# A 7-nm FinFET 1.2-TB/s/mm$^2$ 3D-Stacked SRAM Module With 0.7-pJ/b Inductive Coupling Interface Using Over-SRAM Coil and Manchester-Encoded Synchronous Transceiver

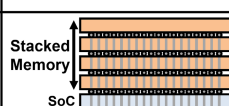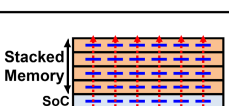Kota Shiba, *Graduate Student Member, IEEE*, Mitsuji Okada, *Member, IEEE*,
Atsutake Kosuge, *Member, IEEE*, Mototsugu Hamada, *Member, IEEE*, and Tadahiro Kuroda, *Fellow, IEEE*

*Abstract*—A 0.7-pJ/bit, 8.5-Gb/s/link inductive coupling inter-chip wireless communication interface for a 3D-stacked static-random access memory (SRAM) has been developed in a 7-nm FinFET process. A new physical placement method that allows coils to be placed over off-the-shelf SRAM macros with small magnetic field attenuation, together with the use of synchronous communication using Manchester encoding and a clocked comparator to enable the detection of small-swing signals, achieves a 26% reduction in SRAM die area compared to through-silicon via (TSV)-based stacking. Interchip communication at 0.7 pJ/bit, 8.5 Gb/s/link was confirmed using test chips. A 4-hi 3D-stacked SRAM module using the proposed interface achieves a 1.2-TB/s/mm$^2$ area efficiency, representing a two orders-of-magnitude improvement over the state-of-the-art 3D-stacked SRAM.

*Index Terms*—3D integration, 3D memory, 7-nm FinFET, clocked comparator, inductive coupling, Manchester encoding, static-random access memory (SRAM), through-silicon via (TSV), ThruChip Interface (TCI).

| | TSV | TCI |
|---|---|---|
| Overview | Stacked Memory / SoC | Stacked Memory / SoC |
| Process | Additional steps needed | Standard CMOS |
| Additional Cost | > 40% | A few % |
| Yield | Low | High |
| Reliability | Low | High |
| ESD Protection | Needed | Not needed |
| Keep *Out* Zone | Needed | Not needed |
| Level Shifter | Needed* | Not needed |

*Between chips with different supply voltages

Fig. 1.   Comparison of TSV and TCI.

## I. INTRODUCTION

**T**HE 3D-stacked memory plays an important role in achieving high-throughput and low-power operation in a variety of devices [1], [2], [3], [4], [5]. The 3D integration of memory dies, such as static-random access memories (SRAMs) [6], [7] and dynamic-random access memories (DRAMs) [8], [9], enables high bandwidth, large capacity, and low-power memory. In particular, 3D-stacked SRAM is used as a large cache in CPUs [3] and as memory in DNN inference accelerators [4], [5], where its low-power and low-latency features contribute to high energy efficiency and high

performance. The 3D-stacked SRAM will continue to be important as the demand for higher capacity SRAMs increases. However, the conventional wired integration using through-silicon vias (TSVs) and $\mu$-bumps poses challenges in terms of added TSV process costs, area overhead costs (TSVs must be placed outside of SRAM macros), and yield loss [6], [7]. A wireless interface using inductive coupling known as ThruChip Interface (TCI) has low cost and high yield due to its compatibility with standard CMOS fabrication processes [10], [11], [12], [13]. Furthermore, as shown in Fig. 1, TCI has the following advantages: high reliability due to the elimination of exposed electrodes, high-speed operation because no electrostatic discharge (ESD) protection circuit is required, high area efficiency because the interface can be placed over custom circuits, and stacking flexibility even on SoCs with supply voltages different from those of the 3D-SRAM because no level shifter circuit is required. In addition, it is possible to supply power to 3D-SRAMs by utilizing highly doped silicon vias (HDSVs) [14] or low-cost, relatively large TSVs while replacing the small TSVs for signals with TCI. However, the conventional TCI-based 3D-stacked SRAM [10] requires large interface areas that limit area efficiency and bandwidth. In this work, we propose an enhancement to TCI for 3D-SRAM

that achieves two orders-of-magnitude higher area efficiency compared to TSV-based stacking.

To solve the area efficiency issue of TCI, this article proposes a method of placing the communication coils over the SRAM macros. The new physical placement method of SRAM macros and TCI coils allows the coil to be placed directly over the SRAMs with only 30% attenuation of the magnetic field, which is much smaller than the conventional 99% attenuation. This reduces the IO area overhead and achieves a two orders-of-magnitude improvement in IO area efficiency compared to the TSV-based method. Furthermore, this work proposes a synchronous transceiver using Manchester encoding, which enables low-power communication even when the magnetic field is attenuated by about 30%. The new physical layout method and synchronous transceiver topology, which enable interchip communication with coils over SRAMs, achieve a 3D-stacked SRAM die area that is two logic process generations smaller than that of a TSV-based 3D-stacked SRAM.

This article is an extended version of our previous HCS'34 presentation [15] and adds extra description of the two proposed methods along with detailed simulation results, experimental results, and performance and scaling discussion. This work is organized as follows. Section II describes an overview and block diagram of the proposed 3D-stacked SRAM. Section III presents the novel over-SRAM coil placement that improves area efficiency by two orders of magnitude and the Manchester-encoded synchronous transceiver. Section IV discusses the experimental results, performance comparisons, and scaling scenarios. Section V concludes this article.

## II. 3D-STACKED SRAM USING INDUCTIVE COUPLING

### A. Overview

The proposed 3D-stacked SRAM channel has a 1-MB capacity per SRAM die and 12 unidirectional wireless links that use 110-$\mu$m coils to cover a maximum communication distance of 32 $\mu$m across four stacked, 8-$\mu$m-thick SRAM dies, as illustrated in Fig. 2(a) [16], [17]. While TCI can support both heterogeneous and homogeneous integrations, in this work, four SRAM dies are homogeneously integrated, which leads to low cost due to the use of common mask sets. A base die core performs memory access to a 4-MB vault spread across the four stacked dies. Since the wireless communication coils are placed over the SRAM macros, the area overhead of the inductive coupling links is negligible because only the transmitter and receiver circuits need an extra area. The maximum data rate of a TCI link is 8.5 Gb/s, which delivers a 4.3 GB/s/vault bandwidth using the memory system architecture of [10]. A die area of 0.52 mm$^2$ is required to achieve the 1-MB capacity and 4.3-GB/s bandwidth, as shown in Fig. 2(b), which is a 26% reduction over the conventional TSV-based interface [6]. Our 3D-SRAM is compared with the conventional 3D-SRAM using $\mu$-bumps and TSVs [6]. While a hybrid bonding-based 3D-SRAM can achieve a tighter pitch [18], [19], it is limited to a stack of 2-hi in commercial applications. The conventional $\mu$-bump implementation is therefore a more practical solution since a stack of 4-hi or more is already commercially available. A hypothetical 3-nm
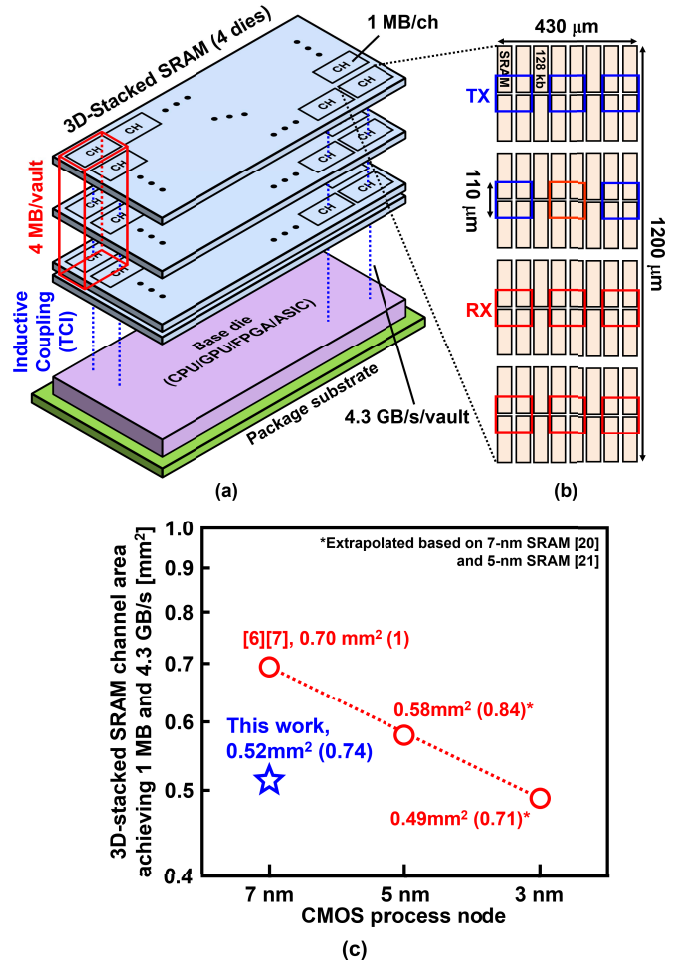


Fig. 2. Proposed 3D-stacked SRAM. (a) Overview, (b) channel floorplan, and (c) comparison with the state-of-the-art 3D-stacked SRAMs.

