

Crosstalk Analysis and Countermeasures of High-Bandwidth 3D-Stacked Memory Using Multi-Hop Inductive Coupling Interface

Kota SHIBA^{†a)}, Atsutake KOSUGE[†], Mototsugu HAMADA[†], *Nonmembers*, and Tadahiro KURODA[†], *Fellow*

SUMMARY This paper describes an in-depth analysis of crosstalk in a high-bandwidth 3D-stacked memory using a multi-hop inductive coupling interface and proposes two countermeasures. This work analyzes the crosstalk among seven stacked chips using a 3D electromagnetic (EM) simulator. The detailed analysis reveals two main crosstalk sources: concentric coils and adjacent coils. To suppress these crosstalks, this paper proposes two corresponding countermeasures: shorted coils and 8-shaped coils. The combination of these coils improves area efficiency by a factor of 4 in simulation. The proposed methods enable an area-efficient inductive coupling interface for high-bandwidth stacked memory.

key words: 3D integration, 3D memory, 8-shaped coil, inductive coupling, through-silicon via (TSV), ThruChip Interface

1. Introduction

The growth of deep neural networks (DNNs) has been increasing the demand for large-capacity and high-bandwidth memory. To meet this demand, high-bandwidth memory (HBM), which is a 3D-stacked DRAM module (3D-DRAM), has attracted much interest [1]–[3]. Communication among 3D-DRAMs and a base die are achieved using through-silicon vias (TSVs), which are mechanical die-to-die electrodes through chips. The HBM has achieved large capacity and high bandwidth thanks to multiple-die stacks and thousands of TSVs, respectively. In addition, a 3D-stacked SRAM module (3D-SRAM) using TSVs and μ -bumps, which is fabricated in a state-of-the-art 7-nm FinFET process and stacked on a system-on-chip (SoC), has been proposed as well [4], [5]. However, a TSV has a high cost, low yield, and low reliability due to the necessity of additional complicated manufacturing processes and the exposure of the electrodes.

To address these TSV issues, a wireless inter-chip communication technology using inductive coupling, namely the ThruChip Interface (TCI), has emerged [6]–[8]. TCI achieves low-power high-speed wireless communication using on-chip coils. Though TSVs need additional manufacturing steps, TCI is compatible with a standard complementary metal-oxide semiconductor (CMOS) process, leading to low costs. TCI enables high reliability and low-power high-speed communication thanks to the elimination of physical contacts and electrostatic discharge (ESD) protection circuits, respectively. Therefore, TCI is a promising low-cost

high-performance inter-chip communication technology to replace TSVs.

A 3D-SRAM stacked on a DNN accelerator using inductive coupling has been proposed [6], [7]. The wireless connection is conducted by using a multi-drop inductive coupling interface (multi-drop TCI). Though the multi-drop TCI is suitable for low-power applications, it has issues with area efficiency [Tb/s/mm²] because it requires large coils. The multi-drop TCI requires 200- μ m-square coils to cover the 64- μ m communication distance (eight 8- μ m-thick chips), while the minimum pitch of μ -bumps connecting TSVs is 40–60 μ m. In [8], an area-efficient multi-hop inductive coupling interface (multi-hop TCI) was proposed. In the multi-hop TCI, the relay transmission is conducted by using about 20- μ m-square coils until the data arrive at a target memory chip.

However, while both horizontal crosstalk discussed in [9] and the newly-introduced vertical crosstalk from coils on multiple stacked-dies need to be evaluated, [8] reported local crosstalk only from coils on the same die. Therefore, this work conducts a detailed analysis of crosstalk and proposes two countermeasures [10].

This paper is organized as follows. Section 2 explains a 3D-stacked memory module using TCI, evaluates a baseline crosstalk model, and raises two problems, followed by the proposal of two countermeasures; shorted coils and 8-shaped coils. Section 3 shows simulated results of the proposed methods and performance comparisons with the baseline. The simulation results show the proposed methods improve the area efficiency by a factor of 4. Section 4 concludes this paper.

