High Speed Data Link for 3DICs

AUNG MYAT THU LINN
SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING
2016
Acknowledgement

I would first like to express my deepest gratitude to my advisor, Professor Tony T. Kim. I am so grateful to him for his endless patience, encouragement and support for the last four years. Thanks for his patient and taking time to scrutinize every single sentence of my manuscripts, I have come to know his writing style and his way of thinking in electronics design and thus, it helps me improved my writing and design idea.

Of course this long journal towards PhD won't be possible without scholarship support from Panasonic Industrial Device, Asia. From there, I come to know Eric Teck Heng Lim, Takefumi Yoshikawa, Wee Sien Hong and Tee Peng Koh whose invaluable contribution, design advice and their at most support throughout my PhD journey could never be taken for granted.

My life in the lab wouldn’t be better than having pet talk once in a while with Prof Siek Liter. His attention and care to all the students in the lab is splendid. My thank also goes to Ms. Gek Eng, who always helps me finishing the lengthy formalities and paper works without a hesitation. I also thanks Mr. David Robert, for helping me to resolve all the software and technical relative issues. There is one person who always comes to my mind whenever Cadence has an issue is Mr. Sia Liang Poo. Without his prompt response, I would have missed the tape-out deadline. Last but not least, I really appreciate all of my teammate, colleagues from lunch group for being part of my life and for their contribution which makes my PhD journey a little more exciting and eventfully. Lastly, I will not forget the unconditional support from my girlfriend, family and friends.
# Table of Contents

Abstract .................................................................................................................................vii

Acknowledgement ....................................................................................................................ii

List of Figures .......................................................................................................................... x

List of Tables ............................................................................................................................ xviii

Chapter 1  Introduction .............................................................................................................. 1

1.1 High Speed Interconnects for 3D ICs Utilizing Inductive Coupling ......................... 4
   1.1.1 Transceiver design and its signalling scheme ......................................................... 4
   1.1.2 Inductor Design and Implementation ................................................................. 9
   1.1.3 Summary ............................................................................................................. 15

1.2 High Speed Interconnects for 3D ICs Utilizing Capacitive Coupling ...................... 17
   1.2.1 Modeling of Capacitive Coupling Interconnect ................................................. 17
   1.2.2 Alignment .......................................................................................................... 18
   1.2.3 Crosstalk .......................................................................................................... 21
   1.2.4 Transceiver Designs .......................................................................................... 24
   1.2.5 Summary .......................................................................................................... 29

1.3 A Dual Mode Transceiver for 3DICs in Face-to-Face Interface ......................... 29

1.4 High Speed Interconnect for Silicon Interposer-based 2.5D ICs ............................ 31

1.5 Summary of Thesis Contribution .................................................................................. 33
Chapter 2  A 3Gbps/ch Simultaneous Bi-Directional Capacitive Coupling Transceiver for 3DICs

2.1 Motivation

2.2 Design of Simultaneous Bi-directional Transceiver

2.2.1 Proposed Capacitive Coupling Interconnect Design

2.2.2 Transceiver Design

2.3 Test Chip Implementation

2.4 Test Chip Measurement Result

2.5 Conclusion

Chapter 3  A 2.31 Gbps/ch Area Efficient Crosstalk Cancelled Hybrid Capacitive Coupling Interconnect for 3D Integration

3.1 Motivation

3.2 Capacitive Coupling Interconnect

3.2.1 The CCI Array

3.2.2 Proposed Hybrid CCIs Array

3.3 Transceiver Design

3.3.1 Transmitter Design

3.3.2 Receiver Design

3.3.3 Simulation

3.4 Measurement

3.5 Conclusion
Chapter 4  Yield Enhancement of Face-to-Face Cu-Cu bonding with Dual-mode Transceivers in 3DICs .................................................................71

4.1 Motivation ..................................................................................71
4.2 Introduction ...............................................................................71
4.3 Dual-Mode Transceiver as BIST/R for Face-to-Face Copper Thermal Fusion Bonding .................................................................74
4.4 Simulation ................................................................................78
4.5 Conclusion ...............................................................................85

Chapter 5 High Speed Low Swing Differential Driver with Post-Cursor RC-Modulated De-emphasis for 2.5D Silicon Interposer .................................86

5.1 Motivation ................................................................................86
5.2 Driver Design ..........................................................................88
5.3 Post Layout Simulation .............................................................90
5.4 Conclusion .............................................................................95

Chapter 6 Conclusion ......................................................................96

References ..................................................................................99
Abstract

The transistor scaling predicted by Moore's Law has driven the development of the semiconductor industries for many decades. As the lithography process approaches to its limit, research and development cost for a new technology node starts to increase as well as the cost of fabrication. As the incurred development time becomes longer, the rate of transistor scaling starts to decline. With 3D integration, there are more rooms to allocate transistors and thus, it improves the transistor counts critical to maintain the scaling trend. Moreover, due to the closed proximity of the circuit blocks in vertically integrated chips, the data bandwidth improves while power consumption and latency reduces. The testament of these benefits can be found in latest AMD's High-Bandwidth-Memory (HBM) achieving three times more bandwidth per watt. The introduction of HBM also solves the performance bottleneck in DRAM bandwidth which limits the GPU performance growth. In HBM, DRAM cores are stacked vertically and connected through through-silicon-via (TSV) and micro-bumps and it provides up to 3.5× bandwidth of the one DRAM chip in the very same footprint. Therefore, 3D integration is proved to be important not only to keep up with the Moore's law but also to catch up with the ever increasing bandwidth requirement from the advanced processor. The interposer technology also known as 2.5D integration provides high-speed channels between a processor and HBM within a short distance which could never achieve in the board level integration.

There are several issues to address to make this 3D and 2.5D integration possible. When more and more dies are integrated together, the total system yield reduces. Moreover, interconnects (especially TSV and direct bondings) reliability
remains a challenge. In 2.5D silicon interposer technology, the line width of high-speed linkages is very small causing high insertion loss to the signal. These are the issues discussed in this thesis.

First of all, alternative interconnects technologies known as proximity communications in vertically stacked dies, opposed to TSV and Cu-Cu direct bonding, are thoroughly discussed and compared in terms of their feasibility, performance and methods used in the inductive coupling interconnect (ICI) and the capacitive coupling interconnect (CCI). Based on the findings, two designs in capacitive coupling scheme are proposed in this thesis. In the chapter 2, the possibility of simultaneous bi-directional signaling with CCI is studied and the novel electrodes and transceiver designs are proposed. This scheme is possible by the proposed interconnect in three cascaded capacitors configuration which ultimately creates multi-level voltages to allow simultaneous bi-directional signaling. The noise margin of the channel is measured to be 200mV when tested in the pseudo-3D interconnect structure. The transceiver is able to transmit and received 1.5Gbps simultaneously achieving the effective data rate of 3Gbps. In the chapter 3, the crosstalk among capacitive coupling electrodes is addressed. The crosstalk issue is well-known in CCI scheme and previously, two designs had been reported in the literature to address the issue. In contrast to the previous designs, the new proposed hybrid-CCI array design achieves not only crosstalk cancellation but also improves interconnect density and energy per bit.

In the chapter 4, it is where the capacitive coupling communication meets with one of the 3D physical channels; Cu-Cu face-to-face thermal compression bonding. Here, the possibility of using the Cu-Cu bonded channel either as an ohmic
or a capacitive coupling is explored. A dual-mode transceiver architecture and design is proposed to switch the transmission mode from ohmic mode to capacitive coupling mode when the Cu-Cu thermal compression bonding failed. In this way, the yield of Cu-Cu bonding can be improved.

Lastly, in 2.5D high-speed data link, multi-taps de-emphasis or pre-emphasis driver is utilized to compensate for the frequency dependent high insertion loss. The conventional multi-taps designs use a lot of silicon area and consume huge power consumption for the emphasis coding. A compact size low power RC-modulated de-emphasis voltage mode driver which has 3-taps equivalent is introduced. Taps extension is done by charging the storage capacitor during the period of consecutive bits is transmitting and the de-emphasis strength is modulated by the resulting voltage in the capacitor itself. By doing so, the amount of hardware required for de-emphasis coding is minimized and the pre-driver switching activity is reduced as well as the dynamic power consumption. The designed de-emphasis strength resolution is up to 5.1dB. It is able to drive up to 10Gbps on a 2cm long micro-strip line on a silicon interposer. The driver, pre-driver and de-emphasis circuits consume 2.37mW at 10Gbps speed with full de-emphasis strength according to the post layout simulation.
List of Figures

Figure 1.1  (a) H-Bridge transmitter and latch based sense-amplifier and (b) single ended current-reuse transmitter arranged in slides to control the transmit power.......................................................... 5

Figure 1.2  Transceiver with pulse width control and precise pulse-shaping........ 6

Figure 1.3  (a) Block diagram of the transceiver for burst data transmission, (b) receiver front-end circuits. ................................................................. 7

Figure 1.4  (a) Explanation of rotary scheme, (b) block diagram of the rotary transceiver. ...................................................................................... 8

Figure 1.5  Illustration of cross-talk cancelled point in Space Division Multiple Access (SDMA), (b) double edged triggered data transmission in Time Division Multiple Access (TDMA).............................................. 10

Figure 1.6  (a) XY coil structure, (b) XY inductor structure with noise from orthogonal logic wire, (c) extended XY inductor to cancel crosstalk noise from (b) ................................................................. 11

Figure 1.7  (a) Spiral stacking of NAND flash memory chips, (b) terraced stacking of NAND flash memory chips. ......................................................... 13

Figure 1.8  The illustration of clover coil for interference cancellation in simultaneous power and data transmission. .................................................. 14

Figure 1.9  (a) Scattered plot of X/D vs. data rate, (b) scattered plot of X/D vs. energy consumption ................................................................. 15

Figure 1.10  The parasitic model of the capacitive coupling interconnect........... 18

Figure 1.11  A rectifier circuit to measure the coupling capacitance Cc.................. 19
Figure 1.12  (a) The interconnect layout of the Vernier alignment system, (b) The electronic alignment system with micro-electrodes to drive signal to RX electrode................................................................. 20

Figure 1.13  Crosstalk components between two interconnects. ............................. 21

Figure 1.14  Shielding of crosstalk components with ground wire between RX electrodes. ................................................................. 22

Figure 1.15  (a) The crosstalk cancellation in butterfly differential interconnects array, (b) the crosstalk cancellation in the hybrid array. ...................... 23

Figure 1.16  (a) the illustration of the typical single-ended receiver for the capacitive coupling interconnect, (b) the circuit in (a) configured as differential transceiver................................................................. 24

Figure 1.17  A bi-stable topology for single-ended receiver with bi-directional capabilities. ................................................................. 26

Figure 1.18  (a) Three cascaded capacitors interconnect configuration for simultaneous bi-directional signaling; (b) four levels signaling at the interconnect (a); (c) The transceiver block diagram for simultaneous bi-directional signaling................................................................. 27

Figure 1.19  (a) The area efficiency vs. technology node; (b) the power efficiency vs. technology node................................................................. 28

Figure 1.20  The cross-section of 2.5D integration using silicon based interposer.... 31

Figure 1.21  (a) 8-bits bus architecture with 2 redundant channels. ......................... 32

Figure 2.1  Cross-section of the face-to-face integration with capacitive coupled transceiver................................................................. 35
Figure 2.2  (a) Schematic model of the proposed simultaneous bi-directional interconnect; (b) equations for signal swings ($V_{SW}$) and dead-zone ($V_{DZ}$); (c) graphical demonstration of 4-level signaling of the proposed interconnect. .................................................................................................................. 35

Figure 2.3  (a) Signal Swings ($V_{SW}$) of the design (a) and (b) in cross-section are shown where M8 is the top metal. Design (b) is proposed due to higher swing. In this work, four interconnects (3D<0:3>) with different capacitance ratios are implemented. Their sizes are: 3D<0> = 21µm×17µm, 3D<1> = 26µm×17µm, 3D<2> = 31µm×17µm, 3D<3> = 33µm×19µm. .................................................................................................................. 36

Figure 2.4  Cascaded capacitor interconnect layout with parasitic shielding by ring-like Metal-6 plate. .................................................................................................................. 37

Figure 2.5  (a) 3×3 array structures where central interconnect is probed at sensing node to measure crosstalk; (b) voltage margins ($V_M$) of the central electrode with respect to different interconnect spacing; (c) signal eye diagram of the 3D<1> with 2.5µm spacing; (d) signal eye diagram of the 3D<1> with 5µm spacing. .................................................................................................................. 39

Figure 2.6  Cross-section of emulated-3D interconnect structure (E3D) and the dimensions of E3D<0:3> are shown on the right......................................................... 40

Figure 2.7  (a) Transceivers architecture of simultaneous bi-directional capacitive coupled interconnect showing only for chip A; (b) voltage clamping circuit; (c) two voltage sensing circuits for higher signal and lower signal swings. .................................................................................................................. 41
Figure 2.8  PVT simulation result of voltage clamp circuit measuring voltage at A’ (V’A). Process variation includes all corners. V_{DD} is varied from 1.1 to 1.2V. Temperature is swept from 0 to 80 degree Celsius. 41

Figure 2.9  (a) Test structure for E3D<0:3> with proposed transceivers; (b) Test structure of 3D<0:3> with proposed transceivers. 42

Figure 2.10  (a) Virtuoso layout of the test chip; (b) die microphotography. 43

Figure 2.11  (a) Testing of E3D<4> with transceivers; (b) its input and output waveform at 28Mbps. 44

Figure 2.12  (a) Measured voltage levels of the E3D<0:3> at 100Mbps/ch; (b) measured V_{M} of the E3D<0:3>; (c) measured bathtub curves of the E3D<0:3> for PRBS16 and PRBS32; (d) transceiver power consumption with respect to frequency. 45

Figure 2.13  Comparison of this work with previous capacitive coupling transceivers and off-chip simultaneous bi-directional transceivers. Labels are citations with respective maximum data rate in Gbps. 47

Figure 3.1  (a) Illustration showing parasitic capacitance of the stand-alone CCI. (b) Illustration of parasitic components (C_{TX}, C_{RX} and C_{TR}) of two CCI in closed vicinity. 50

Figure 3.2  (a) The parasitic capacitances associated between two CCIs with respect to spacing. (b) The coupled noise with respect to spacing between two CCIs within a 3×3 array when V_{TX} swing is 1V. 51

Figure 3.3  (a) The signal and noise model of a single-ended 3×3 array. (b) Signal waveforms of the single-ended transceiver shown in (c). 53
Figure 3.4  Eye-diagram of the recovered pulses from (a) central and (b) peripheral CCIs when the structure in Figure 3.3(a) is used............................54

Figure 3.5  Layout area comparison of crosstalk cancelled CCIs array in forms of (a) differential butterfly CCIs array, (b) single-ended CCIs array with ground shield and (c) the proposed hybrid CCIs array. (d) the graph indicating the percentage of area overhead of two formal array designs in comparison to hybrid array design..................................................55

Figure 3.6  (a) Noise model of common-centroid differential CCI, (b) noise model of single-ended CCI, and (c) the layout of the pseudo-hybrid CCIs array........................................................................................................58

Figure 3.7  The block diagram of the proposed fully-differential transceiver........60

Figure 3.8  (a) A driver with single-ended to differential conversion. (b) A proposed self-biased fully-differential receiver. (c) A hysteric latch for pulse to digital data conversion. (d) Signal waveforms of the receiver.61

Figure 3.9  (a) Half-replica biasing circuit of the proposed receiver. (b) Half-replica biasing with common-feedback to compensate for process variation and mismatch. (c, d) 1000 points Monte Carlo simulation (inclusive of both process variation and mismatch) of the common-mode variation of the receiver with the biasing circuits from (a) and (b) respectively........63

Figure 3.10  The variation in input offset of the receiver in 1000 points Monte Carlo simulation with both process variation and mismatch. .........................64

Figure 3.11  (a, b) the simulated eye-diagram of the recovered data pulses from the peripheral and centre CCIs respectively in the proposed hybrid CCIs array........................................................................................................65
Figure 3.12  The architecture of the implemented Built-In-Self-Test (BIST). ....... 66

Figure 3.13  (a) The micro-photography of the fabricated die in 65nm CMOS technology. (b) The layout of the BIST with test circuits. (c) The layout of the Pseudo-TxRx CCI array. ................................................................. 67

Figure 3.14  The measurement result of (a) The fitted bath-tab curve showing the BER of the central CCI at different scenario of testing, SC<1:3> with Pseudo-TxRx, and (b) The fitted BER bath-tab curve of the peripheral single-ended CCI in SC<2:3> from Pseudo-TxRx........................................... 68

Figure 3.15  The power consumption of 9 transceivers in the proposed 3 × 3 hybrid array from the simulation and measurement. ................................................. 69

