MANCHESTER 1824

The University of Manchester

Energy Efficient Inter-Chip Communication in Heterogeneous Application Domains

Vasilis F. Pavlidis

Advanced Processor Technologies group

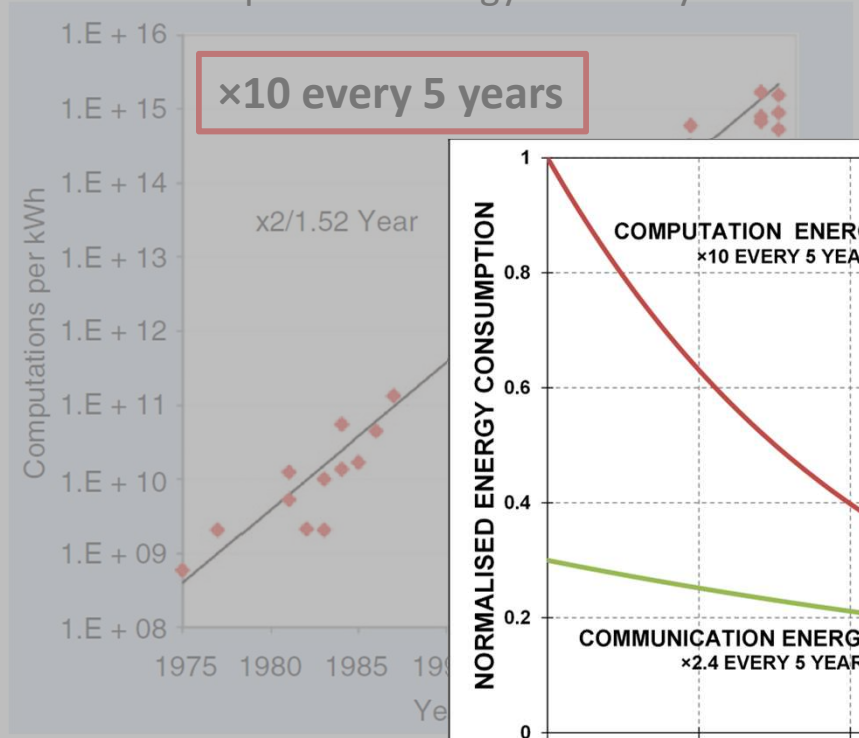
University of Manchester

pavlidis@cs.man.ac.uk

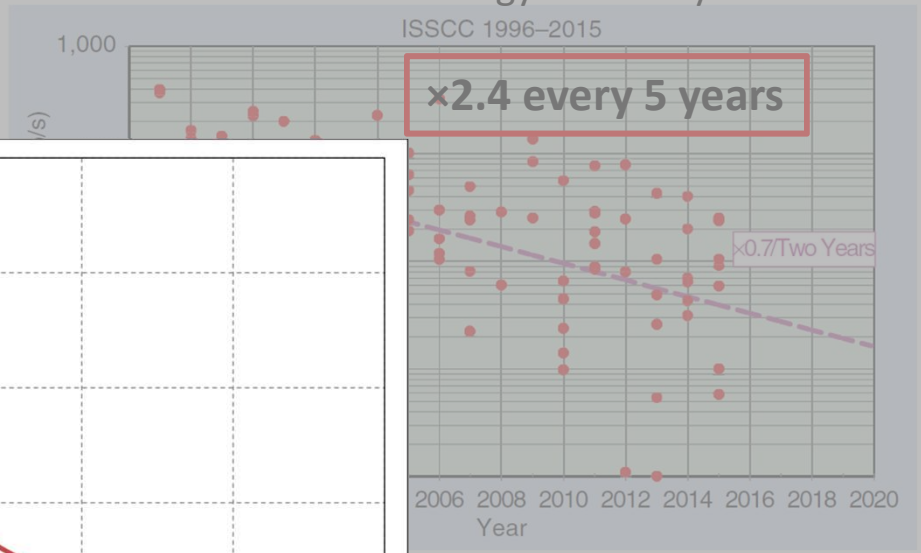
http://www.cs.man.ac.uk/~pavlidiv/

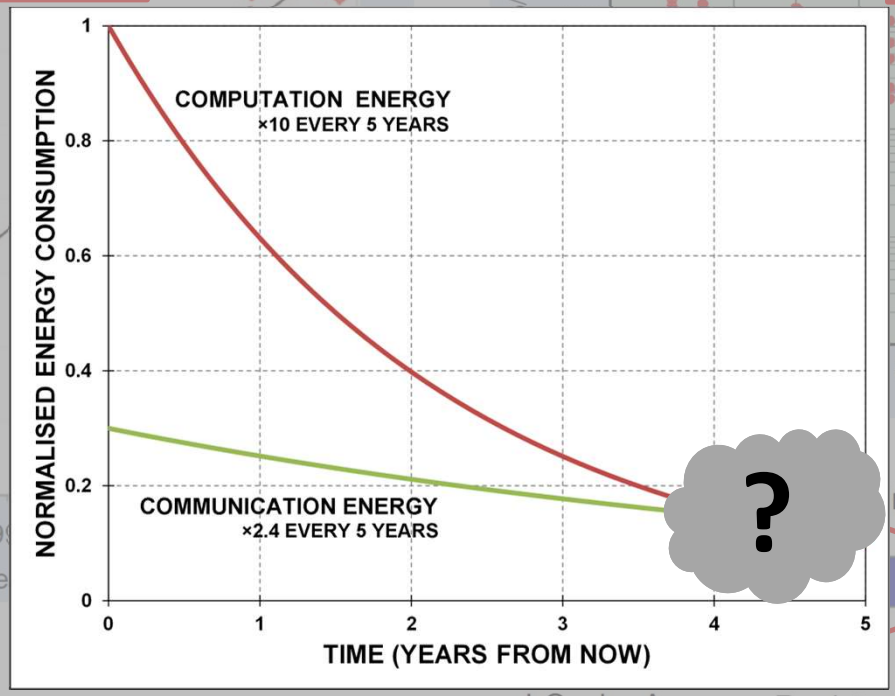
Why Communication Matters?

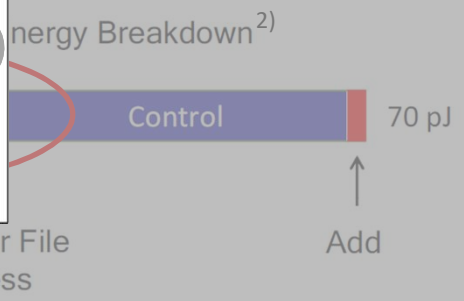
- Computation energy efficiency trend [1)]



- Communication energy efficiency trend [1)]





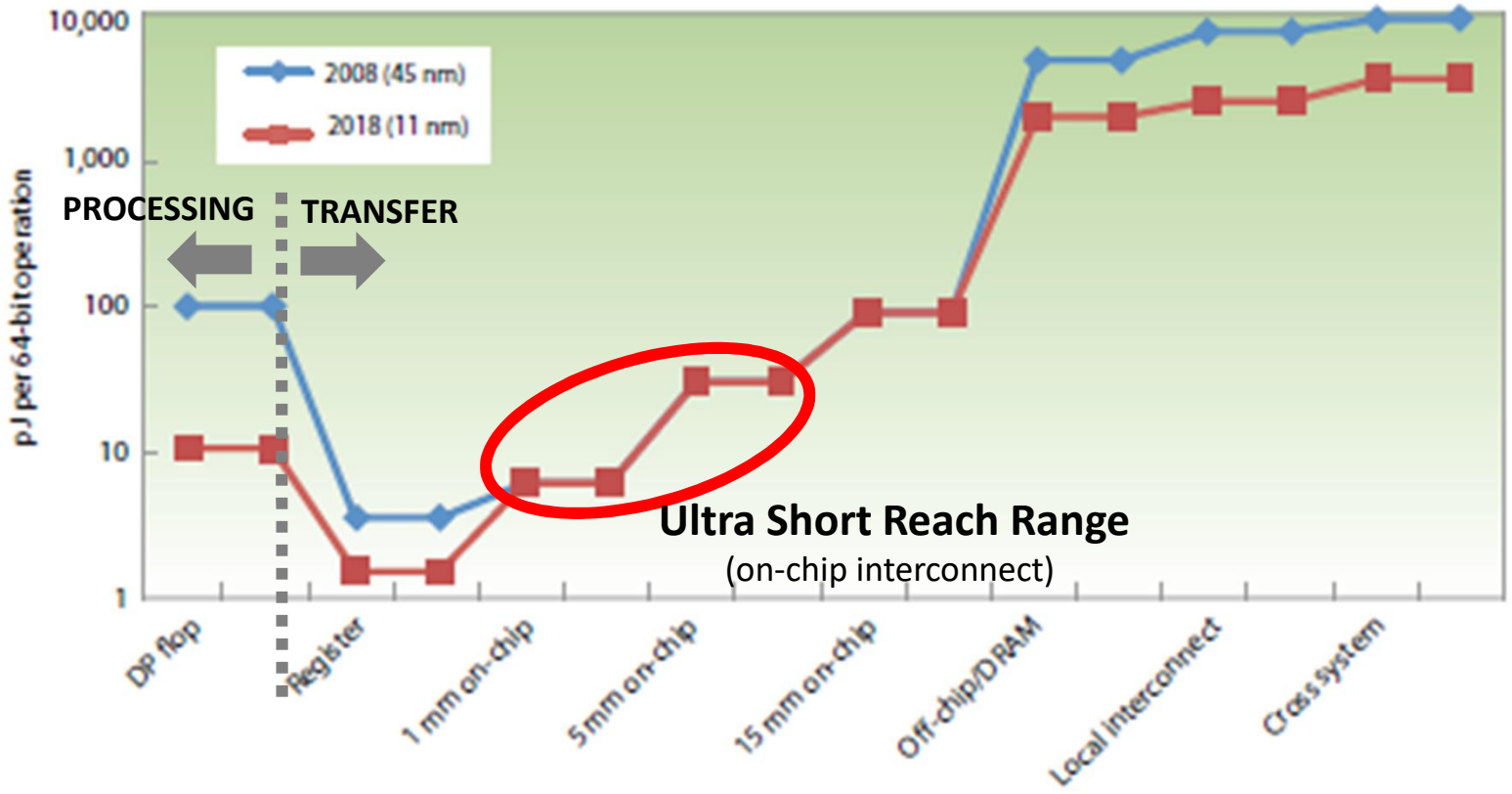


~30% of energy for communication

[1)] H. Tamura, "Looking to the Future: Projected Requirements for Wireline Communications Technology," IEEE Solid-State Circuits Magazine, Vol. 7, No 4, pp. 53 – 62, 2015.

[2)] M. Horowitz, "Computing's energy problem (and what we can do about it)," IEEE International Solid-State Circuits Conference, pp. 10 – 14, Mar. 2014

Energy Cost of Data Processing and Transfer



*P. Kogge and J. Shalf, "Exascale Computing Trends: Adjusting to the "New Normal" for Computer Architecture," *Computing in Science & Engineering,* Vol. 15, No. 6, pp. 16-26, Nov/Dec 2013.

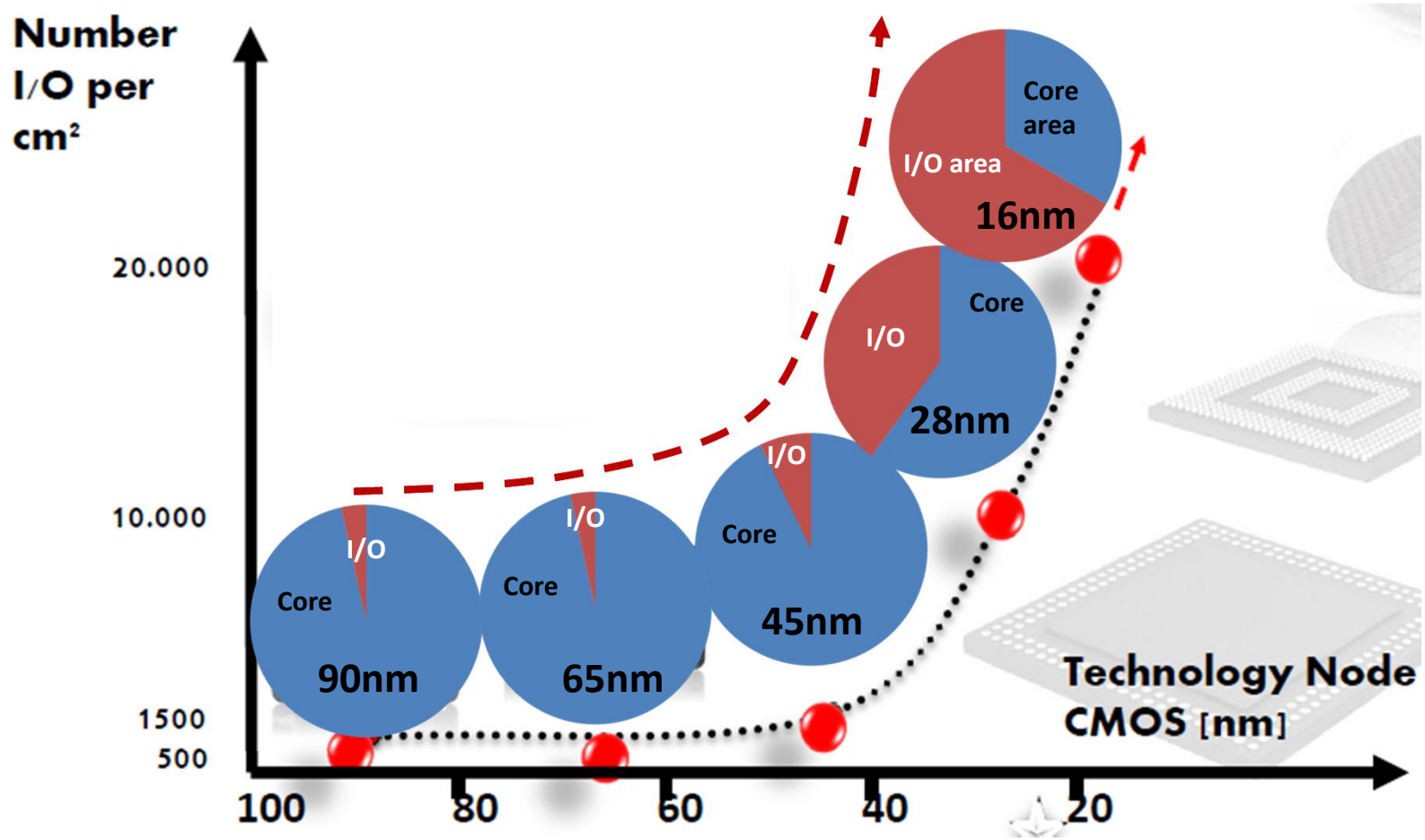
Communication in Ubiquitous Computing

- 20 pJ/Operation
- ADD Op. of two 64-bit operands
- 30% energy for communication

- Total budget: 0.31 pJ/bit
- Communication: 0.1 pJ/bit

[1)]S. Borkar, "Role of Interconnects in the Future of Computing," IEEE Journal of Lightwave Technology, Vol. 31, No. 24, pp. 3927 – 3933, Dec. 2013.