2. Proposed Shorted and 8-Shaped Coils

2.1 Baseline Analysis

Figure 1 illustrates a 3D-stacked memory module along with the arrangement of coils. All memory dies have the same coil patterns and only necessary coils are enabled, which helps stacked-dies to be made from common mask sets. As shown in Fig. 1 (a), relay transmission is done until the data reach a target memory die (memory #4 in the figure) in the downlinks. Then, for a read operation, read data are relayed back to the base die in the same manner in the uplinks. The transmitters are enabled right before they are about to be used. In this work, the chip thickness and coil diameter are set to 8 μ m and 20 μ m, respectively, where the diam-

Manuscript received July 13, 2022.

Manuscript publicized September 30, 2022.

[†]The authors are with The University of Tokyo, Tokyo, 113–8566 Japan.

a) E-mail: shiba@kuroda.t.u-tokyo.ac.jp
DOI: 10.1587/transele.2022CDS0001

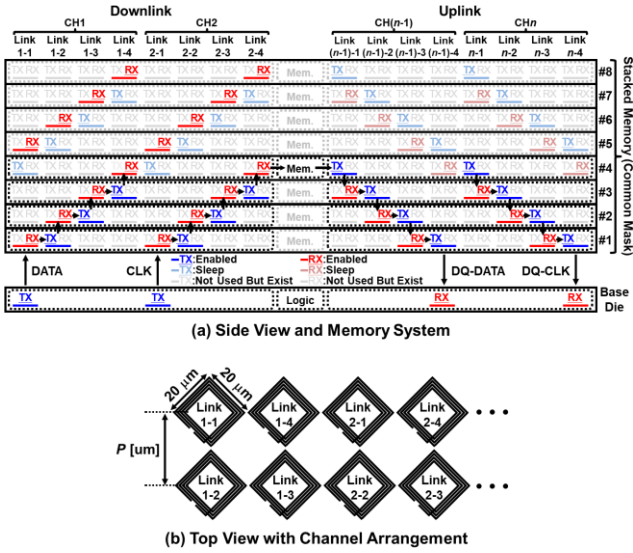


Fig. 1 3D memory using multi-hop inductive coupling interface along with arrangement of coils [8].

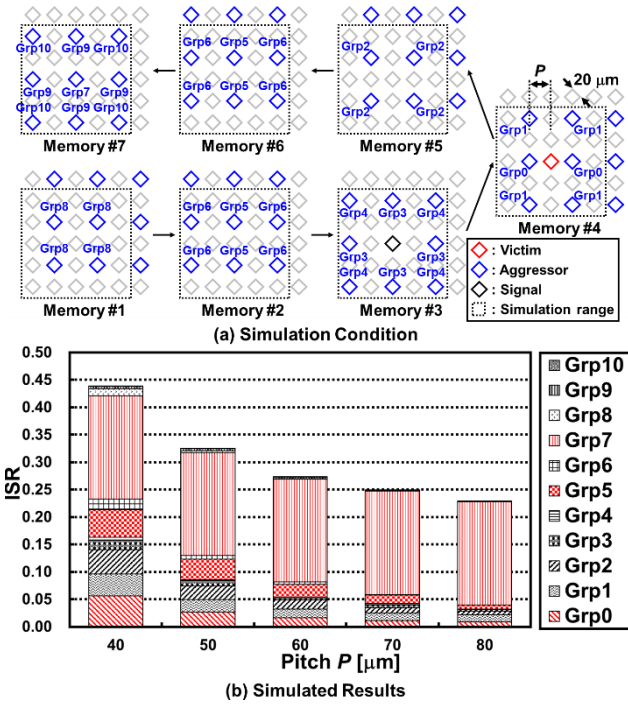


Fig. 2 Baseline crosstalk simulation results.

eter of TCI coils is recommended to be set to 2.5 times as large as the communication distance [11]–[13]. To evaluate crosstalk, the pitch of the coils is regarded as a parametric parameter P , which affects both crosstalk and area efficiency.