Figure 4.1  (a) Cross-section of 3D integrated chip from Tezzaron, (b) cross-section of copper thermal fusion pad with varying pressure. ...................... 72

Figure 4.2  (a) modeling of F2F stacked copper direct bonded pad at near failure condition, (b) an electrical model of (a), and (c) the plot of $R_{12}$ with respect to asperity radius. ................................................................. 75

Figure 4.3  Proposed dual-mode transceiver with resistive sensing circuit for F2F stacked copper thermal fusion pad. ...................................................... 76

Figure 4.4  A typical example of the transceiver for the capacitive coupling interconnect................................................................. 76

Figure 4.5  (a) proposed resistance sensing circuit with reduced transistors count, (b) layout of the dual-mode receiver with the resistive sensing circuit from (a) in dotted line configuration. ............................................. 77

Figure 4.6  (a) 1000 points Monte Carlo simulation of $V_L$, $V_H$ and $V_{RX}$, (b) eye width of the recovered signal from parametric simulation, (c) parametric
simulation of propagation delay and (d) parametric simulation of current drawn at different mode. ................................................................. 79

Figure 4.7  3D plot of one un-fused Cu-Cu bonding surrounded by the fused Cu-Cu bonding. ......................................................................................... 81

Figure 4.8  (a) modeling of F2F stacked copper direct bonded pad at near failure condition, (b) an electrical model of (a), and (c) the plot of $R_{12}$ with respect to asperity radius. ............................................................... 82

Figure 4.9  Shmoo plots of (h) and (d) swept from 10nm to 100nm and 1.5µm to 5µm respectively.  (a) At $R_{12}=\infty$, (b) at $R_{12}=100k\Omega$, (c) at $R_{12}=10k\Omega$. Here VNout<50 mV is a pass; otherwise it is a fail ......................... 84

Figure 5.1  (a) Illustration of 2.5D integration on Silicon Interposer, (b) Virtex-7 FPGA with Silicon Interposer from Xilinx. ........................................ 86

Figure 5.2  (a) The schematic of the proposed driver with RC-modulated post-cursor de-emphasis, (b) the sketched waveform demonstrating how the proposed de-emphasis works. ......................................................... 89

Figure 5.3  (a) a frequency response of the proposed system with and without proposed de-emphasis, (b) impulse response of the proposed system with 8 steps de-emphasis strength, (c) a data eye diagram of the 20mm length micro-strip line without de-emphasis, (d) a data eye diagram of 20mm length micro-strip line with EMS<7>. .............................................. 91

Figure 5.4  (a) eye height with respect to de-emphasis steps, (b) eye SNR with respect to de-emphasis steps, (c) eye width with respect to de-emphasis steps, (d) power consumption of the driver, and de-emphasis circuit with respect to de-emphasis steps. ................................................. 92
Figure 5.5 (a) The layout of the test macro, (b) driver layout.............................. 95
List of Tables

Table 1. Performance Summary and Comparison...........................................................70
Table 2 Performance comparison of the proposed driver with previous works from literature.................................................................................................................94
Chapter 1 Introduction

The exponential increase in transistor counts in the integrated circuits (ICs) as predicted by Moore’s Law has been a driving force behind the advancement of the semiconductor industry. One of the most significant advancements is the technology scaling, which not only allows more transistors in a given space, but also improves IC performance. However, the improved ICs require moving huge amount of data from one chip to another through wires and buses whose performance does not follow the technology scaling trend. Therefore, input and output (I/O) bandwidth has become a performance bottleneck in numerous gigahertz circuits and systems. Three dimensional integrated circuits (3DICs) have been proved to be a solution to the aforementioned issue providing high interconnects densities, heterogeneous integration, improved I/O performance, and small form factors [1-3].

Various I/O techniques have been developed to increase the bandwidth in transmission lines but they come with huge area and power penalties [4]. System-on-Chips (SoCs) have been considered as a solution to relieve the I/O bandwidth limitation; however implementing a large system on a single chip increases the complexity of the design and the time-to-market [5]. Multi-Chip Packages (MCP) allow heterogeneous integration with relatively short wire lengths [6]. Denser System-in-Packages (SiP) offer even shorter signal paths. Advanced MCP comes in Package-on-Package (PoP) technology providing 3D stacking with high I/O density. They utilize bonding wires and soldering balls for interconnection. In the advanced 3D integration, they use micro-bump [7] and Through-Silicon-Via (TSV) technologies [8] allowing smaller form factors and higher I/O densities because the
TSV diameter in the advanced SOI process is just 5µm [1]. Despite that, TSV technology requires additional process steps, which increases the manufacturing cost. In addition, the stringent TSV design rules such as keep-out zone are to be followed so that circuit blocks are not affected by the mechanical stress at high temperature. The TSV wafer-to-wafer integration, much favored for better electrical characteristics, is also expensive due to the yield lost from known-good-dies and TSV related reliability issues [9].

Alternatively, wireless proximity communication such as inductive and capacitive coupling interconnects has been explored in literature to overcome the limitations in planar integration as well as to relax some of the constraints in 3DICs. First of all, these interconnects can be fabricated in conventional CMOS process without special process steps [1, 10, 11]. Moreover, they require less area largely due to the omission of ESD protection [12-14] unlike standard I/Os in off-chip communication and consume less power due to shorter interconnection length. Furthermore, these contactless interconnects are less susceptible to the process variations such as chip misalignment and wafer surface roughness. Known-good dies can be checked before die stacking [15, 16]. This can significantly improve the overall yield. Inductive coupling is a popular method in the multi-tier communication as the amount of flux, which can penetrate through substrate, can be easily controlled. In the capacitive coupling method, only face-to-face integration is feasible due to its proximity requirement. However, it provides higher interconnect density compared to the inductive counterpart; thanks to the communication distance of only a few µm. However, these methods are able to deliver data effectively, their across tiers power delivery is not efficiency as that of wires and physical connections. Moreover,
electromagnetic interference produced by the inductive coupling interconnect is also a concern for system designers.

On one hand, researchers are working to develop proximity transceivers for 3DIC; on the other hand, there are researchers working on to improve the TSV and direct bonding technologies. Tezzaron has adopted the state-of-the-art copper-to-copper thermal compression bonding and TSV technologies for 3DIC implementation [17]. Ziptronix is developing Direct-Bonding-Interconnect (DBI) for 3DIC implementation [18]. Although DBI, which is mainly for Face-to-Face dies interface, presents good reliability, 3DIC still must rely on TSV, which imposes more challenges in fabrication than that of DBI, for Face-to-Back interface. These challenges are also looked into in this thesis.

As 3DIC development is gaining its momentum, one must not over looked the much older 2.5D integration. One of the problems in 3DIC other than the reliability and system yield is the heat dissipation issue. Heat from the core unit especially from a processor could be trapped inside the stacks damaging the unit. Having a processor separated from the auxiliary units such as memory cores would allow much better heat flow of the core processor. With an introduction of silicon interposer for 2.5D integration, the processor can be placed close to memory cores and the copper wires at the BEOL of the interposer provide highly density interconnections between the processor and memory cores. These wires impose different set of issue for the transceiver design compared to that of conventional PCB wire-line. Fine wires from the silicon interposer are very lossy but their short RC-loss wire relaxes impedance matching [19]. These set of requirements are also discussed in this thesis.
1.1 High Speed Interconnects for 3D ICs Utilizing Inductive Coupling

On-chip near-field-communication utilizing inductive coupling interconnect (ICI) has been studied in literature for more than a decade as an alternative to other well-known 3D integration techniques such as micro-bump and Through-Silicon-Via (TSV). Unlike micro-bump and TSV, the inductive coupling scheme is more CMOS compatible and thus requires no additional process steps. In the inductive coupling interconnect, magnetic flux can be sent vertically to neighbouring tiers in 3DIC thanks to the planar structures of the on-chip inductors. Due to the similarity in permeability of silicon substrate and silicon dioxide (SiO₂), the magnetic flux can penetrate multiple tiers and therefore, cross-tier communication can be realized without requiring to relay the signal at every passing tier. The transmission power in this scheme can be controlled according to the communication distance, thanks to controllable current drivability of the transmitter. In other applications, the inductive coupling effect has also been used for delivering power wirelessly to neighbouring tiers.

In the following sections, some of the reported transceiver designs based upon the inductive coupling interconnect is discussed. Inductor design and other related issues such as cross-talk, alignment, and communication distances are also discussed followed by a few system level implementations.

1.1.1 Transceiver design and its signalling scheme

One of the most widely used inductive coupling transmitter is the H-bridge type shown in the Figure 1.1(a) Two inverters are connected to each end of the
inductor. The complementary transmit signal to one inverter is skewed by a delay block to limit the bipolar current pulse passing through the inductor coil. The loop of the transmitting inductor is open when it is not transmitting data because the induced current in the idle transmitting inductor can counteract with the change of the magnetic flux which reduces the induced current in the receiving inductor. The duration of the current pulse can be controlled by adjusting the delay to reduce power consumption; however this reduces the timing margin and thus, the precise timing and duty cycle control of the sampling clock at the receiver is required. At the receiving side, a sense-amplifier is used for voltage sensing and a latch at the subsequence stage is for digital output [20].

Although the H-bridge transmitter is simple to implement, it draws substantial amount of power due to the short current flowing though the inductor during the
magnetic flux generation. One attempt to reduce the short current is to utilize a current-reuse topology shown in the Figure 1.1(b). A single-ended transmitter stores the charge during the pull-up in the capacitor and it will be reused when the pull-down is enabled in the transmitter. In this transmitter, the inverter circuits are arranged in four slides which have different driving strength ratios of 1, 2, 4 and 8. Then, the transmission power is optimized by enabling only certain number of slides combination. The same latch-based sense amplifier is used to sample the receiving data [21].

Another work reported that the power consumption of the H-bridge transmitter could be reduced by reducing the current pulse width (\(\tau\)) while keeping the slew rate unchanged. This is important since the slew rate determines the Bit-Error-Rate (BER) of the signal. When the pulse width is reduced, the transmitting current \((I_P)\) can be scaled down proportionally. As the energy is defined as \(E = V_{DD} \times I_P \times \tau\), three quarters of the energy can be saved by reducing the pulse width by half. The transmitter circuit comprises of a pulse width control circuit using phase interpolators, 4-bits slew rate control with variable capacitors, and a pulse amplitude

![Figure 1.2 Transceiver with pulse width control and precise pulse-shaping.](image-url)
control by manipulating the driver strength. The combination of an analog high gain receiver with high slew rate clock transmission (not shown) helps the clock frequency recover precisely with only 4.8ps-rms jitter and that clock is used to sample the data from the receiving inductor using latch-based sense amplifiers. With the precise pulse control, the transceiver power is reduced to 0.14pJ/b or by 20 times compared to the previously reported design. The transceiver architecture is illustrated in Figure 1.2 [22].

In [23], serial data transmission through inductive coupling is proposed. The benefit of serialization is twofold. An inductor coil is one of the most area consuming parts in the implementation. Therefore, multiplexing parallel data through one link reduces the overall area usage by inductors. Having less number of inductors reduces

![Image of a circuit diagram](image_url)

Figure 1.3 (a) Block diagram of the transceiver for burst data transmission, (b) receiver front-end circuits.
The interferences among them and the effective data rate can be further improved. The block diagram of the transceiver with the serial data transmission is depicted in the Figure 1.3(a). An H-bridge transmitter is utilized without a delay block [Figure 1.1] to increase the timing margin. The receiver is implemented by a half-\( V_{DD} \) biased inverter with a hysteresis stage composed of cross-coupled PMOS transistors to convert short voltage pulses to full swing CMOS logics. It achieves the data rate of up to 11Gbps with 120\( \mu \)m diameter inductor over 15\( \mu \)m communication distance.

ICI using a rotary coding scheme was introduced in [24]. In the rotary coding scheme, ternary symbols are sent instead of binary symbols. A single bit is represented by a shift in phase and the binary bit is indicated by the rotation direction as shown in Figure 1.4(a). The clock can easily be recovered by the rotation period and thus, the power overhead in the Clock Data Recovery (CDR) circuits such as PLL and the symbol overhead in multiple coding systems can be reduced. The block
diagram of the rotary transceivers is shown in Figure 1.4(b). It comprises of three pairs of inductors to allow ternary symbol transmission. Encoder/decoder is required to convert binary to ternary symbols or vice versa.

In summary, the H-bridge transmitter is the most widely used transmitter for the ICI scheme. In relatively low data speed (~1Gbps), the single-ended current-reuse topology and pulse shaping method are employed to enhance the power efficiency of the transmitter while employing the latch-based sense-amplifier as a receiver. However, in high speed link (~10Gbps), analog receivers with high gains is more popular at the receiver front end to improve the signal-to-noise ratio and the jitter performance. Rotary coding can also be integrated into the transceiver design to relax the complexity of the CDR and thus improve the power efficiency of overall I/O systems.

1.1.2 Inductor Design and Implementation

In ICI, voltage at the receiving inductor \( V_R \) is directly proportional to the mutual coupling between two inductors \( M \) and the rate of current change at the transmitting inductor. The voltage at the receiving inductor \( V_R \) can be written as

\[
|V_R| = M \frac{dI_T}{dt} = k \sqrt{L_T L_R} \frac{dI_T}{dt} \quad (1)
\]

where \( L_T \) and \( L_R \) are inductance of transmitting and receiving inductors, respectively. The coupling coefficient, \( k \) can be defined as

\[
k = \int \frac{0.25}{(X / D)^2 + 0.25} \, dt^{1.5} \quad (2)
\]
where $X$ is the distance between two inductors and $D$ is the diameter of the inductors [25]. In general, $D$ is designed to be larger than $X$ so that the received signal level is significantly larger than ambient noise for easy signal recovery.

Cross-talk issues on the inductive coupling interconnect are discussed in [26]. Two ways to minimize the cross-talk are discussed. They are Space Division Multiple Access (SDMA) and Time Division Multiple Access (TDMA). SDMA requires positioning the adjacent receiving inductor in a way which makes both outward and inward magnetic flux pass through the inductor to cancel out the induced current shown in Figure 1.5(a). In TDMA, the adjacent transmitting and receiving inductor pairs are operating with 180-degree phase shift to avoid sampling.
cross-talk induced voltage pulses (Figure 1.5(b)). It is reported that Interference-to-Signal Ratio (ISR) is -14dB with the inductor pitch of 120 µm at the communication distance of 60 µm. At the communication distances of 90 µm and 120 µm, ISR becomes -13dB and -5dB, respectively. Therefore, precise control of the inductor pitch and communication distance is required to achieve good cross-talk cancellation with SDMA. With the combination of SDMA and TDMA, ISR remains between -15dB to -16dB across all three communication distance.

The research work in [21] presents the parallel data transmission through a 3 × 65 inductive channel array. This implementation applies SDMA and TDMA to
reduce crosstalk noise in the inductive array. Transmission power control is built in the transceiver to optimize the power consumption according to the communication distances. Total power consumption of the inductor with 50 µm diameter at the communication distance of 15 µm, 30 µm and 45 µm are at 1.2, 2.2 and 4.1 W, respectively. It also reports that the crosstalk noise is further reduced with the transmission power control and the aggregate data rate of the channels is improved by 80%.

The inductor structure called XY inductor which can be embedded in the logic routing is introduced in [25] (Figure 1.6(a)). In the conventional inductor design, dedicated chip area is required for an inductor because the wire routing is not permitted in the inductor area. The reported XY inductor structure implemented with two different metal layers can be blended into the logic routing, which crosses the inductor layout parallelly or orthogonally. This substantially reduces the area requirement of the inductors at the expense of magnetic flux loss from the eddy current at the logic wires. The power increase of 23% was necessary in [25] to compensate for such loss. The parallel and orthogonal logic wires also impose differential crosstalk to the inductor. Ground shield around the inductor isolates the parallel crosstalk. The orthogonal crosstalk or the differential noise (NI) (Figure 1.6(b)) is converted into the common-mode noise by supplementing with another noise source (N2) at the inductor extension as illustrated in Figure 1.6(c). Due to the propagation delay between two noise sources, the amount of cancellation will vary. The length of the inductor extension also affects the noise. In [25], the optimal extension is found to be 20% of the 4×D and the differential noise from the orthogonal wire is suppressed up to 70%. With the inductor diameter of 300 µm, the
communication distance of 100 µm was achieved. In this implementation, the inductor pitch is twice of its diameter to reduce the crosstalk interference. The unused inductors are left open to suppress its induced current; otherwise it will hinder the change of magnetic flux in the receiving inductor.