I/O Scaling Limitation



*D. Dutoit, "3D System Design: opportunities, challenges, enabling solutions and methodologies," *Proceedings of the 3D IC Conference*, December 2014.

PART I – WIRELINE COMMUNICATION

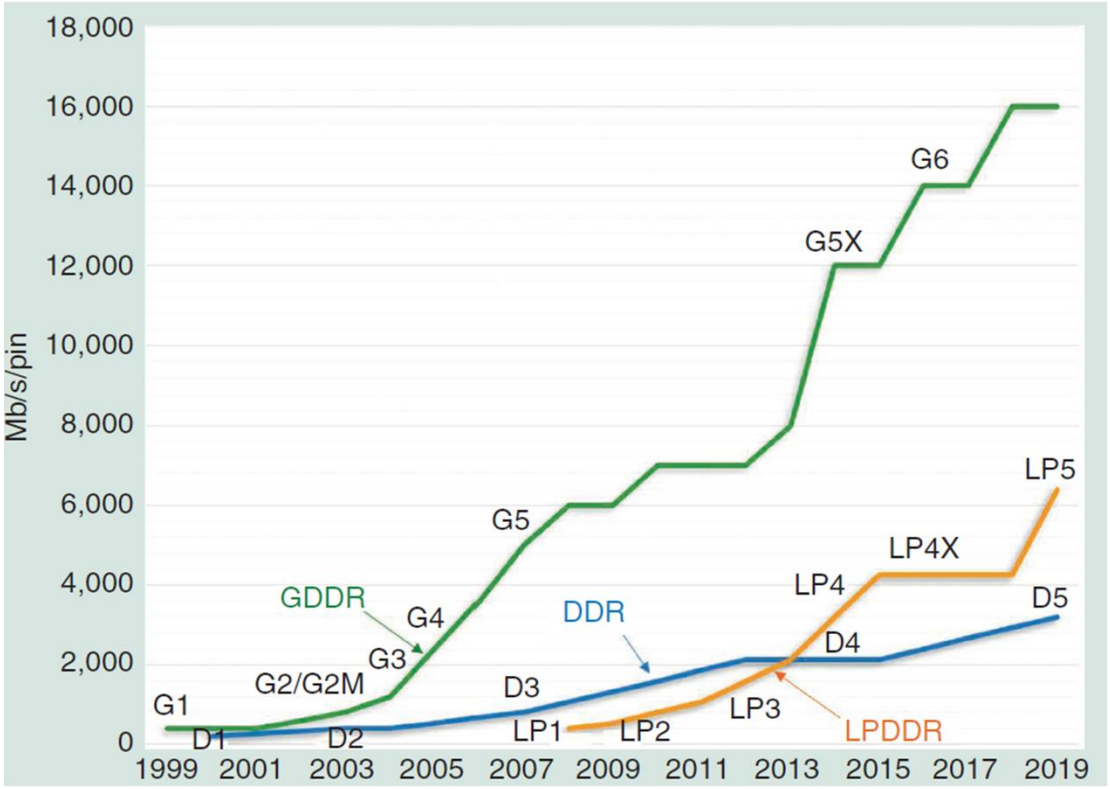
MANCHESTER 1824

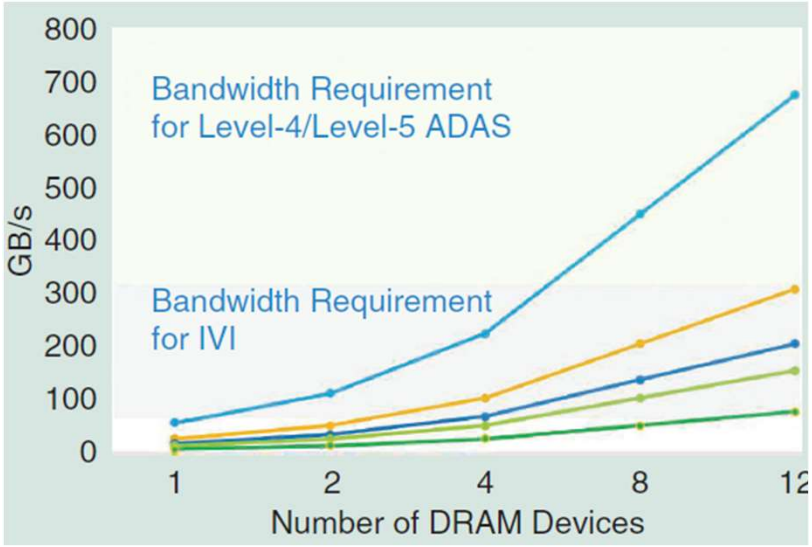
The University of Manchester

Part I - Outline

- Single-Ended Chip-to-Chip Communication
- Low-Swing Signaling for Energy Efficiency
- Data Encoding for Energy Efficiency
- Summary

Serial or Parallel Communication

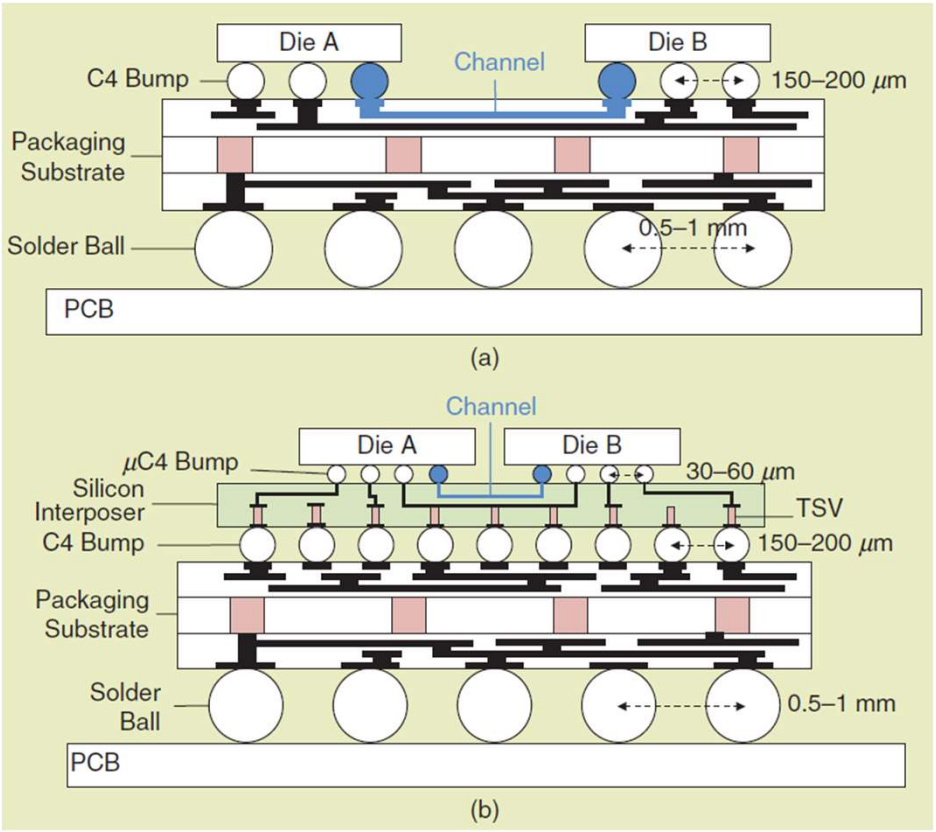


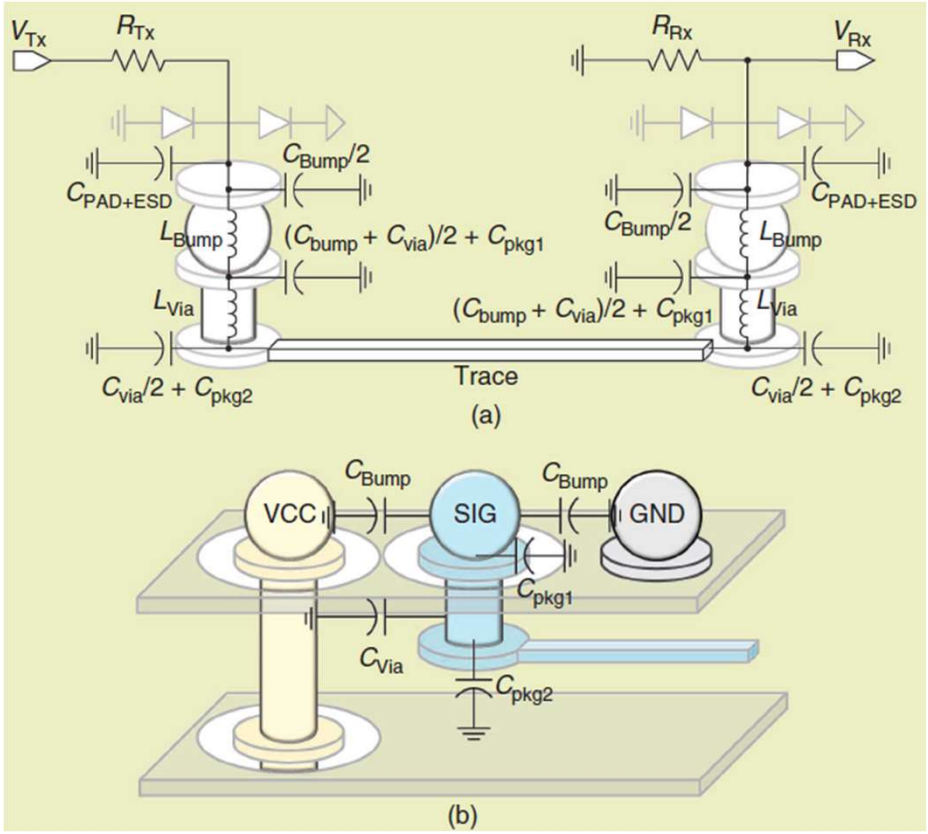


GDDR6 at 14 Gb/s (x32)
LPDDR5 at 6.4 Gb/s (x32)
LPDDR4x at 4.266 Gb/s (x32)
DDR5 at 6.4 Gb/s (x16)
DDR4 at 3.2 Gb/s (x16)

*T. M. Hollis *et al.,* "Recent Evolution in the DRAM Interface," *IEEE Solid State Circuits Magazine*, Vol. 11, No. 2, pp. 14-30, Spring 2019.