Figure 2 shows simulated results of crosstalk in the aforementioned baseline. Figure 2(a) illustrates simulation conditions with the coil arrangement. The coil receiving maximum crosstalk is on the center chip (memory #4). Therefore, a victim coil is designated to a coil in memory #4, and aggressor coils are on the same chip (memory #4), three upper chips (memory #5–7), and three lower chips (memory #1–3). The size of the simulated plane is set to 5×5 . In

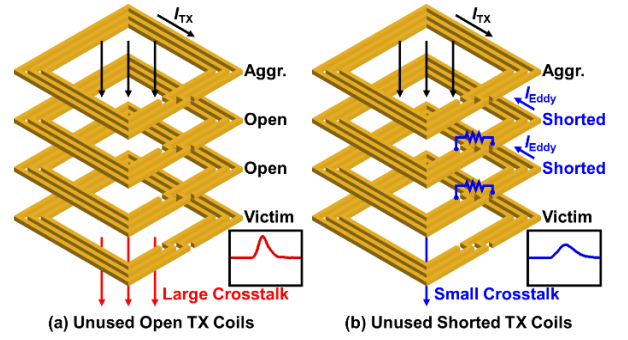


Fig. 3 Proposed shorted coils.

short, the crosstalk from $5 \times 5 \times 7$ cubic dimensions is considered in simulation. To make it easier to see the contributions of each aggressor, the coils, which are relatively located in the same position seen from the victim, are grouped as you can see in the figure.

Figure 2 (b) illustrates the simulated crosstalk results of the baseline. The crosstalk is evaluated in the interference-to-signal ratio (ISR), defined as the crosstalk voltage amplitude level normalized by the signal voltage amplitude level, based on the 3D electromagnetic (EM) simulation and SPICE simulation. The larger the pitch, the smaller the crosstalk and the worse the area efficiency. Hence, there is a trade-off between the ISR and area efficiency. At a pitch of $40 \mu\text{m}$, the three largest crosstalks come from Grp7, Grp0, and Grp5 in order. Grp7 is concentrically arranged with the victim. Grp0 and Grp5 are arranged right beside the victim in the same chip and two chips away, respectively. Therefore, the crosstalk from a concentric coil and adjacent coils has to be reduced for a high-density coil’s arrangement.

2.2 Shorted Coil

First, we propose shorted coils to reduce the crosstalk from a concentric coil (Grp7). When a magnetic field penetrates the looped conductive region, the current, namely the eddy current, flows through that loop and the magnetic field attenuates. The looped conductors between the transmitter (Tx) and receiver (Rx) coils lead to a small received signal and low energy-efficiency. In this work, we intentionally make looped conductors between the victim and aggressor to reduce the crosstalk without any change of the floorplan and packaging method.

As seen in Fig. 1 (a), there are many unused coils in the memory chips. If those unused coils are at an open state as shown in Fig. 3 (a), the magnetic field easily penetrates them. However, if those unused coils are shorted with the resistance as shown in Fig. 3 (b), the magnetic field attenuates because the eddy current flows through the shorted coils. This is how the crosstalk from a concentric coil is reduced.

Though a similar method to reduce the crosstalk by intentionally exploiting the eddy current is discussed in [14], the method requires additional metal plate layers and its additional packaging steps, which results in a larger coil di-

ameter and higher cost, respectively. In this paper, since we use unused coils to reduce the crosstalk, we can keep the size of the coils and require no additional manufacturing steps, leading to high-density I/O and low costs.

2.3 8-Shaped Coil

Second, we propose 8-shaped coils to reduce the crosstalk from adjacent coils (Grp0&5). Figure 4 illustrates the basic concept of 8-shaped coils, which are composed of two inversely-turned rectangular coils. When the current I_L and I_R flows into the Tx coil, magnetic fields B_L and B_R are generated in opposite directions to each other. The transition in the magnetic field induces voltages V_L and V_R in the Rx coil. Therefore, the total received voltage is $V_{RX} = V_L + V_R$.

Since 8-shaped coils wirelessly communicate by using differential magnetic fields, a differential Rx coil can cancel the common-mode magnetic field. In this work, the adjacent coils are rotated 90 degrees as shown in Fig. 5. When the currents I_L and I_R flow into the adjacent Tx coil, the magnetic fields B_L and B_R are generated in the opposite direction. Note that B_L and B_R have the same absolute magnitude and are in opposite directions. Although the left-sided Rx coil receives $B_{L,L}$ and $B_{R,L}$, no voltage is induced because $B_{L,L}$ and $B_{R,L}$ have the same absolute magnitude and are in opposite directions; hence, $B_{L,L} + B_{R,L} = 0$. The right-sided Rx has no induced voltage for the same reason. This