Data transmission using inductive coupling across 128 stacked NAND flash memory chips is presented in [27]. To reduce the area penalty of large inductors, the inductors are placed on top of the memory cores. By placing square inductor diagonal to bit/word lines, the capacitive and inductive interference to the accessed memory core can be significantly reduced. In this work, each die has 3 selectable inductor pairs with transceivers to relay the signal every 8\textsuperscript{th} chip in a top-down manner and all the dies are stacked in a spiral stair manner (Figure 1.7(a)). Like in a spiral staircase where headroom is available, bonding pad area at each step has enough room to perform wire-bonding with a package. The controller chip is at the top of the stack controlling the signal relay to the destination chip. Since mutual inductance is degraded by the eddy current in eight memory cores in the relay path,

![Figure 1.7](image-url) (a) Spiral stacking of NAND flash memory chips, (b) terraced stacking of NAND flash memory chips.
the inductor diameter is increased to 1.1mm and the distance between every relay inductor is 240µm. With spiral stair stacking opposed to terraced stacking [28], higher coupling coefficient is achieved due to less alignment offset among inductors and the communication distance is also reduced due to the lack of spacer (Figure 1.7(b)).

In all the aforementioned implementations, the chip power is supplied through bond wires. If the power supply bond wires are replaced with wireless power transmission, the whole systems can be completely sealed without exposing any bonding pads that are prone to erosion. In [29], inductors for power transferring and inductors for data transmitting are allocated side-by-side, which can incur relatively large interferences between the power inductors and the data inductors. To avoid that, a time-interleaved data and power transmission is adopted. To tackle the above interference issue, a nested clover inductor is proposed in [30] to simultaneously transfer data and power without needing to time-interleaved between data and power transmission. Two differential clock and data inductors are rotated by 90 degree and embedded inside the power inductor as shown in Figure 1.8. The induced current in the clock and data links due to the power inductor is in the form of common-mode

Figure 1.8  The illustration of clover coil for interference cancellation in simultaneous power and data transmission.
interference, which does not interfere with differential mode signal (Figure 1.8(a)). Moreover, the magnetic flux interference between the data and clock inductors are cancelled out resulting in better isolation of the two inductors (Figure 1.8(b)). It is reported that the proposed design is able to achieve the data rate of 6 Gbps and the power transfer of 10mW simultaneously to the distance of 20 µm. The dimension of the data and clock inductors is 300 µm × 300 µm while that of the power inductor is 700 µm × 700 µm.

1.1.3 Summary

As a summary, several inductor design parameters such as alignment tolerance, the effect of $X/D$ ratio to data rate and transceiver power are statically analysed. Due to the magnetic flux spreading from the centre of the transmitting inductor, the alignment tolerance of inductive coupling interconnect is relaxed unlike micro-bump and TSV technology. The alignment tolerance of up to 50µm is reported in [1] with 150µm diameter inductor coil. [31] reports that alignment tolerance is as closed as half the inductor diameter. However, [1] reports is based on measurement.
of a single inductor whereas in [31] inductor pitch is about twice of the inductor
diameter (based on visual clue) and thus, crosstalk interference among them is
minimal. Although there is no experimental result are presented in [21, 25, 26, 32],
the alignment tolerance of those closely tied inductor channels is likely to be less
than that of [1, 31] for the sake of crosstalk interference.

Figure 1.9(a) is the scattered plot of data rate against $X/D$ ratio. The scattered
points are categorized into two groups based on of nature of implemented channel;
i.e., single link and array of links. In general, majority of the single links in Figure
1.9(a) are found to be higher data rate compared to array of links due to the minimum
amount of interference in the system. The trend line for single link also suggests that
exponential improvement in data rate can be obtained by reducing the $X/D$ ratio. However, for the array of links implementation, the relationship between data rate
and $X/D$ ratio is linear and data rate improvement at lower $X/D$ is found to be
insignificant.

The relationship between $X/D$ ratio and transceiver power consumption is
illustrated in Figure 1.9(b) based on the available data. The trend line in the figure
supports the theory of more transmission power for longer communication distance
regardless of whether the links are single or array type.

According to both Figure 1.9(a and b), $X/D$ ratio in many implementation are
hugely varied depending on the power budget and the required data rate. The optimal
$X/D$ ratio for both power and data rate can be found by intersecting two trend lines
from two figures. These two trend lines from Figure 1.9(a and b) are expressed in
Equation (3 & 4) respectively and the equation 4 is normalized according to equation 3.

\[ y = 5.41 \times e^{-1.79x} \quad (3) \]
\[ y = 0.48 \times e^{2.21x} \quad (4) \]

The intersection of two trend lines reveals the optimal \(X/D\) ratio of 0.6 which can provide both reasonable energy efficiency (6.1pJ/bit) and data rate (1.8Gbps)

1.2 High Speed Interconnects for 3D ICs Utilizing Capacitive Coupling

Capacitive coupling interconnect (CCI) is a voltage-driven method unlike a current–driven ICI. Due to the proximity requirement, CCI can be realized only in face-to-face die stacking. It has relatively low parasitic capacitance due to its relatively small electrode sizes and the lack of ESD structure. Therefore, the transceivers can operate with low power consumption. In addition, the interconnect density of the capacitive coupling scheme is higher than that of the wire bonding, the micro-bump and the inductive coupling scheme [10]. In this section, the modeling of CCI and various capacitive coupling transceivers are discussed.

1.2.1 Modeling of Capacitive Coupling Interconnect

CCI is formed when two dies are stacked together in a face-to-face manner [11]. The top metal layer from each die becomes an electrode to realize a coupling capacitor. Signal is transferred from one electrode to the other through capacitive coupling and the amount of coupling is determined depending on the coupling ratio (H) between the coupling capacitance \(C_C\) and parasitic capacitance \(C_P\) at the
receiving electrode [12]. The electrical model of the capacitive coupling is illustrated in Figure 1.10. The coupling ratio of the interconnect ($H$) is defined as follows.

$$H = \frac{v_{RX}}{v_{TX}} = \frac{C_C}{C_C + C_P}$$

($5$)

$C_P$ includes both the electrode to substrate capacitance and the gate capacitance of the receiver’s input devices. For better signal integrity, a larger coupling ratio ($C_C / C_P$) is preferred. This can be achieved by increasing $C_C$ and minimizing $C_P$. Deep sub-micron CMOS process is beneficial not only because it can switch faster but also because it has small gate capacitance achieving higher $H$ value. $C_C$ can also vary by alignment mismatch in die stacking, which directly affects the coupling ratio as well. The misalignment issues are discussed in next section.

1.2.2 Alignment

Alignment is critical in CCI since it directly affects the coupling ratio and accordingly, the signal integrity. In [13], misalignment caused by six degrees of freedom in die stacking are reported. They include three degrees of translation along X, Y and Z-axis and three degrees of rotation at 3 axes such as yawing ($\theta$), rolling ($\beta$) and pitching ($\alpha$). Among them, the capacitive coupling interconnect, in particular, is most sensitive to Z-axis variation due to thickness variation in dielectric material.
between electrodes. This directly changes the coupling ratio of the interconnect [13]. Although the impact of the dielectric thickness variation can be mitigated by using over-designed coupling capacitance values, the variations can be measured by sensing the capacitance of the interconnect using a rectifier circuit as presented in [33] (Figure 1.11). Measuring the capacitance at each corner of a die can reveal the rotation caused by \( \beta \) and \( \alpha \). Several design methods to tackle the misalignment in X, Y and \( \theta \) are also proposed in literature [13, 34, 35].

The offset in the X, Y and \( \theta \) degrades the coupling ratio. When it is large enough, overlapping of electrodes with unintended neighboring electrodes jeopardizes the signal integrity by the degraded coupling ratio and the crosstalk [13]. Moreover, the thermal expansion of dies at high operating temperature is another issue to be tackled when the process technologies of two stacked dies are different [35]. One way to deal with the alignment offset and the thermal expansion is to design the transmitting electrode larger than the receiving electrode so that the overlapping area is maintained under X and Y offsets. For example, if the area of the receiving electrode is \( A_{RX} = x^2 \mu m^2 \), the transmitting electrode area should be greater than or equal to \( A_{TX} = [x \times (1 + (\alpha_{TX} \cdot \alpha_{RX}) \times \Delta T) + \sigma]^2 \mu m^2 \) across the operating temperature range. \( \alpha_{TX} \) and \( \alpha_{RX} \) are the coefficient of linear expansion of \( TX \) and \( RX \).
Thesis

dies respectively and $\sigma$ is the alignment tolerance of the die aligner. Larger receiving electrode is not preferred since they increase parasitic capacitance and accordingly, degrade the coupling ratio.

Figure 1.12 (a) depicts an on-chip Vernier measurement system for confirming chip alignment [13]. Each system consists of 10 transmitting electrode bars with 12.6 $\mu$m pitch and 9 receiving electrode bars (shown in grey color) with 14 $\mu$m pitch. The difference in the pitch of the transmitting bars and that of the receiving bars is the resolution of the system. Although adjacent transmitting bars are alternatively sending complementary data bits, the data pattern at the receivers will be varied depending on the offset in two dies. In perfectly aligned stacking, the central (5th) receiving bar shows null detection in which no signal transition is detected. By employing two Vernier systems horizontally and vertically, it is possible to detect not only the lateral shift in X and Y-axis but also the rotation in $\sigma$.

An on-chip electronic alignment system to correct the offset after die stacking is proposed in [34] (Figure 1.12(b)). The system is designed to correct X and Y
transition. The main idea is to use a number of micro-electrodes instead of one electrode to direct the signal to a receiving electrode. A group of selectable \((4 \times 4)\) micro-electrodes serves as a transmitting electrode and suitable micro-electrodes are selected through a network of multiplexers to correct the offset. The misalignment information can be obtained through the Vernier measurement system. In the implemented system, each micro-electrode has a pitch of 12.5 \(\mu m\), which is a quarter of the receiving electrode and the alignment offset can be corrected up to 6.25 \(\mu m\).

In summary, misalignment in X and Y directions can be evaluated and corrected even after die stacking. However, Z, \(\theta\), \(\beta\) and \(\alpha\) misalignment can only be measured without being able to correct. Thus, overdesigning is still necessary to provide an acceptable yield from die stacking under misalignment.

### 1.2.3 Crosstalk

Capacitive coupling is prone to crosstalk when electrodes are misaligned and neighboring interconnects are in close proximity. The issue of misalignment is already addressed in the previous section and the crosstalk due to proximity is the focus of this section.

![Crosstalk components between two interconnects.](image)

Figure 1.13 Crosstalk components between two interconnects.
Crosstalk components in the capacitive coupling interconnect is illustrated in Figure 1.13. When two or more interconnects are closely placed to improve the interconnect density, the effect of cross-coupling components such as $C_{TR}$ and $C_{RX}$ becomes apparent [36]. They not only reduces the coupling ratio of the interconnect but also causes more crosstalk. The noise from the crosstalk modulates the amplitude of the signal and subsequently it causes the receiver output to jitter. In worst case, the signal recovery can fail entirely. In this section, various countermeasures to the crosstalk are described.

One way to deal with the crosstalk in single-ended interconnect array is to block the cross-coupling components with grounded wires as shown in the Figure 1.14. In this arrangement, the grounded wires are directly inserted between the $RX$ electrodes to minimize $C_{RX}$. Using extended $TX$ electrodes over the grounded wires also significantly reduce the fringing capacitance or $C_{TR}$ from the neighboring $TX$ electrodes [13].

Another way of mitigating the crosstalk issue is to use a differential interconnect array structure. Figure 1.15(a) describes the butterfly differential CCIs array structure [34] illustrating how the crosstalk in channel A is cancelled while surrounded by channels (B, C, D and E). Among them, the channel B and the channel
E do not contribute any noise to the channel A because the coupled noise from B+ and E+ are cancelled out by their respective complementary noise. The remaining channels (C and D) introduce common-mode noise to the channel A which can be rejected by adopting differential receiver.

Another approach is to use a hybrid CCIs array reported in [36, 37]. The hybrid interconnect array has both single-ended and differential interconnects interleaved together as illustrated in Figure 1.15(b) and thus, its area consumption is reduced improving the interconnect density (please refers to Chapter 3 ). Moreover, the hybrid CCIs array is more area efficient than adding ground wires to the single-ended CCIs array because these wires implemented in thick metal imposed area overhead due their larger minimum width and clearance requirement (please refers to Chapter 3 ). The differential interconnect A is in the common-centroid form which exposes equal area of differential electrodes’ edges to every surrounding single-ended ones (B, C, D and E). In this array arrangement, all the single-ended
interconnects do not receive any crosstalk noise as their received noises are always complementary and are cancelled out on the spot. Moreover, the crosstalk from the single-ended interconnects to the differential interconnects is common-mode in nature which can also be rejected by differential receiver.

In summary, the methods of crosstalk countermeasure can be put into two groups; i.e., passive approach and active approach. In the passive approach, grounded wires are inserted between the RX electrodes to shield the cross-coupled components while using the conventional single-ended receivers. In the active approaches, some crosstalk noise is cancelled out by their respective complementary noise components and the residual common-mode noise is rejected by the differential receiver.

1.2.4 Transceiver Designs

The capacitive coupling interconnect is voltage-driven and a two-staged static inverter can be employed as a transmitter. This facilitates the simple transmitter

![Transceiver Diagram](image)

Figure 1.16 (a) the illustration of the typical single-ended receiver for the capacitive coupling interconnect, (b) the circuit in (a) configured as differential transceiver.
design with less power consumption compared to H-bridge transmitters from the inductive coupling. For differential signal transmission, single-ended signal is converted to differential signal. Three stages of inverters are used to generate the inverted signal; whereas the non-inverted signal itself is gone through two staged inverters with a delay inserted in between so that the complementary outputs will possess a phase shift closest to 180 degree.

Figure 1.16(a) explains an inverter-based amplifying receiver with resistive feedback. The gain of the 1st inverter is high because its input is biased at its switching threshold through the resistive feedback implemented with a transmission gate. When it is integrated to capacitive coupling interconnect, it produces voltage pulses at its output whenever there is a data transition at the input. The pulses are further amplified by another inverter before registering it to a latch at the subsequent stage. This design is resilient to Inter-Symbol-Interference (ISI) as long as the duration of Non-Return-to-Zero (NRZ) pulses from the receiver output is less than the symbol period. Its high speed capability is demonstrated in differential architecture achieving up to 10Gbps in 0.18µm CMOS technology (Figure 1.16(b)) [38]. Another variant of this receiver is presented in [39]. It uses a diode-connected MOS feedback rather than the resistive one. Its operating frequency is less than resistive feedback type because it directly converts the coupled signal close to the full-swing digital logic.

A bi-stable receiver topology (Figure 1.17) is proposed to reduce the static current consumption at the front-end inverter [11]. This is particularly important in parallel data transmission where the data rate is not very high. In this case, static power overwhelms dynamic power in the aforementioned topology. Instead of using
resistive feedback, the input of the front-end inverter is attached to diode-connected PMOS and NMOS to level the received signal at $V_{TH}$ for “0” or $V_{DD} – V_{TH}$ for “1” after data transition. Because the diode-connected MOS has higher threshold compared to that of the transistors at the inverter, either PMOS or NMOS of the inverter is always biased in the sub-threshold region. Therefore, the static power at the inverter is dramatically reduced. Moreover, the dual feedback path prevents static current from flowing in the diode-connected devices after the data recovery. In this design, the speed of the signal recovery is primarily limited by the delay in the dual feedback path.

Transceivers can also be configured for bi-directional transmission (Figure 1.17) by employing a transceiver at each side of interconnect and a control is used to enable or disable an appropriate set of a transmitter and a receiver according to the direction of data transmission. In this approach, the power consumption is higher than that of uni-directional transmission due to larger parasitic capacitance associated
at the TX and RX electrodes. Its maximum data rate is also degraded due to the reduction of coupling ratio from the additional parasitic capacitance.

Going further, a simultaneous bi-directional transmission scheme using three cascaded capacitors is explored in [40-42] (Figure 1.18(a)). In this scheme, 4-level of voltages are formed (Figure 1.18(b)) at the receiving electrodes depending on the transmitting bits (TX1, TX2) from both side of the interconnect (namely “00”, “01”, “10” and “11”). Two receivers (H and L) are used to track the lower voltage swings (00 and 01) and the upper voltage swings (10 and 11), respectively. When the transmitter transmits logic “1”, the upper receiver (H) at the same side is activated; when logic “0” is transmitted, the lower receiver (L) is enabled. Both receiver outputs are wired to 2×1MUX which is controlled by the transmitting bit to select the

Figure 1.18 (a) Three cascaded capacitors interconnect configuration for simultaneous bi-directional signaling; (b) four levels signaling at the interconnect (a); (c) The transceiver block diagram for simultaneous bi-directional signaling.
Thesis

Correct received data (Figure 1.18(c)). The voltage clamp circuit is to maintain the voltage level at the receiving electrode when the data bit does not change for a long period. More details of the design and implementation is explained in Chapter 2.

