenna within an over-mold of material 2, a 25% improvement in radiation efficiency is observed. The -10-dB $|S_{11}|$ bandwidth of the 300-GHz patch antenna improves 6X from ~3 GHz to 18 GHz upon encapsulation within a 140 μm thick EMC made of material 3. The peak gain with this over-mold also increases by ~3 dB over the entire band of operation. Factors contributing to the reduction in radiation efficiency have also been discussed. It is determined that the EMC material causes ~30% reduction in antenna efficiency. In contrast, the metallic losses are the major contributors to up to 70% reduction in radiation efficiency of the antenna simulated in this work. The results presented in this paper, along with that shown in [5] and [6], demonstrate a technique to co-design an on-chip patch antenna incorporating packaging effects.

REFERENCES

- [1] P. H. Siegel, "THz technology," IEEE Trans. Microw. Theory Tech., vol. 50, no. 3, pp. 910–928, Mar. 2002.
- [2] T. W. Crow, W. L. Bishop, D. W. Porterfield, J. L. Hesler, and R.M. Weikle, "Opening the THz window with integrated diode circuits," IEEE J. Solid-State Circuits, vol. 40, no. 10, pp. 2104–2110, Oct. 2005.
- [3] D. L. Woolard, E. R. Brown, M. Pepper, and M. Kemp, "THz frequency sensing and imaging, a time of reckoning future applications?," IEEE Proc., vol. 93, no. 10, pp. 1722–1743, Oct. 2005.
- [4] C. E. Shannon, "A Mathematical Theory of Communication," 1948. <http://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf>.
- [5] H. S. Bakshi et al., "Low-Cost Packaging of 300 GHz Integrated Circuits With an On-Chip Patch Antenna," in IEEE Antennas and Wireless Propagation Letters, vol. 18, no. 11, pp. 2444–2448, Nov. 2019, doi: 10.1109/LAWP.2019.2943371.
- [6] H. S. Bakshi, "200-400 GHz Antennas in Integrated Circuits Incorporating Packaging Effects," Ph.D. dissertation, The University of Texas at Dallas, Richardson, TX, 2022.

Design and Optimization of High-Speed Digital Bus Over RF Channel

Nikhita Baladari and Robert Wenzel

NXP Semiconductors

Email: nikhita.baladari@nxp.com, robert.wenzel@nxp.com

Abstract— Integration of high-speed digital electronics and high-frequency radar channels in a package with limited layers can result in compromised return paths and degraded signal integrity. In this paper, we designed a digital block over a 77 GHz RF channel and used different package models to evaluate the digital coupling between the digital and RF blocks and the effect on their signal integrity.

Keywords— Signal integrity, Mixed-signal, RF, RADAR, 5G, 6G packaging, Waveguide, Launcher

I. INTRODUCTION

Mixed digital and RF packages for automotive RADAR, 5G and 6G applications require careful attention to digital and RF signal integrity as well as cost sensitivity. Therefore, it is desirable to use fewer substrate layers. A package such as 4-layer 1-2-1 instead of 8-layer 2-2-2 can offer significant cost savings in the long run. This generally means a main digital signal routing layer on M1 as microstrip where M2 is the ground plane. However, due to density requirements of digital and when a waveguide launcher is used, the M2 area above the launcher must remain RF ground, not the digital ground, even in large areas under the signal microstrips. This raises concern over the signal integrity of the digital bus as well as potential coupling of digital harmonics into the RF.

In this paper we analyze two variants of a waveguide-launcher-in-package design with M1 digital signal microstrip lines running partially over the 77GHz RF M2 ground plane. We examine the digital signal integrity and assess the noise coupled from the digital to RF Loop. The signal integrity is compared to a reference design (without RF) that uses standard microstrip with solid digital GND plane on M2.

In the first variant, the existing M1 coplanar digital ground fingers between the signals do not connect all the way through to the die cage. Whereas in proposed second variant, the fingers are squeezed-in and carried through all the way into the die cage. Results show that the complete connection of these fingers bridges the digital return current over the RF ground patch well-enough to repair the signal integrity to achieve quality similar to the reference microstrip design.

II. DESIGN AND TEST CASES

The overall module (Fig. 1) packages a fully-integrated 77GHz RFCMOS automotive RADAR transceiver chip and provides eight waveguide-based RF IO. It consists of a 1-2-1 4-layer flip-chip chip-scale-package substrate using solder ball grid array (BGA) for most signals but provides for mating of the eight waveguides in a downward-firing arrangement.

The digital block with high-speed data lines runs partially over the M2 RF ground waveguide backer on the top metal layer (M1) as illustrated in Fig. 2. The normally-conducted RF signal from the chip is wired to eight line-to-waveguide transducers that are located on the bottom two layers of the package. These must be backed with RF ground on M2.

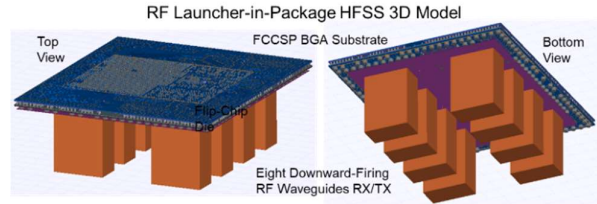


Fig. 1. 3D view of the overall module. A Flip-chip RF die is mounted to a FCCSP BGA having launchers with down-firing waveguides attached.

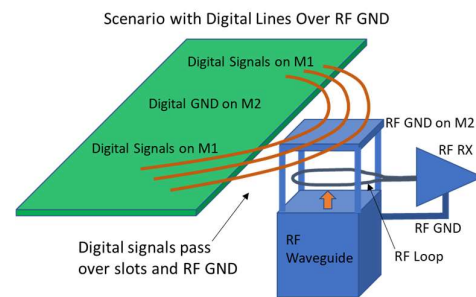


Fig. 2. Illustration of issue with digital lines travelling over RF ground.

The no-RF reference design (A) using standard microstrip, where the data lines are over their own ground, is shown in Fig. 3 (left), while the first variant of the mixed-signal module (B), having split grounds, is shown in Fig. 3 (right). The first variant has a coplanar digital ground on M1, but the ground fingers end in open-circuit a distance away from the die edge and do not make it to the die cage. This disrupts the return path and the ground current need to find a long-way around.

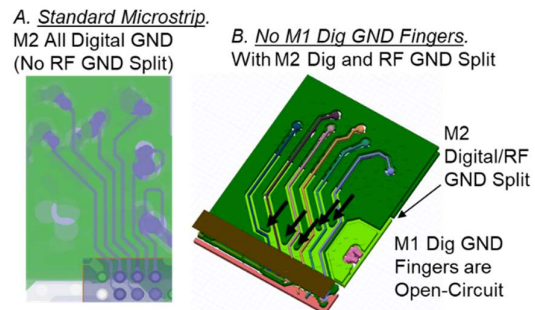


Fig. 3. Reference standard microstrip design (left) compared to the mixed-signal digital/RF module (right) having no M1 GND fingers

In the second variant of the mixed signal module (C) we used the M1 digital coplanar ground finger and bridge traces that are on M1 layer running alongside the data lines as shown in Fig. 4. The ground fingers are taken all the way into the die cage and connect to the digital ground die bumps. This approach maintains an adjacent ground return path within the M1 plane for these digital signals. Fig. 4 shows how the return path with ground finger traces has been designed on layer M1.

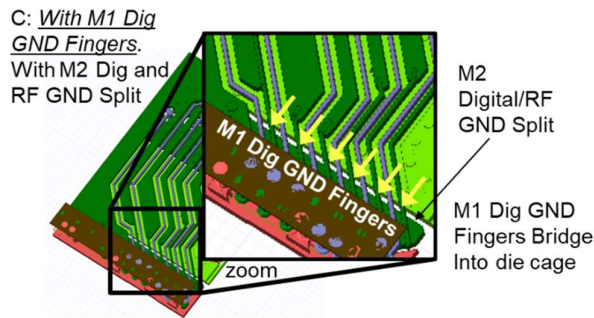


Fig.4. Mixed-signal digital/RF module having M1 GND fingers

The above three cases were modeled in HFSS (High-Frequency Structure Simulator, from Ansys) to generate wideband S-parameter models. These models were used in the ADS (Advanced Design System, from Keysight) circuit bench to evaluate the digital signal integrity and RF coupled noise.

III. MODELING

A. S-parameter models

The S-parameters for the three package test cases have been extracted using HFSS modeling. A frequency sweep from DC to 100GHz was used for the simulation. The adaptive meshing process at multiple frequencies is used to efficiently adapt the mesh for a more accurate and reliable solution across the entire frequency range. A suitably-high number of frequency points are used including DC point and low-frequency points as-needed for accurate PISI simulation and a fine interpolation convergence setting is applied.

B. ADS Testbench

ADS has been used to carry out the channel simulations to evaluate the signal integrity of the digital and RF blocks using the S-parameter models extracted in HFSS. Simplified driver, board models and pseudorandom data patterns are used for the simulations for simplicity. Settings are made in ADS that ensure all devices are characterized through 100GHz. The testbench designed in ADS for carrying out the simulations is shown in Fig. 5.