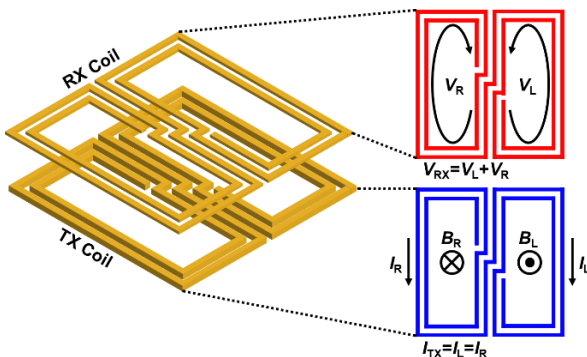


Fig. 4 Proposed 8-shaped coils.

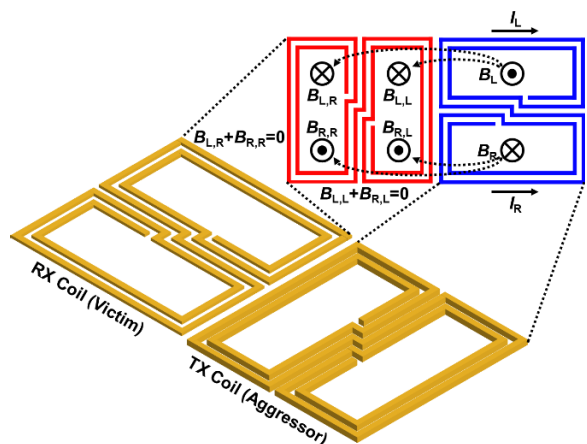


Fig. 5 Crosstalk tolerance of proposed 8-shaped coils.

is how the crosstalk from adjacent coils is reduced thanks to the 8-shaped coils.

Although the 8-shaped coils are used to demonstrate wireless full-duplex communication in [15], this paper exploits it to reduce the adjacent crosstalk and achieve high-density coil placement.

3. Results and Comparisons

3.1 Simulation Results

The crosstalk analysis results with the proposed shorted coils and 8-shaped coils are shown in Fig. 6. Figure 6(a) illustrates the 8-shaped-coil arrangement with shorted coils. To make a fair comparison with the baseline, the 8-shaped coils have a 36- μm diameter, where 20- μm standard coils and 36- μm 8-shaped coils have the same mutual-inductance because small 8-shaped coils have a low coupling coefficient.

Figure 6(b) shows the crosstalk simulation results. The short resistance is set to 150 Ω with the transistor's on-resistance taken into account. The crosstalk from Grp7 and Grp0&5 is reduced thanks to the introduction of the shorted coils and 8-shaped coils, respectively. The crosstalk from others is also reduced thanks to the differential nature.

3.2 Performance Comparison

Figure 7 illustrates the crosstalk analysis summary of the baseline and the proposed models. Thanks to the shorted coils and 8-shaped coils, the pitch can be shrunk down from

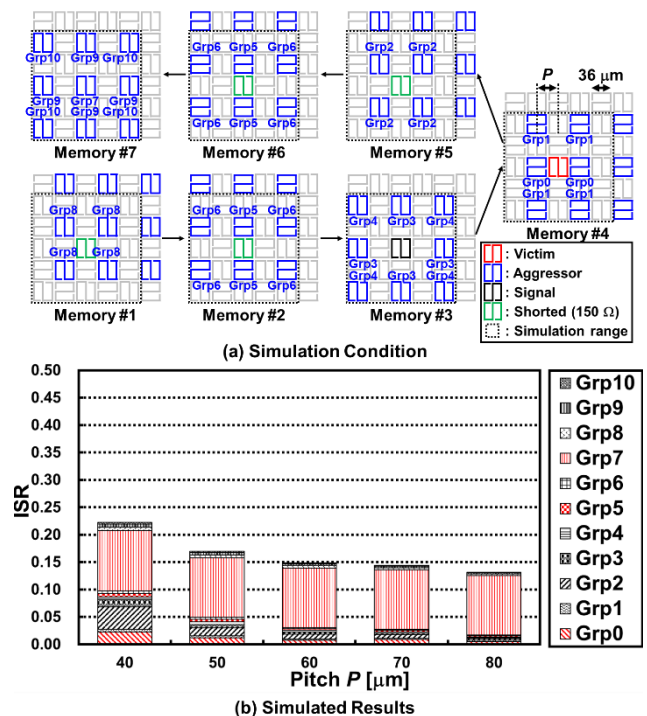


Fig. 6 Crosstalk simulation results with proposed shorted coils and 8-shaped coils.