TSV-based 3D-stacked SRAM, where the SRAM blocks are scaled using an SRAM scaling rate of 0.78 extracted from an actual 7-nm SRAM [20] and 5-nm SRAM [21] and the IO area remains the same as 7 nm, requires a total die area of 0.49 mm$^2$. Therefore, the area achieved in this work using a 7-nm process is almost equivalent to that of a TSV-based solution implemented in a logic process that is two generations more advanced [22].

### B. Block Diagram

Fig. 3 shows a block diagram of the proposed 3D-SRAM module. As seen in [10], each channel in the base die can access any die in the vault or enter into sleep mode independent of the other channels. This architecture enables capacity scalability and low-power operation. Data communication is carried out in a source-synchronous manner. SRAM macros and TCI links operate at a frequency of 260 MHz and 8.5 GHz, respectively, where the 8.5-GHz clock is generated using an internal ring oscillator (RO) [23]. A 32:1 SerDes (serializer/deserializer) is used to exchange data between the TCI links and the SRAM macros and base die logic. Since the TCI transmitter drivers and receiver differential analog circuits,
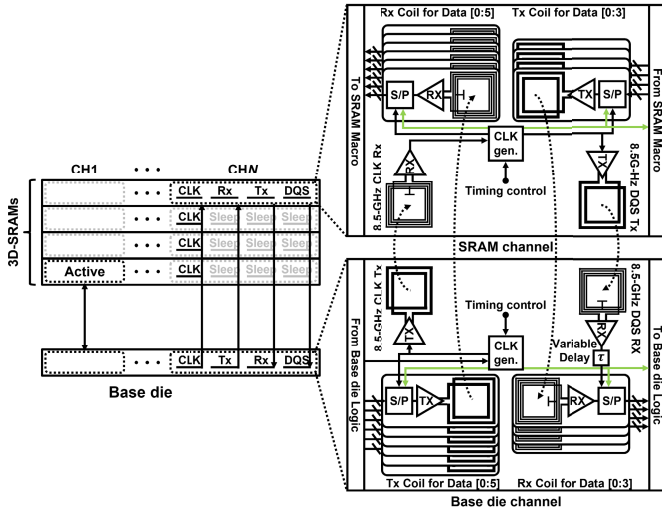
Fig. 3.   Proposed 3D-stacked SRAM block diagram.



Fig. 4.   Comparison of an inductive coupling interface (a) off SRAMs [10], (b) over a legacy-node SRAM, and (c) over 7-nm SRAMs.

which account for most of the TCI power consumption, consume the same dc power independent of the data rate, 32:1 SerDes reduces energy per bit and layout area by a factor of 32. The read and write latencies, including SerDes and SRAM operations, are three and two cycles, respectively, as discussed in [10], the same as that of the conventional TSV-based 3D-SRAM.

## III. PROPOSALS

This section proposes two key techniques to achieve a highly area-efficient wireless interface for the high-bandwidth 3D-SRAM. First, we propose a new method of placing coils over SRAM macros that improves area efficiency by two orders of magnitude over TSV methods while keeping the magnetic field attenuation due to the macros to only 30%. Second, we propose a synchronous wireless communication method using Manchester encoding that enables low-power communication even when the magnetic field is attenuated by 30%.

### A. Over-SRAM Coil

In the conventional TCI-based 3D-SRAM [10], TCI coils are placed outside of the SRAM hard macros area to prevent signal attenuation due to the power supply network. This results in low area efficiency and hence low channel bandwidth, as shown in Fig. 4(a). Placing the coils over the SRAM macros dramatically improves area efficiency, but it results in magnetic field attenuation due to eddy current loops of the same diameter as the coils being generated in the power mesh of the macros, as shown in Fig. 4(b). There will also be added signal ringing due to increased parasitic capacitance, which lowers the self-resonance frequency. This makes wireless communication impractical. The eddy current on the SRAM macros, both on the chip being accessed and those in the communication path, has adverse effect on the TCI communication. The metal layer structure of the SRAM and custom power mesh is shown in Fig. 5. In addition to
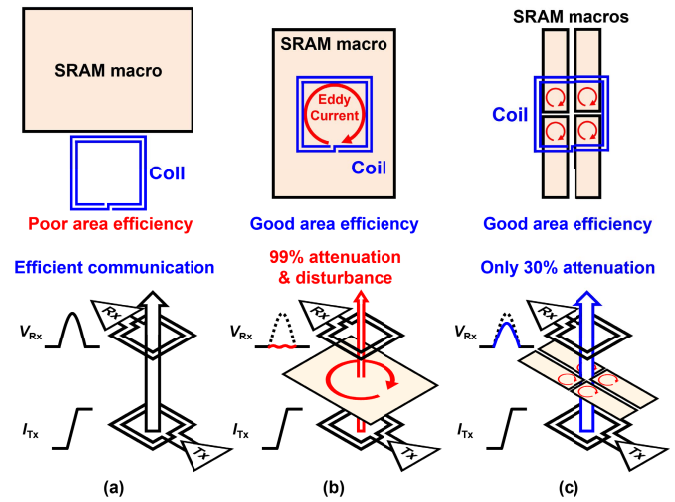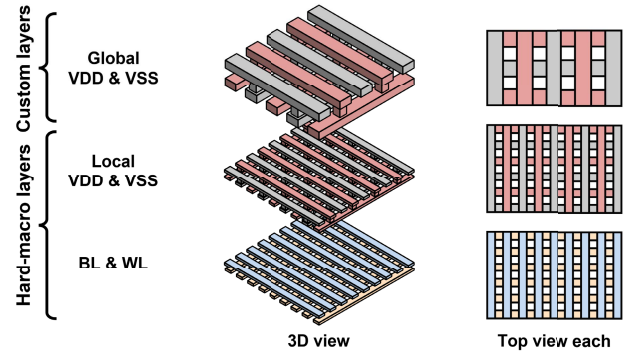


Fig. 5.   Metal layer structure of an SRAM and power mesh.

bit-lines (BLs) and word-lines (WLs), local power mesh is formed in the SRAM hard macro. To further enhance the SRAM power supply, a custom global power mesh can be formed on the SRAM macros. Although it is possible to design the custom global power mesh to eliminate eddy current loops by making fishbone-like mesh structure, modifying the local power mesh inside the SRAM hard macro is very time-consuming and costly both in design and verification and is impractical for the average designer [24]. Therefore, in this article, we propose a new physical layout method, where the coil is placed across four 7-nm off-the-shelf SRAM macros as illustrated in Fig. 4(c). Placing coils over SRAM macros with minimized eddy current loops without redesigning the off-the-shelf macros is a key innovation of this work. The generated eddy current loops become small relative to the coil diameter, and the magnetic field attenuation can be suppressed to 30%, which is much smaller than the 99% attenuation of the conventional placement. It will be explained later how we quantify attenuation.

Fig. 6 shows the simulation results of the effect of the coil position relative to the macros on the magnetic field attenuation. As shown in Fig. 6(a), four 128-kb SRAM macros are arranged under a coil, whose outer and inner diameters are
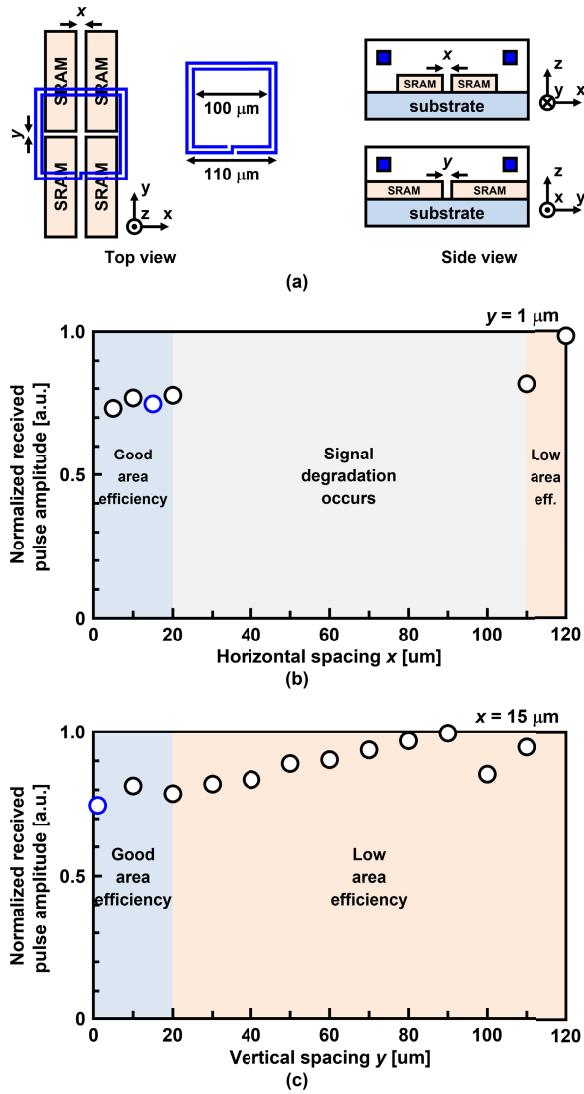
Fig. 6. Simulated effect of SRAM and coil positions on magnetic field attenuation (a) simulation conditions, (b) results of horizontal spacing, and (c) results of vertical spacing.
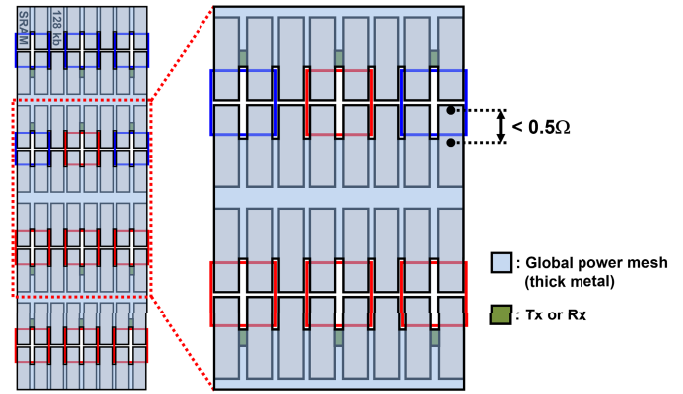