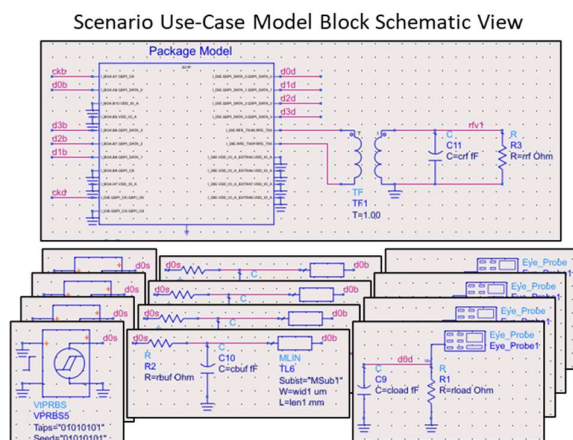


Fig.5. Simplified ADS testbench showing extracted package s-parameter model, data/clock sources, PCB transmission lines, and load terminations.

For the digital signal integrity portion of the simulations, eye diagrams are used to visualize the quality. The eyes are further characterized to pull out the maximum peak-peak

jitter, reduced eye width and reduced eye height for the different package models to quantify the differences in signal integrity. It was chosen to use a 4 Gbps DDR-clocked (clocked on both rising and falling edges) setup with four data lines and a clock line. Signals are driven into the package in RX mode. For SI simulation we used a transient simulation time of 1000ns worth of bits and automated timestep of up to 25ps to get reasonable run time and data size for the eye diagrams. For realism and slightly tempering of the results a small 1 to 5 ps worth of random jitter is applied. Each signal is uncorrelated to the others. A 50-ohm source termination strategy is used at the remote driving device. All signals use a 50-ohm board line of 4cm length (though the models allow for sweeping the values if desired). All lines are terminated inside the module with a small capacitive load at the die pad.

For the digital-to-RF coupling simulations, a shorter set of 25ns worth of bits is used to capture the mm-wave noise spectrum which is obtained through chirp-z time-to-frequency transform. A much finer maximum timestep of 1ps is used to get the needed bandwidth in the transform.

IV. RESULTS

A. Signal integrity of the digital block

We first analyzed the signal integrity of the digital signals based on package design choices in the three test cases. The simulations were carried out at different rise and fall times (Trf) to also understand how the signal integrity is being affected and the extent to which a better package design helps. Fig. 6 shows the comparison between the eye diagrams obtained for mixed signal variant package designs B (without fingers) and C (with fingers). While the signal integrity of both the packages have deteriorated with reduced rise and fall times, the package with ground fingers showed substantially better results more in-line with the reference design A.

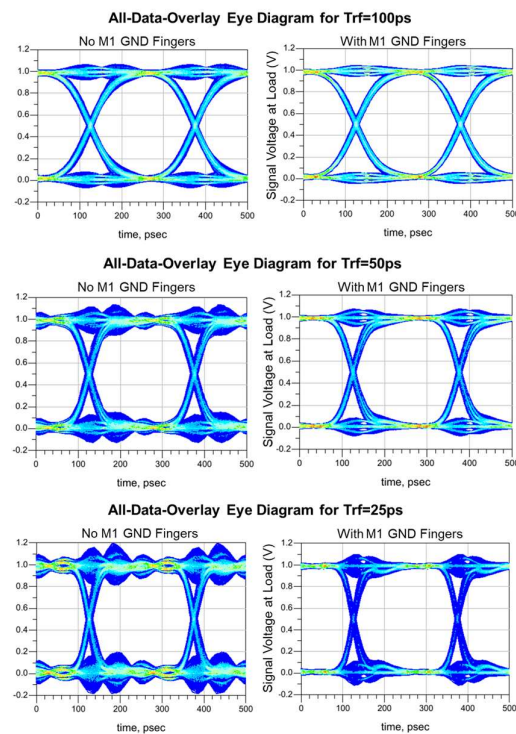


Fig.6. Eye diagrams for different Trf.

Fig. 7 shows a comparison between the eye diagram obtained for the reference design A (left), versus the fingered design B, with a rise time of 25ps.

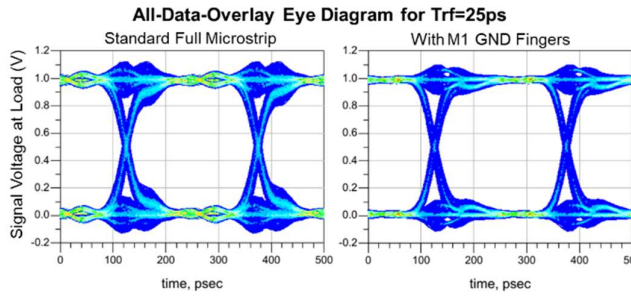


Fig.7. Eye diagram Std. vs fingered design for Trf=25ps.

Summary charts comparing results for the three packages are given in Fig. 8. The proximity between the reference model A and the with-ground-finger model C can be noted while the model B using no ground fingers shows higher deterioration.

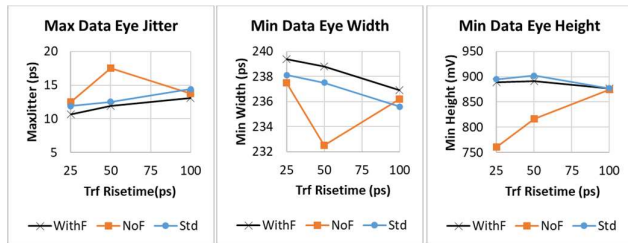


Fig.8. Eye diagram properties showing With finger is about like or a little better than Standard while No-finger is degraded

B. Digital coupling to RF

To analyze the effect of digital lines on the RF block we compared the RF loop noise of package B vs. C. As expected, the rise and fall times (Trf) played the key role in the amount of coupled RF loop noise at mm-wave frequencies. The with-fingers C design stood up better to reduced risetime than the non-finger variant. The time-domain noise plots in Fig. 9 for a 100ps risetime show noise to be cut in half using the fingers.

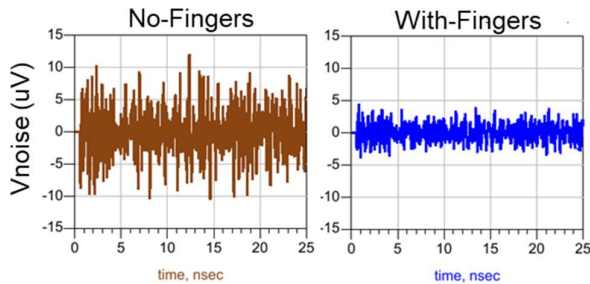


Fig.9. RF Loop Noise voltage for Trf=100ps.

The spectrum of the RF noise for Trf of 100ps, 50ps, and 25ps, are shown in Figs. 10, 11 and 12. The faster risetimes are seen to push the power out more into the RF range. Faster future bitrates may require the faster risetime. The results are summarized by the charts in Fig. 13, where the fingered design shows potential for 5 to 10 dB less coupling than no-fingers approach in both sub-10GHz band and mm-wave bands.

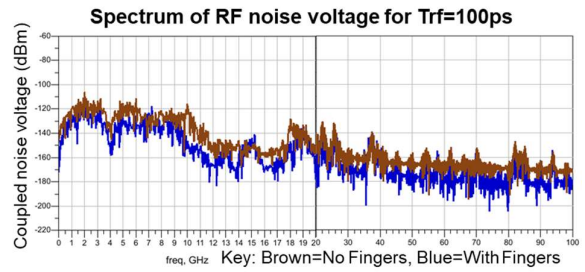


Fig.10. RF Loop Noise voltage spectrum for Trf=100ps.

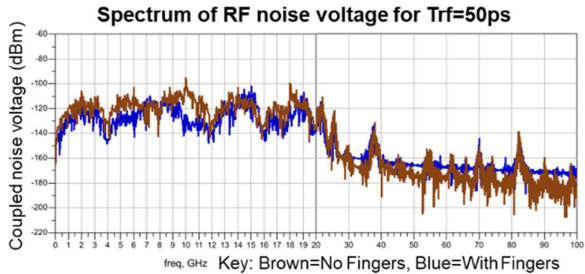


Fig.11. RF Loop Noise voltage spectrum for Trf=50ps.

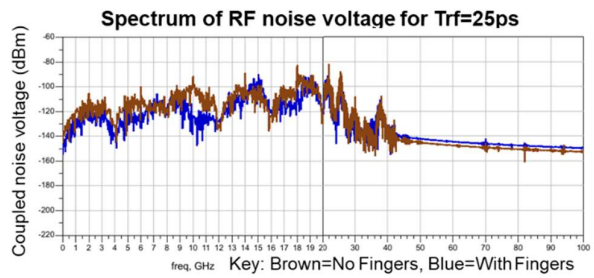


Fig.12. RF Loop Noise voltage spectrum for Trf=25ps.