However, all the aforementioned transceivers designed are not suitable for butterfly differential CCIIs array and hybrid CCIIs array due to their poor CMRR. In [34], a differential sense-amplifier is used as a receiver for butterfly differential CCIIs array and it had achieved 1.8Gbps/ch in 180µm CMOS technology. On the other hand, self-biased fully differential receiver is reported to use with hybrid CCIIs array [36]. This transceiver is tested with pseudo-hybrid CCIIs array achieving 2.31Gbps in worst case crosstalk scenario in 65nm CMOS technology. The detail of it is covered in the Chapter 3.

In general, the capacitive coupling interconnect schemes are more beneficial in parallel data transmission because their electrode sizes are relatively smaller than those of the inductive coupling ones. Therefore, systems requiring multi-bits data transmission in parallel such as memory systems are promising applications. Single-
ended interconnect is mainly used where the data rate is not high enough to cause significant signal integrity issue. Differential interconnect is indispensable when cross-talk cancellation is required due to the design constraints in area.

1.2.5 Summary

In face-to-face integration, the capacitive coupling interconnect (CCI) has several advantages over the inductive coupling interconnect (ICI). First of all, the interconnect size of CCI is smaller than that of ICI in general. Secondly, the crosstalk issue in CCI is less prominent compared to that of ICI and thus, CCI can achieve higher interconnect density providing more communication parallelism. CCI is also scalable with technology scaling as shown in Figure 1.19(a). The exponential improvement in data rate per square area of the interconnect size is observed when the technology scales. Moreover, its transceiver power consumption is scaled down dramatically (Figure 1.19(b)) as it moves from one technology node to another due to supply voltage scaling and gate capacitance scaling which results in higher coupling ratio of the interconnect relaxing the voltage gain requirement at the receiver [43, 44].

1.3 A Dual Mode Transceiver for 3DICs in Face-to-Face Interface

As mentioned in previous two sections (1.2 and 1.3), there are two techniques (ICI and CCI) which have been explored in the literature alternative to ohmic interconnects in 3DICs, such direct Cu-Cu Face-to-Face bonding (F2F) and TSV assisted copper Face-to-Back bonding (F2B). ICI which has a capability to establish cross-tier communication can possibly be a possible replacement for TSV while CCI can be a good alternative for the direct copper Face-to-Face bonding. What has driven for the development of the ICI and CCI is mainly due to their better reliability compared to the ohmic interconnects. However, one cannot omit the unprecedented
reliability improvement of these ohmic interconnects throughout the years. For example, Direct Bonding Interconnect (DBI) from Ziptronic has a failure rate of only 1ppm[18]. Although TSV still has reliability issues, the researchers have learnt to deal with BIST/R techniques [45-47]. Moreover, ICI and CCI have their own drawbacks. Firstly, they still rely on ohmic contacts for power delivery and secondly, they drain more or less amount of static current due to the involvement of analog circuitry which becomes a concern for 3DICs with hundreds of interconnects [44]. Therefore, in Chapter 4 [48], a new scheme is proposed to utilize the advantages of both CCI and Cu-Cu thermal compression bonding while minimizing their own disadvantages. To enable this, a Cu-Cu thermal compression bonding pad specification is proposed in such a way that when the bonding failed, the two copper pads produce enough coupling between the two. Therefore, the thermally fused Cu-Cu bonding forms an ohmic channel; whereas, these two Cu-Cu pads function as a coupling channel when the Cu-Cu bonding failed due to the wafer surface contamination and warpage. A self-configurable dual-mode transceiver is introduced to work with either of the channel conditions. In the transceiver, the driver is nothing but an inverter chain and the transceiver’s self-configurability is enabled at the receiver side where it is configurable to be a static receiver for the ohmic channel; otherwise, it is arranged as trans-impedance amplifier for the coupling channel. Since the majority of the bonding are assumed to be successfully bonded, most of the receives are working as static receivers saving a lot of static current compared to the case where all the channels are designed for the coupling mode only. Only a minority of the channels, where the bondings are not successful, forces the receivers to work
as the trans-impedance amplifiers resulted in literally boosting the yield of the Cu-Cu Face-to-Face bonding.

The condition of the channel is assessed by the resistance sensor at the receiver side during chip start up based on the channel resistance and the assessment result is retained at the one bit memory at the receiver. More details are explained in Chapter 4.

1.4 High Speed Interconnect for Silicon Interposer-based 2.5D ICs

The 2.5D ICs has been considered as an intermediate technology before moving towards the true 3D ICs integration with TSV. With an introduction of silicon based interposer, its dense back end of line (BEOL) copper wires provide fine-pitch high data rate signal lines for local chip-to-chip communication. The bump size can be reduced down to 50µm diameter or less since there is no mismatch in the coefficient of thermal expansion between the interposer and chips. Figure 1.20 depicts such an example where two flip-chips are mounted on a silicon carrier. However, such fine-pitch wire has inherent channel losses at high data rate leading to significant inter-symbol interference (ISI). Therefore, channel equalization is necessary.

The channel insertion loss is 17dB at 5GHz [19] when striped line has length = 20mm, width = 1.2µm, thickness = 1.2µm, spacing = 1.2µm and distance to ground
is 1.6\(\mu\text{m}\). To increase signal swings, the low impedance (10-20\(\Omega\)) transmitter together with the high impedance (2k\(\Omega\)) receiver termination is used for the I/O transceiver. Although the high impedance termination allows reducing the static power consumption, the impedance mismatch creates the signal reflection; however, in this lossy channel, such reflection is found to be insignificant when its propagation delay is below 0.5UI. At the receiver side, combination of one discrete tap Decision-Feedback-Equalizer (DFE) and the adjustable continuous-time Infinite-Impulse-Response (IIR) filter is adopted instead due to the higher taps requirement in conventional DFE architecture. 8 bits bus-level architecture is presented in Figure 1.21 and the system consumes 4.3pJ/bit over 20mm strip line having 6\(\mu\text{m}\) width, 3\(\mu\text{m}\) thickness, 4\(\mu\text{m}\) spacing and 3\(\mu\text{m}\) distance to ground [49].

Other than applying equalization at the receiver, pre/de-emphasis can be implemented at the driver or transmitter to compensate for the insertion loss at high
frequencies. Especially for the lossy transmission line like the one in the interposer, multi-tap pre- or de-emphasis are required. There are many pre- and de-emphasis circuits presented in the literature [50, 51] [52, 53] which involves circuitries to calculate pre or post cursors strength while maintaining the impedance matching. Such emphasis circuitries alone can consumes a lot of power which increases with increase in number of taps. To reduce the power consumption at the multiple-taps design, a RC modulation de-emphasis scheme is introduced in this thesis. As the capacitor is charged through a resistor, the voltage profile at the capacitor is exponential which then, modulate the de-emphasis. The exponentially increasing de-emphasis can then compensate the exponentially rising signal tails due to the frequency dependent loss. This idea is clearly explained in Chapter 5.

1.5 Summary of Thesis Contribution

In this thesis, a wide variety of topics are covered. More details of the thesis contributions are yet to discuss in next four chapters. Basically, four technical Chapters (2-5) can be categorized into two parts. The first part (Chapter (2-4)) is dealing with the various inter-tier communication schemes through Face-to-Face interface in 3DIC. At the last part (Chapter 5) is regarding with the communication between two chips (not just limited to 3DICs) on Through-Silicon-Interposer (TSI). In Chapter 2, a simultaneous bi-directional CCI and its transceiver design are presented. In Chapter 3, a hybrid CCIs array structure is proposed to address the prominent issues such as crosstalk and interconnect density. The reliability of the copper direct bonding is addressed in Chapter 4. To improve the reliability of the copper direct bonding in Face-to-Face stacked configuration, a dual-mode transceiver is proposed to work either as a ohmic mode for low resistance contact or a capacitive
coupling mode for high resistance contact. Such mode switching is enabled by the proposed resistance sensor. Chapter 5 touches on the transceiver design for 2.5D silicon interposer. A RC-modulated de-emphasis for the drive is presented for the lossy transmission line in the silicon interposer. More details discussion is to be followed in the subsequent chapters.
Chapter 2  A  3Gbps/ch  Simultaneous  Bi-Directional

Capacitive Coupling Transceiver for 3DICs

Figure 2.1  Cross-section of the face-to-face integration with capacitive coupled transceiver.

Figure 2.2  (a) Schematic model of the proposed simultaneous bi-directional interconnect; (b) equations for signal swings ($V_{SW}$) and dead-zone ($V_{DZ}$); (c) graphical demonstration of 4-level signaling of the proposed interconnect.

2.1 Motivation

As explained early in the Chapter 1 that the capacitive coupling interconnect (CCI) can be configured as uni- and bi-directional signaling. A cross-section of the Face-to-Face stacked dies is shown in Figure 2.1. Moving on, this chapter explicitly analyze a novel interconnect structure and its transceiver to transmit and receive data simultaneously through a single CCI from. The rest of the discussion is organized as
Figure 2.3 (a) Signal Swings ($V_{SW}$) of the design (a) and (b) in cross-section are shown where M8 is the top metal. Design (b) is proposed due to higher swing. In this work, four interconnects (3D<0:3>) with different capacitance ratios are implemented. Their sizes are: 3D<0> = 21µm×17µm, 3D<1> = 26µm×17µm, 3D<2> = 31µm×17µm, 3D<3> = 33µm×19µm.

follows. Section 2.2 describes the proposed simultaneous bi-directional signaling scheme and the design of the transceiver. The test chip implementation is explained in Section 2.3. The measurement results are presented in Section 2.4, followed by the conclusion in Section 2.5.

### 2.2 Design of Simultaneous Bi-directional Transceiver

A conventional uni-directional signaling interconnect is composed of only one coupling capacitor, $C_C$ acting like a channel for signal transmission, and one transceiver pair. In the proposed structure, we introduce two additional driving capacitors, $C_D$ on both sides of the $C_C$ (Figure 2.2(a)), resulting in 4 voltage levels at
the sensing nodes (A and B) through charge sharing among the capacitors (Figure 2.2(c)). When TxL is transmitting '0', the signal level at A will be at either '00' or '01'. On the other hand, the level at A will be at either '10' or '11' when TxL is transmitting '1'. Therefore, depending on the transmitting data bit, the receiver can be configured to sense a specific voltage range of either ‘00’-‘01’ or ‘10’-‘11’, which will be explained in detail in Section 2.2.2. The equations in Figure 2.2(b) show that parasitic capacitance, $C_P$, significantly degrades the signal swing ($V_{SW}$) and increases the dead zone ($V_{DZ}$). In the next section, we will discuss the proposed interconnect structure and its optimization to increase $V_{SW}$.

### 2.2.1 Proposed Capacitive Coupling Interconnect Design

The transceiver performance depends on $V_{SW}$ since it has to be sensed by the receiver. Larger $V_{SW}$ is desirable for better transceiver performance. To maximize $V_{SW}$, two possible interconnect structures with parasitic shielding are considered as shown in Figure 2.3(a) and (b) where highlighted lines indicate the unshielded area.

![Figure 2.4 Cascaded capacitor interconnect layout with parasitic shielding by ring-like Metal-6 plate.](image-url)
of interconnect exposed to nearby wires and substrate. To form the same $C_D$ value in the two designs, M6 in Figure 2.3(b) has to be larger than M7 in Figure 2.3(a) due to the reciprocal relationship between amount of capacitance and the distance between two plates; hence, larger M6 with the extension beyond the top metal provides better parasitic shielding. $V_{SW}$ improvement of 10% is achieved in the latter design (Figure 2.3). The digital 3D mock up of Figure 2.3(b) is shown in Figure 2.4. Four interconnects (3D<0:3>) with different capacitance ratios are designed according to the dimension shown in Figure 2.3(b) and they provide $V_{SW}$ of ~200 mV when they are simulated in post-layout level. Since $V_{SW}$ is proportional to $C_C \times C_D$ based upon the equations in Figure 2.2(b), $C_D$ is set comparable to $C_C$ so that $V_{SW}$ is not highly affected by the misalignment in die stacking which affect the $C_C$ value. Simulation proves that unnoticeable variable variation in $V_{SW}$ is observed when $C_C$ varies by $\pm 20\%$ [42].

In the typical layout where interconnects are placed in an array form, crosstalk effects between interconnects can degrade the voltage margin ($V_M$). To simulate the crosstalk, a number of 3×3 interconnect arrays (Figure 2.5(a)) are designed and simulated with various interconnect spacing ranging from 1 μm to 5 μm (Figure 2.5(b)). The central interconnect is analyzed using PRBS signal since it undergoes the highest level of crosstalk from the neighboring interconnects and the simulated eye-diagram in shown in (Figure 2.5(c) and (d)). Based on the simulation results, the crosstalk is not tolerable with the spacing of 1 μm to 2.5 μm; however, the spacing of 3 μm and above improves $V_M$ steadily as the spacing increases. Therefore, the minimum spacing of 3 μm can be chosen for reliable signaling. If the chip alignment accuracy is $\pm \beta \mu m$, the spacing between interconnects should be at
Figure 2.5  (a) 3×3 array structures where central interconnect is probed at sensing node to measure crosstalk; (b) voltage margins (VM) of the central electrode with respect to different interconnect spacing; (c) signal eye diagram of the 3D<1> with 2.5µm spacing; (d) signal eye diagram of the 3D<1> with 5µm spacing.

least 3 + β µm. Moreover, the interconnect size does matter to the crosstalk as well. Simulation shows that a larger interconnect (3D<3>) is more tolerant to noise compared to smaller interconnects such as 3D<0>.

To mimicking the proposed 3D interconnect shown in Figure 2.3(b), the emulated-3D interconnects (E3D<0:3>) shown in Figure 2.6 are fabricated in a single die for analyzing the interconnect and transceiver performance without die stacking. It is realized by splitting C_C into two C_C/2 capacitors and connecting them in a cross-coupled way. By doing so, the performance of the proposed transceiver can be characterized without actual 3D integration which requires expensive equipment and additional process steps. The fabricated E3D<0:3> have the same capacitance values
Electrode Sizes

\begin{align*}
E3D<0> &= 20\mu m \times 11\mu m \\
E3D<1> &= 21\mu m \times 12\mu m \\
E3D<2> &= 21\mu m \times 13\mu m \\
E3D<3> &= 21\mu m \times 14\mu m \\
E3D<4> &= E3D<3>
\end{align*}

Figure 2.6  Cross-section of emulated-3D interconnect structure (E3D) and the dimensions of E3D<0:3> are shown on the right.

as 3D<0:3>; thus, they also provide \( V_{SW} \) of \( \sim 200mV \), which is similar to that of 3D<0:3> respectively. Due to the differences in proximity of the two electrodes and the dielectric materials, the dimensions of E3D<0:3> are relatively smaller than that of 3D<0:3>.

2.2.2 Transceiver Design

One side (Chip A) of the proposed transceiver architecture is illustrated in Figure 2.7(a). The transceiver consists of a driver, two receivers, a voltage clamping circuit and a multiplexer to select either OutH or OutL from the receivers. The driver is implemented with two inverters in series to drive the transmitting digital signal to \( C_D \). Two receivers (L and H) designed with PMOS and NMOS differential pairs (Figure 2.7(c)) respectively sense the voltage level at the sensing nodes (A and B) with respect to the reference voltage levels \( V_{RefL} \) and \( V_{RefH} \) generated by a biasing circuit. Receiver L senses “00” or “01” while receiver H detects “10” or “11”. The use of differential amplifiers as receivers relaxes the amount of charge coupling to the sensing node compared with latch based design. The outputs of the receivers are fed to the multiplexer where only one output is selected based the transmitting signal, TxL. The clamp circuit illustrated in Figure 2.7(b) holds the sensing node voltage...
levels by the leakage current through the off NMOS switch during data transmission. At low frequency, the leakage from A’ to A (I_{leak}) compensates the charge lost at A. The voltage at A’ (V’_A) has three different levels depending on the three control signals (UP, DN, CT) shown in Figure 2.7(b).

The PVT sensitivity of the voltage at A’ is simulated with various corners. Figure 2.8 illustrates the voltage levels at A’ with 104 samples for each level. The spread in the “Down” level is mainly due to the variation in the threshold voltage of

Figure 2.7 (a) Transceivers architecture of simultaneous bi-directional capacitive coupled interconnect showing only for chip A; (b) voltage clamping circuit; (c) two voltage sensing circuits for higher signal and lower signal swings.