Fig. 7. Global power delivery network.

power supply wiring, which greatly lowers the self-resonance frequency of the coil, which causes signal degradation. If the macros are placed inside the coil ($x < 20$ $\mu$m), the influence of the macros is mitigated, and the magnetic field attenuation is only 30%. The influence of the macros remains almost the same as long as they are inside the coil. In this work, $x$ is set to 15 $\mu$m to allow either the transmitter or receiver circuit to be placed between the SRAM macros. Next, Fig. 6(c) shows the simulation results of the normalized pulse amplitude when $y$ is a variable and $x$ is set to 15 $\mu$m. Similarly, when the macros are placed outside the shadow of the coil ($y > 110$ $\mu$m), the eddy currents have little effect, so little attenuation of the magnetic field occurs. When $y < 20$ $\mu$m, the attenuation of the magnetic field saturates. In this work, $y$ was set to 1 $\mu$m to maximize area efficiency because no circuit is placed between the top and bottom SRAM macros.

Based on this analysis, it will be discussed how to select the proper size of SRAM macros. The analysis assumes that two SRAMs lined up side by side are completely inside the coil. Therefore, by selecting an SRAM with the short side of less than half the inner diameter of the coil, the magnetic field attenuation can be reduced to less than 30%. Since the inner diameter of the coil in this work is 100 $\mu$m, any SRAM with its short side <50 $\mu$m can be selected to place under the coil. The 64-bit single-bank SRAM with the largest number of words is selected in this work. Therefore, any 64-bit single-bank SRAM can be placed under coils in this work. On the logic side, this means any SRAM less than 64-bit in width can be used as cache placed under coils in the base chip.

It is also important to avoid magnetic field attenuation due to the global power delivery network as well as due to the local one inside the SRAM macros. Fig. 7 illustrates how to design the global power delivery network to minimize the attenuation. If the global power mesh is stretched over the entire chip or channel, a large attenuation due to eddy current will occur. To avoid the attenuation, the part of the mesh between SRAMs is cut around and inside the coils. Note that the impact of the global power network is taken into account in the results shown in Fig. 6. Although cutting the mesh has a negative impact on IR drop, the power delivery strength is still sufficient because the power supply resistance from the cut to the endpoint of the SRAM is less than 0.5 $\Omega$. Although

110 and 100 $\mu$m, respectively. $x$ and $y$ are spacing between the four SRAM macros in the $x$- and $y$-directions, respectively, which measure how much the coil overlaps with the macros. Attenuation can be evaluated by measuring the amplitude of the received pulse signal. The attenuation rate is expressed as the received pulse amplitude normalized to the case where no SRAM macro is placed. The pulse amplitude is obtained by using 7-nm SPICE simulation along with 3D electromagnetic (3D-EM) simulation to extract the S-parameters of the coils. Fig. 6(b) shows the simulation results of the normalized received pulse amplitude when $x$ is a variable and $y$ is set to 1 $\mu$m. When the macros are placed outside the shadow of the coil ($x > 110$ $\mu$m), there is negligible eddy currents in the macros and hence almost no attenuation of the magnetic field, but the area efficiency is poor due to a large blank area. On the other hand, when the macros are placed between the coil wiring and the substrate (20 $\mu$m $< x < 110$ $\mu$m), it is imprac- tical because the coil wiring is completely covered by the

Fig. 8. Floor plan comparison among (a) TSV + $\mu$-bump [6], (b) conventional TCI [10], and (c) proposed method.
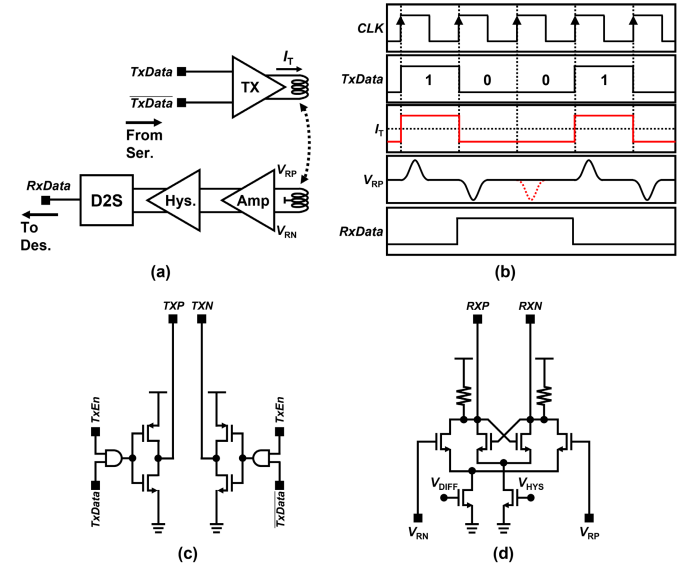


Fig. 9. (a) Block diagram, (b) waveforms of conventional TCI, (c) transmitter driver schematic, and (d) receiver hysteresis comparator driver schematic.

the space between the SRAMs seems to be dead space, it is not because it is possible to place circuits such as buffers and decoupling capacitors in addition to transmitters and receivers there. In addition, the coils are placed at a pitch that satisfies the crosstalk-to-signal ratio of $< -10$ dB as discussed in [25], where the crosstalk limit determines the maximum number of neighboring coils.

The SRAM macro placement method and the global power supply delivery technique can reduce the magnetic field attenuation to 30%. The 30% attenuation does not have adverse effect on the interface yield, but it forces the current of the conventional driver to be 1.43 $(=1/0.70)$ times larger with the associated driver circuit area increase to compensate for the lower received pulse amplitude. However, because the new transceiver circuit proposed in Section III-B allows wireless communication using a reduced, received amplitude, the driver current does not need to be increased and the 30% attenuation does not affect the yield, cost, power, and area.

If the same 4.3-GB/s bandwidth and 1-MB/ch capacity were to be achieved in the 3D-SRAM using TSVs and $\mu$-bumps fabricated in a 7-nm FinFET process [6], [7], 112 $\mu$-bumps with 0.76 Gb/s/link at 40-$\mu$m pitch would be required, resulting in a die area of 0.70 mm$^2$, as illustrated in Fig. 8(a). On the other hand, if coils were to be placed on the outside of the SRAM macros, as illustrated in Fig. 8(b), a large interface area overhead would be required, resulting in a large die area of 0.78 mm$^2$. Whereas both the conventional wired and wireless methods require a large chip area, by using the over-SRAM TCI, the channel area can be reduced to 0.52 mm$^2$, and the total IO area overhead is only 0.0037 mm$^2$ for the transceiver circuits, resulting in an IO area efficiency of 1162 GB/s/mm$^2$, as shown in Fig. 8(c). The same area efficiency is achievable not only for SRAM chips but also for the base chip. Since the base chip consists mainly of logic, coils can be easily placed over circuits with minimized attenuation by using the fishbone power supply structure proposed in [24]. Even if the base chip contains SRAMs, coils can be placed over the SRAMs as discussed in this article. In this

way, the IO area efficiency can be improved by two orders of magnitude and the die area can be reduced by 26% compared to the conventional method. To achieve the same IO area efficiency with $\mu$-bumps as our 3D SRAM, the $\mu$-bump pitch would need to be smaller than 5.7 $\mu$m, an improvement of one order of magnitude from the 40 $\mu$m of [6]. Compared to the conventional TCI-based 3D-SRAM [10], this work achieves a 34% reduction in chip area, i.e., chip cost, without major system-level changes. Compared to the TSV-based method [6], this work reduces the chip cost by 26% as well as the integration cost due to the wireless nature of our interface.