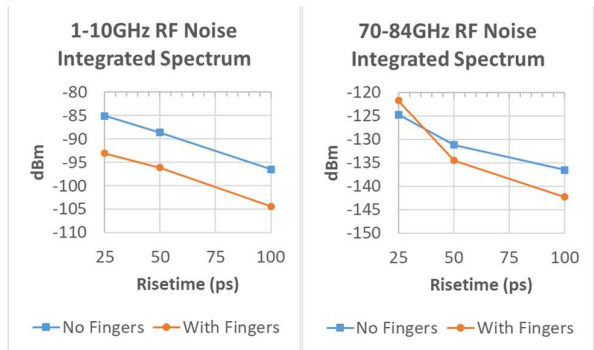


Fig.13. RF Loop Integrated Noise in 1-10GHz and 70-84GHz bands.

V. CONCLUSION

Mixed-signal RF and mm-wave packages with waveguide launchers require careful attention when density drives digital and RF domains into closer proximity. This paper describes the modeling used to conclude that digital bus could pass over an RF ground and maintain quality as long as coplanar M1 ground fingers are used to bridge sufficient return current.

REFERENCES

- [1] Stephen H. Hall; Howard L. Heck, "Advanced Signal Integrity for High-Speed Digital Designs, IEEE, 2009.
- [2] Seler, Ernst et al., "Chip-to-rectangular waveguide transition realized in embedded Wafer Level Ball Grid Array (eWLB) package," WAMICON 2014, pp. 1-4, 2014.

2D Spectral Transposed Convolutional Neural Network for S-Parameter Predictions

Yiliang Guo, Xingchen Li and Madhavan Swaminathan

3D Systems Packaging Research Center (PRC), School of Electrical and Computer Engineering

Georgia Institute of Technology, Atlanta, GA, USA

madhavan.swaminathan@ece.gatech.edu

Abstract—In packaging problems, S-parameter predictions are necessary. Machine learning methods lead to dimensionality related challenges which we address here through spectral transposed convolutional neural network using 2D kernels. Results show that Normalized Mean-squared Error (NMSE) dropped 0.002 by using 53.7% of the parameters.

Index Terms—microvia interconnection, convolutional neural network, up-sampling

I. INTRODUCTION

Circuit designs often require the use of scattering (s) parameters for signal integrity analysis. S-parameters represent frequency responses. Based on a combination of input geometries for a structure, groups of S-parameters can be computed. These responses can be learned through machine learning methods. Learning S-parameters as an input parameter for a neural network can create dimensionality problems since each frequency point represents a dimension. An alternate approach is moving the S-parameters to the output, which requires up-sampling strategies. Traditional unsampling methods include fully connected neural network (FCNN), unpooling and autoencoder [1]. Recently we demonstrated the use of Spectral Transposed Convolution Neural Network (S-TCNN) [2] for these problems, where the design parameters are learned using a fully connected feed forward neural network (FFNN), followed by the learning of the features using a 1D convolutional kernel. By creating a latent space, the neural network is used to upsample the frequency responses thereby constructing the S-parameters.

In this paper, we use 2D kernels to complete the up-sampling process [3] instead of 1D kernels as well as optimizing the structure of S-TCNN to enable the movement of 2D kernels. Based on 2D kernels, we are able to achieve better prediction accuracy with a significant reduction in the number of trainable parameters. In this way, the computational resources can be reduced using fewer parameters, with the reduced risk associated with over-fitting. In this paper, we apply STCNN with 2D kernels for learning and predicting S-parameters of interconnects in glass interposers, as shown in Fig.1.

II. TECHNICAL APPROACHES

A. Convolution and Transposed Convolution

In machine learning, we often construct multiple connected convolutional layers followed by activation function to map

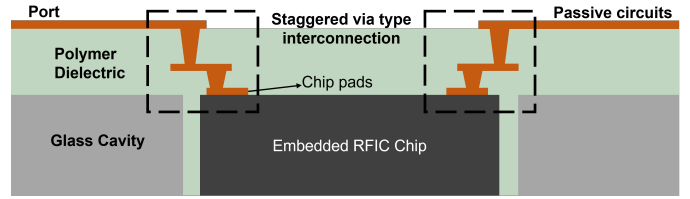


Fig. 1. A typical chip embedding model containing embedded chip, interconnection and passive components

data from higher dimensional input space into lower dimensional feature space. In Toeplitz form, the output y given input x and kernel function h can be written as:

$$y = f(h * x) = f(Hx) \quad (1)$$

$$\text{with } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{H} = \begin{bmatrix} w_1 & w_2 & \cdots & w_k & 0 & \cdots & 0 \\ 0 & w_1 & w_2 & \cdots & w_k & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & w_1 & w_2 & \cdots & w_k \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \quad (2)$$

where $*$ denotes the convolution operation, $f(\cdot)$ is the non-linear activation function, \mathbf{H} is the convolution matrix of size $n \times m$ and \mathbf{y} is the downsampled output of size $n = m - k + 1$.

Instead of extracting feature patterns from input data, we need to make predictions based on design parameters, which requires the information flow to be reversed. This can be achieved using transposed convolution operation. Similarly, for an n dimensional input vector \mathbf{x} , transposed convolution operation can be expressed as:

$$y = f(h *^T x) = f(H^T x) \quad (3)$$

where $*^T$ is the transposed convolution operation and \mathbf{y} is the upsampled output of size $m = n + k - 1$. Both convolution and transposed convolution operations share some common parameters, such as stride which controls the length of kernel movement in one step while padding is used to change the size of the output matrix. Different kernels aim to capture various local patterns, thus the choice of kernel becomes important.

B. 1D Kernel

The mechanism of 1D kernel is illustrated in Fig. 2. The width of kernel must be equal to the width of the input matrix, thus the kernel can only move in a single direction. 1D kernels are generally applied in the processing of sequential data.

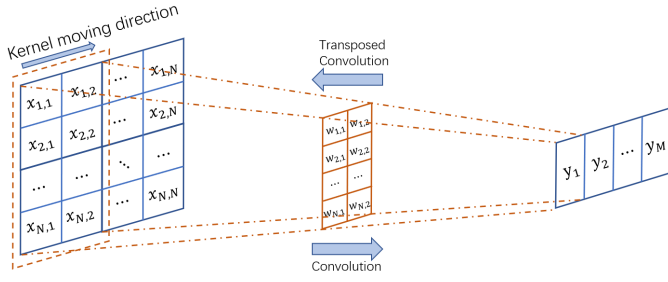


Fig. 2. 1D kernel

C. 2D Kernel

Mathematically, the 2D kernel is written as:

$$y[m, n] = x[m, n] * h[m, n] = \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} x[i, j] h[m - i, n - j] \quad (4)$$

where x and y are input and output functions and h is the kernel function. The mechanism of 2D kernel is illustrated in Fig. 3. The main difference compared to 1D kernel is that the size of 2D kernel can be defined by the user and it can move in both directions. Each filter in the convolutional layers can consist of multiple 2D kernels, resulting in a size of $[C_{in}, H, W]$ where C_{in} denotes the input channels and H, W represent the spatial width and height. Each filter yields a single channel output. Thus the 2D kernel shares the learnable weights for the entire input matrix and it can be used to capture the spatial information elegantly.

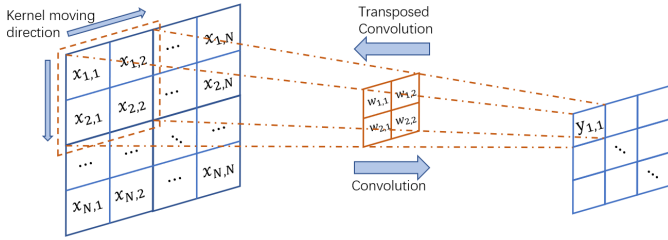


Fig. 3. 2D kernel

III. MODEL SETUP

A. Staggered Via Setup

The staggered via is modeled using Ansys HFSS as shown in Fig.4. The model incorporates an embedded co-planar waveguide (CPW) chip inside a glass cavity and two copper RDL layers (M1 and M2) plated on ABF, a polymer dielectric, laminated above the glass core. The model is set up with one side connecting to a microstrip line on the M2 layer and another side connecting to the embedded CPW chip and interconnected through the staggered via. Thicknesses of glass, chip, ABF layer 1, ABF layer 2 and plated copper are $150 \mu\text{m}$, $150 \mu\text{m}$, $30 \mu\text{m}$, $72.5 \mu\text{m}$ and $6 \mu\text{m}$, respectively. Via sizes are set to have an aspect ratio of 1:1, which means the diameter of via 2 is $72.5 \mu\text{m}$ and via 1 is $30 \mu\text{m}$. To tune the performance of the staggered via structure in terms of s-parameters, there are ten parameters that can be changed as shown in the second

subplot of Fig.4 (b). Ranges of these parameters are listed in TABLE I.