Figure 2.8 PVT simulation result of voltage clamp circuit measuring voltage at A’ (V’A). Process variation includes all corners. V_{DD} is varied from 1.1 to 1.2V. Temperature is swept from 0 to 80 degree Celsius.
Figure 2.9 (a) Test structure for E3D<0:3> with proposed transceivers; (b) Test structure of 3D<0:3> with proposed transceivers.

the transistor closest to DN due to P and T whereas the larger spread of the “Up” level is affected by all P, T and V which is due to UP transistor’s cut-off $V_{gs}$ and the “Center” level variation is caused only by VDD variation as long as two diode-connected transistors are matched. However, process variations have no effect on $V_A$ at the high frequency signaling due to the insignificant $I_{leak}$. The wider spread in the “Up” level of $V'_A$ will not affect the voltage clamping operation as $V_A$ will track the VDD variation. The reset switch (Rst) is disabled during normal operation so that the parasitic capacitance at the node A’ is decoupled from the node A without performance degradation.
2.3 Test Chip Implementation

Figure 2.9 illustrates the overall architecture of the test chip consisting of 4 transceivers with the emulated-3D interconnects (E3D<0:3>) and 4 transceivers with 3D interconnects (3D<0:3>). In testing the emulated-3D interconnects, pseudo random signals, PRBS16 and PRBS32 are applied to TxL and TxR of the selected transceiver. The recovered signals, RxL and RxR are directly compared with the delayed TxR and TxL respectively for checking the operation of the interconnects. The results from bit-by-bit comparison are stored in the BER COUNT block to be readout. However, during testing 3D<0:3>, only one PRBS16 is utilized to send TxL to a selected transceiver at chip A. The recovered data RxR at chip B is resent to chip A through TxR which will be recovered at RxL to compare with the delayed TxL at the data checking block. Likewise, the comparison results are stored in BER COUNT. This paper only presents the test results of E3D<0:3> due to the die stacking issue for 3D<0:3>. The bathtub curve of the proposed interconnect was measured by controlling the delay (DLY<0:4>) for data checking with the time step of 40ps. The crosstalk effect was not analyzed in this test chip for simplicity.
In addition, the implemented interconnects are not arranged in an array structure. Moreover, all interconnects are spaced with more than 10 µm where crosstalk noise is insignificant in the interconnect performance measurement. All the biasing voltages ($V_{\text{biasL}}$, $V_{\text{biasH}}$, $V_{\text{RefL}}$ and $V_{\text{RefH}}$) are supplied from off-chip for easy calibration. The layout and the microphotography of the test chip are presented in Figure 2.10. The overall area of the test chip is 0.0312 mm$^2$.

2.4 Test Chip Measurement Result

Figure 2.11 shows the measured waveforms of ‘Data Input’ and 'Data Output' from the transceiver with the E3D<4> interconnect. A pseudo-random data pattern is sent to TxL of the transceiver. Its recovered data at the RxR is fed into TxR and subsequently it is recovered at RxL which is connected to 'Data Output' I/O (Figure 2.11(a)). Due to the bandwidth limitation of the I/O, the signal probing was conducted at 28Mbps. Figure 2.11(b) demonstrates that the data pattern at ‘Data Output’ is identical to that of 'Data Input'. Note that such a slow data rate is possible due to the function of the clamping circuit.
Figure 2.12 (a) Measured voltage levels of the E3D<0:3> at 100Mbps/ch; (b) measured $V_M$ of the E3D<0:3>; (c) measured bathtub curves of the E3D<0:3> for PRBS16 and PRBS32; (d) transceiver power consumption with respect to frequency.

The voltage levels at the sensing nodes are measured by sweeping $V_{RefL}$ and $V_{RefH}$ and observing at the output of the data checking block. The error-free ranges of $V_{RefL}$ and $V_{RefH}$ represent the voltage levels of 00-01 and 10-11, respectively. The measurement result (Figure 2.12(a)) confirms that $V_{SW}$ for all E3D<0:3> is roughly 200 mV at 100Mbps/ch which is consistent with the simulated result. The $V_{DZ}$ variation across different interconnects is mainly caused by the variation in the $C_D$ value of the respective interconnects (referring to $V_{DZ}$ equation) and it is also found to be frequency independent. Figure 2.12(b) shows the degradation of average $V_M$ in the upper and lower swings. Due to the coupled high frequency switching noise from
near-by test circuits to the transceivers and interconnects, the measured \( V_M \) is degraded down to 50 mV at 3Gbps data rate. No positive \( V_M \) is observed from 3.2Gbps and above. At 3Gbps, E3D<0, 3> have the timing margin of 400ps while the timing margin of 440ps was measured at E3D<1, 2> (Figure 2.12(c)). Therefore, it can be explained that the maximum data rate of this transceiver is limited by the degradation of \( V_M \) rather than the timing constraints such as jitter and inter-symbol-interference.

Unnoticeable variation in transceiver power consumption is observed regardless of the interconnect sizes. This is mainly due to the fact that the receiver power consumption is not highly affected by \( V_{SW} \) and/or \( V_{DZ} \) unlike [12]. Each transceiver pair consumes 117 \( \mu \)W when two 1.5Gbps pseudo-random patterns are transmitted and received simultaneously; effectively at 3Gbps/ch. When a periodic data pattern is given to the transceivers, the power consumption at 3Gbps increases up to 140 \( \mu \)W which is 23 \( \mu \)W higher than the simulated result shown in Figure 2.12 (d).

Figure 2.13 compares this work with other state-of-the-art works. Capacitive coupling integration has been improved over the past two decades in terms of power (pJ/bit) and area (Mbps/\( \mu \)m\(^2\)) efficiency due to the advancement in circuit techniques, scaling and integration methods. Our work is not as area efficient as some of the latest works due to trade-off of capacitor size with \( V_M \); however, its power consumption is the smallest among the reported designs. When it compared with other off-chip bi-directional and simultaneous bi-directional transceivers [54], our design with 3D<0> consumes 1.5 order of magnitude less power and 60× better
This work in E3D<0> [6] 1.4
[7] 1.2
[17] 10
[15] 1
Estimation for 3D<0>
This work in E3D<0>

Figure 2.13 Comparison of this work with previous capacitive coupling transceivers and off-chip simultaneous bi-directional transceivers. Labels are citations with respective maximum data rate in Gbps.

It also outperforms the inductive coupling transceiver [22] by 7.6× in Mbps/µm² and 3× in pJ/b.

2.5 Conclusion

In this chapter, the design challenges of the simultaneous bi-directional transceiver are discussed in details. Various interconnect structures are studied and finally a structure with enhanced parasitic shielding for larger signal swing is proposed. The proposed simultaneous bi-directional transceivers are implemented in a commercial 65nm CMOS technology. The performance of the proposed interconnect is successfully tested in the emulated-3D interconnects and the measurement results confirm that it can operate up to 3Gbps/ch with the voltage margin (V_M) of 50 mV. The measured V_M from the signal eye is as large as 200 mV at 100Mbps/ch and the timing window is as wide as 400 ps at the maximum data rate. The total power consumption of the transceiver is as low as 140 µW at the maximum data rate. In summary, this work proves that the proposed capacitive coupling face-
to-face integration is more power and area efficient compared to other inductive coupling 3D integration and other off-chip simultaneous bi-directional signaling schemes. However, its performance is not as good as its CCI counterparts.

In general, there are two kinds of receivers presented in the literature for CCI such as voltage and current sensing receivers. The one proposed in this chapter works based on voltage sensing and thus, the voltage margin plays an important role in the channel performance. The voltage margin in the proposed design is apparently not as good as other uni-/bi-directional scheme and that is mainly because the transceiver is sending two bits simultaneously from both ends of the channel forcing two bits to share the available voltage margin. What makes the issue even worst is that three coupling capacitances are put in series to send two bits in the channel and thus, all individual coupling capacitors in the series must be large enough to maintain the effective capacitance but yet, the growth of parasitic capacitance along with the bigger coupling capacitance are not in favour to the already degraded voltage margin. Lastly, the process variation of these three series capacitor causes the voltage level to shift which is almost impossible to track and adjust the receivers sensing threshold accordingly. Even if such process variation can be tracked, the overall system overhead would definitely overwhelm the benefits obtained by sending two bits simultaneously. Therefore, in subsequence chapters (3, 4), the works had been focused on uni-directional signaling only.
Chapter 3 A 2.31 Gbps/ch Area Efficient Crosstalk Cancelled Hybrid Capacitive Coupling Interconnect for 3D Integration

3.1 Motivation

Earlier in Chapter 1, it stages the issues in CCI such as alignment and crosstalk and presents the various solutions available in the literature. In previous solutions to mis-alignment such as the Vernier alignment system and the electronic alignment system with micro-electrodes presented in section 1.2.2 are found to have area and power overhead. The alignment resolution achievable in these methods is outdated by today standard because the state-of-the-art wafer aligner can align much better than that. Therefore, there is a need to improve the alignment accuracy beyond what wafer aligner provides.

The crosstalk among CCIs challenges the signal integrity. Up to day, two solutions have been presented in the literature. One method calls upon the ground shielding wire which is placed in between two single-ended CCIs to shield the crosstalk. In second method, it mainly utilizes the CMRR capability to cancel the crosstalk. All of two solutions work well but the method one comes with area penalty from ground wires and the second method causes both area and power overhead due to the used of all differential CCIs. These issues have to be address to improve the interconnect density and power efficiency.

Here in this chapter, a sample, yet elegant solution to improve alignment accuracy is presented. With the introduction of hybrid CCI array, the power and area overhead issues are addressed since the single-ended CCIs are interleaved with
differential CCIs. The remaining sections of the paper are as follows. In section 3.2, the design methodology of the stand-alone and array type CCIs are discussed followed by that of the proposed hybrid CCIs array. Section 3.3 explains the proposed transceiver design, and the measurement results and the conclusion are detailed in section 3.4 and 3.5, respectively.

3.2 Capacitive Coupling Interconnect

A capacitive coupling interconnect (CCI) can be formed by stacking two dies with thinned or removed passivation in the face-to-face configuration where two top metal pads from each die are aligned and separated by a thin layer of adhesive serving as a dielectric medium as well as a mechanical support (Figure 2.1). A coupling channel is established between the pads due to proximity. Due to the present of parasitic capacitance, the amount of coupled voltage is attenuated and the coupling ratio of the CCI (H) is defined as

\[
H = \frac{V_{RX}}{V_{TX}} = \frac{C_C}{C_C + C_P}
\]
where $V_{TX}$ and $V_{RX}$ are transmitted and received voltage swings, respectively. $C_C$ is the coupling capacitance between two pads and $C_P$ is the combination of transceiver and pad related parasitic capacitance (Figure 3.1(a)). In the test environment, a small electrode size is preferable for less parasitic capacitance but it tends to perform poorly in noisy environment like SoCs. Therefore, to achieve a good signal-to-noise ratio of the received signal, the CCI must possess not only high $H$ value, but also high $C_C$. Although misalignment in RX and TX pads can compromise the $C_C$ value, having TX pad relatively larger than that of RX’s helps to make sure that overlap area between TX and RX pads are well maintained as well as the $C_C$ value itself. A gate capacitance of the receiver also degrades the $H$ value. Therefore, deep sub-micron CMOS process has more advantageous not only because it can switch faster but also because smaller gate capacitance helps to improve the $H$ value.

### 3.2.1 The CCI Array

When more than one CCIs are allocated in the closed vicinity, the present of cross-coupling components between CCIs (Figure 3.1(b)), such as $C_{TR}$ and $C_{RX}$,
increases $C_P$ degrading the coupling ratio ($H$). The actual parasitic capacitance ($C_P'$) in this case can be expressed as

$$C_P' = n \times \left( \frac{C_{TR} C_P}{C_{TR} + C_P} + \frac{C_{RX} C_P}{C_{RX} + C_P} \right) + C_P; \quad (0 \leq n \leq 4)$$

where $n$ is the number of interconnects in the vicinity. Moreover, these cross-coupling components introduce crosstalk noise to the CCI as well. To demonstrate that, FEM simulation is performed on two $10 \, \mu m \times 10 \, \mu m$ CCIs to extract all the coupling capacitances with the spacing from 1 to $5 \, \mu m$ between two CCIs. Dielectric of adhesive is set to $\epsilon_r = 2.4$ and its thickness is fixed at $1 \, \mu m$ to mimic the integration reported in [11]. The simulation result in Figure 3.2(a) shows that the value of $C_{TR}$ doesn’t change much within the specified spacing; whereas, $C_{RX}$ value goes up exponentially as the spacing get smaller. The value of $C_{TR}$ is much less than that of $C_{RX}$ but it plays a significant role in crosstalk because of the direct coupling from the transmitting electrode. Figure 3.2(b) illustrates the relationship between the spacing and the amount of the coupled noise caused by both $C_{RX}$ and $C_{TX}$. At $1 \, \mu m$ spacing, $C_{TR}$ contributes 21% of the total crosstalk noise. The spacing of $5 \, \mu m$ raises the contribution of noise from $C_{TR}$ up to 56%.

A typical $3 \times 3$ CCIs array layout with noise model is shown in Figure 3.3(a) where all the interconnects are single-ended CCIs. Among all the 9 CCIs, the center CCI is the area of interest where the crosstalk noise is at the maximum. $N$ and $N^*$ are the crosstalk transfer function from the peripheral CCIs to the central CCI<5> but the insignificant $N^*$ is ignored in this study. The crosstalk transfer function ($N$) which can be derived as
The received signal of the central CCI (\(V_{RX5}\)) can be defined as

\[
N = \frac{C_c C_{RX} + C_{TR} C_p + C_{IR} C_{RX}}{2C_{RX} C_p + C_{IR} (C_{RX} + C_p) + C_p (C_c + C_p)}
\]

The received signal of the central CCI (\(V_{RX5}\)) can be defined as

\[
v_{RX,OUT} = G \times v_{RX5}; G \text{ is receiver gain}
\]

\[
v_{RX5} = v_{SIGNAL} + v_{NOISE} = H \times (V_{TX5} \angle \alpha) + \sum_{n=1}^{4} [N \times (V_{TXn} \angle \theta)]_{\theta = \alpha, \alpha + 180^\circ}
\]

i.e., \(v_{RX5}\) is the superposition of the received signal and the coupled crosstalk noise. Depending on the phase of \(V_{TXn}\) from the surrounding CCIs, the amplitude of \(V_{RX5}\) will be modulated, which in turns varies the rise and fall time of the receiver output, \(V_{RX,OUT}\), causing jitter. To analyze the jitter, post layout simulation is performed with the parasitic model of 3 x 3 single-ended CCIs array, extracted from FEM.
simulation. The CCI dimension is $10 \times 10 \, \mu m^2$ with $1 \, \mu m$ thick dielectric ($\epsilon_r = 2.4$) and all the CCIs are spaced by $2 \, \mu m$ from one another. The transceiver consists of a driver, a pulse receiver and a hysteretic latch as shown in Figure 3.3(c). The driver drives the TX signal which then, couples to RX in the form of pulses due to the resistive feedback at the pulse receiver. The pulse at RX is amplified by the inverter and generates RX_OUT, which is fed into the latch for pulse to digital conversion producing “RX_D”. These waveforms are summarized in Figure 3.3(b). Figure 3.4(a) and (b) show the simulated eye-diagrams of the recovered signals at the receiver outputs ($V_{RX\_OUT5}, V_{RX\_OUT1}$) from the central and peripheral CCIs respectively in the CCIs array. The result shows that $V_{RX\_OUT5}$ (Figure 3.4(a)) is affected not only by the jitter but also by the error-bits recovery when the transceivers are signaling with Pseudo-Random-Binary-Sequences (PRBS) signal. The $V_{RX\_OUT1}$ eye-diagram is also degraded similarly but its jitter is less than that of $V_{RX\_OUT5}$ by 25% theoretically (Figure 3.4(b)). An error can occur when the amplified noise ($G \times V_{NOISE}$) without data transition (i.e. $v_{SIGNAL} = 0$) goes beyond the hysteretic window of the latch as shown in Figure 3.4. To minimize the jitter and the bit-error rate, the CCIs in the
array should be placed far apart as much as 5µm in some reported designs [11, 12] or a stand-alone CCI is used to achieve high speed signaling [38]. Thus, substantial amount of area is wasted due to the isolation requirement reducing the area efficiency and density of the CCIs.

In [34], the butterfly differential CCIs array is proposed to combat the crosstalk issue (Figure 3.5(a)). In this design, the signal at CCI<1+> is left undisturbed by the complementary coupled noises from CCI<2+> and CCI<2-> and
similarly, for CCI<1-> which is also not affected by the noises from CCI<5+> and CCI<5->. CCI<4+> and CCI<3-> introduces noises to CCI<1+> and CCI<1-> equally causing common-mode shift but it will not affect the differential signaling. Although the butterfly structure tackles the crosstalk issue significantly, it requires large area overhead due to the empty spaces in the 9 CCIs array configuration. Another issue is load imbalance at the differential pairs of CCI<4:9>. Apparently, CCI<4-, 5-, 6-, 7-, 8-, 9-> have less parasitic loading compared to their respective complementary pairs. Dummy CCIs can be inserted to balance the loading but it further increases the total area. Another simple solution to reduce crosstalk comes as the single-ended CCIs array but with added ground shield for crosstalk isolation (Figure 3.5(b)) [13]. Although it consumes less area compared to the butterfly CCI array by 21%, using thick metal as ground shield reduces the CCI density due to the minimum width and clearance requirement of the thick metal. Therefore, the next generation CCI array should not only crosstalk tolerant but also more area efficient.