### B. Manchester-Encoded Synchronous Transceiver

We also propose a low-power transceiver topology using synchronous receivers to detect 30% lower amplitude pulses. The block diagram and operating waveforms of the conventional communication method are shown in Fig. 9(a) and (b), respectively. The non-return-to-zero (NRZ) signal to be transmitted determines the direction of the current flowing through the transmitter coil. During data transitions, the direction of the current changes and pulse-shaped signals are induced in the receiver coil. These pulses are detected by an asynchronous hysteresis comparator to complete communication. The schematic diagrams of the transmitter driver and the receiver hysteresis comparator are shown in Fig. 9(c) and (d), respectively. The driver is called a CMOS H-bridge transmitter, which makes current flow through the coil when the diagonally located transistors are turned on. The hysteresis comparator consists of a differential amplifier stage and a latch stage, and the threshold for pulse signal detection is adjustable by controlling the tail current of the latch stage. This topology is widely used in many TCI-based applications. However, in this method, the driver current must be increased by 30% to compensate for the 30% pulse signal attenuation to maintain the receiver margin, which leads to large power consumption.
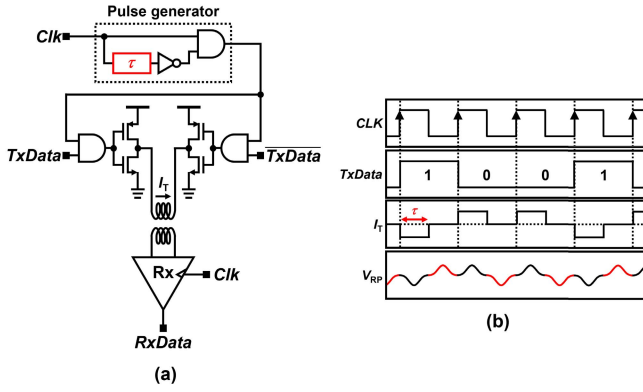
Fig. 10.  (a) Block diagram and (b) waveforms of conventional pulse transmitter [26].
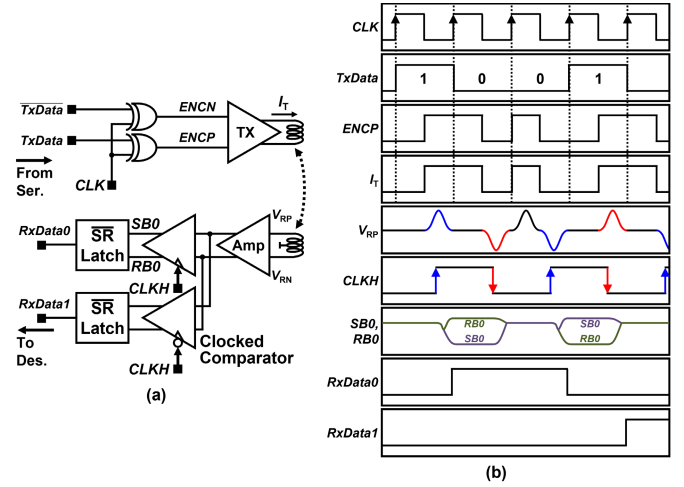


Fig. 11.  (a) Block diagram and (b) waveforms of the proposed Manchester-encoded synchronous TCI.



Fig. 12.  Schematic of the proposed (a) transmitter and (b) receiver.

To achieve reliable wireless communication without sacrificing energy efficiency, a new method to detect small-amplitude pulses is required.

Clocked comparators are effective for detecting low-swing signals because they can achieve higher gains than that of the asynchronous comparators. However, as illustrated in Fig. 9(b), pulse signals are generated in the receiver coil only during data transitions. Therefore, no signal is induced when a constant value is transmitted, resulting in an indeterminate output that is susceptible to noise.

A transmitter circuit topology to generate a pulse signal in the receiver coil in every cycle was proposed in [26], as shown in Fig. 10(a), which is called a pulse transmitter because the transmitter current is pulse shaped. The pulse generator generates a positive pulse on every rising edge of the clock. While the pulse is high, the input of one and only one of the two inverters go high because of the differential data inputs, allowing current to flow through the transmitter coil in a direction determined by the polarity of the transmitted data. Thus, a receive pulse signal is generated in every cycle. However, this method has two drawbacks. First, the pulse generator circuit requires a delay circuit, and the delay must be optimized. Second, the change in the current flowing through the transmitter coil is half that of the conventional CMOS H-bridge transmitter, so the received pulse signal is halved as well. Therefore, in this article, to solve the delay adjustment and pulse halving issues of the conventional pulse transmitter, we propose a new transmitter utilizing Manchester encoding to generate a bit transition every cycle on a synchronous basis.

Fig. 11 shows the block diagram and waveforms of the synchronous transceiver topology using Manchester encoding and a clocked comparator. The conventional CMOS H-bridge transmitter is used to drive the coil, and a Manchester encoding circuit is used in the preceding stage. Manchester-encoded data always generate a data transition in every cycle. Therefore, a pulse signal is induced in the receiver coil in every cycle, enabling detection by the clocked comparator. The clocked comparator is used to detect small-amplitude pulse signals utilizing positive feedback, allowing communication using the over-SRAM coils, whose signals are still attenuated although to a lesser extent. However, the smaller the input differential
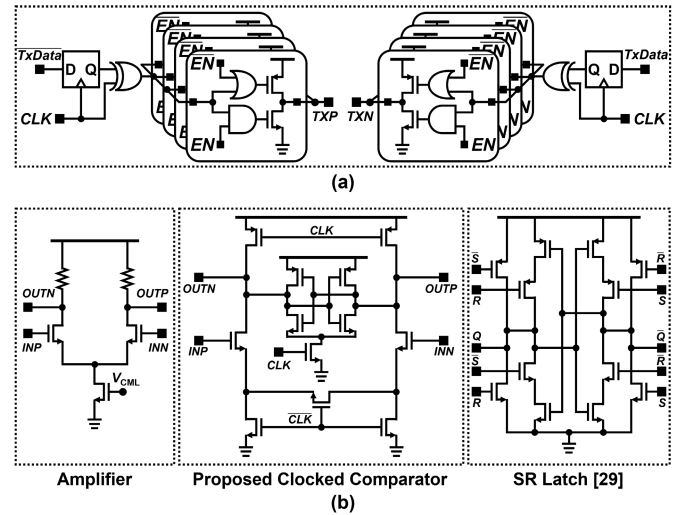
signal is, the logarithmically longer the time between the rise of the clock and the determination of the comparator output, limiting the operating frequency. In other words, data rate is traded off for receiver sensitivity in detecting small pulse signals. As a countermeasure, this work used two clocked comparators in parallel to halve the required operating speed of each clocked comparator, thereby simultaneously achieving low-amplitude pulse detection and high-speed communication.

Fig. 12(a) shows the schematic of the transceiver. The Manchester encoding circuit of the transmitter consists of a simple D-FF and XOR circuit, leading to a small area and low power consumption. In addition, the driver in the final stage can control the amount of current flowing through the coil with binary-weighted enable signals. Thus, the transmitter circuit is easy to design because it consists of only standard cells. In the receiver, a current-mode logic (CML) buffer is used in the first stage to amplify the pulse signal and reduce kickback noise from the clocked comparators. Even though the clocked comparators are parallelized, the basic clocked
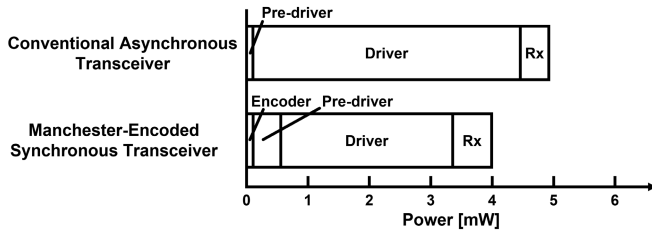
Fig. 13.   Simulation comparison between conventional asynchronous and proposed synchronous transceivers.



Fig. 14.   (a) Microphotograph of test chip, (b) transmitter layout, and (a) receiver layout.

comparator [27] cannot support high-speed operation. A high-speed clocked comparator is proposed in [28], but it has the drawback that its output cannot achieve full swing. Therefore, a modification is made to separate the two legs of the tail current of the high-speed clocked comparator to enable full-swing output. For the SR latch, we used [29], which is capable of high-speed operation and has good duty cycle.

Fig. 13 shows a simulated comparison between the conventional asynchronous transceiver circuit and the proposed synchronous transceiver circuit. When the conventional transceiver is used to communicate using the over-SRAM coil, the driver current must be increased to compensate for the attenuation of the magnetic field, which results in increased power consumption. On the other hand, the proposed synchronous transceiver topology can detect low-amplitude pulses, thus keeping the driver power consumption low. Furthermore, the Manchester encoding circuit consists of only D-FF and XOR, so the power consumption required for encoding is only 0.1 mW.