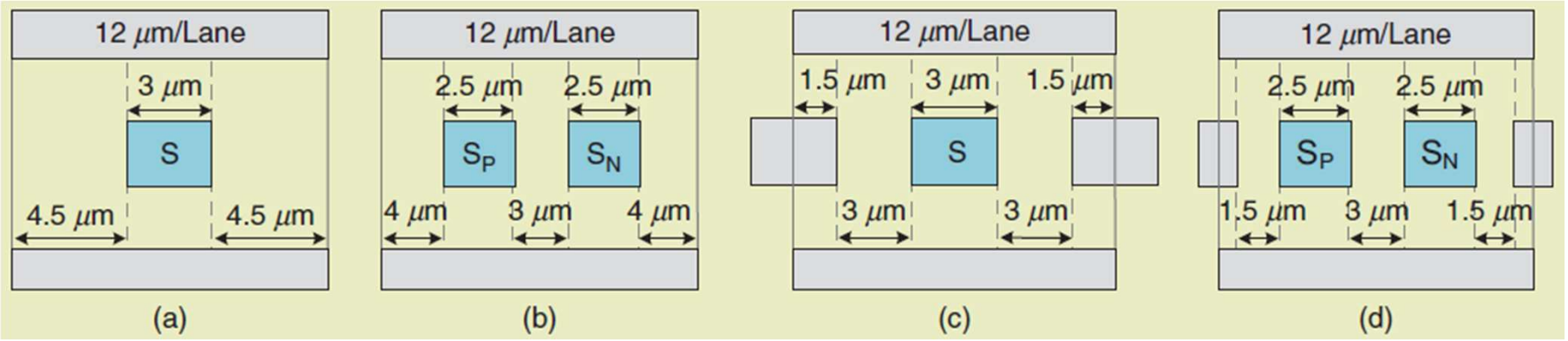
Modern Packaging Solutions

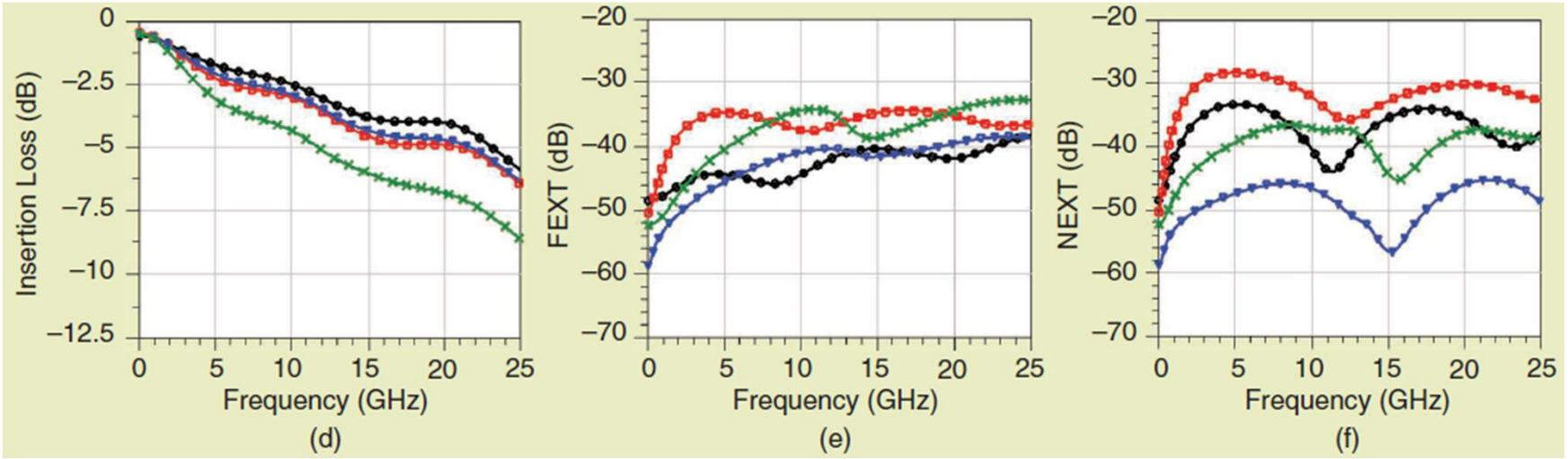




*B. Dehlaghi, N. Wary, and T. C. Carusone, "Ultra-Short Reach Interconnects for Die-to-Die Links," *IEEE Solid State Circuits Magazine*, Vol. 11, No. 2, pp. 42-53, Spring 2019.

Differential or Single-Ended Signaling



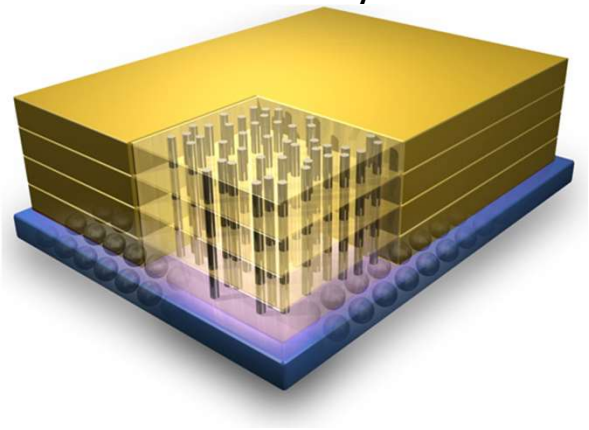


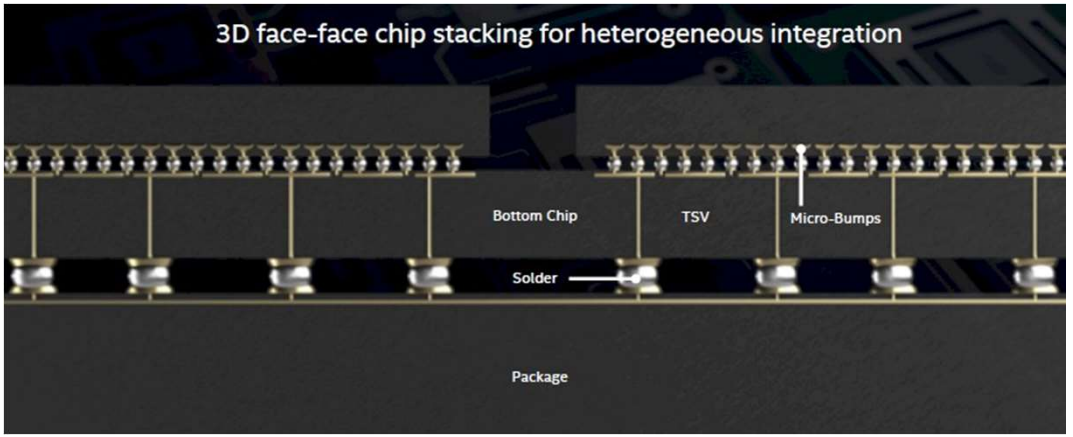
- Single ended with ground shield
- Single ended
- Differential
- Differential with ground shield

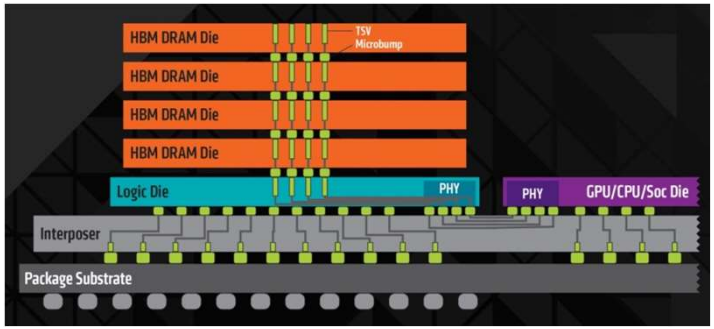
*B. Dehlaghi, N. Wary, and T. C. Carusone, "Ultra-Short Reach Interconnects for Die-to-Die Links," *IEEE Solid State Circuits Magazine*, Vol. 11, No. 2, pp. 42-53, Spring 2019.

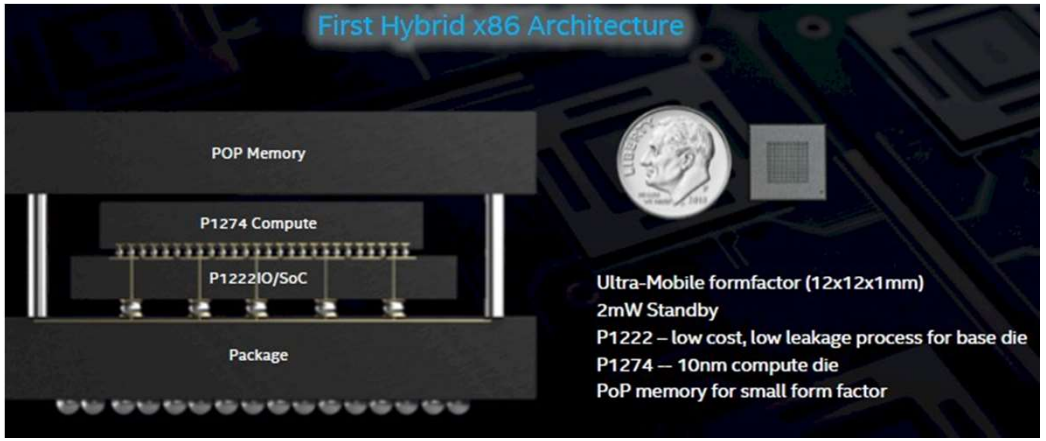
Commercial 2.5-D/3-D Integrated Systems

- Micron HMC (hybrid memory cube)
  - 15x bandwidth of DDR3
  - 70% less energy per bit
  - Lower latency









- AMD GPU– R9 NANO 4G
- Foveros Intel's packaging architecture

MANCHESTER 1824

The University of Manchester

Part I - Outline

- Single-Ended Chip-to-Chip Communication
- Low-Swing Signaling for Energy Efficiency
- Data Encoding for Energy Efficiency
- Summary

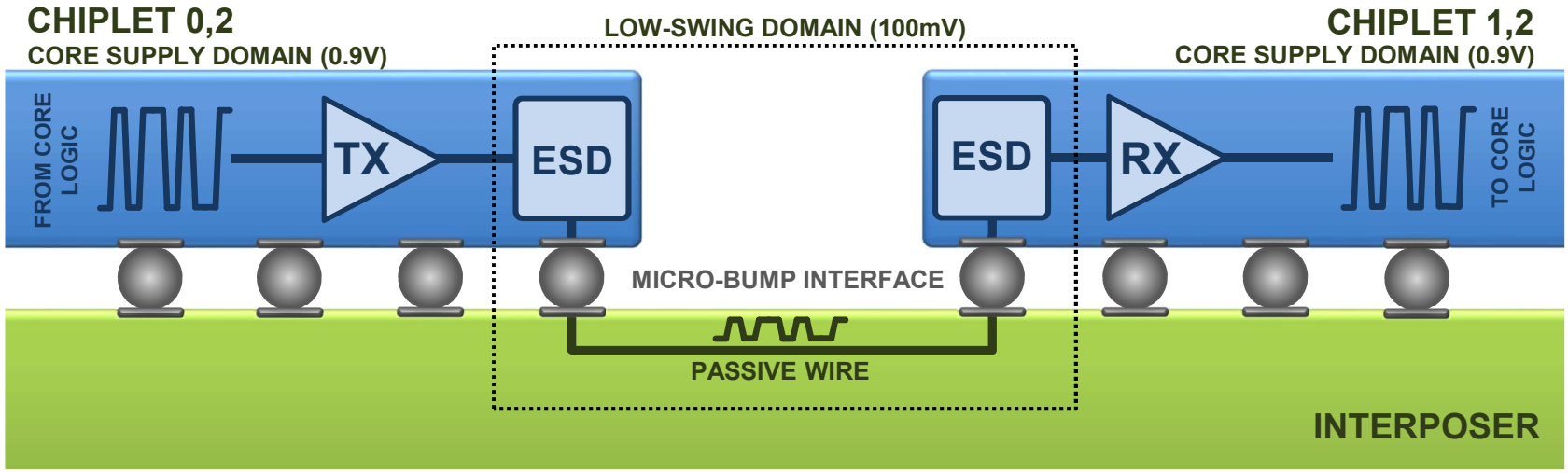
MANCHESTER 1824

The University of Manchester

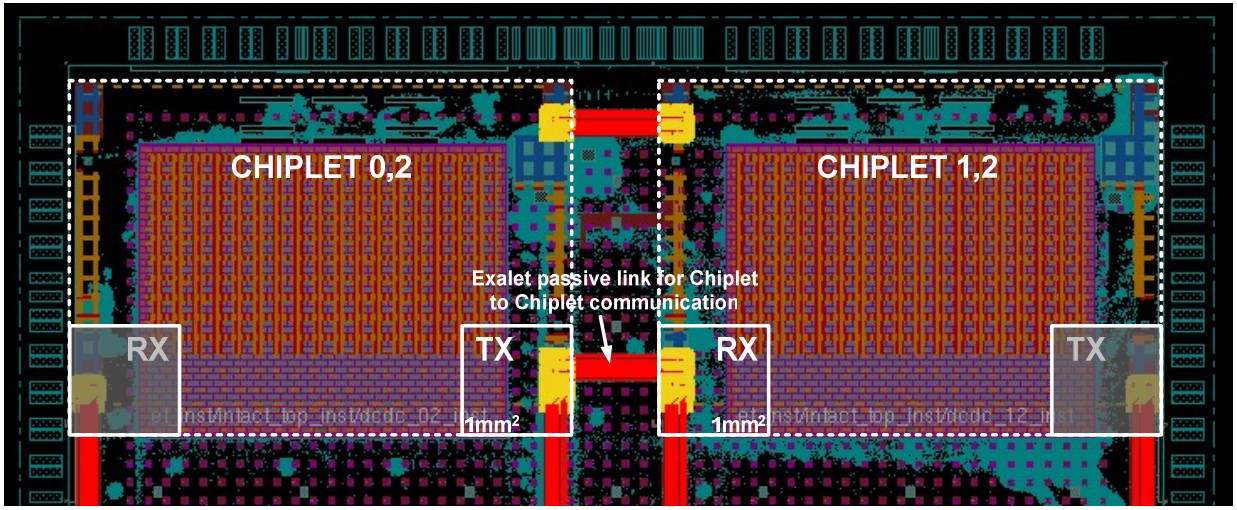
Part I - Outline

- Single-Ended Chip-to-Chip Communication
- Low-Swing Signaling for Energy Efficiency
  - Low-Swing Transceiver Design
  - Simulation Results
- Data Encoding for Energy Efficiency
- Summary