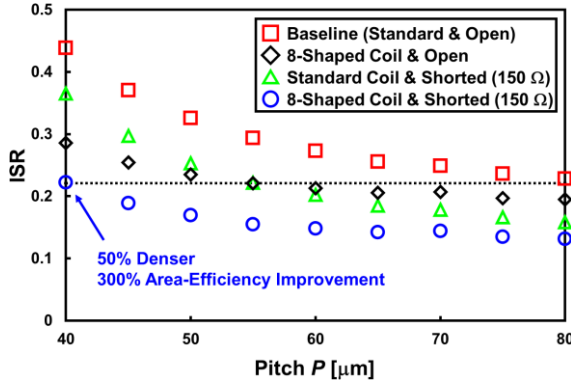


Fig. 7 Crosstalk analysis summary.

Table 1 Performance comparison.

	JSSC'17 ^[1]	JSSC'19 ^[6] TCAS-I'21 ^[7]	This work (Baseline)	This work (Proposed method)	This work (7 nm)
Coil Diameter D [μm]	N/A	200	20	36	36
Communication Distance z [μm]	N/A	64	8.0	8.0	8.0
I/O Pitch P [μm]	48 / 55	400	80 (1)	40 (1/2)	40
Data-Rate [Gb/s/pin]	2.4	3.6	3.6	3.6	64
Area Efficiency [Tb/s/mm ²]	0.91	0.0225	0.141 (1)	0.563 (4)	10
I/O Interface	μ -Bump +TSV	Coil	Coil		
Process Node	20-nm DRAM	40-nm CMOS	40-nm CMOS		7-nm CMOS

80 μm to 40 μm when the ISR is 0.22 (−13 dB). In short, the proposed methods improve the area efficiency by a factor of 4 compared with the baseline. Therefore, the proposed crosstalk countermeasures are key enablers for a high-density inductive coupling memory interface.

Table 1 summarizes the performance comparison of the conventional and proposed models. In a 7-nm CMOS process, a 10-Tb/s/mm² wireless interface is achievable thanks to the shorted coils and 8-shaped coils, overwhelming the area efficiency of TSVs.

4. Conclusion

This paper analyzes the crosstalk of high-bandwidth 3D-stacked memory using a multi-hop inductive coupling interface, where two main crosstalk sources are exposed. This work proposes two corresponding countermeasures, shorted coils and 8-shaped coils, to reduce the crosstalk from a concentric coil and adjacent coils, respectively. The area efficiency of the multi-hop TCI is improved by a factor of 4 thanks to the shorted coils and 8-shaped coils, which achieves high-bandwidth 3D-stacked memory.

Acknowledgments

This work was supported by JST, ACT-X Grant Number JPMJAX210A and JSPS KAKENHI Grant Number 21J11729.

References

[1] K. Sohn, W.-J. Yun, R. Oh, C.-S. Oh, S.-Y. Seo, M.-S. Park, D.-H.

Shin, W.-C. Jung, S.-H. Shin, J.-M. Ryu, H.-S. Yu, J.-H. Jung, H. Lee, S.-Y. Kang, Y.-S. Sohn, J.-H. Choi, Y.-C. Bae, S.-J. Jang, and G. Jin, "A 1.2 V 20 nm 307 GB/s HBM DRAM with at-speed wafer-level IO test scheme and adaptive refresh considering temperature distribution," *IEEE J. Solid-State Circuits*, vol.52, no.1, pp.250–260, Jan. 2017.

[2] D. Kim, J. Kung, S. Chai, S. Yalamanchili, and S. Mukhopadhyay, "Neurocube: A programmable digital neuromorphic architecture with high-density 3D memory," *Proc. IEEE Int. Symp. Comput. Archit. (ISCA)*, vol.44, no.3, pp.380–392, June 2016.