TABLE I
CHARACTERIZATION PARAMETERS (μm)

Parameters	Min	Max	Parameters	Min	Max
G_{stub}	20	60	l_{ost}	30	80
l_{ms}	500	700	W_{cpw}	45	65
W_{gnd}	300	700	G_{cpw}	20	80
W_{stub}	60	120	G_{fill}	30	60
l_{stub}	80	200	W_{ms}	70	200

B. Network Structure

Before all the data is fed into the transposed convolutional neural network, we need to eliminate the bias that may mislead the neurons, which is called normalization. At this time, 2D kernels can not be applied directly since the dimension of input is still relatively low. We construct 3 fully connected linear layers to up-sample the input matrix as the initial step. As a next step, 2 extra dimensions are added in order to reshape the up-sampled input into batches of the matrix: [batches, channels, rows, columns], where the channel determines how many 2D transposed convolution filters are created for each batch. In our example, we used six 2D convolutional layers to learn the frequency patterns each followed by a \tanh activation function. The data is then flattened and a $CoordConv$ layer [4] is utilized to maintain the frequency axis information. Further, the Bayesian dropout [5] technique is used to evaluate the model related uncertainty during the inference loop. The model is trained on PyTorch using CUDA to accelerate the training process. 600 model simulations based on the Latin hypercube sampling (LHS) method within the range specified by TABLE. I are derived from HFSS amongst which 500 groups of simulations are used for training while the rest of them are used for testing. The frequency axis in our application

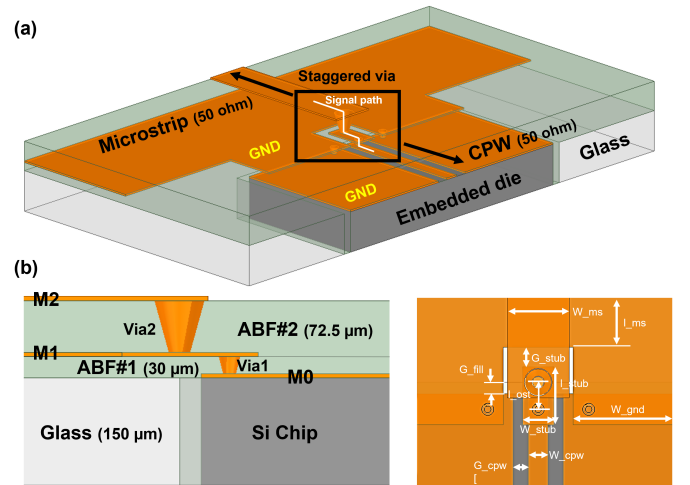


Fig. 4. Modeling and size of a staggered via.(a) Overview of the model. (b) Sizes of the structure.

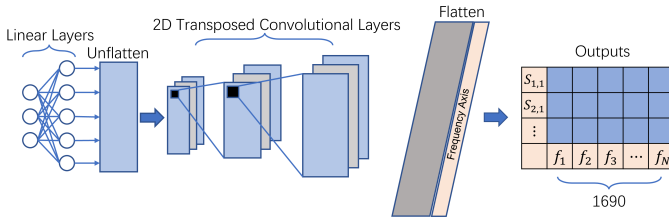


Fig. 5. NN structure

ranges from 1GHz to 170GHz in a linearly spaced step of 0.1GHz, which corresponds to 1690 points in total.

IV. RESULTS

The goal of our experiment is to utilize 2D kernels to achieve better prediction accuracy while reducing the required computational resources compared with 1D S-TCNN. The loss function used for back propagation is described in [2]:

$$L_{\text{freq}} = \frac{1}{N} \sum_{n=1}^N \sqrt{\frac{1}{K} \sum_{k=1}^K (y_{n,k} - \hat{y}_{n,k})^2} \quad (5)$$

After a certain number of iterations, we use normalized mean-squared error (NMSE), which can provide a normalized scale for different outputs, to evaluate the quality of predictions:

$$NMSE = \frac{1}{N} \sum_{n=1}^N \left(\frac{\sum_{k=1}^K (\hat{y}_{n,k} - y_{n,k})^2}{\sum_{k=1}^K (y_{n,k} - \frac{1}{K} \sum_{k=1}^K y_{n,k})^2} \right) \quad (6)$$

The comparisons between 1D and 2D S-TCNN are listed in Table II. The trainable parameters include all the weights and biases in the corresponding network. The 1D S-TCNN has 4 transposed convolutional layers consisting of more than 80K parameters while the 2D S-TCNN has 2 more layers but contains much fewer trainable parameters to be optimized. The required RAM for 2D S-TCNN is also smaller than 1D S-TCNN which means it saves more computational resources. Based on 2000 iterations, the training time for 2D S-TCNN is longer which is expected due to the bi-directional movement of the kernels. Both 1D S-TCNN and 2D S-TCNN can achieve promising accuracy in the case of frequency response predictions. 2D S-TCNN however has a lower final loss as well as NMSE with the number of parameters being only 53.7% of 1D S-TCNN. Examples of predictions on testing set are shown in Fig. 6. Both 1D and 2D S-TCNN can learn the S-parameter patterns and make predictions for new design parameters. 2D S-TCNN yields relatively smoother curves and better captures the frequency resonance due to the bi-directional movements.

V. CONCLUSION

We show that with 2D kernels the number of trainable parameters of S-TCNN is reduced significantly without the loss of prediction accuracy. Both the width and length of the 2D kernel can be defined thus it can fit the output dimension flexibly. In our model, we use 53.7% of the total number of parameters of 1D S-TCNN and the final loss is reduced from 0.535 to 0.311. The GPU RAM required to train the neural network is also reduced by 1.5 GB. Future work may

TABLE II
1D AND 2D KERNELS COMPARISON

	1D S-TCNN	2D S-TCNN
Trainable Parameters	85908	46142
Linear Layers	3	3
Convolutional Layers	4	6
Channels	20,50,50,50	10,30,30,20,15,12
Strides	1, 2, 2, 3, 2	(2, 2), (2, 2), (2, 2) (1, 2), (2, 4), (1, 2)
Activation Function	ReLU	tanh
Training Time	4.2 min	5.6 min
RAM	8.7 GB	7.2 GB
Final Loss	0.535	0.311
Validation NMSE	0.024	0.022

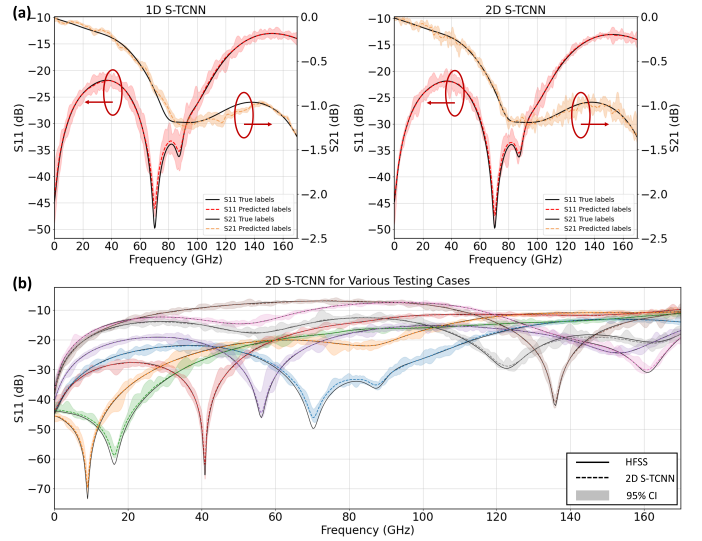


Fig. 6. (a) 1D and 2D S-TCNN prediction comparison (b) 2D S-TCNN predictions for complicated frequency patterns

include optimizing the structure of 2D convolutional layers to accelerate the training time.

ACKNOWLEDGMENT

This work was supported by DARPA under the Warden program (Project Number GR00013386).

REFERENCES

- [1] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proceedings of ICML workshop on unsupervised and transfer learning*, pp. 37–49, JMLR Workshop and Conference Proceedings, 2012.
- [2] H. M. Torun, H. Yu, N. Dasari, V. C. K. Chekuri, A. Singh, J. Kim, S. K. Lim, S. Mukhopadhyay, and M. Swaminathan, "A spectral convolutional net for co-optimization of integrated voltage regulators and embedded inductors," in *2019 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pp. 1–8, IEEE, 2019.
- [3] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," *arXiv preprint arXiv:1603.07285*, 2016.
- [4] R. Liu, J. Lehman, P. Molino, F. Petroski Such, E. Frank, A. Sergeev, and J. Yosinski, "An intriguing failing of convolutional neural networks and the coordconv solution," *Advances in neural information processing systems*, vol. 31, 2018.
- [5] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*, pp. 1050–1059, PMLR, 2016.