Chip-to-Chip Communication Link in ExaNoDe

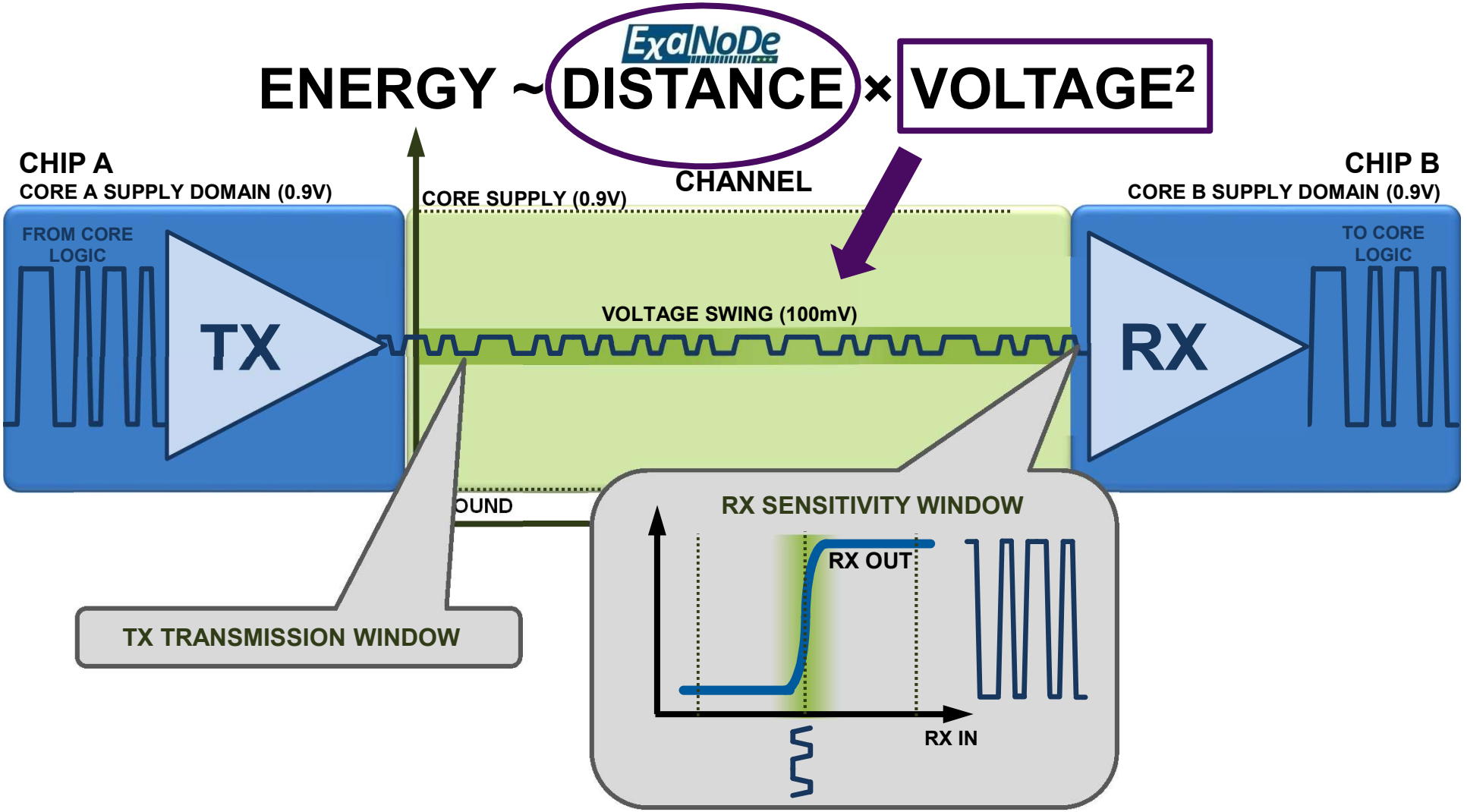


Cross section of the interconnect with single-ended low swing I/O interface

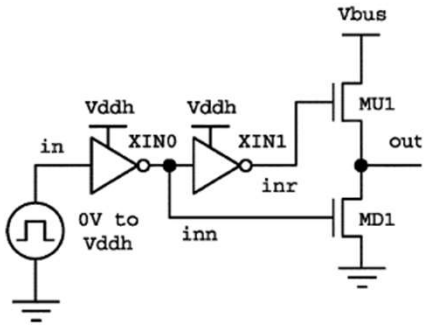


Physical view of the interposer with the projected location of the transceiver and the passive link

Low Swing Signaling Scheme

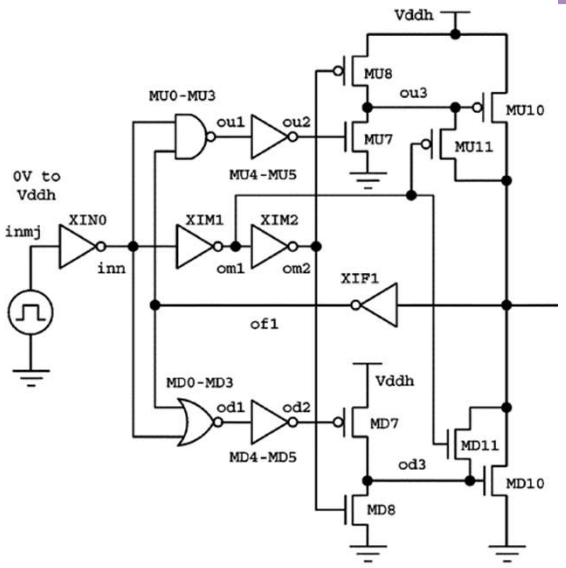


State-of-the-Art Low Swing Transmitters



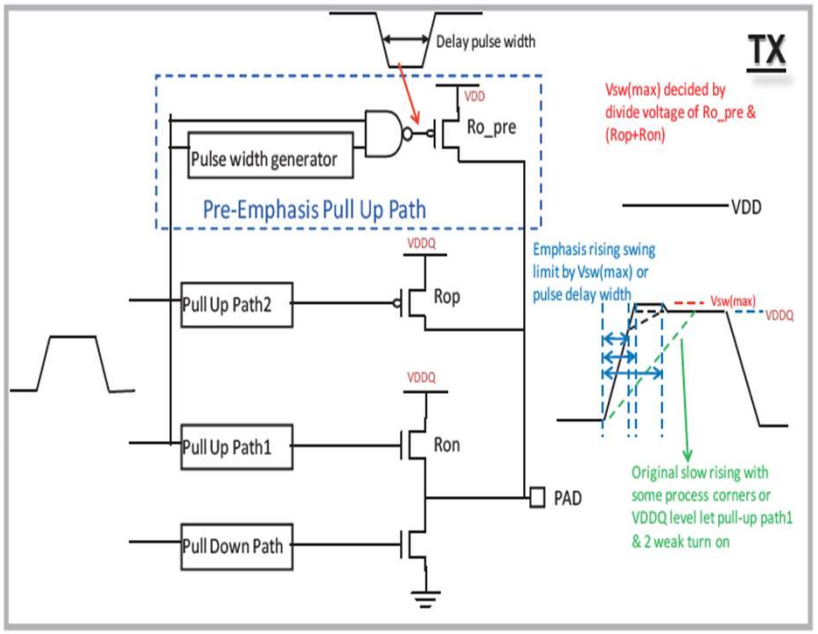
Threshold voltage based asymmetric buffer[1)]

- ✔ Compact design
- ✖ Two voltage domains



Threshold voltage based symmetri buffer[1)]

- ✔ One voltage domain
- ✖ "Large" low swing



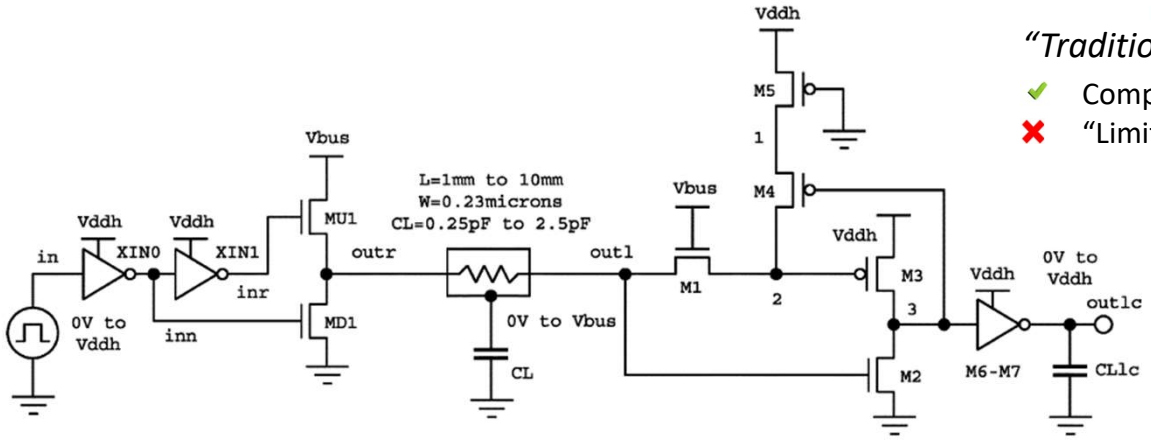
Self-timed delay based buffer[2)]

- ✔ One voltage domain
- ✔ Adjustable output swing
- ✖ Requires bus weak keepers
- ✖ Sensitive to parameter variability

[1)] J. C. Garcia Montesdeoca, "CMOS Driver-Receiver Pair for Low-Swing Signalling for Low-Energy On-Chip Interconnects," IEEE Transactions on VLSI Systems, Vol. 17, No. 2. Feb 2009

[2)] M. S. Lin, et al., "An extra low-power 1Tbit/s bandwidth PLL/DLL-less eDRAM PHY using 0.3V low-swing IO for 2.5D CoWoS application," IEEE Symposium on VLSI Technology, Jun. 2013

State-of-the-Art Low Swing Receivers

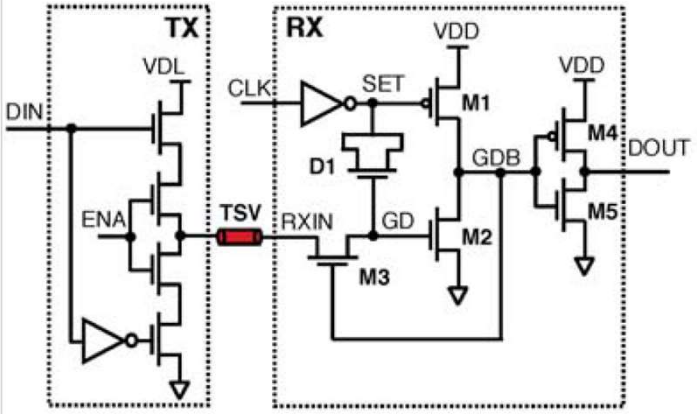


"Traditional" level shifter[1)]

- ✓ Compact design
- ✗ "Limited" low swing (min V_{TH} of M2)

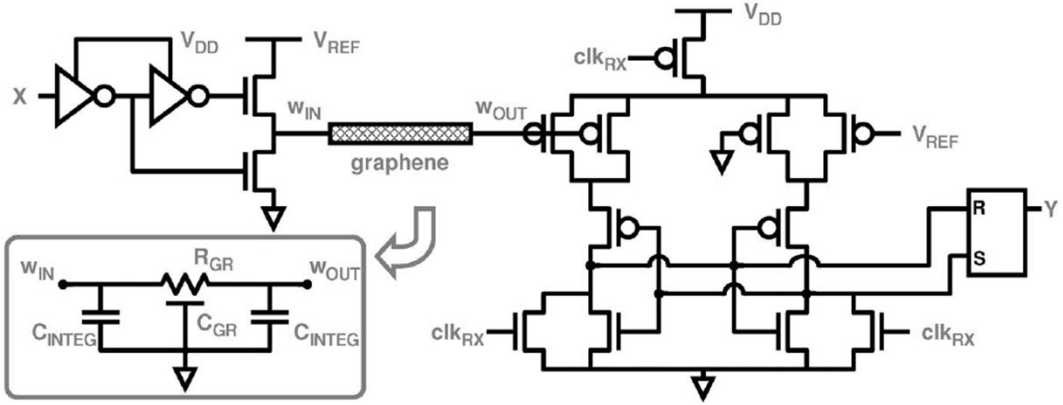
Comparator-based RX[3)]

- ✓ Differential operation
- ✓ High sensitivity and speed
- ✗ Requires clock and reference



Diode-gated sense amplifier shifter[2)]

- ✓ Compact design
- ✗ Requires precise clock alignment
- ✗ Sensitive to parameter variability