## IV. EXPERIMENTAL RESULTS AND COMPARISONS

### A. Experimental Results

Fig. 14 shows the test chip fabricated in a 7-nm FinFET process and layout of the transmitter and receiver blocks. The sizes of the transmitter and receiver are $12 \times 11$ and $13 \times 33$ $\mu$m, respectively. As shown in Fig. 15, a test module consisting of two chips connected face-to-face was used for evaluation. Power supply and control signals are provided to the bottom chip via bonded wires and to the top chip via bumps between the chips.

Fig. 16 shows a block diagram of the test module measurement environment. An 8.5-GHz clock, required to verify 8.5-Gb/s wireless communication, was generated by an internal RO due to the difficulty of bringing in an external high-speed clock. The bit error rate (BER) bathtub curve was measured using an internal PRBS9 signal generator, checker, and variable delay (VD) circuits. The tunable parameters of the TCI transceiver circuit, RO, and VDs were set by writing to internal registers using a serial peripheral interface (SPI). The actual oscillation frequency of the RO and the delay time of the VDs were measured by using an external oscilloscope.

Fig. 17(a) shows the measured energy consumption breakdown of 8.5-Gb/s wireless communication using the clocked comparator, Manchester encoding, and over-SRAM coil. The clocked comparator, which is capable of detecting
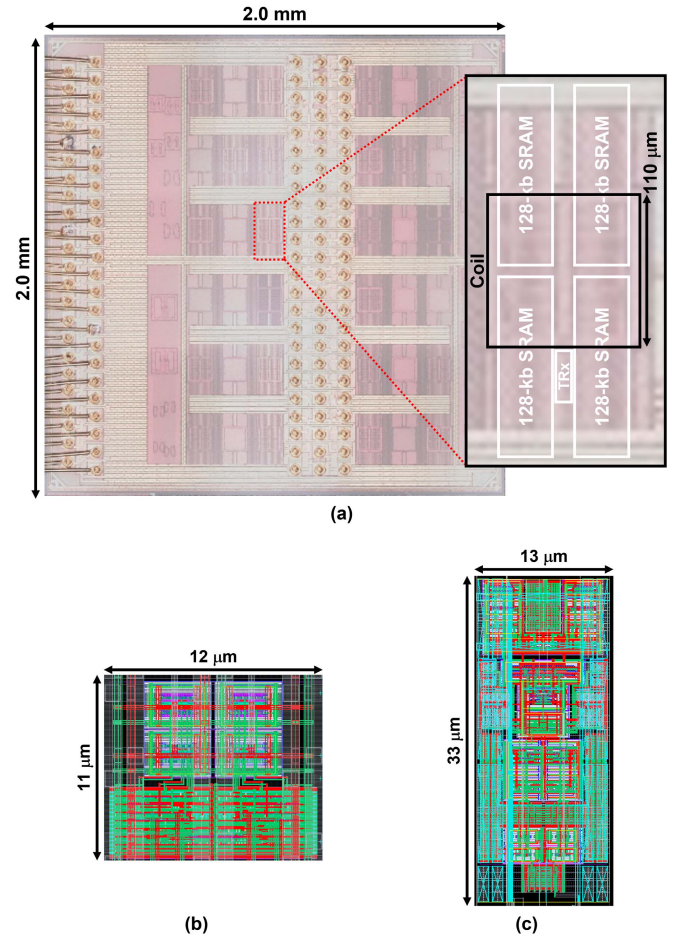


Fig. 15.   (a) Top- and (b) side-view photographs of test module.
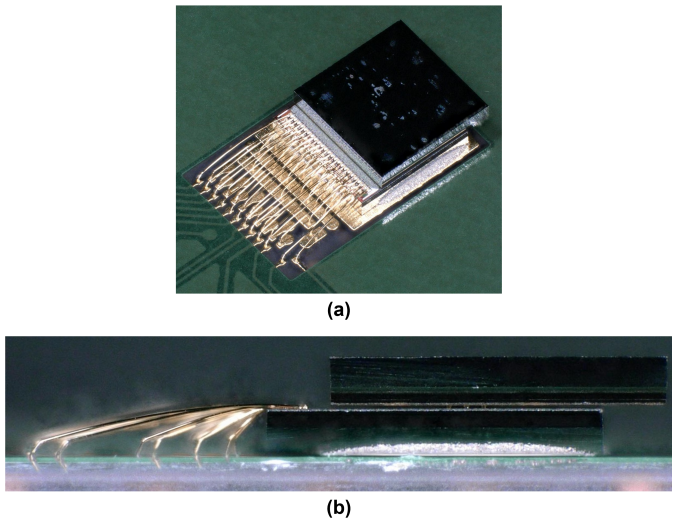
low-amplitude pulses, allows the driver power to be reduced to achieve 0.7-pJ/b wireless communication.

Fig. 17(b) shows the measured BER results using 8.5-Gb/s PRBS9. The burst length is independent of the PRBS length
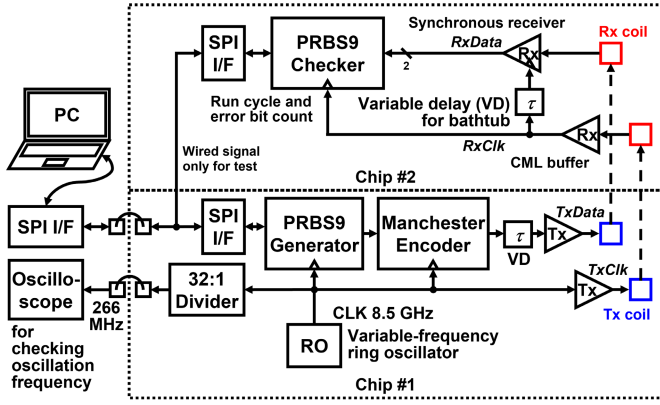
Fig. 16.   Block diagram of experimental setup.

since this work utilizes Manchester encoding, where a pulse signal is always induced in every cycle. Therefore, considering the operating speed and circuit area, PRBS9 was adopted for the test. When the SRAMs under the coils are active, there is concern that noise generated in the receiver coil from the SRAMs may result in bit errors. However, it can be seen that the bathtub curves are almost the same whether the SRAMs are operating or not. Conversely, there is also concern that noise generated on the internal nodes of the SRAMs from the transmitter coils may adversely affect the operation of the SRAMs. However, it was reported in [30] that the minimum operating voltage of the SRAM remains almost unchanged even during TCI operation, so TCI operation does not affect SRAM operation either.

The weakness of synchronous reception using the clocked comparator is that it is susceptible to clock jitter and requires fine phase adjustment. In the evaluation environment of this work, clock jitter is generated because an on-chip current-controlled delay adjustment circuit is used to measure the bathtub curve, but even in this measurement environment with clock jitter, a BER of $10^{-12}$ was achieved with a timing margin of 11 ps (0.09 UI). Since the VD circuit can be adjusted in 1.1-ps increments, a BER of $10^{-12}$ was achieved at a setting of ten steps, indicating that the communication is sufficiently tolerant of process, voltage, and temperature (PVT) variations. To further confirm that the communication is reliable, the bathtub was extrapolated to a BER of $10^{-16}$. Since bit errors occur because of variations in the timing of triggering of the pulse due to jitter, the bathtub curve can be approximated by an error function [31]. Based on the extrapolated results, a BER of $10^{-16}$ is achievable with a timing margin of 3.5 ps. A BER of $10^{-16}$ can therefore be achieved at a setting of four steps of the VD circuit.