Methods to Characterize Radiation Patterns of WR5 Band Integrated Antennas in a Flip-Chip Enhanced QFN Package

Aditya N. Jogalekar^[1], Oscar F. Medina^[1], Andrew Blanchard^[1], Rashaunda Henderson^[1], Mahadevan K. Iyer^{[1][3]}, Hassan Ali^[2], Rajen Murugan^[2], Tony Tang^[4]

^[1] The University of Texas at Dallas, Richardson, TX, USA; ^[2] Texas Instruments Incorporated, Dallas, TX, USA; ^[3] Amkor Technologies Inc, AZ, USA; ^[4] Astera Labs, Santa Clara, USA.
anj170004@utdallas.edu

Abstract— Radiation pattern measurements of a millimeter wave (mmWave) antenna integrated in a package possesses several challenges due to its miniaturized size, available feeding methods, and inherent structural limitations of the package that calls for an innovative solution. This paper discusses two novel approaches to feed a wideband antenna fabricated inside a flip-chip enhanced QFN (FCeQFN) using a standard waveguide in the frequency range of 140GHz to 220GHz. The two approaches support the proposed antenna characterization methodology by achieving a 41.25% and 100% of -10dB bandwidth with a maximum attenuation of 3dB and 3.34dB, respectively. Further, we discuss the performance comparison of these transitions along with their implementation feasibility. A brief description of measurement structures and antenna radiation pattern analysis for a slot bow-tie antenna is reported.

Keywords— Antenna-in-package (AiP), flip-chip enhanced QFN (FC-eQFN), millimeter-wave (mmWave) planar antenna radiation pattern characterization methodology using waveguides.

I. INTRODUCTION

Antenna-in-package (AiP) is emerging as a standard practice for high-speed radio frequency (RF) transceivers and being recognized as universal solution. With the surging demand for compact and low-cost solutions in mobile communication, industrial and automotive sensors, AiP is becoming the popular option with an ability to realize a balance between the requirement of miniaturizing RF front-end modules (RF FEMs) and optimal performance. This is achieved by targeting millimeter wave (mmWave) frequencies, utilizing the benefits of higher bandwidth and smaller real estate [1]. By implementing this approach, the area consumed by RF FEMs which is dominated by antennas, has been reduced and in the range of millimeter squares, challenging the measurement and characterization process [2].

The most popular approach to measure these small elements is by a probe-based technique that requires a high frequency ground-signal-ground (GSG) probe to feed the antenna [3]. This method possesses several challenges including, (a) proximity effect of probes causing surface waves to travel towards the antenna infiltrating the radiation patterns [4], (b) effect of measurement setup that brings the components like the waveguide closer to the antenna in the direction of polarization [3], and (c) with AiP, problems associated with package substrates such as surface roughness affecting the proper contact of probe and the possibility of damaging them in the process. The wavelength shrinks and has minimum variation at these high frequencies. Thus, increasing the antenna feed length by $3 \sim 4\lambda$ does not help to reduce the effect of (a) and (b).

This research is funded by the Semiconductor Research Corporation (SRC) through the Texas Analog Center of Excellence (TxACE). SRC Task #2810.056

In this work, two new approaches to feed the planar antenna structure inside a flip-chip enhanced QFN (FCeQFN) are introduced for the first time that supports the improvised measurement methodology. The transitions for these approaches are designed using a three-layer (3L) FCeQFN substrate that enables the realization of different types of transmission lines and substrate integrated waveguides (SIW). A study on implementation feasibility of these transitions is provided along with the initial fabricated structures. Further, a detailed discussion on the impact of the transition on antenna radiation patterns is discussed using a comparison of an ideal and transition fed slot bow-tie (SBT) in FCeQFN package.

II. PACKAGE STRUCTURE

A. FCeQFN Package and Modified Version of FCeQFN

The base FCeQFN structure shown in [5] provides the ability to build multiple copper layers using the additive process of copper plating into mold compound to form a substrate as shown in Fig.1. The silicon die is flipped on this routable substrate forming the entire base FCeQFN package. This package provides the benefit of implementing solid wall structures that ensures end-to-end metal continuity, providing excellent isolation and no observed inter-layer and inter-structure radiation leakage.

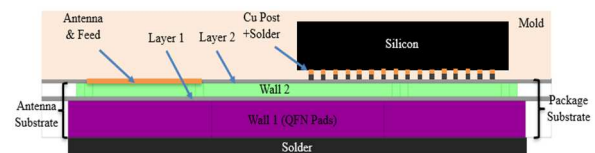


Fig. 1. Initial design of FCeQFN package with two-layer substrate.

The modified version of FCeQFN package is realized with a 3L substrate as shown in Fig. 2. This provides benefit to both the antenna and transitions due to increased substrate thickness from $175\mu\text{m}$ [5] to $200\mu\text{m}$, extending the modularity for implanting different structures inside the package that will be discussed in the following sections.



Fig. 2. Modified three-layer FCeQFN package substrate.

III. MEASUREMENT METHODOLOGY AND TRANSITION DESIGN

The methodology is based on feeding the device under test (DUT) i.e., a planar SBT antenna structure, through an electromagnetic radiation barrier, isolating the DUT from the feed as shown in Fig. 3. The package is soldered to a printed

circuit board (PCB) that has a ground plane and a plated through hole via, connecting a standard waveguide from underneath of PCB. With the benefits of minimal leakage that can be achieved through this package structure, a completely isolated transition can be realized limiting any radiation leakage on the PCB ground that can translate into surface waves and infiltrate the measurements.

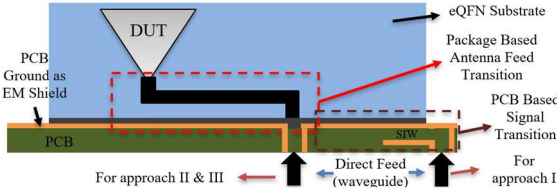


Fig. 3. Measurement methodology for mmWave antenna in eQFN package.

The external signal feed used for this methodology is a waveguide, similar to the one used in [6]. The signal from waveguide passes through the plated and then into the package with a transition that turns the signal 90 degrees in the direction of the DUT. The approaches taken for realizing these transitions are described in the following sub-sections.

A. Package Based Transition Design – Approach-I

Approach-I is designed using the multi-level transmission lines interconnected through vias that drive the signal from PCB to the antenna. The package is connected to a PCB through a via that is completely surrounded by wall 1 present in the bottom most layer of the package as shown in Fig. 4.

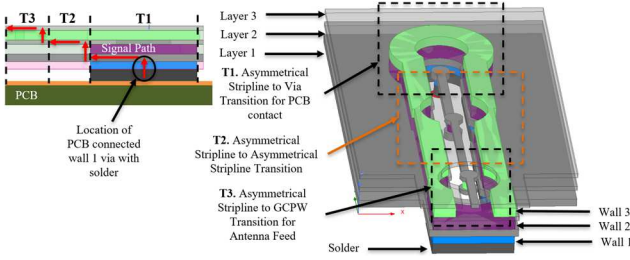


Fig. 4. Approach-I with multi-level transmission line based antenna feed.

Considering the measurement setup for approach-I shown in Fig. 3, wall 1 via is soldered to a buried via on PCB that connects to an embedded stripline followed by a horizontal SIW. The PCB integrated SIW is excited using a standard air-filled waveguide that is connected to a vector network analyzer (VNA) through WR5 band frequency extenders.

Inside the FCEQFN package, the signal travels through the wall 1 via that is connected to the asymmetrical stripline present in layer 1 directing it towards DUT. An asymmetrical-to-asymmetrical stripline transition is required to bring the signal from layer 1 to layer 2. The SBT antenna is part of layer 3 as shown in Fig.1, that is fed by a conductor backed coplanar waveguide (CBCPW) to avoid the balun implementation. An asymmetrical stripline to grounded CPW (GCPW) step is accomplished using wall 1 via followed by GCPW to CBCPW transition that delivers the signal to the antenna.

B. Package Based Transition Design – Approach-II

Approach-II works with a direct feed that does not require PCB-based embedded transitions. The package is soldered on a plated through hole via that connects an external waveguide to the vertical SIW inside the FCEQFN package as described in [5]. Fig. 5 shows the signal propagation path that starts with a vertical SIW which is derived using both the layer and wall

1 and 2, respectively. A vertical to horizontal SIW transition is realized by creating an opening in layer 2 that couples the signal without using any traditional methods such as, extending a similar dimension waveguide, using back-short and, using a coupling element such as a patch. The dimensions of the vertical SIW for this solution aligns to the standard WR5 band waveguide, i.e., $1295\mu\text{m} \times 648\mu\text{m}$ whereas, the horizontal SIW has dimensions of $\sim 600\mu\text{m} \times 45\mu\text{m}$. Further, a horizontal SIW to GCPW transition is required to get the signal transitioned in the required antenna feed format.

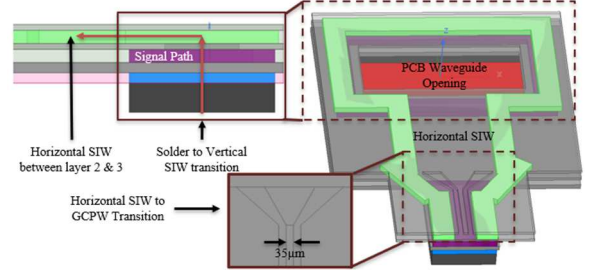


Fig. 5. SIW based antenna feed inside package for approach-II.

C. Package Based Transition Design – Approach-III

Approach-II is an extension of the method described in [5], calling it as approach-III, that utilizes a coupled patch mechanism to translate the signal from waveguide to an asymmetrical stripline.