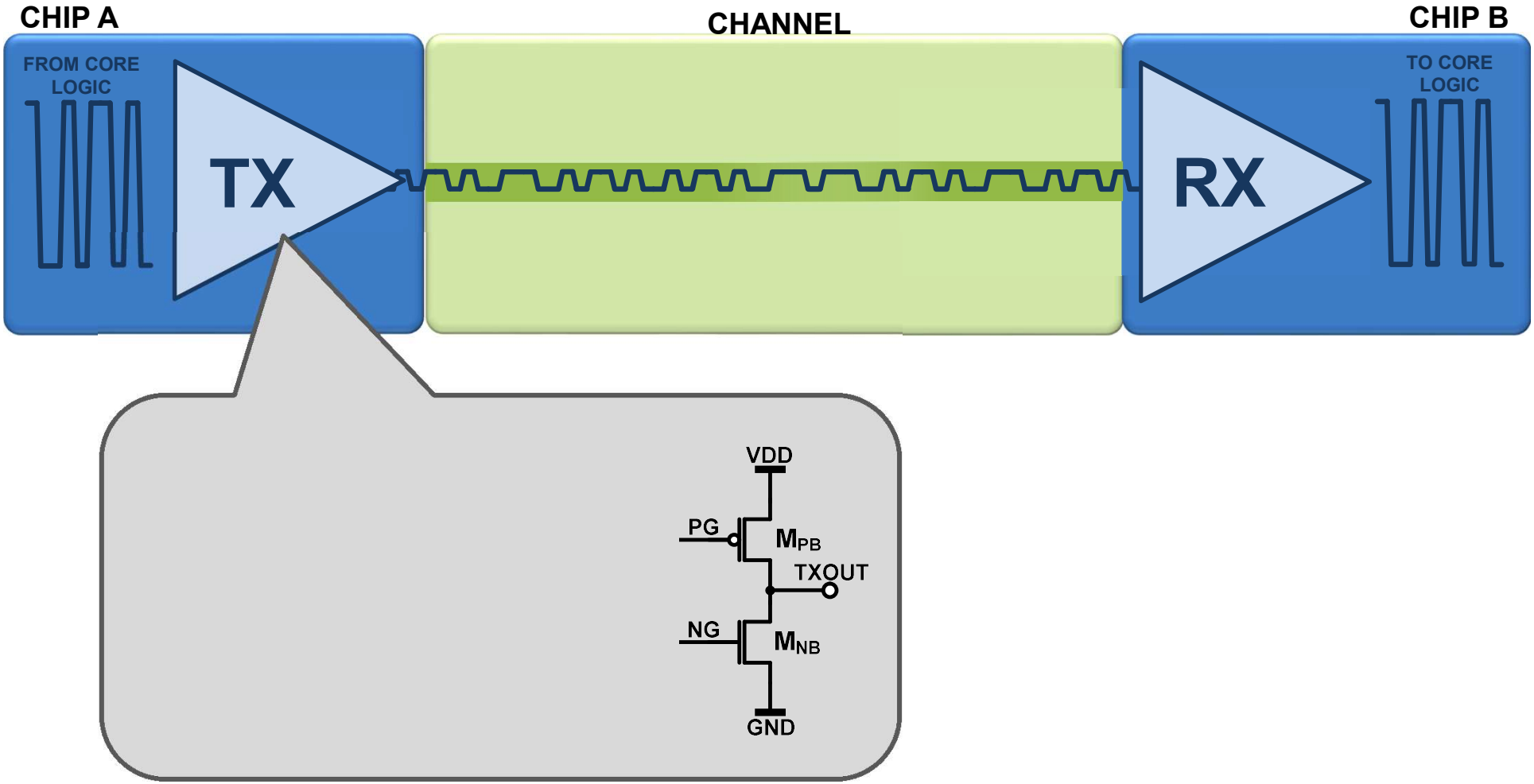


[1)] J. C. Garcia Montesdeoca, "CMOS Driver-Receiver Pair for Low-Swing Signalling for Low-Energy On-Chip Interconnects," IEEE Transactions on VLSI Systems, Vol. 17, No. 2. Feb 2009

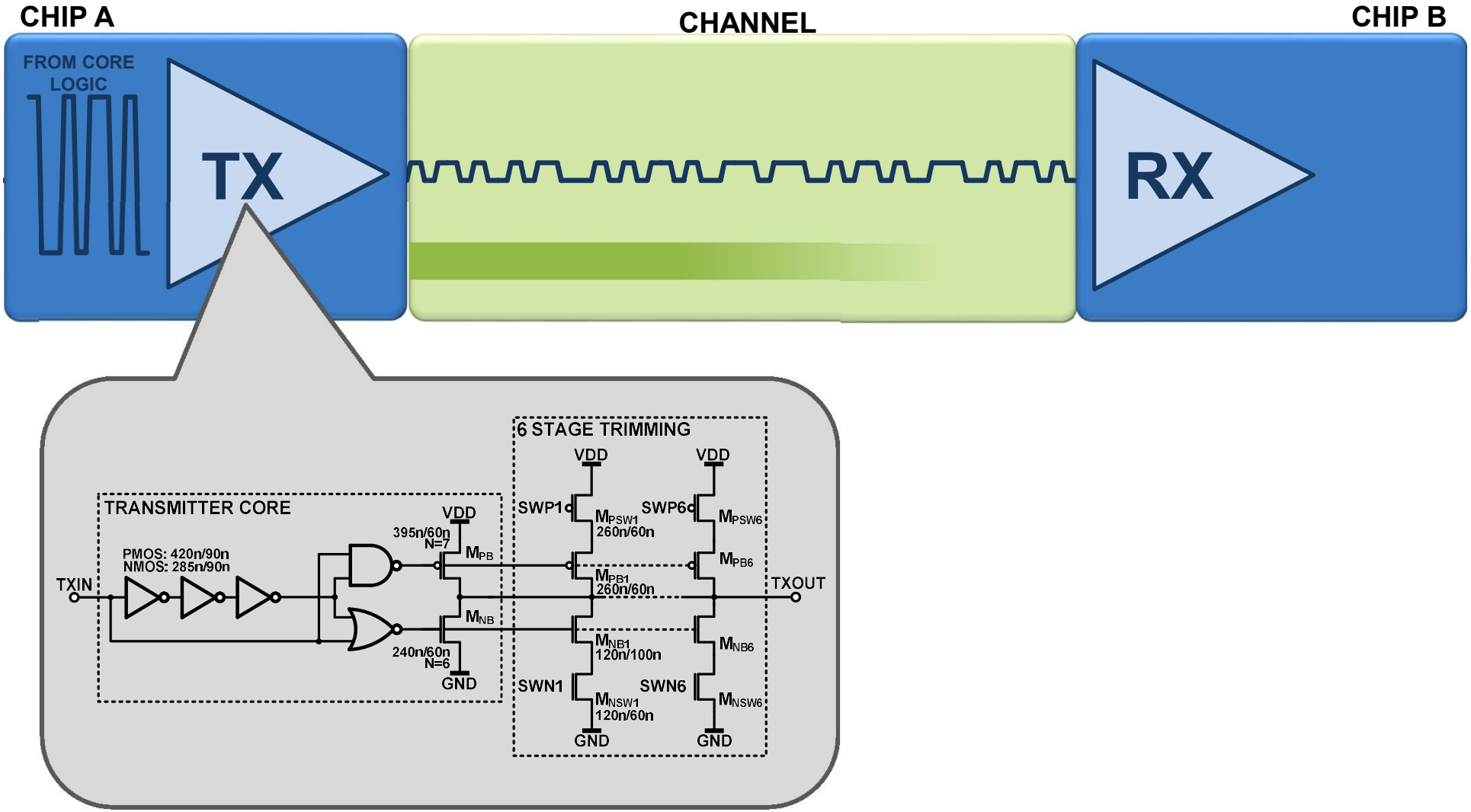
[2)] Y. Liu, et al., "A Compact Low Power 3D I/O in 45nm CMOS," ISSCC 2012

[3)] K. J. Lee, et al., "Low-Swing Signaling on Monolithically Integrated Global Graphene Interconnects," IEEE Transactions on Electron Devices, Vol. 57, No.12, Dec. 2010