3.2.2 Proposed Hybrid CCIs Array

In the proposed 3 × 3 hybrid CCIs array, the differential CCIs are interleaved with the single-ended CCIs as depicted in Figure 3.5(c). The common-centroid structure of the differential CCIs not only provides good matching between the differential pairs but also balances the parasitic loading. The area comparison of three different CCIs array designs is made in Figure 3.5. For a fair comparison, each CCI of all designs is given an overlap area of 100µm² between the TX and RX pads regardless of whether it is single or differential CCI. The dimension of the TX pad is 1 µm larger than that of the RX pad so that the RX pads are completely overlapped
with the TX pads by the SmartView®NT aligner that has ±0.5 µm 3σ alignment tolerance. In FEM simulation, the misalignment of ±0.5 µm generates only 0.98% variation in the $C_C$ value; while 4.7% $C_C$ variation occurs when the TX pad is not extended. All the layouts of the three CCIs array are carried out in a commercial 65nm CMOS technology using the top metal layer and CCIs are put together closely according to the minimum allowable distance. As of the layouts shown in Figure 3.5, the hybrid CCIs is the most area efficient design among all; thanks to its efficient layout without using ground shield. The CCIs array with the ground shield design suffers the area overhead of 19.5%; whereas the butterfly CCIs array also presents 52.3% area overhead in comparison to the proposed hybrid CCIs array. Figure 3.5(d) depicts the area overhead of the previously reported CCI arrays compared to hybrid CCIs. It shows that as the number of CCIs in the array increases, the percentage of area overhead for butterfly array decreases from 52.3% and saturates around 10% whereas, in the array with ground shield, the area overhead increases with the increase in the number of CCIs in the array. It is also found that the butterfly array is more area efficient than the shielded array in a larger array size. However, the proposed hybrid array is the most area efficient regardless of the array size. The jitter performance is simulated at 2Gbps PRBS signals for all three different crosstalk cancelled channels in Figure 3.5. The absolute jitter for butterfly scheme (Figure 3.5 (a)) is at 23ps whereas in the shielded single-ended scheme, its jitter is standing at 18ps. In the proposed scheme, the jitter for the differential channel is 25ps and for the single-ended channel, its jitter is only at 20ps. Therefore, the average jitter 3x3 array channels is 22.7ps. The jitter result also implies that the small footprint of the hybrid CCIs array can be obtained without sacrificing the jitter performance.
The common-centroid CCIs are also effective in the crosstalk cancellation. Since it exposes equal area of its pad edges to each and every surrounding single-ended CCI, the crosstalk interference to the differential CCI (Figure 3.6 (a)) is in the form of common-mode noise which is cancelled in differential signaling as shown in the equation shown below.

\[
\frac{1}{2} H \times v_{TX1} \angle \theta - \frac{1}{2} H \times v_{TX5} \angle \alpha - \frac{1}{2} H \times v_{TX5} \angle (\alpha + 180) - \frac{1}{2} H \times v_{TX6} \angle (\alpha + 180) + \frac{1}{2} H \times v_{TX6} \angle \alpha - \frac{1}{2} H \times v_{TX7} \angle (\alpha + 180) + \frac{1}{2} H \times v_{TX7} \angle \alpha \]

Figure 3.6 (a) Noise model of common-centroid differential CCI, (b) noise model of single-ended CCI, and (c) the layout of the pseudo-hybrid CCIs array.
$\Delta V_{RX5} = \frac{(V_{SIGNAL} + V_{NOISE})}{2} - \frac{(\overline{V}_{SIGNAL} + V_{NOISE})}{2}$

$= \left[ \frac{H^*}{2} V_{TXS} \angle \alpha + \sum_{n=1}^{4} \frac{N}{2} V_{TXn} \angle \theta \right]_{\theta=\alpha, \alpha+180^\circ}$

$\left[ -\frac{H^*}{2} V_{TXS} \angle (\alpha+180^\circ) + \sum N \frac{N}{2} V_{TXn} \angle \theta \right]_{\theta=\alpha, \alpha+180^\circ}$

$= H^* V_{TXS} \angle \alpha$  \quad ($H^* = H - 2 \times N$)

Here, $H^*$ is the coupling ratio of the differential CCI whose value is equal to $H$ but attenuated by mutual coupling of its own differential pairs which is the equivalent of $(2 \times N)$. At the single-ended CCI, two crosstalk components from each of the surrounding differential CCI are equal in amplitude and opposite in phase to each other (Figure 3.6(b)) which results in on-the-spot crosstalk cancellation as shown in the equation below.

$V_{RX9} = V_{SIGNAL} + \sum \left( \frac{V_{NOISE}}{2} + \overline{V}_{NOISE} \right)$

$= HV_{TX9} \angle \theta$

$+ \sum_{n=6}^{8} \left[ \frac{N}{2} V_{TXn} \angle \alpha + \frac{N}{2} V_{TXn} \angle (\alpha+180^\circ) \right]_{\alpha=0, \theta+180^\circ}$

$= HV_{TX9} \angle \theta$

Therefore, in theory, none of the CCI gets interfered by any crosstalk noise. In practice, due to mismatch and process variations especially from the electrodes layout, crosstalk noise may not be cancelled perfectly but the significant amount of improvement over the conventional array design (Figure 3.3(a) for example) can be expected.

As for fabrication and testing purpose, a pseudo-hybrid CCI arrays are designed as shown in Figure 3.6(c). The CCI are realized in M6 layer for TX pad and M4 for
RX pad. TX pads are smaller than RX pads to achieve the actual coupling capacitance ($C_C$) as if it would have been implemented in 3D stacked die. In this design, the crosstalk component, $C_{TR}$, is less due to smaller TX pads and thus, the amount of N is less than that of what would have been in 3D face-to-face stacking. Therefore, the T-shaped crosstalk aggressors are added to boost the N to achieve more realistic crosstalk affect.

3.3 Transceiver Design

This section explains the proposed differential transceiver design. It consists of a transmitter, a receiver and a latch as shown in Figure 3.7.

3.3.1 Transmitter Design

The proposed transmitter (Figure 3.8 (a)) is composed of digital inverters and a RC delay realized by transmission gates. $\overline{TX}$ is generated by inverting the “IN” with three inverters chain and non-inverting TX is produced by two inverters with the RC delay inserted in between. This delay stage is tuned to make sure that TX and $\overline{TX}$ are complementary to each other.
Due to the unique structure of the hybrid CCIs array, the receiver has to meet some requirements. First of all, the receiver must have high common-mode-rejection ratio (CMRR). This is critical as the crosstalk noise is induced into its electrodes in the form of common-mode noise. Moreover, as $H$ is attenuated by mutual coupling ($2N$), a high gain receiver is preferable to compensate for the attenuation.

The receiver shown in Figure 3.8(b) is proposed to meet the above mentioned requirements. In the proposed receiver, two differential amplifiers’ inputs ($RX$ and $RX$) are cross-coupled to each other to ensure the minimum input offset. The current
mirror loads are used in the receiver instead of the resistive load to achieve high CMRR. Instead of using a separate biasing circuit for ac coupled inputs, RX_OUT and RX_OUT̅ are connected to RX and RX through transmission gates so that both inputs and outputs are biased at the common-mode voltage \( V_{CM} \) level of the receiver itself. The receiver produces complementary output pulses whenever there is a data transition at the differential inputs. They return back to \( V_{CM} \) after the transition (Figure 3.8(d)) due to its resistive feedback. The voltage output of the receiver can be written as below.

\[
\Delta V_{RX\text{-OUT}} = RC \frac{d\Delta V_{TX}}{dt}
\]

It is also important to note that the pulse width \( t_{PULSE} \) of the receiver outputs has to be shorter than one unit-interval \( t_{UI} \) of a symbol to prevent inter-symbol-interference (ISI). The pulses at RX_OUT and RX_OUT̅ are subsequently registered at the hysteretic latch (Figure 3.8(c)) at the following stage to hold the data until another data transition occurs at the receiver input. The \( V_{CM} \) level should lie well within the hysteretic window of the latch \( V_L < V_{CM} < V_H \) to avoid false data latching. The latch is realized by cross-coupled inverters similar to an SRAM cell circuit.
Thesis

Figure 3.9  (a) Half-replica biasing circuit of the proposed receiver. (b) Half-replica biasing with common-feedback to compensate for process variation and mismatch. (c, d) 1000 points Monte Carlo simulation (inclusive of both process variation and mismatch) of the common-mode variation of the receiver with the biasing circuits from (a) and (b) respectively.

Figure 3.9(a) presents the typical half-replica biasing circuit for the receiver. It biases the receiver’s tail transistors in order to maintain $V_{CM}$ at VDD/2 which falls in between $V_L$ and $V_H$. In an actual case, $V_{CM}$, $V_L$ and $V_H$ are affected by mismatch and process variation. The Monte Carlo simulation illustrated in Figure 3.9(c) shows that 376 points out of 1000 points failed to meet the condition ($V_L < V_{CM} < V_H$). Huge variation of $V_L$ and $V_H$ are due to the use of the minimally sized transistors at the latch to reduce its input loading and static current during data latching. To make
The improved biasing circuit reduces the $V_{CM}$ variation as low as ±18 mV as depicted in Figure 3.9(d). This also guarantees that $V_{CM}$ always lies within the hysteretic window of $\min(V_H - V_L) = 110$ mV and $\max(V_L - V_H) = -110$ mV. The receiver input offset is also an important aspect to verify when its input signal swing is just over a few hundred mV. With the new biasing circuit applied, the input offset of the receiver is as low as ±30 mV, which is confirmed by Monte Carlo simulation, with the minimum channel length drawn at the differential input gates (Figure 3.10). Even if the $V_{CM}$ variation and input offset are combined (±48mV), the overall variation is still well within the differential hysteretic window of ±110mV; thanks to cross-coupled receiver configuration and proposed biasing circuit.
Simulation

The proposed transceivers in a $3 \times 3$ hybrid CCIs array are integrated and simulated in the post-layout level. The transceiver from Figure 3.3(c) is used for single-ended CCI. To compare fare and square with the results from Figure 3.4(a) and (b), the dimension specified in Figure 3.5(c), which provides the same pads overlap area and spacing of its comparison peer, is used to extract the coupling model from FEM simulation. The simulated eye diagrams of RX_OUT from single-ended and differential receivers are shown in Figure 3.11(a) and (b), respectively. Both eye diagrams are very clean compared to their counterparts in Figure 3.4 and the jitter is as small as 30 ps when the simulation is done with PRBS signals. The noise level from both eye diagrams is well controlled within the hysteretic windows and thus, bit error rate from data recovery will be significantly improved.

Figure 3.11 (a, b) the simulated eye-diagram of the recovered data pulses from the peripheral and centre CCIs respectively in the proposed hybrid CCIs array.

**3.3.3 Simulation**

The proposed transceivers in a $3 \times 3$ hybrid CCIs array are integrated and simulated in the post-layout level. The transceiver from Figure 3.3(c) is used for single-ended CCI. To compare fare and square with the results from Figure 3.4(a) and (b), the dimension specified in Figure 3.5(c), which provides the same pads overlap area and spacing of its comparison peer, is used to extract the coupling model from FEM simulation. The simulated eye diagrams of RX_OUT from single-ended and differential receivers are shown in Figure 3.11(a) and (b), respectively. Both eye diagrams are very clean compared to their counterparts in Figure 3.4 and the jitter is as small as 30 ps when the simulation is done with PRBS signals. The noise level from both eye diagrams is well controlled within the hysteretic windows and thus, bit error rate from data recovery will be significantly improved.
3.4 Measurement

The architecture of the Build-In Self-Test (BIST) circuit is depicted in Figure 3.12. The key function of the BIST is to check the bit-error-rate (BER) of the central transceiver in three different scenarios (SC<1:3>). When SC<1> or best test scenario is enabled, only the central transceiver is operating. During SC<2>, all the transceivers are activated and transmitting PRBS signals with different seeding. The worst case scenario occurs at SC<3> when all the peripheral transceivers are transmitting the same set of data at which the highest amount of common-mode noise is introduced to the central CCI. The SC selection is done at MUX and the selected transmitting signals are sent to the drivers. Subsequently, the transmitting signals are transferred through the pseudo-hybrid CCIs array whose configuration is shown in Figure 3.6(c) to mimic the performance of the actual 3D integrated hybrid CCIs array as illustrated in Figure 3.5(c).
The BER block will check the data transmission error by comparing the recovered data from receivers with the transmitted data which is delayed to match the timing. The shift register stores all the control signals of each and every block. Figure 3.13(a) shows the microphotography of the test chip. The I/Os of the test macro are placed on one side of the die to allow face-to-face integration. However, actual face-to-face stacking-based testing was not conducted due to the alignment and contamination issue during face-to-face die bonding. In the Pseudo-TxRx block, the transmitters are located at the side of the pseudo-hybrid CCIs array and the receivers are placed right under the pseudo-hybrid CCIs (Figure 3.13(c)).
The Figure 3.14 shows the measured BER bathtub curves from the central and peripheral transceivers at 2.31 Gbps. During the measurement with the SC<1>, the eye opening of 0.6UI is observed at the central receiver. Providing that there is only one transceiver operating, 0.4UI jitter is largely due to the VCO jitter, power supply and substrate noise. At the SC<2>, its eye-opening is dropped to 0.5UI by 0.1UI which is mainly caused by the random common-mode shift due to the crosstalk. During the SC<3> which is the worst case, its eye-opening is found to be at 0.4UI dropped from 0.5UI at the SC<2>. 100% increase in jitter from SC<2> to SC<3> is not only because of the worst case crosstalk but also due to the power supply noise imposed by simultaneous switching of all the peripheral transceivers.
Nonetheless, the proposed receiver performs reasonably well with crosstalk related jitter no greater than 86ps or 0.2UI at the worst condition. In simulation with an ideal voltage source, the jitter increases only by 43% from SC<2> to SC<3>. On the other hand, all the peripheral single-ended transceivers perform much better at SC<3> where all the transceivers transmit the same data set than at SC<2> where the random data are transmitted (Figure 3.14(b)). Although the measurement is done only at 2.31 Gbps due to the speed limitation of the BIST, the maximum frequency of the transceivers is expected to be higher. The static power of the Pseudo-TxRx block is found to be larger than that of simulation by 73 µW due to process variation. The total power of the Pseudo-TxRx is 1.11 mW at 20.79 Gbps or 0.053 mW/Gbps. The power of the face-to-face stacked 3D-TxRx block is expected to be quite close to that of the pseudo-TxRx block since the front-end receivers carry fixed dc-current.

Table 1 summarized the performance of the overall design. In comparison with, the proposed design shows the best power efficiency per data bit. The design
also scores the highest rating in the area efficiency (Mb/s/µm²) because only 2 µm spacing between CCIs is expected in the 3D CCIs array.

### 3.5 Conclusion

This chapter proposes a hybrid CCIs array structure to achieve crosstalk cancellation and smaller footprint of its predecessors. In the proposed array, CCIs can be placed as close as with 2 µm distance generating the maximum jitter of 86 ps. The interleaved integration of both single-ended and differential CCIs in one array reduces the layout area by as much as 52.3%. We also proposed a variation tolerant receiver with a bias circuit for recovering data even under a large common-mode level shift from the crosstalk. The fabricated pseudo-TxRx block with 9 channels achieves total data rate of 20.79 Gbps and the power of 53 µW/Gbps in 65nm CMOS technology.