IV. RESULTS AND DISCUSSION

The simulated S-parameters of approach-I are shared in Fig. 6 that shows a full band coverage (80GHz) with a solder thickness of $100\mu\text{m}$. The results are from Ansys HFSS with a preliminary surface roughness model providing a maximum attenuation of 3.34dB at 220GHz. The impedance looking from the GCPW side is 50Ω whereas, from the PCB connected via is $\sim 27\Omega$. Fig. 6 shows both $\sim 27\Omega$ and 50Ω normalized impedance at PCB connected via displaying the impedance mismatch tolerance of the transition.

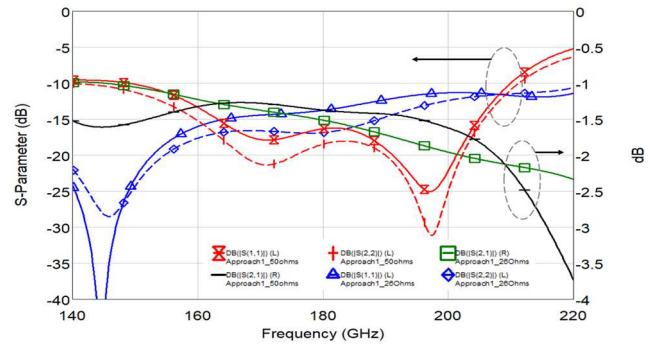


Fig. 6. S-parameters of approach-I based antenna feed transition.

The package to PCB interface via i.e., part of wall 1 has a $55\mu\text{m}$ diameter with a pitch of $160\mu\text{m}$. Though, it aligns with a few PCB manufacturing process, demands for strict process control. The variation of solder thickness impacts both via 1 and the asymmetrical stripline on layer 1 due to proximity to the closet ground i.e., present on PCB. The inductance of via 1 increases with the thickness shifting the resonant frequency and impacting the performance of the transition. Further, to realize this as a complete solution, a PCB with a stripline or a CPW to horizontal SIW transition is required adding more complexity. This makes the transition dependent on the assembly process that can be difficult to control.

Fig. 7. illustrates the S-parameters of approach-II showing a limited -10dB bandwidth of 33GHz i.e., covering 41.25% of WR5 band. Radiation leakage observed at the SIW to GCPW junction impacts the performance of approach-II, providing a maximum attenuation of ~3dB at 213GHz. A performance comparison of all three approaches is provided in table I.

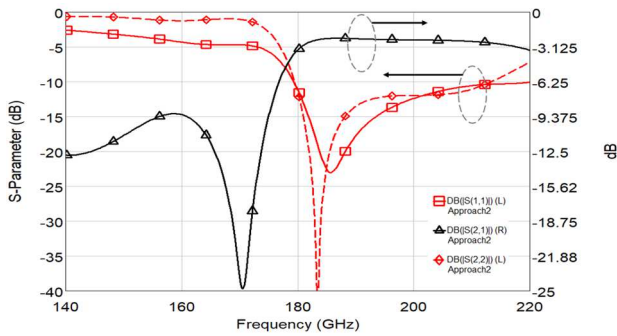


Fig. 7. S-parameters of approach-II based antenna feed transition.

TABLE I. TRANSITION PERFORMANCE COMPARISON

Specifications	Approach I	Approach II	Approach III*[5]
-10dB Bandwidth	80GHz	33GHz	17GHz
Min. Attenuation	1.33dB	2.42dB	1.7dB
Max. Attenuation	3.34dB	3.4dB	2.5dB

* Restricted two-layer FcEQFN package with 175 μ m substrate thickness

In approach-II, the solder thickness acts like a waveguide extension connecting the electroplated via on PCB with a standard waveguide and making it tolerant towards the thickness variation, exhibiting less dependency on assembly process. The manufacturability of PCB based via is confirmed through the initial fabricated PCB that matches the dimensions of a standard WR5 band waveguide as shown in Fig. 8 (a). In approach-III, the closest ground to the asymmetrical stripline is present inside the package i.e., Layer 1, confirming minimal impact on the performance due to solder thickness variation.

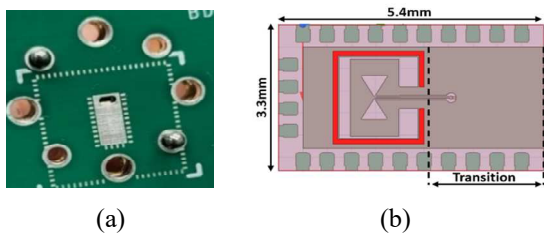


Fig. 8. (a) Preliminary fabricated PCB for validation of via dimensions, (b) transition fed antenna using approach I.

A transition integrated SBT antenna with approach-I is shown in Fig. 8 (b). The radiation patterns of both cuts, phi 0deg and 90deg, are depicted in Fig. 9 and compared with an ideal fed antenna. The comparison of radiation patterns in Fig. 9 confirms the negligible impact of the transition.

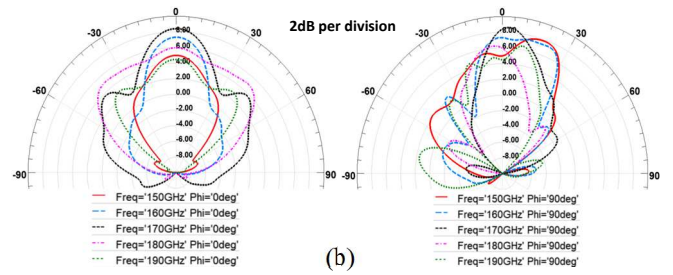
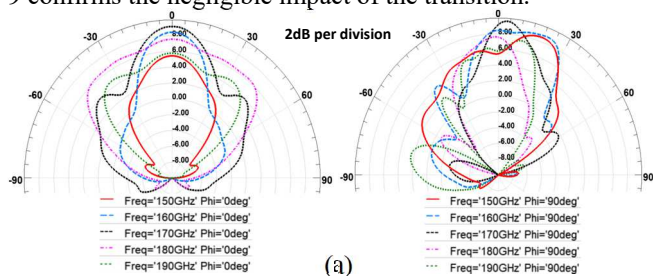


Fig. 9. Radiation pattern of SBT antenna – (a) ideal fed, (b) approach-I based transition fed.

The only observed effect is in peak realized gain due to a minimum signal attenuation of 1.33dB in the transition. Similar effect is observed in case of other approaches as well.

CONCLUSIONS

The paper introduces two new antenna feed approaches that support the proposed measurement methodology to characterize the radiation patterns of an antenna integrated in a FcEQFN package for WR5 frequency band applications. With a modified 3L FcEQFN configuration, approach-I offers a prospective solution providing a full band coverage with maximum attenuation of 2.23dB and better isolation. A detailed study on end-to-end implementation feasibility of both approaches highlights the strict fabrication and assembly process control requirements for realizing approach-I, bringing approach-II a viable option. A brief comparison of antenna radiation patterns is also provided showing minimal impact of the transition, providing the proof of concept.

ACKNOWLEDGMENT

This research is supported by the Semiconductor Research Corporation (SRC) and member companies in collaboration with The University of Texas at Dallas. The authors would like thank Dr. S. Sankaran, Ms. V. Khanolkar and Mr. P. Thompson, of Texas Instruments, and Dr. M. Lee, Dr. M. McGarry, Dr. H. S. Bakshi, Dr. N. Mahjabeen of UT Dallas for their support.

REFERENCES

- [1] Y. Zhang and J. Mao, "An Overview of the Development of Antenna-in-Package Technology for Highly Integrated Wireless Devices," in Proceedings of the IEEE, vol. 107, no. 11, pp. 2265-2280, Nov. 2019, doi: 10.1109/JPROC.2019.2933267.
- [2] H. Gulan et al., "Probe based antenna measurements up to 325 GHz for upcoming millimeter-wave applications," 2013 Int. Workshop on Ant. Tech. (IWAT), 2013, pp. 228-231, doi: 10.1109/IWAT.2013.6518338.
- [3] D. Titz, F. Ferrero and C. Luxey, "Development of a Millimeter-Wave Measurement Setup and Dedicated Techniques to Characterize the Matching and Radiation Performance of Probe-Fed Antennas [Measurements Corner]," in IEEE Ant. and Prop. Magazine, vol. 54, no. 4, pp. 188-203, Aug. 2012, doi: 10.1109/MAP.2012.6309179.
- [4] Z. Zheng and Y. P. Zhang, "A Study on the Radiation Characteristics of Microelectronic Probes," in IEEE Open Journal of Antennas and Propagation, vol. 3, pp. 4-11, 2022, doi: 10.1109/OJAP.2021.3131239.
- [5] A. Jogalekar et al., "Slot Bow-Tie Antenna Integration in Flip-Chip and Embedded Die Enhanced QFN Package for WR8 and WR5 Frequency Bands," 2022 IEEE 72nd Electronic Components and Technology Conference (ECTC), 2022, presented, June 1st, 2022, unpublished.
- [6] A. Jam and K. Sarabandi, "A Submillimeter-Wave Near-Field Measurement Setup for On-Wafer Pattern and Gain Characterization of Antennas and Arrays," in IEEE Tran. on Instru. and Meas., vol. 66, no. 4, pp. 802-811, April 2017, doi: 10.1109/TIM.2017.2654128

A Low EMI Board-to-board Connector Design for 5G mmWave and High-speed Signaling

Keunwoo Kim¹, Junghyun Lee¹, Seokwoo Hong¹, Hyunwoo Kim¹, Boogyo Sim¹, Kyungjune Son¹, Taemin Shin¹, Keeyoung Son¹, Jinyoung Kim², Kyubong Kong³, and Joungho Kim¹

¹School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

²Korea Electric Terminal (KET), Incheon, Republic of Korea

³SK Hynix, Icheon, Republic of Korea

Abstract— In this paper, we propose a new board-to-board connector design for 5G mmWave and high-speed signaling. The proposed board-to-board connector has a socket shield and mid-plate between pins. The socket shell reduces electromagnetic interference (EMI), emitted to the outside and the mid-plate reduces crosstalk between terminals. We verified the signal integrity performance and EMI reduction effect of the newly added ground structures through EM simulation. However, due to the increased ground structure, the return current path of the signal is split, and resonances are generated. We analyzed the resonances through the J-field change according to the frequencies.