Fig. 17(c) shows a measured shmoo graph of data rate versus driver peak current along with the supply voltage information since the received pulse amplitude, proportional to the driver peak current, is an important factor for correct communication and data rate of TCI. Since changing the driver current by sweeping the supply voltage would result in inaccurate measurement due to changes in factors other than the pulse amplitude, such as the RO and VD circuits, the driver current was adjusted by using the driver current adjustment circuit.
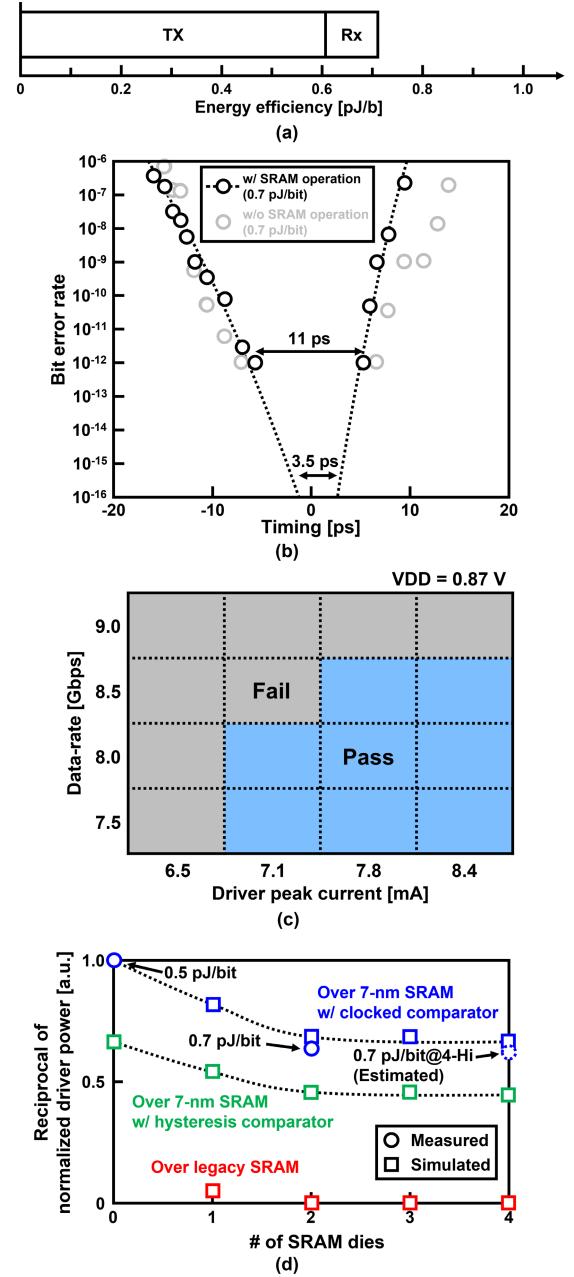


Fig. 17.   Measured results of (a) power consumption, (b) bathtub curve, (c) shmoo, and (d) effect of sandwiched SRAMs.

Sweeping the driver peak current at the driver supply voltage of 0.87-V, 8.5-Gb/s wireless communication was measured at 7.8 mA.

Fig. 17(d) shows the relationship between the number of stacked SRAM chips and the reciprocal of the required normalized transmission driver power. The required driver power is proportional to the amount of magnetic field attenuation and the minimum detectable pulse amplitude of the receiver. Zero SRAM die refers to the case of conventional placement of coils away from the macros. Due to the new physical layout of the over-SRAM coil, a significant improvement in area efficiency is achieved, while the magnetic field attenuation is kept to 30% even when four SRAM chips are stacked. According to

TABLE I

PERFORMANCE COMPARISONS

| | MICRO'17 [8] | ISSCC'20 [9] | TCAS-I'21 [10] | Hot Chips'20 [6] | Hot Chips'20 [6] (Extrapolated to 4 Hi) | This work | This work |
|---|---|---|---|---|---|---|---|
| Results | Measured | Measured | Measured | Measured | Measured | Measured | Estimated |
| Technology | 20-nm DRAM | 1y-nm DRAM | 40-nm CMOS | 7-nm FinFET | 7-nm FinFET | 7-nm FinFET | 7-nm FinFET |
| Memory type | HBM2 DRAM | HBM2E DRAM | SRAM | SRAM | SRAM | SRAM | SRAM |
| Data bus | Bi-directional | Bi-directional | Uni-directional | Uni-directional | Uni-directional | Uni-directional | Uni-directional |
| Stack # | 8 | 12 | 8 | 1 | 4 | 2 | 4 |
| Bandwidth | 256 GB/s | 640 GB/s | 28.8 GB/s | 24.3 GB/s | 24.3 GB/s | 4.3 GB/s | 4.3 GB/s |
| μ-bump pitch | 48 / 55 μm | 48 / 55 μm | - | 40 μm | 40 μm | - | - |
| IO area overhead (*1) | 2.8 mm² | 2.8 mm² | 11.52 mm² | 0.92 mm² | 0.92 mm² | 0.0037 mm² | 0.0037 mm² |
| Bandwidth per IO area overhead | 92 GB/s/mm² | 231 GB/s/mm² | 2.5 GB/s/mm² | 26 GB/s/mm² | 26 GB/s/mm² | 1162 GB/s/mm² | 1162 GB/s/mm² |
| Data-rate | 2.0 Gb/s | 5.0 Gb/s | 3.6 Gb/s | 0.76 Gb/s | 0.76 Gb/s | 8.5 Gb/s | 8.5 Gb/s |
| I/O energy consumption | ~ 2 pJ/bit | N/A(~2.5 pJ/bit(*2)) | 1.5pJ/bit | 0.1 pJ/bit | 0.4 pJ/bit (*3) | 0.7 pJ/bit | 0.7 pJ/bit |
| Interface type | TSV + μ-bump | TSV + μ-bump | TCI | TSV + μ-bump | TSV + μ-bump | TCI | TCI |
| Chip size | 12 mm × 8 mm | 11 mm × 10 mm | 14.3 mm × 8.5 mm | 9.0 mm × 9.0 mm | - | 2.0 mm × 2.0 mm | 2.0 mm × 2.0 mm |

*1: IO area only for signal excluding power
*2: Estimated from ratio of the squared voltage and stack # of HBM2 (1.2 V, 8 Hi, [8]) and HBM2E (1.1 V, 12 Hi, [9])
*3: Capacitance load of 4 × # of Rx's, μ-bumps and TSVs driven by Tx compared with 1 Hi

simulation, the magnetic field attenuation for a 4-hi stack is almost equivalent to that for a 2-hi stack. Therefore, an energy efficiency of 0.7 pJ/b is expected to be achieved for a 4-hi stack, the same as that measured using a 2-hi stack. Furthermore, the synchronous transceiver topology using Manchester encoding and the clocked comparator enables low-amplitude pulse detection, which requires less driver power than conventional asynchronous communication using the hysteresis comparator. Thus, the new physical placement technique and the new synchronous transceiver topology enable high area efficiency and low-power wireless communication.

### B. Performance Comparison

A comparison with previous works is shown in Table I. As discussed in Fig. 17(d), a 4-hi 3D-SRAM is estimated to achieve the same specification as the measured 2-hi 3D-SRAM. This article improves the IO area efficiency by one to two orders of magnitude compared to previous works [6], [8], [9], where TSVs and μ-bumps have a large area overhead. Compared to the estimated four-stack version of [6] fabricated in the same 7-nm FinFET process, the proposed 3D-stacked SRAM using inductive coupling improves the area efficiency by two orders of magnitude. Our 4-hi 3D-SRAM is estimated to achieve 0.7 pJ/b, an energy efficiency competitive with the 0.4 pJ/b estimated for a TSV-based 4-hi 3D-SRAM. The energy consumption of the TSV-based 3D-SRAM is proportional to the number of chips in the stack because it depends on the load capacitance of the TSVs and μ-bumps. Fig. 18 shows a graph plotting energy efficiency versus IO area efficiency. The proposed inductive coupling interface utilizing the over-SRAM coil and synchronous transceiver achieves an overwhelmingly higher area efficiency while showing a competitive energy efficiency. As illustrated in Fig. 18, there is a tradeoff between area efficiency and energy efficiency for the
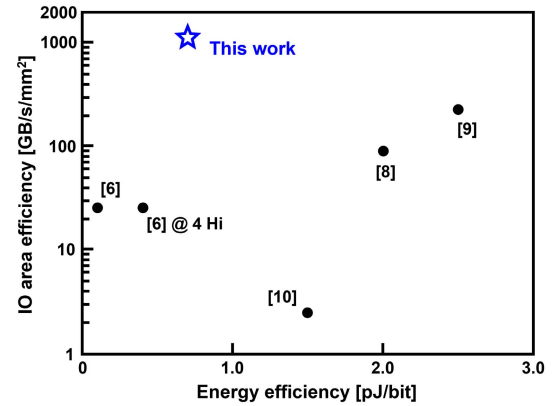


Fig. 18. Performance comparisons.

wireless and wired interfaces. The chip cost of our 3D-SRAM is lower due to its high area efficiency. Furthermore, the integration cost is also lower due to the wireless nature of the interface. Therefore, while the TSV-based 3D-SRAM is suitable for low-power applications, the proposed 3D-SRAM is suitable for low-cost applications.

### C. Scaling

This section discusses scaling scenarios for the 3D-SRAM module as summarized in Table II. The proposed 3D-SRAM module achieves 1 MB of SRAM capacity with a channel area of 0.43 × 1.2 mm. Assuming a 12-mm² SRAM die, 20 channels can be placed, assuming an area overhead of about 10% in addition to the channels. In other words, the proposed 3D-SRAM module achieves a total SRAM capacity of 80 MB by stacking four 12-mm² SRAM chips. The International Roadmap for Devices and Systems (IRDS) estimated an area scaling factor of 1/4 from a 7- to 1.5-nm

TABLE II

SCALING SCENARIO

|  | 7 nm | 1.5 nm |
|---|---|---|
| # of channel | 20 | 40 |
| Stack # | 4 | 4 |
| Capacity per channel | 1 MB | 2 MB |
| Bandwidth per vault | 7.5 GB/s | 30 GB/s |
| Total capacity | 80 MB | 320 MB |
| Total bandwidth | 150 GB/s | 1200 GB/s |
| Chip size | 12 mm$^2$ | 12 mm$^2$ |

SRAM [32]. Therefore, if the 3D-SRAM module is realized in an upcoming 1.5-nm process, it can achieve a total SRAM capacity of 320 MB with the same die area.