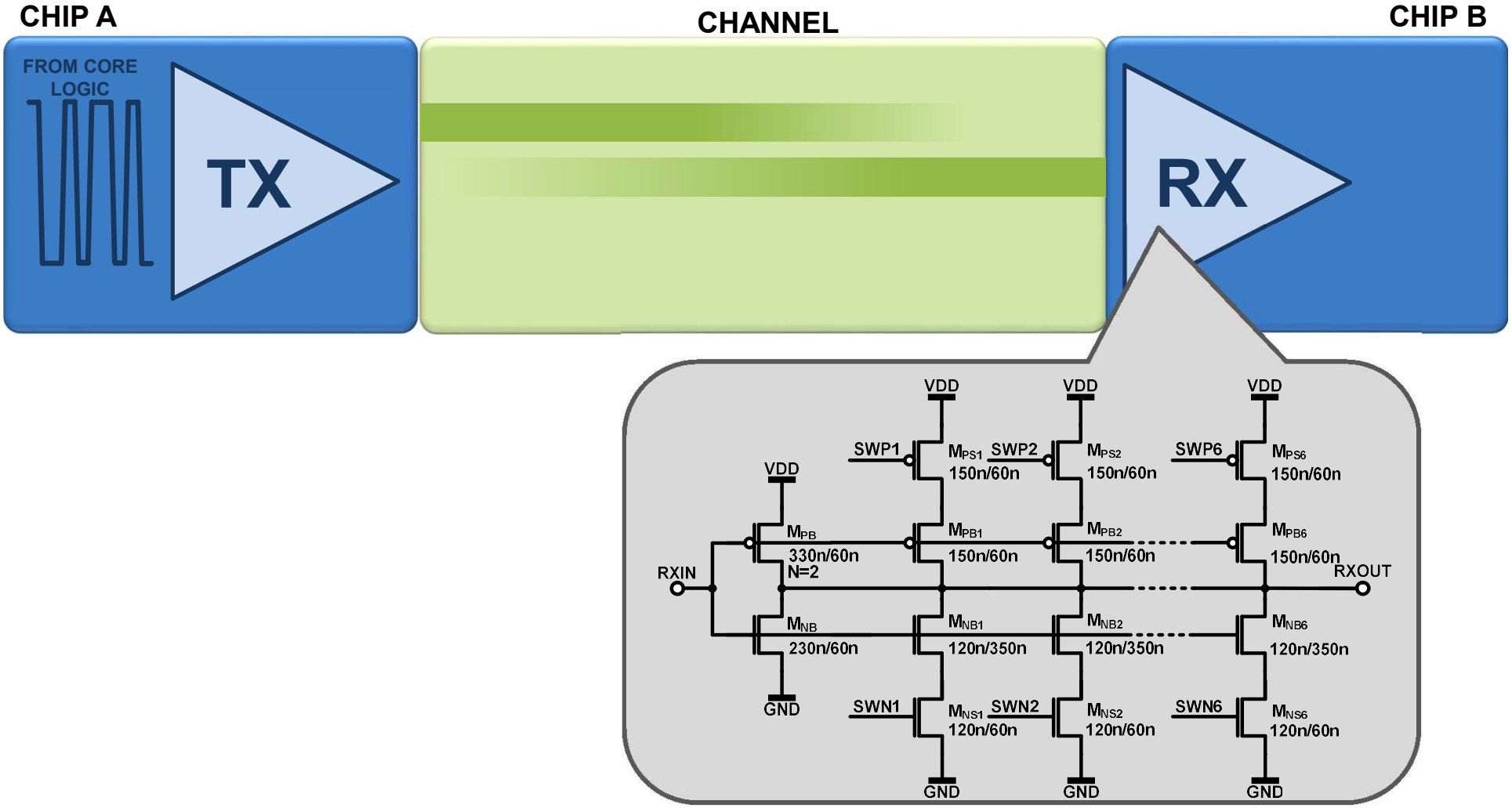
Proposed Low Swing Transmitter



Variability Compensation in Transmitter Circuit

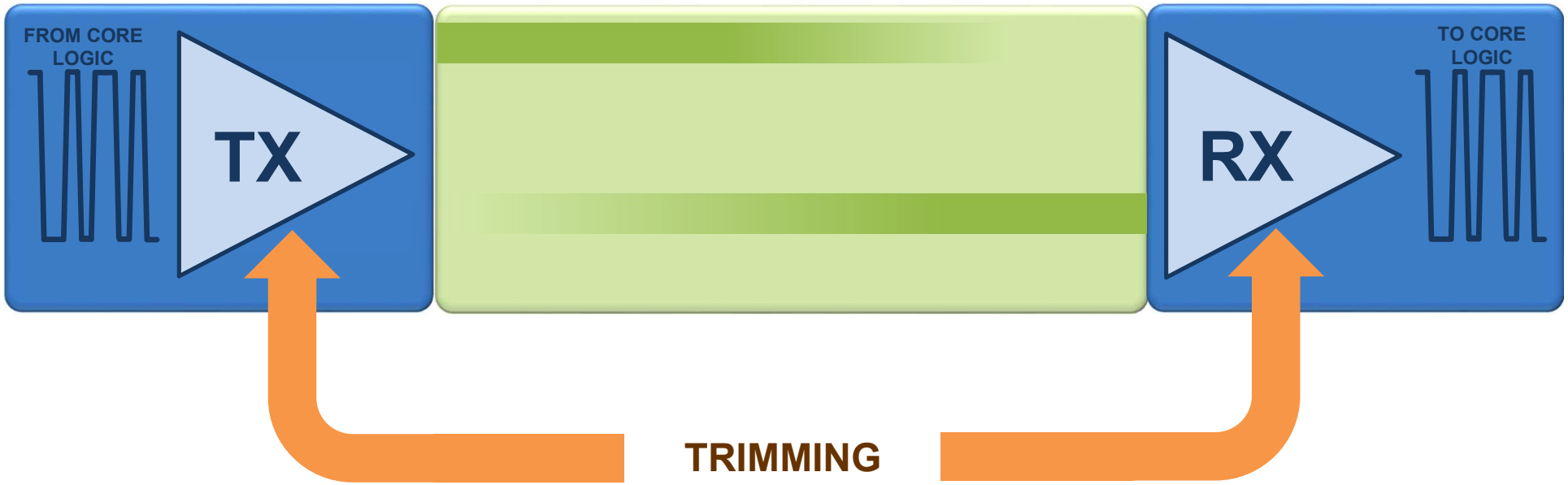


Variability compensation in Receiver Circuit

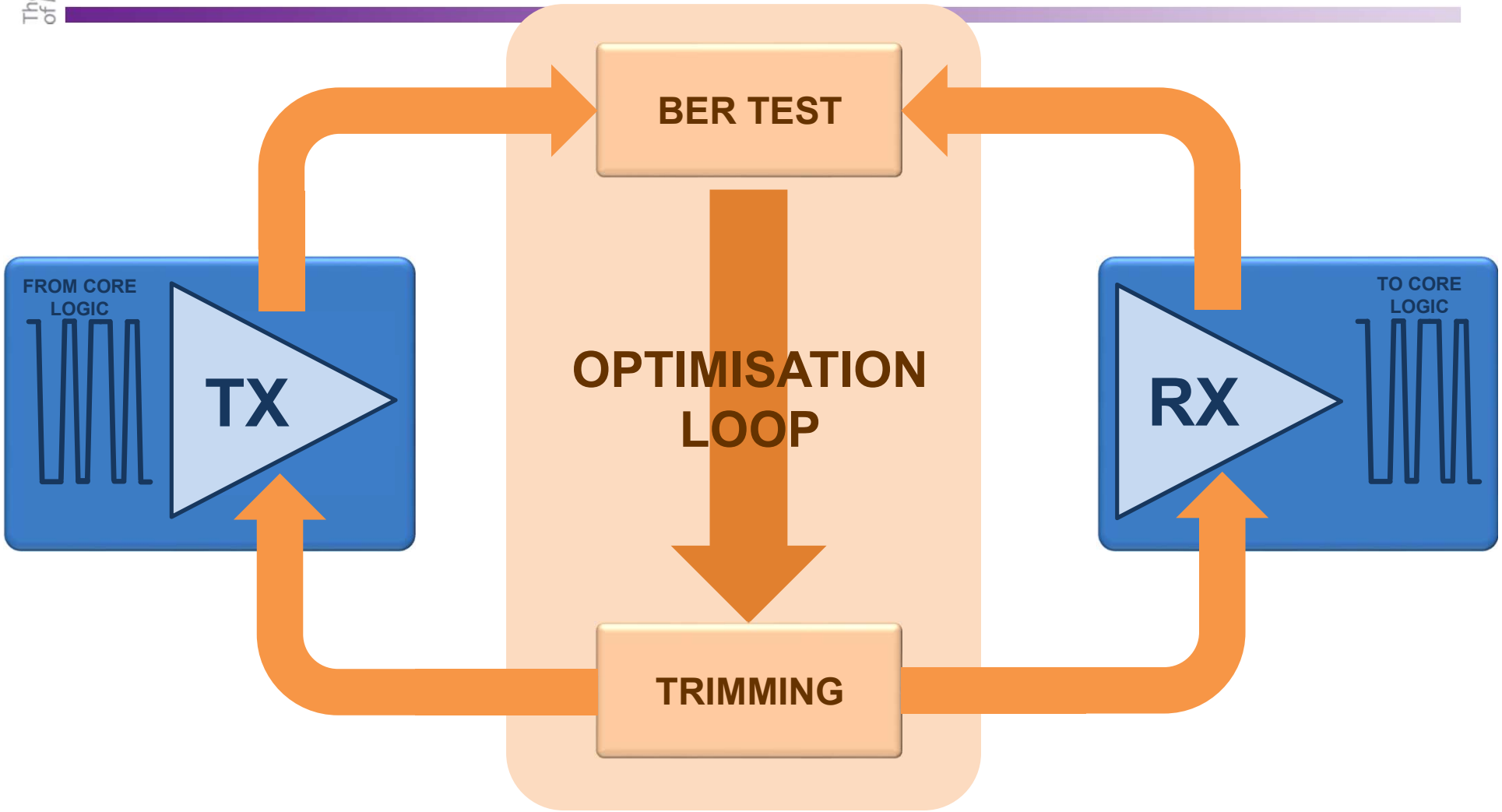


*P. Mroszczyk and V. F. Pavlidis, "Mismatch Compensation Technique for Inverter-Based CMOS Circuits," *Proceedings of the IEEE International Symposium on Circuits and Systems*, May 2018.

Transceiver Trimming



Calibration of Trimming Circuit



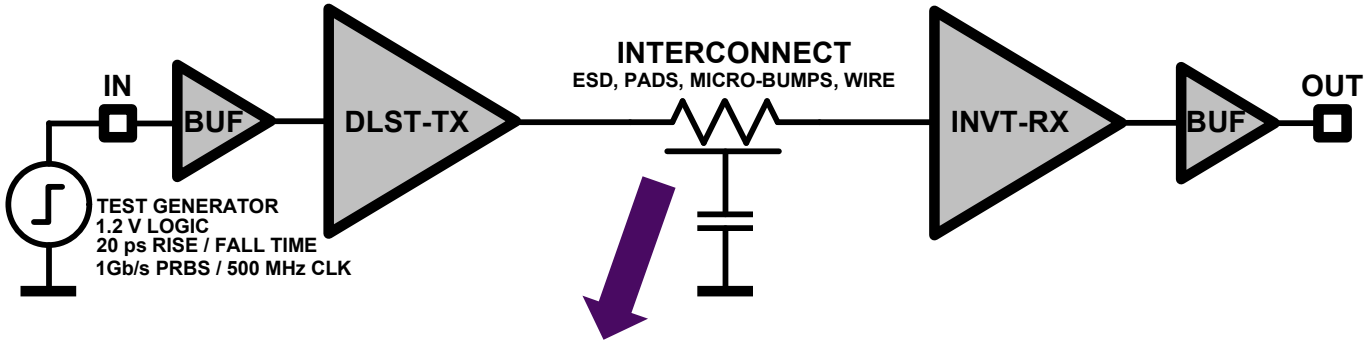
MANCHESTER 1824

The University of Manchester

Part I - Outline

- Single-Ended Chip-to-Chip Communication
- Low-Swing Signaling for Energy Efficiency
  - Low-Swing Transceiver Design
  - Simulation Results
- Data Encoding for Energy Efficiency
- Summary

Experimental Set-up

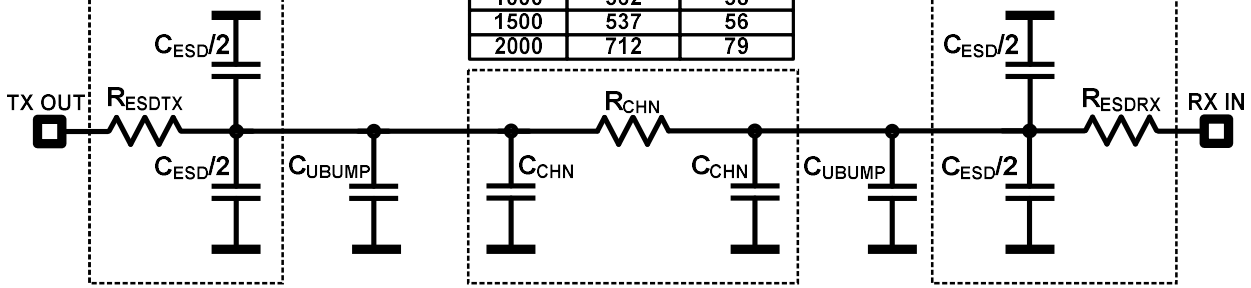


ESD-TX MODEL
R_{ESDTX}: 150 mΩ
C_{ESD}: 115 fF (including PAD)

WIRE MODEL

L [µm]	R_{CHN} [Ω]	C_{CHN} [fF]
500	186	20
1000	362	38
1500	537	56
2000	712	79

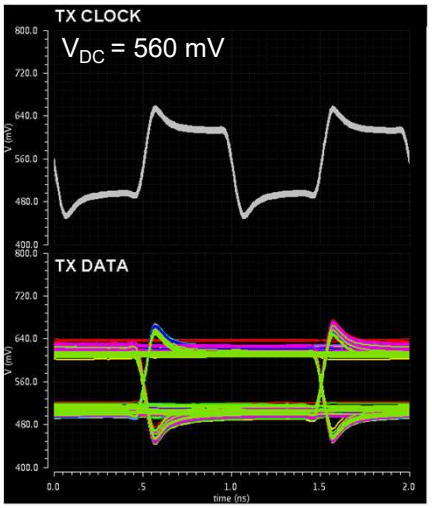
ESD-RX MODEL
R_{ESDRX}: 2 kΩ
C_{ESD}: 115 fF (including PAD)

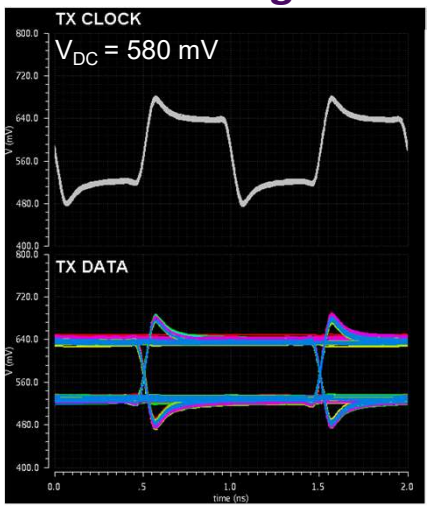


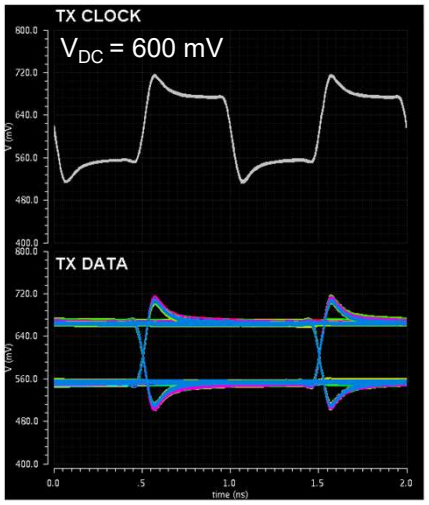
ESD and wire models are based on RC extraction from the layout at the nominal corner.
PAD pitch: 20 µm, PAD size: 10 µm x 10 µm, bump capacitance: 30 fF (worst case estimation)