Keywords— Board-to-board (BtB) connector, electromagnetic interference (EMI), mid-plate, return current path split, resonance, shielding structure

I. INTRODUCTION

In recent years, mobile electronic devices are becoming more compact and integrated. Mobile devices require various sensors, antennas, and chips to perform a set function, and they must be combined into a single system. However, each component is modularized, so a miniaturized connector is needed to connect each component without loss. One of these components is a board-to-board (BtB) connector used to connect two printed circuit boards (PCB) or flexible printed circuits (FPC) [1]

Fig. 1. shows the approximate form of the BtB connector and how it is used in a mobile application. The BtB connector has the advantage of being able to use both analog and digital signals, single-ended and differential signals. For RF, the BtB connector transfers data from the 5GmmWave antenna module or 3G/4G/GPS/Wi-Fi antenna to the main PCB. It is also available for high-speed signaling, which consists of differential pairs for USB or Thunderbolt. It is also possible to transfer the power of the battery to the main PCB.

However, the increasing data rate of mobile devices and the use of high-frequency bands such as 5G mmWave cause signal integrity (SI) and electromagnetic interference (EMI) problems in BtB connectors. First, the EMI problem is serious because there are fewer ground structures than PCB traces in the BtB connector. In addition, there are various components such as RF modules, analog circuits, sensors, etc. on one PCB, and as data rate increases, intra-system noise due to noise generated by each component as well as loss is becoming a bigger problem. In particular, in the case of RF noise, even small noise can be greatly amplified, so it is even more sensitive as Fig. 1. shows.

There are many attempts to add additional shielding structures to reduce EMI or crosstalk on various connectors, including BtB connectors [1-2]. Although there are a wide variety of ways to reduce EMI or crosstalk, the approach of simply adding ground structures is likely to further deepen EMI

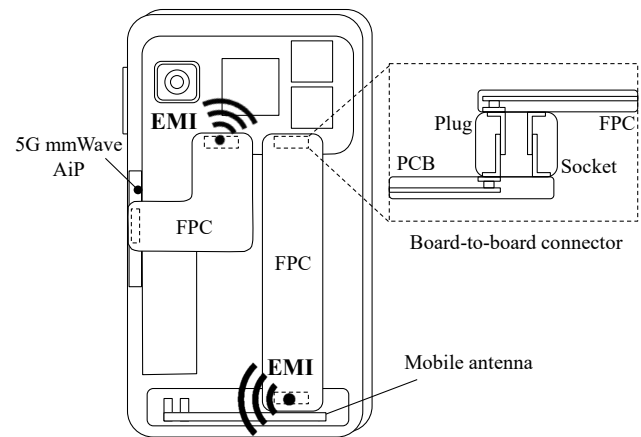


Fig. 1. Board-to-board connector at a mobile device and EMI issues

problems or reduce SI performance. Resonance may occur when a floating conductor exists or has multiple return current paths [3].

In this paper, we proposed a BtB connector with a socket shield and mid-plate located between plug terminals. The added ground structures are effective in reducing the increased crosstalk and EMI due to the increased data rate. The proposed structure was verified through an E-field obtained through 3-D EM simulation. In addition, how the additional side effect of the proposed structure was analyzed in terms of return current path and the design guide for better performance was presented

II. PROPOSAL OF BOARD-TO-BOARD CONNECTOR WITH SOCKET SHELL AND MID-PLATE

Fig. 2. and Table I show the appearance and dimension information of the proposed BtB connector. The BtB connector is divided into the plug and socket parts. The plug and socket have terminals that transmit signals or act as references, consisting of a horizontal part that meets the solder mask defined (SMD) type pads, and a vertical part that allow the plug and socket to contact stably in a clip form. Each terminal is spaced at a certain distance to the specification. As shown in Fig. 2. (b), the mid-plate for reducing crosstalk between terminals belongs to the plug part and is assembled in two considering manufacturability. The socket has a shell structure to prevent EMI emitted to the outside. The mid-plate and terminals are composed of copper alloy, while the socket is composed of stainless. The lower part of the socket shell is slightly bent outward to minimize the coupling effect with the terminal.

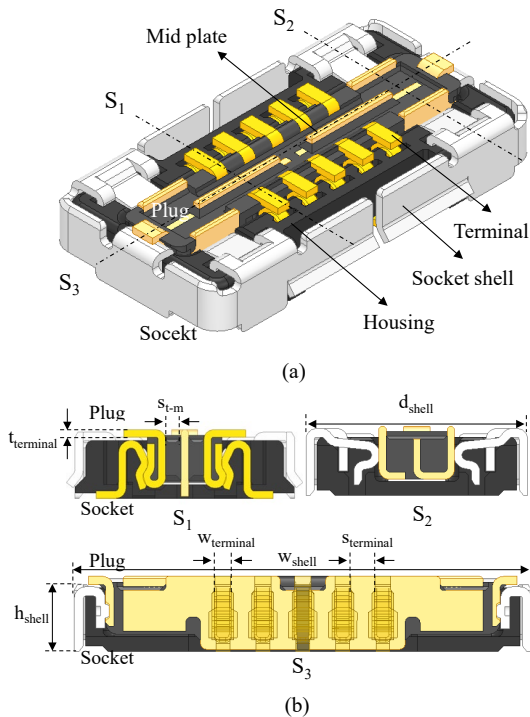


Fig. 2. (a) Proposed board-to-board (BtB) connector (b) and its side view

TABLE I

MATERIAL PROPERTIES AND DIMENSIONS OF THE PROPOSED BOARD-TO-BOARD CONNECTOR

Parameter	Value
Material of the housing	LCP S475 (Dk=3.7, Df=1e-3)
Material of the socket shell	STS 301(cond=1.38e6 S/m)
Material of the terminal and mid-plate	C7025 ($\sigma=2.33e7$ S/m)
Width of the socket shell (w_{shell})	4 mm
Depth of the socket shell (d_{shell})	2.11 mm
Height of the socket shell (h_{shell})	0.63 mm
Width of the terminal ($w_{terminal}$)	0.12 mm
Space between terminals ($s_{terminal}$)	0.23 mm
Thickness of the terminal ($t_{terminal}$)	0.07 mm
Space between a terminal and a mid-plate (s_{t-m})	0.15 mm

The impedance of conventional BtB connectors is defined by adjacent ground terminals, while shell and mid-plate additionally affect the impedance of the proposed BtB connectors. In addition, the mid-plate is short only on the plug PCB side and the shell is short only on the socket PCB side.

The proposed BtB connector is designed to be used simultaneously for single-ended RF such as 5G mmWave and high-speed signaling using differential pairs such as USB/Thunderbolt. Fig. 3. shows an example pin map for each situation. The proposed BtB connector structure is applicable not only to the 10-pin BtB connector shown in Fig. 2 but also to all BtB connectors with various pin counts.

III. SIMULATION RESULT AND ANALYSIS OF PROPOSED BTB CONNECTOR

Fig. 4. shows E-field at 1 GHz of the proposed BtB connector with shell and mid-plate and conventional BtB connector. In the BtB connector without a shell, the E-field formed by the terminal is emitted to the outside without any restraint. On the other hand, a BtB connector with a shell

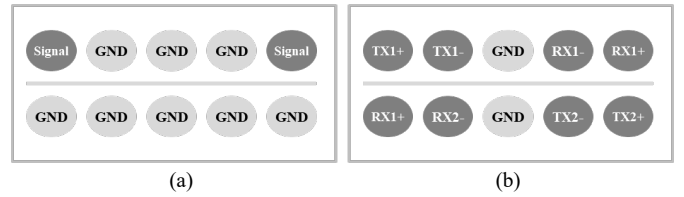


Fig. 3. (a) pin map of 10 pin BtB connector for 5G mmwave and (b) pin map for high-speed signaling

blocked EMI generated from the terminal. Moreover, The lower side of the shell is bent outward so it maintains distance from the terminal, and accordingly, an unnecessary coupling effect with the terminal is removed. The mid-plate shorted to the plug side PCB significantly reduced the crosstalk to the opposite terminal. However, the E-field of the socket side was still emitted to the opposite side, because the mid-plate is opened with socket side PCB ground.