Regarding bandwidth, while an 8.5-Gb/s wireless communication data rate was measured in the evaluation by using the test module, a 15-Gb/s maximum data rate was observed in the postlayout simulation with parasitic extraction. If a 15-Gb/s data rate is realized, a bandwidth of 7.5 GB/s/vault is achievable. Thus, the 3D-SRAM module in 7 nm with a die area of 12 mm$^2$ is expected to achieve 150 GB/s. According to the TCI scaling scenario discussed in [33], when the transistor speed is improved by a factor of $\alpha$ following process scaling, $\alpha$ times faster wireless communication can be achieved by designing the self-resonance frequency of the coils to be $\alpha$ times higher. Therefore, assuming that transistors in the 1.5-nm process operate twice as fast as those in the 7-nm process, a 30-Gb/s data rate and 15-GB/s/vault bandwidth can be expected for the 1.5-nm process. Furthermore, if the SRAM macros become smaller in the 1.5-nm process, the constraints on coil placement can be further relaxed to allow the coils to be arranged in a more area-efficient checkerboard manner to double the number of channels to 40 [25]. In addition, although the read and write data paths are separated in this work to simplify system verification, it is expected that 30 GB/s/vault can be achieved by making the data paths bidirectional. In other words, a total bandwidth of 1200 GB/s can possibly be achieved when the 4-hi 3D-SRAM module is realized in the 1.5-nm process.

Thus, the proposed 3D-stacked SRAM module using inductive coupling can improve in both SRAM capacity and bandwidth following process scaling and is expected to achieve a 320-MB capacity and 1200-GB/s bandwidth by stacking four 12-mm$^2$ SRAM chips in the 1.5-nm process, which is expected to be available by 2030. Such capacity and bandwidth meet the requirements of CPUs used in future high-performance computing and are expected to improve system performance by a factor of approximately 10 [34]. Thus, in addition to being currently two logic generations ahead of TSV-based stacking, the proposed 3D-stacked SRAM module is expected to scale well into the future and meet future requirements.

## V. CONCLUSION

A 0.7-pJ/bit, 8.5-Gb/s/link inductive coupling interchip wireless communication interface is proposed. A new physical placement method allows coils to be placed over off-the-shelf SRAM macros with only 30% magnetic field attenuation. Synchronous wireless communication using Manchester encoding and a clocked comparator enables the detection of small-swing signals to achieve low-power operation. The proposed 3D-SRAM using inductive coupling achieves a 26% reduction in SRAM die area compared to TSV-based stacking. The test chip was fabricated in a 7-nm FinFET process and inter-chip communication at 0.7-pJ/bit, 8.5-Gb/s/link was confirmed using a test module consisting of two chips. A 4-hi 3D-stacked SRAM module using the proposed interface will achieve a 1.2-TB/s/mm$^2$ area efficiency and 0.7-pJ/bit energy efficiency, representing a two orders-of-magnitude improvement in area efficiency over the state-of-the-art 3D-stacked SRAM.

## REFERENCES

[1] D. Kim, J. Kung, S. Chai, S. Yalamanchili, and S. Mukhopadhyay, "Neurocube: A programmable digital neuromorphic architecture with high-density 3D memory," in *Proc. IEEE Int. Symp. Comput. Archit. (ISCA)*, Jun. 2016, pp. 380–392.

[2] M. Gao, J. Pu, X. Yang, M. Horowitz, and C. Kozyrakis, "TETRIS: Scalable and efficient neural network acceleration with 3D memory," *SIGARCH Comput. Archit. News*, vol. 45, no. 1, pp. 751–764, Apr. 2017.

[3] M. Evers, L. Barnes, and M. Clark, "The AMD next-generation 'Zen 3' core," *IEEE Micro*, vol. 42, no. 3, pp. 7–12, Aug. 2021.

[4] K. Ueyoshi et al., "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.

[5] K. Shiba, M. Okada, A. Kosuge, M. Hamada, and T. Kuroda, "Poly-omino: A 3D-SRAM-centric architecture for randomly pruned matrix multiplication with simple rearrangement algorithm and X0.37 compression format," in *Proc. 20th IEEE Interregional NEWCAS Conf. (NEWCAS)*, Jun. 2022, pp. 99–103.

[6] K. Cho et al., "SAINT-S: 3D SRAM Stacking Solution based on 7 nm TSV technology," in *Proc. IEEE Hot Chips 32nd Symp. (HCS)*, Aug. 2020, pp. 1–13.

[7] 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," in *Proc. IEEE 71st Electron. Compon. Technol. Conf. (ECTC)*, Jun. 2021, pp. 1115–1118.

[8] M. O'Connor et al., "Fine-grained DRAM: Energy-efficient DRAM for extreme bandwidth systems," in *Proc. 50th Annual IEEE/ACM Int. Symp. Microarchitecture*, Oct. 2017, pp. 41–54.

[9] C.-S. Oh et al., "22.1 A 1.1 V 16GB 640GB/s HBM2E DRAM with a data-bus window-extension technique and a synergetic on-die ECC scheme," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2020, pp. 330–332.

[10] K. Shiba et al., "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 Systems I, Reg. Papers*, vol. 68, no. 2, pp. 692–703, Feb. 2021.

[11] D. Ditzel, T. Kuroda, and S. Lee, "Low-cost 3D chip stacking with ThruChip wireless connections," in *Proc. IEEE Hot Chips Symp. (HCS)*, Aug. 2014, pp. 1–37.

[12] T. Kuroda, "Circuit and device interactions for 3D integration using inductive coupling," in *IEDM Tech. Dig.*, Dec. 2014, pp. 18.

[13] K. Niitsu et al., "An inductive-coupling link for 3D integration of a 90 nm CMOS processor and a 65 nm CMOS SRAM," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2009, pp. 480–481.

[14] K. Shiba, M. Hamada, and T. Kuroda, "3D system-on-a-chip design with through-silicon-via-less power supply using highly doped silicon via," *Jpn. J. Appl. Phys.*, vol. 59, no. SG, Apr. 2020, Art. no. SGGL04.

[15] K. Shiba, M. Okada, A. Kosuge, M. Hamada, and T. Kuroda, "A 7-nm FinFET 1.2-TB/s/mm² 3D-stacked SRAM with an inductive coupling interface using over-SRAM coils and manchester-encoded synchronous transceivers," in *Proc. IEEE Hot Chips 34 Symp. (HCS)*, Aug. 2022, pp. 1–14.

[16] Y. S. Kim et al., "Ultra thinning down to 4-$\mu$m using 300-mm wafer proven by 40-nm node 2Gb DRAM for 3D multi-stack WOW applications," in *Proc. Symp. VLSI Technol. (VLSI-Technology)*, 2014, pp. 1–2.

[17] L.-C. Hsu et al., "Analytical ThruChip inductive coupling channel design optimization," in *Proc. 21st Asia South Pacific Design Automat. Conf. (ASP-DAC)*, 2016, pp. 731–736.

[18] R. Chen et al., "3D-optimized SRAM macro design and application to memory-on-logic 3D-IC at advanced nodes," in *IEDM Tech. Dig.*, Dec. 2020, p. 15.

[19] X. Jiang et al., "A 1596-GB/s 48-Gb stacked embedded DRAM 384-core SoC with hybrid bonding integration," *IEEE Solid-State Circuits Lett.*, vol. 5, pp. 110–113, 2022.

[20] J. Chang et al., "12.1 A 7 nm 256Mb SRAM in high-K metal-gate FinFET technology with write-assist circuitry for low-VMIN applications," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2017, pp. 206–207.

[21] J. Chang et al., "15.1 A 5 nm 135Mb SRAM in EUV and high-mobility-channel FinFET technology with metal coupling and charge-sharing write-assist circuitry schemes for high-density and low-VMIN applications," in *IEEE Int. Solid-State Circuits Conf. (ISSCC)*, Feb. 2020, pp. 238–240.

[22] T. Song et al., "24.3 A 3 nm gate-all-around SRAM featuring an adaptive dual-BL and an adaptive cell-power assist circuit," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2021, pp. 338–340.

[23] N. Miura, Y. Kohama, Y. Sugimori, H. Ishikuro, T. Sakurai, and T. Kuroda, "A high-speed inductive-coupling link with burst transmission," *IEEE J. Solid-State Circuits*, vol. 44, no. 3, pp. 947–955, Mar. 2009.

[24] 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," *Trans. Fundam. Electron., Commun. Comput. Sci.*, vols. E98–A, no. 12, pp. 2584–2591, Dec. 2015.