[3] M. O'Connor, N. Chatterjee, D. Lee, J. Wilson, A. Agrawal, S.W. Keckler, and W.J. Dally, "Fine-grained DRAM: energy-efficient DRAM for extreme bandwidth systems," *MICRO-50*, pp.41–54, Oct. 2017.

[4] K. Cho, et al., "SAINT-S: 3D SRAM stacking solution based on 7nm TSV technology," *IEEE Hot Chips 32 Symposium (HCS)*, Aug. 2020.

[5] S.-K. Seo, C. Jo, M. Choi, T. Kim, and H.-e. Kim, "CoW package solution for improving thermal characteristic of TSV-SiP for AI-inference," *2021 IEEE 71st Electronic Components and Technology Conference (ECTC)*, pp.1115–1118, 2021.

[6] K. Ueyoshi, K. Ando, K. Hirose, S. Takamaeda-Yamazaki, M. Hamada, T. Kuroda, and M. Motomura, "QUEST: Multi-purpose log-quantized DNN inference engine stacked on 96-MB 3-D SRAM using inductive coupling technology in 40-nm CMOS," *IEEE J. Solid-State Circuits*, vol.54, no.1, pp.186–196, Jan. 2019.

[7] K. Shiba, T. Omori, K. Ueyoshi, S. Takamaeda-Yamazaki, M. Motomura, M. Hamada, and T. Kuroda, "A 96-MB 3D-stacked SRAM using inductive coupling with 0.4-V transmitter, termination scheme and 12:1 SerDes in 40-nm CMOS," *IEEE Trans. Circuits Syst.-I: Regular Papers (TCAS-I)*, vol.68, no.2, pp.692–703, Feb. 2021.

[8] K. Shiba, T. Omori, M. Usui, M. Hamada, and T. Kuroda, "Area-efficient multihop inductive coupling interface for 3D-stacked memory with 0.23-V transmitter and sub-10- μm coil design," *IEEE Solid-State Circuits Letters (SSC-L)*, vol.3, pp.370–373, 2020.

[9] N. Miura, T. Sakurai, and T. Kuroda, "Crosstalk countermeasures for high-density inductive-coupling channel array," *IEEE J. Solid-State Circuits (JSSC)*, vol.42, no.2, pp.410–421, Feb. 2007.

[10] K. Shiba, T. Omori, M. Okada, M. Hamada, and T. Kuroda, "Crosstalk analysis and countermeasures of high-density multi-hop inductive coupling interface for 3D-stacked memory," *IEEE Electrical Design of Advanced Packaging and Systems (EDAPS)*, pp.1–3, Dec. 2020.

[11] L.-C. Hsu, J. Kadomoto, S. Hasegawa, A. Kosuge, Y. Take, and T. Kuroda, "Analytical thruchip inductive coupling channel design optimization," *ACM ASP-DAC*, pp.731–736, 2016.

[12] L.-C. Hsu, J. Kadomoto, S. Hasegawa, A. Kosuge, Y. Take, and T. Kuroda, "A study of physical design guidelines in thruchip inductive coupling channel," *IEICE Trans. on Fundamentals*, vol.E98-A, no.12, pp.2584–2591, Dec. 2015.

[13] Y.S. Kim, S. Kodama, Y. Mizushima, N. Maeda, H. Kitada, K. Fujimoto, T. Nakamura, D. Suzuki, A. Kawai, K. Arai, and T. Ohba, "Ultra thinning down to 4- μm using 300-mm wafer proven by 40-nm node 2Gb DRAM for 3D multi-stack WOW applications," *IEEE Symp. VLSI Technology*, pp.1–2, 2014.

[14] M. Saito, Y. Sugimori, Y. Kohama, Y. Yoshida, N. Miura, H. Ishikuro, T. Sakurai, and T. Kuroda, "2Gb/s 15pJ/b/chip inductive-coupling programmable bus for NAND flash memory stacking," *IEEE J. Solid-State Circuits (JSSC)*, vol.45, no.1, pp.134–141, Jan. 2010.

[15] Y. Yoshida, N. Miura, and T. Kuroda, "A 2 Gb/s bi-directional inter-chip data transceiver with differential inductors for high density inductive channel array," *IEEE J. Solid-State Circuits (JSSC)*, vol.43, no.11, pp.2363–2369, Nov. 2008.