---

**Table 1. Performance Summary and Comparison**

<table>
<thead>
<tr>
<th></th>
<th>This Work</th>
<th>[15]</th>
<th>[12]</th>
<th>[25]</th>
<th>[16]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CMOS Tech</strong></td>
<td>65nm</td>
<td>130nm</td>
<td>130nm</td>
<td>180nm</td>
<td>350nm</td>
</tr>
<tr>
<td><strong>Supply Voltage</strong></td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
<td>N.A</td>
<td>N.A</td>
</tr>
<tr>
<td><strong>Gbps/ch</strong></td>
<td>&gt;2.31</td>
<td>1.23</td>
<td>1.7**</td>
<td>1.8</td>
<td>1.35</td>
</tr>
<tr>
<td><strong>Total Power(mW)</strong></td>
<td>0.122</td>
<td>0.134</td>
<td>0.256</td>
<td>5.4</td>
<td>8.8</td>
</tr>
<tr>
<td><strong>mW/Gbps</strong></td>
<td><strong>0.053</strong></td>
<td><strong>0.108</strong></td>
<td><strong>0.151</strong></td>
<td>3</td>
<td><strong>6.57</strong></td>
</tr>
<tr>
<td><strong>CCI size(µm²)</strong></td>
<td>11×11*</td>
<td>8×8</td>
<td>8×8</td>
<td>N.A</td>
<td>50×50</td>
</tr>
<tr>
<td><strong>CCI spacing (µm)</strong></td>
<td>2*</td>
<td>~8</td>
<td>8×8</td>
<td>36×36</td>
<td>50×50</td>
</tr>
<tr>
<td><strong>Mb/s/µm²</strong>*</td>
<td><strong>12.49</strong></td>
<td><strong>4.8</strong></td>
<td>N.A</td>
<td><strong>0.53</strong></td>
<td><strong>0.54</strong></td>
</tr>
</tbody>
</table>

*the estimated value based on Pseudo-Hybrid CCIs array characteristic, **clock transmission, ***the area calculation includes both CCI size and spacing between the CCIs.
Chapter 4 Yield Enhancement of Face-to-Face Cu-Cu bonding with Dual-mode Transceivers in 3DICs

4.1 Motivation

The ICI and CCI presented in the Chapter 1-3 show their throughput capability, ease of manufacturing and impeccable reliability. The fundamental idea behind the ICI and CCI development is to provide alternative to its no so reliable counterparts (especially TSV). However, nothing is perfect. One of their limitations is that they still rely on ohmic contact (direct bonding or TSV) for power delivery. The power delivery using ICI shows great promise but its efficiency is incomparable to what ohmic contact can provide. Furthermore, both ICI and CCI transceivers consume static power from analog blocks; not a desirable trait in ultra low-power application. By now, one becomes aware that pros and cons of the proximity communications and ohmic contact are complementary to each other - meaning whatever that comes as pros for proximity communications becomes cons for ohmic contact and vice versa. Therefore, this chapter focuses on the merging of both CCI and ohmic contact into one system to utilize the bests from both approaches.

4.2 Introduction

Three dimensional (3D) integration technology has been considered as a solution to complex system integration making heterogeneous integration viable. With the 3D integration, vertical interconnects become possible, which improve interconnect density, communication distance, propagation delay, dynamic power consumption, and form factors[1]. There are a number of compelling bonding technologies enabling the 3D integration. The combination of bonding and Through-Silicon-Via (TSV) is usually employed when stacking dies Face-to-Back or Back-to-
Back. Otherwise, just bonding is used when stacking is performed in the Face-to-Face configuration. Compared to the primitive solder-based bonding such as Cu/Sn bonding, Cu-Cu bonding has better scalability and excellent mechanical strength to support the stacked layers and electrical conductivity due to the absence of intermetallic compound \cite{55}. The state-of-the-art Cu-Cu thermal fusion bonding technology is employed by Tezzaron for wafer-to-wafer stacking either in Face-to-Face or Face-to-Back \cite{17}. In the Face-to-Face bonding, the exposed top copper pads are aligned and hard pressed for fusion at high temperature. In the latter, the copper pads are deposited on the back side of the substrate for thermal fusion with copper pads from facing die and their electrical connection is completed by Through-Silicon-Via (TSV). Figure 4.1(a) illustrates an example of such bonding interface in 3DICs. With the alignment inaccuracy of less than 1 µm in the wafer level bonding, fine-pitch bonding for high level of interconnects density can be achieved. These bonded pads provide mechanical support as well as electrical connectivity.
According to Tezzaron, only over 50% of the interconnects in stacked CMOS image sensors are utilized for electrical; whereas, the rests are for mechanical support [18].

However, these direct fusion technology and TSVs are not highly regarded as reliable. Copper surface imperfections such as micro-scratches from the Chemical Mechanical Polishing (CMP) [56], micro-roughness, dishing, wafer war-page and contamination [57] are the main culprits of the copper fusion failures (Fig 1(b)) and cracks, and undercuts in TSVs [58] undermine the electrical connectivity of the Face-to-Back stacked tiers. An example of the failed copper fusion interface due to insufficient pressure is shown in Figure 4.1(b) [59]. Its bonding interface has only a few contact points when insufficient pressure is applied and thus, the resistance of the bonding interfaces increases significantly affecting the signal integrity. Even if the entirely separated pads exhibit strong coupling between the two due to their closed proximity, the receiver designed for ohmic contact is not able to recover the signal due to the undefined DC level at the receiver input. In order to recover the data, the receiver should recognize the interface condition and adapt accordingly. The addition of redundancy to signal paths is a sample solution to improve the bonding reliability; however, the reliability improvement is based on statistics and more redundancy is required to achieve near 100% reliability. This solution might be acceptable for DC and low frequency signals but not preferable for digital signaling because the parasitic capacitance from the redundant paths increases the dynamic power consumption [46, 47].

To address this issue, we propose a dual mode transceiver with the capability of built-in-self-test/repair (BIST/R) for Face-to-Face stacked Cu-Cu bonding without adding any redundant paths. A resistance sensor embedded in the proposed receiver
monitors the bonding interface condition during the system start-up and autonomously reconfigures the receiver to work as either the ohmic contact mode when two pads are intact or the capacitive coupling mode when the pads are separated. The sensor is self-contained and not requiring any other peripheral scanners and complex control circuitry which are usually required in many reported TSV BIST/R [46, 47]. However, this transceiver is only suited for the Face-to-Face stacked Cu-Cu bonding where the Cu pads can provide necessary coupling required for the signaling.

In this brief, we explicitly analyze the proposed dual-mode transceiver for Cu-Cu bonding in the Face-to-Face stacked 3D ICs. In section II, we describe the BIST/R system for the Face-to-Face Cu-Cu bonding utilizing the dual-mode transceiver. The analysis of the proposed transceiver with simulations is presented in section III. Finally, in section IV, the conclusion is drawn.

4.3 Dual-Mode Transceiver as BIST/R for Face-to-Face Copper Thermal Fusion Bonding

Figure 4.2(a) represents the model of copper thermal fusion interface at the near failure condition [60]. Two rough surfaces of both TX and RX pads are modelled as one perfectly smooth pad (RX) and equivalent rough pad (TX). The gap between two pads is (h) which is the distance from RX pad to mean surface of TX pad. Considering the worst case scenario, only a single contact point presents whose asperity diameter is given as 2a. Due to the small constriction shown in Figure 4.2(a), the electrons travel in ballistic motion and thus, the constriction effective resistance is \( R_{12} \) whose equation is given in Figure 4.2(b). The value of \( R_{12} \) deviates from the
ohmic resistance when the constriction radius is less than the electron mean path of the conductor. As shown in Figure 4.2(c), $R_{12}$ become higher than its ohmic resistance when the contact radius is less than its electron mean path which is 38nm for copper. In the $R_{12}$ equation, $k$ is the Knudsen ratio and $\Gamma(k)$ is a function decreasing from 1 to 0.694 as $k$ increase from 0 to $\infty$. The gap between TX and RX pads forms a coupling capacitor, $C_C$ which can be extracted from FEM simulation and is in parallel with $R_{12}$. $R_W$ is the wire resistance from the Back-End-of-the-Line (BEOL) of the die itself and $C_P$ is the parasitic capacitance. The overall model of the interface is shown in Figure 4.2(b). This model is used in simulation with the proposed dual-mode transceiver.
The design of the dual-mode transceiver is based on the capacitive coupling transceiver [38] shown in Figure 4.4. The driver is nothing but an inverter chain. At the receiver side, the input and the output of I1 are connected through a pass-gate resistor (TG) as a feedback to detect the data transition or voltage change at TX and pulses are produced at the I1 output. When pull up and pull down network of I1 are having same strength, RX is biased at VDD/2. Pulses from I1 are amplified by I2 and latch at L1 to recover full swing digital signal. I3 acts as a buffer to drive the load.

To make the transceiver working in ohmic contact, the pass-gate should be removed or disable. Therefore, in the proposed dual-mode transceiver, a resistance sensor is added in addition to transceiver at Figure 4.4 so as to enable and disable the pass-gate in accordance with the interface resistance (Figure 4.3). The sensor is
Figure 4.5 (a) proposed resistance sensing circuit with reduced transistors count, (b) layout of the dual-mode receiver with the resistive sensing circuit from (a) in dotted line configuration.

composed of sense enable transistors (M1 and M2), write transistors (W1 and W2), a latch (L2), a reset transistor (M3) and a load balancing transistor (M4). During the chip start-up, the node (P) from L2 is reset to GND and TX will be reset either to GND or VDD depending on the system configuration. At that point, the pass-gate is already enabled and $V_{RX}$ is weakly biased at VDD/2. During the sensing period after the resetting, W1 and W2 compete each other to overwrite P and N. When RX is connected to GND, W2 wins over W1 and P is pulled up to VDD disabling the pass-gate and then, the transceiver is configured to operate in ohmic mode. Similarly, when RX is closed to VDD, W1 beats W2 and N is pulled to GND disabling the
pass-gate. When the interface resistance is high, the RX node will be biased around VDD/2 regardless of whether TX is reset to GND or VDD. In this case, both W1 and W2 are not strong enough to overwrite the initial value in L2. Therefore, the transceiver’s pass-gate remains enabled for the rest of the operation to serve as a feedback resistor for I1.

When the initial state of TX during reset is known, the resistance sensor can be reconfigured as shown in Figure 4.5(a) to reduce the transistors count in the transceiver block. The configuration in solid line (includes W1, M2, M3, TG and L2) is for TX when the data is low during chip start-up. When TX is high during chip start-up, the dotted line configuration, which includes M1, W2, M3, M4, TG and L2, can be employed. Figure 4.5(b) is the layout of the dual-mode receiver block in 65nm CMOS technology including the resistance sensor design in dotted line configuration. The layout has proved that the circuit overhead is not significant and the receiver fits under the copper pad dimension of 5 µm × 5 µm which is later used for comprehensive analysis and simulation.

### 4.4 Simulation

Generally, the transceiver has one zero due to $C_C$ and two complex poles due to the parasitic capacitance at RX and OUT1 (see Figure 4.3). Therefore, the highest response occurs at the poles location and the highest operational frequency should be centre at the poles for the maximum effectiveness. Knowing the minimum $C_C$ value available to the system, the transistors at (I1) and the feedback transmission gate are designed based on the iteration of ac simulation and the parasitic extraction.
The proposed resistance sensor which sense $V_{RX}$ exhibits a hysteric window between $V_L$ and $V_H$. In other words, if $V_{RX}$ is fallen between $V_L$ and $V_H$ during the sensing period, the receiver is configured in the capacitive coupling mode; otherwise,
it is configured in the ohmic mode as depicted in Figure 4.6(a). With (I1) and (TG) designed, the driver impedance and operation mode threshold of R_{12} can be found if the values for V_L and V_H are known. For this experiment, we choose V_L and V_H to be 200mV and 1V respectively and 1kΩ driver impedance for fast switching. This would eventually leads to R_{12} value of 3kΩ. The sigma of V_L and V_H is ~30 mV obtained from 1000 points Monte Carlo simulation (Figure 4.6(a)) and that subsequently varies the trigger R_{12} values from 1 kΩ to 5.6 kΩ. However, this variation is acceptable and does not cause any signal integrity issue. Taking into account the V_L and V_H variation, the resistance sensor has the supply noise margin of 100 mV in the worst case corner (Figure 4.6(a)).

The parametric post-layout simulation of the dual-mode transceiver is carried out with the two 5 µm × 5 µm thermally fused pads model shown in Figure 4.2(b) whose gap (h) and the asperity radius (a) are swept from 10 nm to 100 nm and from 1nm to 0.001nm, respectively. The chosen pad dimension is to achieve enough coupling when h is at 100 nm. The respective C_C values at h = 10 nm, 25 nm, 50 nm and 100 nm are 22 fF, 9.2 fF, 4.7 fF and 2.2 fF. These values are extracted by a FEM simulator using air as a dielectric medium to emulate worst case condition. According to the simulation result shown in Figure 4.6(b), the jitter performance of both ohmic mode and capacitive coupling mode are excellent across all the parametric points at 2Gbps. The maximum propagation delay different across all the points is just 35ps (Figure 4.6(c)), which is good enough for many applications requiring wide I/O signalling without causing significant clock and data skew. The power consumption of the transceiver is peaked at the capacitive coupling mode when h is 10nm, where C_C is the largest (Figure 4.6(d)). The larger C_C value is, the
higher the power consumption is as more charges are coupled to RX node and discharged. When the direct bonding failure rate is 1 ppm reported in Ziptronix process [18] which is done in low temperature unlike high temperature fusion by Tezzaron, the power peaking in the capacitive coupling mode imposes negligible power overhead in overall system. Moreover, the resistance sensor is only activated once during chip start-up and it won’t draw any current during normal operation.

In more realistic scenario, Cu pads are laid across the die in an array manner. The distance between pads can be placed as close as 1.5µm in some commercial 65nm CMOS process with the given the pad dimension of 5µm × 5µm. When properly fused pads are placed at the minimum distance, the short channel length in this Face-to-Face bonding would not cause any crosstalk issue. The crosstalk issue starts to arise only when one of the Cu-Cu bondings fails forcing the receiver to
operate at the capacitive coupling mode. The worst case scenario is when the capacitive coupling channel is surrounded by the ohmic channels as depicted in Figure 4.7. As ohmic channels are signaling at rail-to-rail voltage, it is very likely that the receiver in the capacitive coupling mode picks up the coupling noise and register false data at the latch (L1 in Figure 4.3). We investigated the above issue and quantified the impact of crosstalk on the signaling. To do so, we assume that there is no data transition at TX of the capacitive coupling channel and yet, all the surrounding channels operating at ohmic mode switch data at the same frequency in phase to induce maximum crosstalk to the RX node. Doing so with varying the distance between channels (d) in the simulation, we can find out the proper isolation distance between pads to avoid data error.

The equivalent crosstalk model is shown in Figure 4.8 in which $g_m$ belongs to PMOS and NMOS’s at the receiver front end (I1), $r_{01}$ is the receiver’s (I1) output resistance and $C_L$ is the capacitive load at (I1) output. $R_{TG}$ is the transmission gate resistance, $R'_{TG}$ is $R_{TG}$ in parallel with $R_{12}$, $C_C$ is the capacitance between TX and RX as shown in Figure 4.7 and $C_P$ is the parasitic capacitance associated with RX.
Finally, $C_N$ is the crosstalk coupling between RX and a nearby channel (Figure 4.7).

The transferred function of the model is shown in equation 1.

$$A(j\omega) = \frac{V_{OUT1}}{V_{Nin}} = \frac{8r_0r_{T1}R_{T1}g_mC_N}{r_o+R_{T1}+2g_mr_oiR_{T1}} \frac{\omega}{C_LC_{RX}R_{T1}R_{T1}r_{oi}+C_{L}R_{T1}r_{oi}+C_{RX}R_{T1}G_{T1}R_{T1}r_{oi}} \frac{\omega}{(\omega+1)}$$

(1)

$$\omega_o = \sqrt{\frac{r_{oi}+R_{T1}+2g_mr_{oi}R_{T1}}{C_LC_{RX}R_{T1}R_{T1}r_{oi}}}$$

(2)

From equation 1, the magnitude of $A(j\omega)$ is directly proportional to $R'_{TG}$ which means that having smaller $R_{12}$ at the coupling channel will definitely reduce the amplitude to coupled noise. According to the equation 2, the complex poles of the model are hovering around $\omega_o$ which is inversely proportional to square root of $C_{RX}$. Therefore, having bigger $C_{RX}$; meaning bigger $C_C$ will shift the complex poles to the left suppressing the coupling noise at the frequency of interest (1GHz square wave).