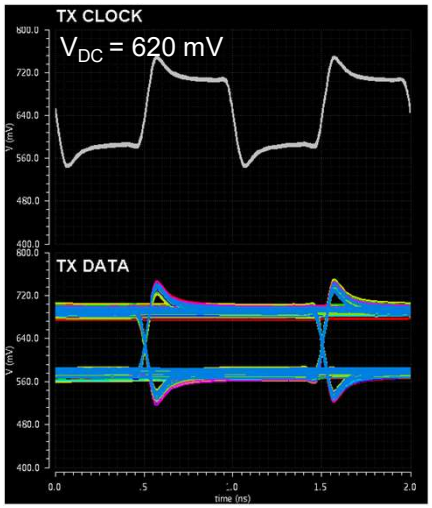
Transceiver Trimming

- **TX transmission window trimming**

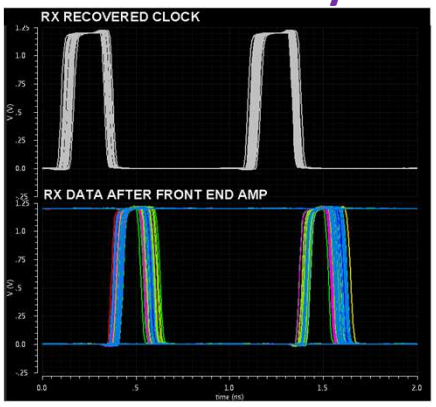


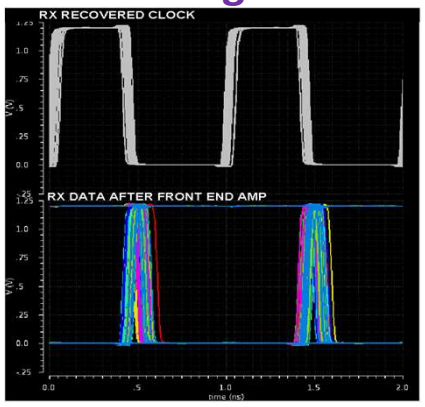


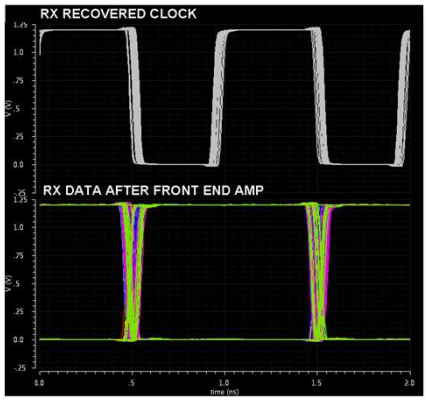


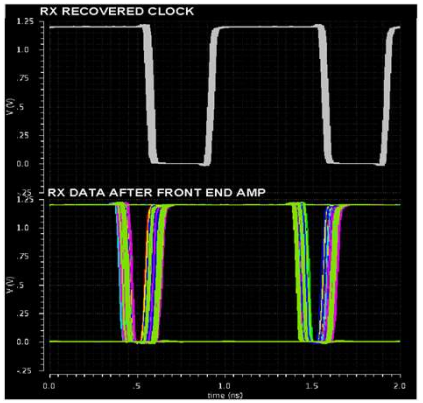


- **RX sensitivity window trimming**



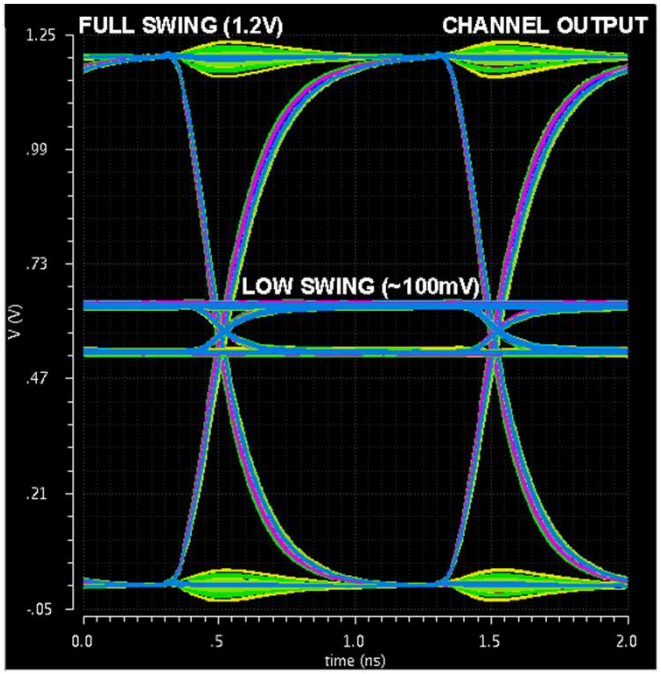


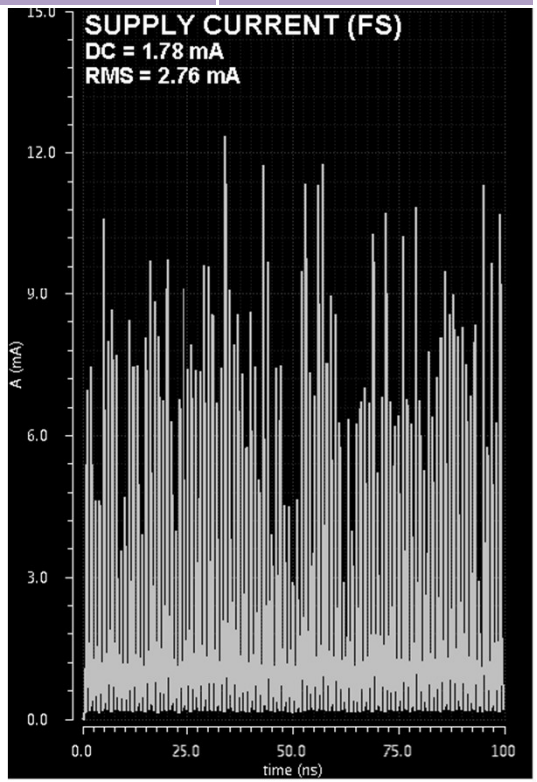


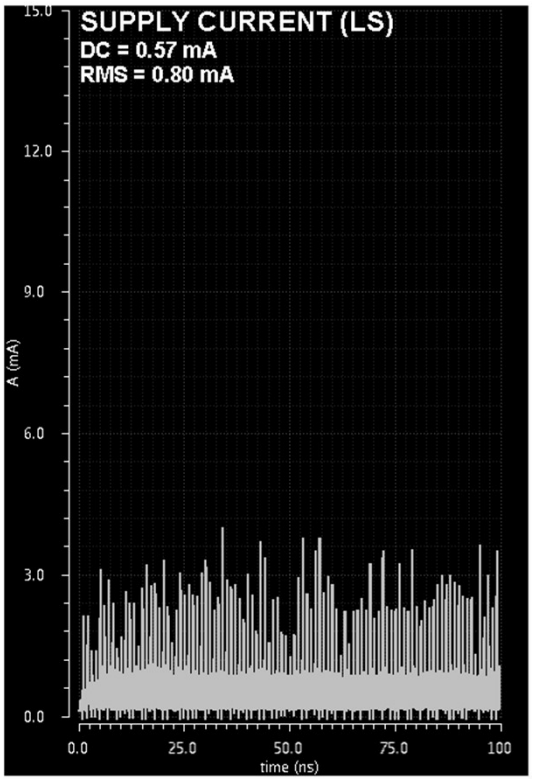


Comparison to Full Swing Transceiver

PARAMETER	FULL SWING	LOW SWING	RATIO (FS/LS)
ENERGY (TX+RX+CDR)	267 fJ/bit	85 fJ/bit	3.2
LATENCY (TX+RX)	390 ps	680 ps	0.6
EDP (ENERGY×LATENCY)	104 fJ·ns	58 fJ·ns	1.8
SW NOISE (RMS)	2.76 mA	0.80 mA	3.5

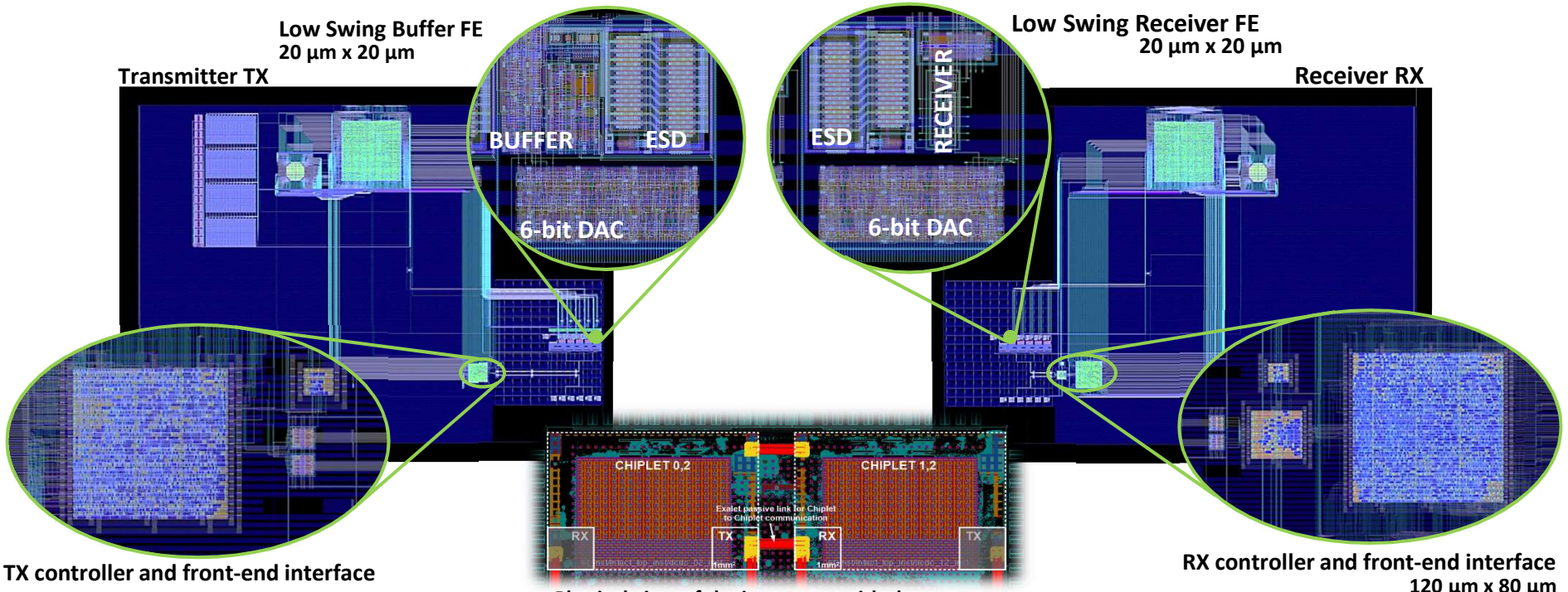






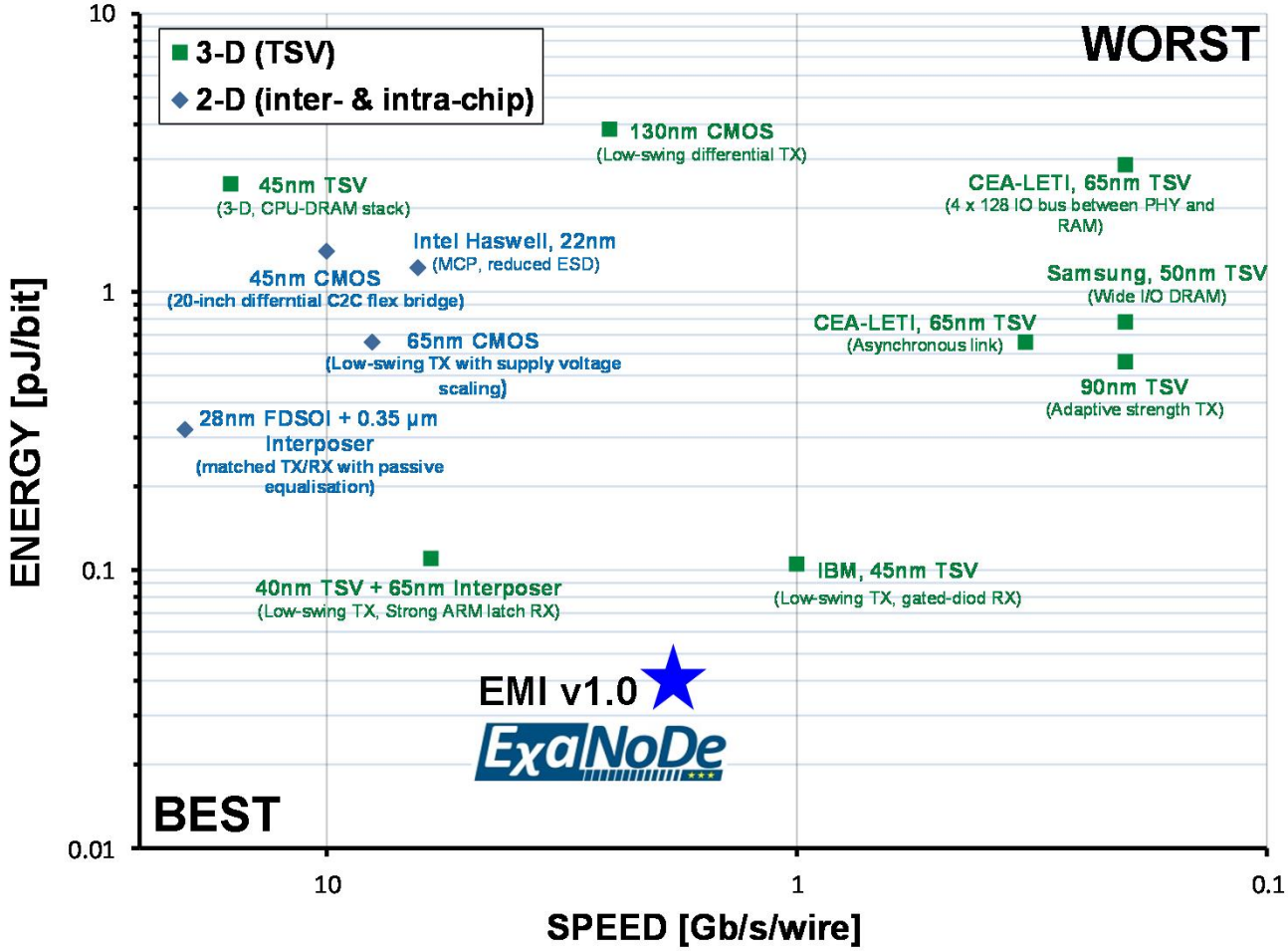
*P. Mroszczyk and V. F. Pavlidis, "Ultra-Low Swing CMOS Transceiver for 2.5-D Integrated Systems," *Proceedings of the IEEE International Symposium on Quality Electronic Design*, pp. 262-267, March 2018.