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

[26] K. Niitsu et al., "Analysis and techniques for mitigating interference from power/signal lines and to SRAM circuits in CMOS inductive-coupling link for low-power 3-D system integration," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 19, no. 10, pp. 1902–1907, Oct. 2011.

[27] D. Schinkel, E. Mensink, E. Klumperink, E. van Tuijl, and B. Nauta, "A double-tail latch-type voltage sense amplifier with 18ps setup+hold time," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2007, pp. 314–605.

[28] M. Abbas, Y. Furukawa, S. Komatsu, J. Y. Takahiro, and K. Asada, "Clocked comparator for high-speed applications in 65 nm technology," in *Proc. IEEE Asian Solid-State Circuits Conf.*, Nov. 2010, pp. 1–4.

[29] B. Nikolic, V. G. Oklobdzija, V. Stojanovic, W. Jia, J. K.-S. Chiu, and M. M.-T. Leung, "Improved sense-amplifier-based flip-flop: Design and measurements," *IEEE J. Solid-State Circuits*, vol. 35, no. 6, pp. 876–884, Jun. 2000.

[30] K. Niitsu et al., "Interference from power/signal lines and to SRAM circuits in 65 nm CMOS inductive-coupling link," in *Proc. IEEE Asian Solid-State Circuits Conf.*, Nov. 2007, pp. 131–134.

[31] N. Miura et al., "A 1Tb/s 3W inductive-coupling transceiver for 3D-stacked inter-chip clock and data link," *IEEE J. Solid-State Circuits (JSSC)*, vol. 42, no. 1, pp. 111–122, Jan. 2007.

[32] *International Roadmap for Devices and Systems 2017 Edition*. Accessed: Oct. 14, 2020. [Online]. Available: https://irds.ieee.org/images/files/pdf/2017/2017IRDS_MM.pdf

[33] T. Kuroda, "Near-field coupling integration technology," *ESC Trans.*, vol. 72, no. 3, pp. 83–91, May 2016.

[34] J. Domke et al., "At the locus of performance: A case study in enhancing CPUs with copious 3D-stacked cache," 2022, *arXiv:2204.02235*.

**Kota Shiba** (Graduate Student Member, IEEE) was born in Tokyo, Japan, in 1995. He received the B.S. and M.S. degrees in electronics and electrical engineering from Keio University, Yokohama, Japan, in 2018 and 2020, respectively. He is currently pursuing the Ph.D. degree with the Graduate School of Engineering, The University of Tokyo, Tokyo.

He is a JSPS Research Fellow (DC2) and a JST ACT-X Researcher. His research interests include inductive coupling wireless communication, high-speed I/O interface, 3D system integration, and low-power static-random access memory (SRAM).

Mr. Shiba was a recipient of the IEEJ Tokyo Branch Student Encouragement Award in 2018 and a co-recipient of the IEEE ICECS Best Student Paper Award in 2020.

**Mitsuji Okada** (Member, IEEE) received the master's degree in the management of science and technology (MOST) from the Tokyo University of Science, Tokyo, Japan, in 2007.

He joined NEC Corporation, in 1983, where he was engaged in the application engineering of Microprocessors. From 1993 to 1998, he was involved in the standardization of IEEE 802.3 and 802.11 as a Voting Member. From 2006 to 2011, he led the device development of an ultralow-power wireless communication systems at Renesas Electronics Corporation, Japan. From 2011 to 2015, he was consulted the research and development of an ultralow power system large-scale integrations (LSIs). In 2019, he joined The University of Tokyo, where he is currently a Project Academic Specialist of the Systems Design Laboratory. He is currently a Researcher of the Research Association for Advanced Systems (RaaS). His research interests include reconfigurable processing architecture, ultralow-power subthreshold, and asynchronous circuit design.

**Atsutake Kosuge** (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Keio University, Yokohama, Japan, in 2012, 2014, and 2016, respectively.

From 2014 to 2017, he was a JSPS Research Fellow with Keio University. From 2017 to 2020, he held research positions at Hitachi Ltd. and Sony Corporation. In 2021, he joined The University of Tokyo, Tokyo, Japan. He is currently an Assistant Professor of Systems Design Laboratory (d.lab) and a Researcher of the Research Association for Advanced Systems (RaaS). His research interests include energy efficient computing, computational sensing, and 3D integration technologies.

Dr. Kosuge was a recipient of the 2013 Nikkei Electronics Japan Wireless Technology Best Award and 2020 IEICE Young Researcher's Award and a co-recipient of the ASP-DAC'15 Special Feature Award. He has served as a Technical Program Committee member for the IEEE A-SSCC (Asian Solid-State Circuits Conference) since 2021, IEICE ICD (Integrated Circuit and Devices), and served as an Organizing Committee Member for the IEEE COOL Chips (Symposium on Low-Power and High-Speed Chips and Systems).

**Mototsugu Hamada** (Member, IEEE) was born in Nara, Japan, in 1968. He received the B.S., M.S., and Ph.D. degrees in electronic engineering from The University of Tokyo, Tokyo, Japan, in 1991, 1993, and 1996, respectively.

In 1996, he joined Toshiba Corporation, where he was engaged in wireless and low-power electronic circuits design with the Toshiba's Center for Semiconductor Research and Development, Kawasaki, Japan. From 2002 to 2004, he was a Visiting Scholar with Stanford University, Stanford, CA, USA. From 2011 to 2016, he was with the Mixed Signal IC Division as a Group Manager of the Power Analog IC Design Group to lead the development of analog mixed signal ICs. In 2016, he joined Keio University and was a Project Professor. In 2020, he joined The University of Tokyo, where he is currently a Project Professor of Systems Design Laboratory. He is a Senior Fellow of the Research Association for Advanced Systems (RaaS). His research interests include low-power, high-speed CMOS design, low-power wireless systems and circuits design, and power management systems design.

Dr. Hamada was a recipient of the 2007 IEEE International Conference on Computer Design (ICCD) Best Paper Award and the Design Automation Conference (DAC) 2010 Best User Track Poster Award. He was also recognized in the list of "authors of ten or more articles in the past ten years" at the International Solid-State Circuits Conference 2013 (ISSCC2013). He has served as a Technical Program Committee Member for the International Solid-State Circuits Conference (2003–2009 and 2011) and VLSI Circuits Symposium (2018–2023) and Asian Solid-State Circuits Conference (2005–2012 and 2017–2022), where he served as the RF Subcommittee Chair, Digital Subcommittee Chair, Student Design Contest Chair, and Technical Program Committee Chair.

**Tadahiro Kuroda** (Fellow, IEEE) received the Ph.D. degree in electrical engineering from The University of Tokyo, Tokyo, Japan, in 1999.

In 1982, he joined Toshiba Corporation. From 1988 to 1990, he was a Visiting Scholar with the University of California at Berkeley, Berkeley, CA, USA, where he conducted research in the field of very large scale integration (VLSI) CAD. In 1990, he joined Toshiba, where he was engaged in the research and development of BiCMOS application-specified integrated circuits (ASICs), emitter-coupled logic (ECL) ASICs, and high-speed low-power CMOS LSIs. He invented a variable threshold-voltage CMOS (VTCMOS) technology to control $V_{TH}$ through substrate bias and applied it to a discrete cosine transform (DCT) core processor in 1995. He also developed a variable supply-voltage scheme to control $V_{DD}$ by an embedded dc–dc converter and employed it to a microprocessor core and an MPEG-4 chip in 1997. He left Toshiba to join Keio University in 2000 and became a full professor in 2002. He was the Mackay Professor with the University of California at Berkeley, in 2007. He invented a ThruChip Interface (TCI) by using magnetic coupling for communications among stacked chips in 2008 and a transmission line coupler (TLC) by using electromagnetic coupling for communications among stacked PCBs in 2012. He left Keio to join The University of Tokyo in 2019. He is the Director of the Systems Design Laboratory (d.lab), The University of Tokyo, and the Chairperson of Research Association for Advanced Systems (RaaS). He has published more than 450 articles, including 38 ISSCC articles, 29 VLSI Symposia articles, 19 CICC papers, and 18 A-SSCC papers. He wrote 30 books/chapters and filed more than 200 patents.

Dr. Kuroda is a fellow of IEICE and the Chair of Symposium on VLSI Technology and Circuits. He was an elected AdCom Member of two terms. He was a recipient of the 2005 P&I Patent of the Year Award, the 2007 ASP-DAC Best Design Award, the 2009 IEICE Achievement Award, and the 2011 IEICE Society Award. He served as a Steering Committee Chair for A-SSCC, a Vice Chair for ASP-DAC, subcommittee chairs for A-SSCC, ICCAD, SSDM, and VLSI-DAT, and TPC members for ISSCC, Symposium on VLSI Circuits, CICC, DAC, ASPDAC, ISLPED, SSDM, ISQED, and other international conferences. He was a Distinguished Lecturer and a representative of Region 10 for the IEEE Solid-State Circuits Society.