First, when applying the pin map of 5G mmWave, all four cases had multiple resonances in insertion loss and FEXT. Since the length of the BtB connector is very short, so it is most important to reduce the influence of resonance at the operating frequency by making the resonance frequency as large as possible rather than reducing the loss of the connector itself. These resonances are caused because the return current path of the signal changes depending on the frequency. The resonance occurs at different frequencies because the ground structure of each case is different.

The resonance frequency due to the return current path split is defined as follows:

$$f_{res} = \frac{c}{\lambda_{res}\sqrt{\epsilon_{eff}}} = \frac{c}{2\Delta L\sqrt{\epsilon_{eff}}} \quad (1)$$

, where f_{res} is resonance frequency, λ_{res} is the wavelength at the resonance frequency, c is the velocity of light, ϵ_{eff} is the effective dielectric constant of the terminal, and ΔL is the difference of the return current path. This is because a resonance occurs at a frequency in which the phase difference of signals transmitted through each return current path becomes 180 degrees.

The return current path according to the frequency of the proposed BtB connector was analyzed through a complex J-field as shown in Fig. 6. The return current path was classified as return path 0 (RP0) to adjacent ground terminals, return path 1 (RP1) to the shell of left under the corner, return path 2 (RP2)

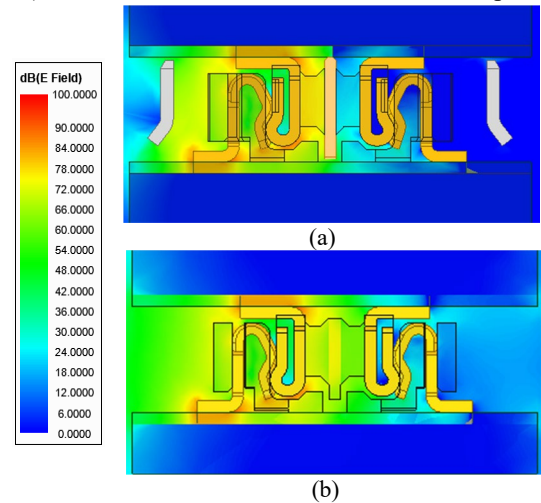


Fig. 4. (a) E-field of BtB connector with socket shell and mid-plate at 1 GHz (b) E-field of BtB connector without socket shell and mid-plate at 1 GHz

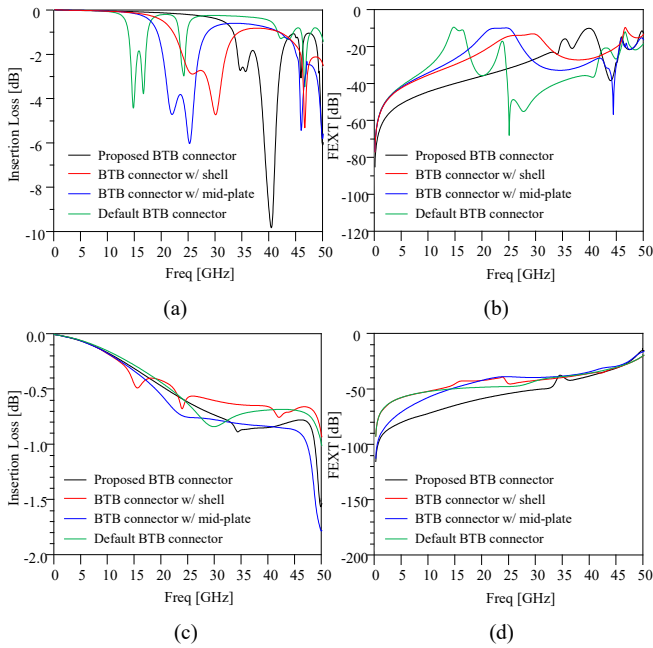


Fig. 5. (a) The insertion loss and (b) far-end crosstalk (FEXT) of BtB connector for 5G mmWave. (c) The insertion loss and (d) far-end crosstalk (FEXT) of BtB connector for high-speed signaling.

to the shell of the right upper corner, return path 3 (RP3) to the shell of right under the corner, and return path 4 (RP4) to the opposite side terminal. The resonances occur at 34.5 GHz, 35.5 GHz, and 40 GHz as shown in Fig. 5. (a). Since the Nyquist frequency when using 64-QAM in the 28 GHz 5Gmmwave band is 4.67GHz, and the Nyquist frequency in the USB4, which has a 20 Gbps data rate per lane, is 10GHz, assuming that up to 7th harmonic frequency affects signal performance, all of the resonances affects performance. Although the propagation velocity and effective dielectric constant for each return current path are different, the path length difference ΔL is 2.56 mm, 2.48 mm, and 2.22 mm based on the micro stripe line effective dielectric constant of the terminal [4]. These values are the length difference between RP1 and RP2, RP1 and RP3, and RP0 and RP4, respectively.

When the pin map of the high-speed signal is applied to the BtB connectors, the resonance size becomes very small because it has different pairs and the number of ground terminals is significantly reduced. Furthermore, thanks to the mid-plate, FEXT was significantly reduced in both pin maps

IV. DISCUSSION AND CONCLUSION

In this paper, we proposed a BtB connector with socket shell and mid-plate for low EMI. Through 3-D EM simulation, we verified the reduced EMI and crosstalk. Moreover, we analyzed resonances that greatly affect the signal performance when the BtB connector is used for 5G mmWave and USB4 in terms of return current path.

In the previous chapter, the magnitude of FEXT decreased and resonance frequency increased as additional ground structures were added. The reason why FEXT decreases is additional ground structures and the reason why resonance frequency increases are the length difference between the return current paths decreases. Therefore, to have better performance than the proposed BtB connector, the mid-plate and socket shell

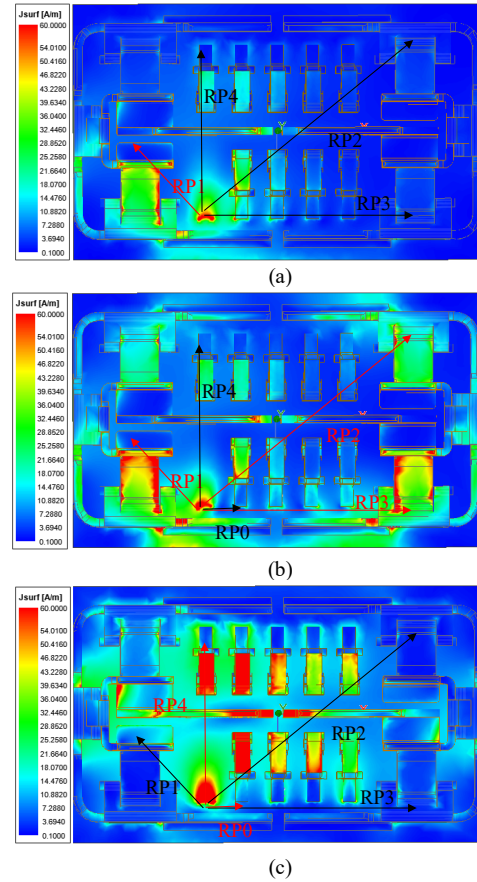


Fig.6. (a) Complex J-field of the proposed BtB connector at 32.5 GHz, (b) 35 GHz, (c) 40 GHz

must be shorted to both sides PCB ground plane, but physically impossible for micro stripe line type PCB and FPC. The only way is to make the difference between the return current path small by simply transforming the position of the PCB pad and the shape of the shell.

ACKNOWLEDGMENT

We would like to acknowledge the technical support from ANSYS Korea. This work was supported in part by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2022M3I7A4072293) and the Technology Innovation Program (20015559) funded by the Ministry of Trade, Industry & Energy(MOTIE, Korea). This work was supported by Samsung Electronics Co., Ltd (IO201207-07813-0

REFERENCES

- [1] Y. -J. Ren, C. -K. Sun and B. Litzlbeck, "Low Leakage RF Coaxial Connectors and Board-to-Board Connectors with Radiation Emission Control," 2022 IEEE Space Hardware and Radio Conference (SHaRC), 2022
- [2] X. Yan et al., "EMI Investigation and Mitigation of Flexible Flat Cables and Connectors," 2021 IEEE International Joint EMC/SI/PI and EMC Europe Symposium, 2021
- [3] Y. Ko, K. Ito, J. Kudo and T. Sudo, "Electromagnetic radiation properties of a printed circuit board with a slot in the ground plane," 1999 International Symposium on Electromagnetic Compatibility (IEEE Cat. No.99EX147), 1999
- [4] Y. J. Yoon and B. Kim, "A new formula for effective dielectric constant in multi-dielectric layer microstrip structure," IEEE 9th Topical Meeting on Electrical Performance of Electronic Packaging (Cat. No.00TH8524), 2000