The noise amplitude at the node OUT1 is defined as,

$$V_{OUT1}=4\times \sum_{n=1,3}^{\infty} \left[ \frac{2^n\sqrt{\nu\omega}}{\nu\omega} A(j\omega_n) \right]$$

where $n$ is the harmonics of the square wave from nearby 4 channels. Figure 4.9 is the shmoo plot showing the result of the (h) and (d) swept from 10nm to 100nm and 1.5µm to 5µm respectively where $|V_{OUT1}| < 50$ mV is considered as pass and otherwise it is a fail. However, the selection process of the pass and fail threshold is arbitrary but is based on the design constrains. If the hysteric windows of data latch (L1) is ($VDD/2 \pm 150$) mV and the required noise margin is ±100mV, it would be reasonable to keep crosstalk noise at $|V_{OUT1}|$ to be less than 50 mV. Figure 4.9(a) is the result when $R_{12}$ is infinity and Figure 4.9(b) and (c) represent when $R_{12}$ are at 100kΩ and 10kΩ respectively. Figure 4.9 also confirms the point made earlier where smaller $R_{12}$ and larger $C_C$ suppress noise level. To select
Figure 4.9  Shmoo plots of (h) and (d) swept from 10nm to 100nm and 1.5µm to 5µm respectively. (a) At R12=∞, (b) at R12=100kΩ, (c) at R12=10kΩ. Here VNout<50 mV is a pass; otherwise it is a fail.

the proper (d) value, we have to look into two parameters; R_{12} and (h). Since R_{12} value is hard to predict, on the safe side, we can assume that R_{12} is always infinite. (h) value can be predicted by the finishing capability of the CMP machine and the
diameter of the dust particle presents in the clean room which can be trapped during thermal fusion process and etc. Once maximum (h) value is known in the given process environment, it is easy to decide the spacing (d) of the channels as far as the signal integrity is the concern. According to the simulation result, if (h) value is guaranteed to be less than 10nm, the 5µm × 5µm pads can be placed as close as d = 1.5µm (a minimum allowable spacing by the process) from one another which results in effective channel pitch of just 6.5µm.

4.5 Conclusion

This paper presents a transceiver design solution to improve the yield of the Cu-Cu bonding in Face-to-Face stacked configuration. The transceiver has a resistive sensor which allows it to adapt to the condition of the Cu-Cu bonding. The transceiver works in the coupling mode utilizing the proximity of two Cu pads which are failed to fuse together. Alternatively, it works normally in ohmic mode when two Cu pads are fused together as one. The transceiver performs well in both modes with only 35ps propagation delay different across the two different modes. The power consumption of the transceiver can be higher at the coupling mode than that of in ohmic mode; however, it won’t significantly affect the overall power consumption from the transceivers as a couple of Cu-Cu bonding failure is expected. The crosstalk affect is also found to be insignificant if the gap (h) between two unfused pads is under 10nm. Moreover, this chapter also serves a guideline (atleast in my perspective) for a possible adoption of CCI technology in the industry to supplement the widely adopted Cu-Cu bonding technology.
Chapter 5 High Speed Low Swing Differential Driver with Post-Cursor RC-Modulated De-emphasis for 2.5D Silicon Interposer.

5.1 Motivation

As Moose’s Law continues due to technological scaling, the transistor size has been shrinking down to 14nm feature size allowing processor to go faster and more powerful. A tremendous amount of data is, then, produced from the computation and moving the data between processors or processor to memory remains a challenge because Input/Output (I/O) technology does not scale along with the transistor scaling. Moreover, the advance in mobile computing demands low power I/O due to limited power budget. Therefore, it is a challenging task to design low power I/O for ever increasing data rate requirement. The introduction of silicon carrier or Through-Silicon-Interposer (TSI) allows fine pitch wire-line with tens of thousands of high speed I/O supported by the dense back end of line (BEOL) copper wires (Figure 5.1).

Due to a good matching of thermal coefficient of expansion between the TSI and the active chips, smaller micro-bumps can be used to connect the two resulting in better
impedance continuity compared to bigger solder bumps. The active chips can be placed closer on the TSI and thus, it brings down the wire length to cm range from several inches in backplane communication. However, the fine-pitch thin wire-line in TSI suffers significant losses at high frequency causing inter-symbol-interference (ISI). As of [19], a channel loss at high frequency (10Gbps) is becoming apparent at 5mm length and above. In commercial product such as Virtex-7 whose die size is more than 10mm × 10mm [61], the wire to bridge between FPGA and memory on the interposer can be longer than 5mm. At the wire-length of 10mm and above, the signal is also suffered from long tail effect and thus, multi-taps de- or pre-emphasis scheme is essentially required in I/O driver to compensate it as well as the high frequency losses presented in 2.5D integration on TSI. The encoding circuit for de- or pre-emphasis needs current-mode-logic (CML) gates rather than static logic gates to improve supply-noise immunity and jitter performance. CML gate takes more area and power consumption presumably than the static logic gate. Only a couple of CML gates are needed for 2-tap pre-emphasis encoding [50, 51] while in 4-tap case, the number of CML gates employed increases tremendously in [52, 53]. Furthermore, they are intended for backplane communication, these designs are not well suited for mobile chips which require compact sized low power I/Os. The popular capacitive pre-emphasis driver for Network-on-Chip (NoC) [62] is a very good tender for TSI due to their similar RC losses in the transmission line. However, the transmission line in TSI is likely to be an order of magnitude of longer than the length of NoC (2mm in [62]) and such capacitive pre-emphasis is not enough to compensate for high frequency losses at the TSI channel. In [63], 3-taps capacitively coupled Finite-Impulse-Response (FIR) filter for transmitter is presented for longer on-chip channel
length of 5mm. The capacitance weight for each taps is optimized and pre-fixed based on the transmission line characteristic and it managed to achieve 4.9Gbps/ch. Since the tap weights cannot be re-configurable for various channel length and characteristic of the TSI defined by the system designers, this architecture is again not suitable for TSI channel. A single-tap de-emphasis voltage-mode driver is presented in [64] for 10mm channel but due to less number of taps compared to that [63], yet for longer channel length, it achieves only 2Gbps/ch.

One thing in common among the transceivers for on chip channel is that their taps are implemented with delay element rather than pipe-lined taps found in the transceivers [50-53] for off chip channel. The main reason is that the impedance matching is relaxed in on-chip channel but not in off-chip channel at which precise impedance matching can only be obtained with pipe-lined taps de/pre-emphasis driver. The results from [62-64] also imply that multi-taps emphasis driver is a necessity for longer TSI channel. Moreover, it is more power efficient to do channel equalization at the driver rather than at the receiver side at which pipe-lined Decision-Feedback-Equalization are commonly used as witnessed in [19].

In general, there is a need to develop adjustable tap-weight multi-taps de/pre-emphasis driver especially for TSI channel. Therefore, in this chapter, a RC-modulated post-cursor de-emphasis is proposed for the low swing high speed differential driver.

5.2 Driver Design

The schematic of the driver with de-emphasis is shown in Figure 5.2(a). It consists of three parts; pre-driver, de-emphasis circuit and driver itself. The driver
circuit is made up of four NMOS transistors; above two transistors formed pull-up network and other two for pull-down network. The driver is supplied with lower supply voltage so that both pull up and pull down transistors are in linear region as well as to provide low swing signal. Unlike any other de- or pre-emphasis [50-53], the proposed de-emphasis is far less complicated because it consists only one XOR gate (G1) for de-emphasis encoding. “EN” or enable signal is to on/off the G1 as well as the de-emphasis circuit. The delay block (DLY) is responsible for $t_{DLY}$ shown in Figure 5.2(b). The combination of G1 and DLY blocks produces the signal pulse.
Thesis (DE) which rises exponentially after time delay of $t_{DLY}$ from every data transition of 0 to 1 or 1 to 0 and then after, goes back to zero very quickly at next data transition. This whole process repeats again at the next data period as shown in Figure 5.2(b).

The exponential rising of DE pulse is possible by weakening the pull-up network of the G1 so that G1 will charge the gate capacitance ($C_g$) of $M_{<2:4>}$ slowly (with time constant $RC$) whereas, the pull-down network of the G1 remains strong so that the “DE” pulse can quickly return to ground. The crossbar is composed of transistors $M_{<2:7>}$.

The output from the XOR gate (DE) modulates the crossbar strength through $M_{<2:4>}$ and $EM_{<0:2>}$ signal is to pre-select the de-emphasis strength. The transistors pairs ($M_2$, $M_5$), ($M_3$, $M_6$) and ($M_4$, $M_7$) are arranged in series configuration. Their relative strength ratio is (1:2:4) respectively and therefore, de-emphasis strength is adjustable up to 8 steps $EMS_{<0:7>}$.

As the signal long tail effect is exponential in nature, the exponentially varied crossbar strength due to “DE” perfectly suppress the long tail effect if “DE” signal is well characterized. Since the de-emphasis strength is modulated continuously unlike being discretely modulated in [50-53], the proposed driver with de-emphasis introduces less noise to the transmission line. The driver output impedance is designed to be 30Ω and it reaches to 22.5Ω when the de-emphasis crossbar is fully turned on. However, such impedance variation has little or no impact by the signal reflection since the transmission line is very lossy at 20mm length.

5.3 Post Layout Simulation

Figure 5.3 (a) presents the frequency response of the de-emphasis system which schematic representation is shown in the figure itself. The schematic is only the half circuit of the differential system where $R_{T1}$ and $R_{T2}$ are the termination resistances.
and the rest (dly, R, C and RV) are components related to de-emphasis system. Among them, C itself is the gate capacitance of M<2:4> and the variable resistor (RV) is the crossbar, M<2:7> themselves. The frequency response of the proposed de-emphasis system suggests that with full de-emphasis strength at EMS<7>, it is able to reduce the DC level by ~5dB and leaving only the 3dB cut off at 5GHz which is the desire operation frequency. The variation of the de-emphasis system is analyzed by applying 5% variation to dly, R, C and RV. DC level is found to be varied from -4.98dB to -5.67dB and unnoticeable frequency response variation occurs at the 5GHz.

Figure 5.3  (a) a frequency response of the proposed system with and without proposed de-emphasis, (b) impulse response of the proposed system with 8 steps de-emphasis strength, (c) a data eye diagram of the 20mm length micro-strip line without de-emphasis, (d) a data eye diagram of 20mm length micro-strip line with EMS<7>. 
The pulse response of the implemented de-emphasis with various de-emphasis strength is shown in Figure 5.3(b). The white arrow indicates the direction of which de-emphasis strength is increased from EMS<0> to EMS<7>. At EMS<0>, the response tail presents up to 3rd post cursors (C3). Such post cursors tail is eliminated completely at EMS<7>. Therefore, this de-emphasis method can achieve up to 3-taps post-cursor equivalent de-emphasis with just one delay element. The cross section of the 20mm long transmission line used for simulation is shown in Figure 5.3(b). The eye diagram of the 10Gbps system with and without de-emphasis is presented in
Figure 5.3(c and d) and the visual clue suggests that there is a significant improvement with the proposed de-emphasis applied.

The system performance with de-emphasis is summarized in Figure 5.4. The eye height improves significantly from -15.56mV to 57mV over the course of EMS<0> to EMS<7>. SNR improves by 210% at the EMS<7> compared to that of EMS<0>. At EMS<7>, the eye width is found to be widened by 100% compared to EMS<0>. Along with the aforementioned improvement, the total power consumption of the driver and de-emphasis circuit increases from just 1.92mW at EMS<0> to 2.37mW at EMS<7>. The main cause of this increment is due to the reduction of crossbar (M<2:7>) resistance as EMS goes from 0 to 7 and thus, more current flows from VDDL and GND. Moreover, due to the modulation of crossbar resistance due to de-emphasis, the driver’s differential termination resistance varies from 30Ω at EMS<0> to 22.5Ω at EMS<7>. As mentioned previously, such impedance modulation won’t have significant impact on the signal integrity due to the RC-domination of the transmission line [19]. The receiving end of the transmission line is terminated with 70Ω.
Table 2 Performance comparison of the proposed driver with previous works from literature.

<table>
<thead>
<tr>
<th>Technology CMOS</th>
<th>[53]</th>
<th>[51]</th>
<th>[54]</th>
<th>[52]</th>
<th>This Work**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed (Gbps)</td>
<td>16</td>
<td>12.5</td>
<td>4</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Taps</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3*</td>
</tr>
<tr>
<td>Pow (Drv+EM) (mW)</td>
<td>57.5</td>
<td>4.58</td>
<td>~5</td>
<td>3</td>
<td>2.37</td>
</tr>
<tr>
<td>Pow Efficiency (pJ/bit)</td>
<td>3.3</td>
<td>0.36</td>
<td>1.2</td>
<td>0.3</td>
<td>0.225</td>
</tr>
<tr>
<td>Area (Drv.+EM) (µm²)</td>
<td>65×180</td>
<td>~50×50</td>
<td>~100×100</td>
<td>~150×50</td>
<td>17×19</td>
</tr>
</tbody>
</table>

* equivalent taps.  ** Post-Layout Simulation

The performance comparison between the proposed driver and others competing driver’s designs is made in Table 2. It clearly shows that the proposed driver design surplus all the competing designs in both power and area efficiency. It is noted that although the proposed driver has 3-taps equivalent, its power consumption is less than 2-taps driver. The most significant improvement over the comparison works is in the footprint of the driver. The proposed driver footprint is 6.7× less than the 2-taps driver reported in [49]. Figure 5.5 shows the fabricated test chip layout in 65nm CMOS technology. The test chip includes DLL, serializer, de-serializer and BER counter for testing and measurement.
5.4 Conclusion

This chapter presents RC-modulated de-emphasis for low swing differential driver for low power chips integrated with the silicon interposer. Instead of digitally trimmed the post-cursor strength, the charge accumulated over the post-cursor is used to modulate the de-emphasis strength. Therefore, the amount of computation required for de-emphasis encoding is relaxed. The post-layout simulation has proved that the power consumption of the proposed drive with de-emphasis consumes less power and yet occupies far less area than its counterparts.
Chapter 6 Conclusion

In this thesis, a wide range of interconnect technologies for 3DICs are explored ranging from the ohmic contacts such as TSV and Cu-Cu bonding to the wireless contacts links such as CCI and ICI. However, the key focus of this thesis is given to CCI which is for Face-to-Face stacked chips and the goal is find out where the CCI technology can be fitted into the industrial usage. The CCI technology was first introduced by S. A. Kuhn in 1995 and was intensively promoted by R. Drost and his affiliates in 2000s. However, CCI technology never took off to industry adoption mainly due to the advanced in the flip chip technologies such as micro-bump and Cu-Cu bonding technologies. However, it never stops me from doing research on CCI. The key motivation of the whole work is because CCI is more reliable than its counterparts and it can be cheaply fabricated. The work begins with the simultaneous bi-directional CCI signaling (Chapter 2) which is an extension of other works based on the uni- and bi-directional signaling for CCI. The concept of sending data from two ends of the channel simultaneously is proved in emulated-CCI channel but is faced with several challenges as well. The channel parasitic capacitance and sharing of available voltage margin to two bits are contributed to the degradation of overall voltage margin and channel performance. The process variation of the three series capacitors in the channel also causes the voltage offset and the size of these capacitors are significantly larger than what would have been in uni-directional signaling. Therefore, the subsequence work is focus on to improve the signal integrity of uni-directional CCI instead. Hence, in Chapter 3, the prominent issue in CCI signaling, crosstalk, is addressed by introducing the hybrid CCI array in conjunction with the high Common-Mode-Rejection-Ratio receiver. Its capability in
the crosstalk cancellation is proved in 65nm CMOS test chip running at 2.31Gbps/ch. What makes it special compared to other crosstalk cancellation is that it can provide the same level of performance like others do in a smaller channel pitch. Despite the crosstalk issue addressed, CCI must rely on the flip chip technology for the power and low frequency signals. Moreover, the static power at the analog front-end receiver for CCI is not desirable trait when there are many CCI in the 3DICs. These two issues are the biggest drawbacks of CCI and they cannot be mitigated. One technology closely resemble to CCI is the Cu-Cu bonding except for the fact that two top-level Cu pads from two dies are brought together and are fused under high temperature and pressure. The Cu-Cu bonding is highly adopted in advanced CMOS chip but it also suffers from occasional bonding failure due to contamination and wafer warpage. Although such bonding failure makes the Cu-Cu bonding useless to the conventional voltage driver, this unbounded channel is still usable if the transceiver would have been designed for CCI. Therefore, the next step for pushing the CCI technology to a new level is to combine it with the Cu-Cu bonding technology. In Chapter 4 the strength of CCI technology is incorporated into the Cu-Cu thermal compression Face-to-Face bonding to improve the yield of the compression bonding. This is possible with the introduction of self-configurable dual-mode transceiver which can work either at the ohmic channel or the capacitive channel. The transceiver robustness is proved in any possible bonding failure scenarios.

When it comes to 3DIC, one must not ignore the 2.5D Integration which is regarded as an intermediate technology that eventually enable 3DIC integration possible. Without 2.5D integration, the processor would not be able to fully utilize
the through-put capability of 3D stacked HBM as mentioned in Chapter 1.4. Therefore, in Chapter 5 three taps equivalent RC-modulated reconfigurable de-emphasis driver for 2.5D Through-Silicon-Interposer (TSI) is proposed. TSI interconnect appears to have same RC loss as of on chip interconnect but it is longer and its length can be varied depending on the system requirement. Here, RC-modulation is utilized to perform de-emphasis rather than the data pipe-lining to achieve multi-taps equivalent de-emphasis without much power and area overhead. Since its RC-modulation is reconfigurable based on the different insertion loss in the channel, this proposed de-emphasis driver is much useful for any length of TSI interconnect up to 20mm.
References