Design of Transceiver in 28 nm FDSOI

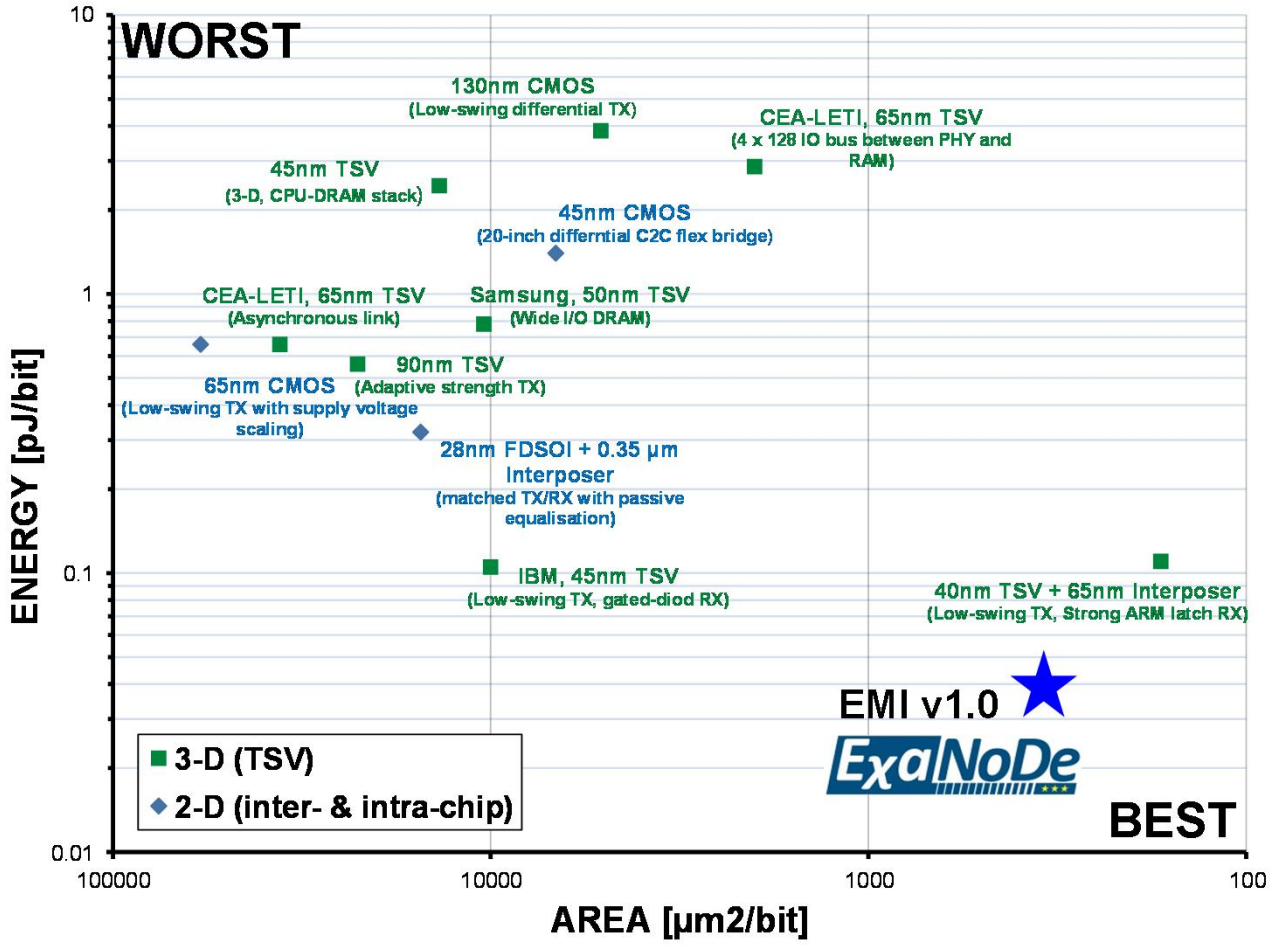


- Exascale Manchester Interconnect (EMI) v1.0
  - Energy: **44.5 fJ/bit**, Speed: **2 Gb/s/wire** (SDR), bandwidth: **256 Gb/s** (128-wire link), **5 Tb/s/mm²**
  - Advanced body biasing scheme for parameter variability trimming
  - Up to **3× less power consumption** compared to a standard full swing solution (< 0.1 pJ/bit)
  - Over **5× less switching noise** compared to a standard full swing solution
  - Latency: 2 clock cycles from TX to RX (0.41 ns for level conversion and signal propagation)

Energy *versus* Speed Comparison



Energy *versus* Area Comparison



MANCHESTER 1824

The University of Manchester

Part I – Outline

- Single-Ended Chip-to-Chip Communication
- Low-Swing Signaling for Energy Efficiency
- Data Encoding for Energy Efficiency
- Summary

MANCHESTER 1824

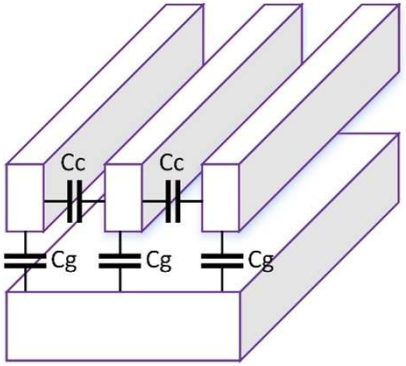
The University of Manchester

Part I – Outline

- Single-Ended Chip-to-Chip Communication
- Low-Swing Signaling for Energy Efficiency
- Data Encoding for Energy Efficiency
  - Data Encoding Approaches
  - Adaptive Word Reordering
  - Simulation Results
- Summary

Static Encoding Schemes

- *A-priori* knowledge of data statistics
  - Gray code
    - Single transition in case of sequential data words
  - T0 code
    - Prevents transitions in case of sequential data words
  - Beach Solution
    - Application oriented, prior analysis of data stream is required
  - Working Zone
    - Assumes that only a subgroup of the address space is used
  - Probability based Mapping
    - Frequent words are mapped to words with low Hamming weight
  - Partial Bus Invert (PBI)
    - Subgroup of lines is formed according to transition probabilities of bus lines



MANCHESTER 1824

The University of Manchester

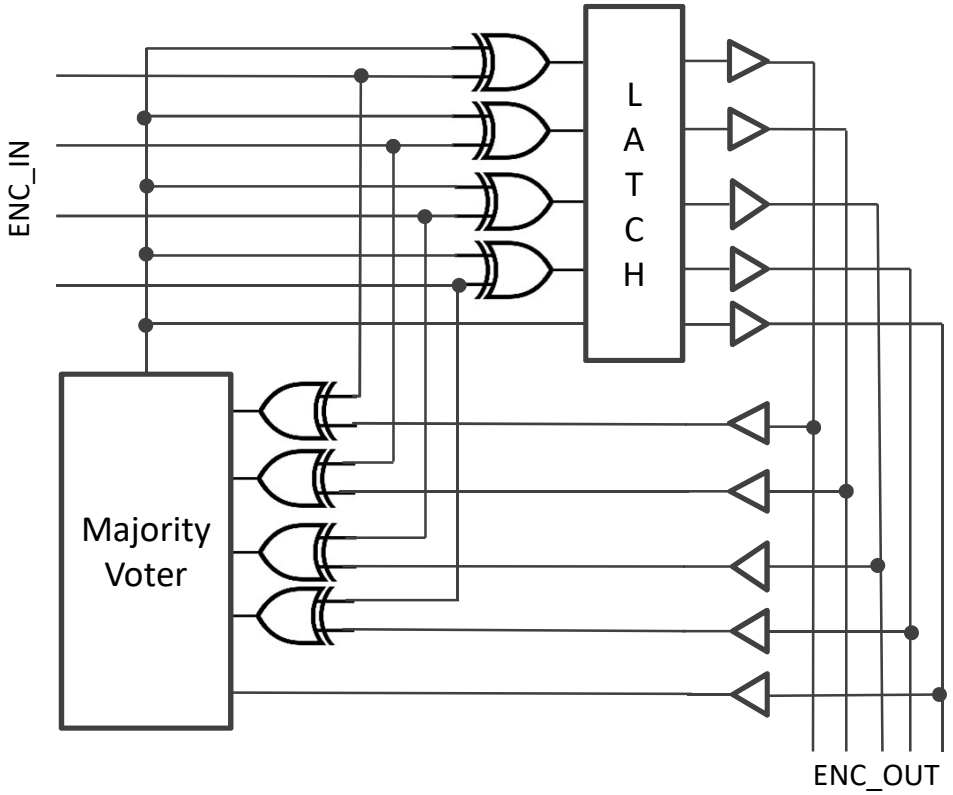
Adaptive Encoding Schemes

- NO prior knowledge of data statistics
  - Bus Invert (BI)
    - Word is inverted when more than half bits would switch
  - Adaptive Partial Bus Invert (APBI)
    - A subgroup of bus lines is inverted, which is changed periodically
  - Frequent Value
    - Encodes the frequent words which are stored in memory
  - Adaptive Dictionary Encoding
    - Number of bits is reduced using a dictionary to store recurring patterns
  - Adaptive Bus Encoding (ABE)
    - Highly correlated lines are encoded
  - Coupling-based schemes

Bus Invert (BI)

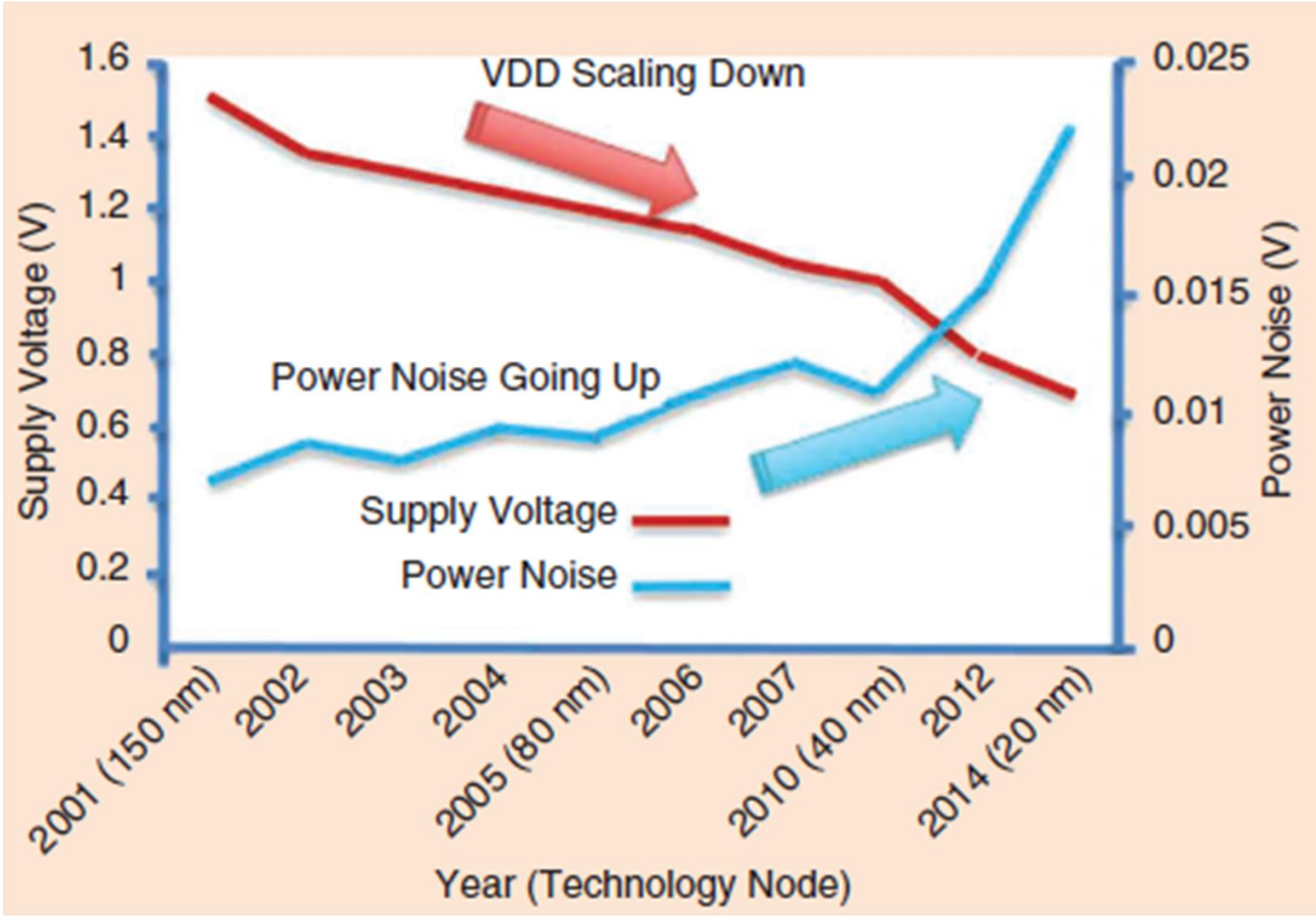
- Calculate number of transitions
- Invert the data word if more than half of the bus lines switch
- 1 extra bus line

0 0 0 0	→	0 0 0 0 0
1 0 1 1		0 1 0 0 1
0 1 1 0		0 1 1 0 0
1 0 0 1		0 1 1 0 1
1 1 1 0		1 1 1 0 0
1 1 0 0		1 1 0 0 0



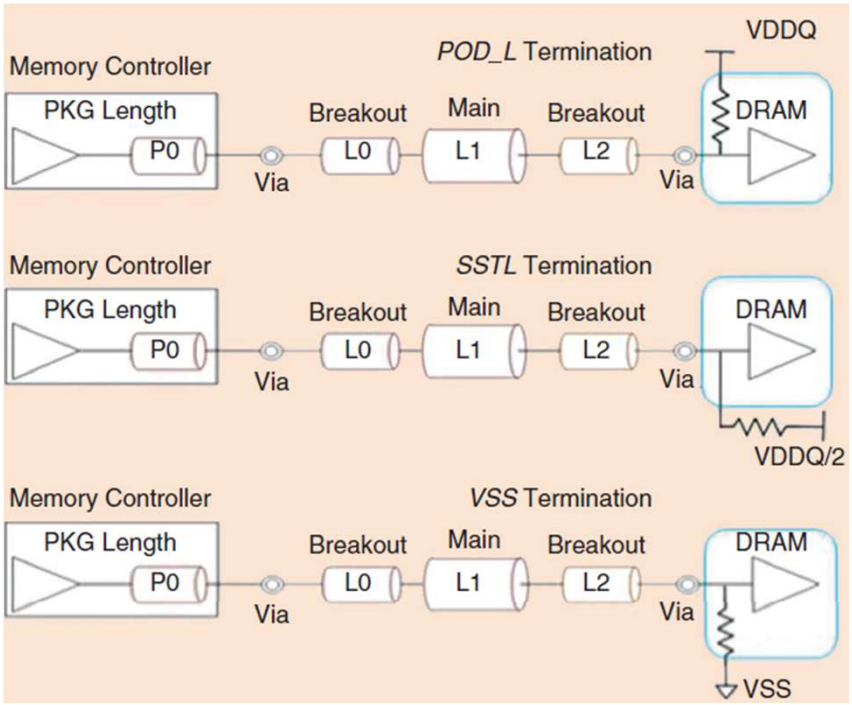
[1] M. R. Stan and W. P. Burleson, "Bus-Invert Coding for Low-Power I/O," *IEEE Trans. on VLSI Systems*, Vol. 3, No. 1, pp. 49–58, March 1995.

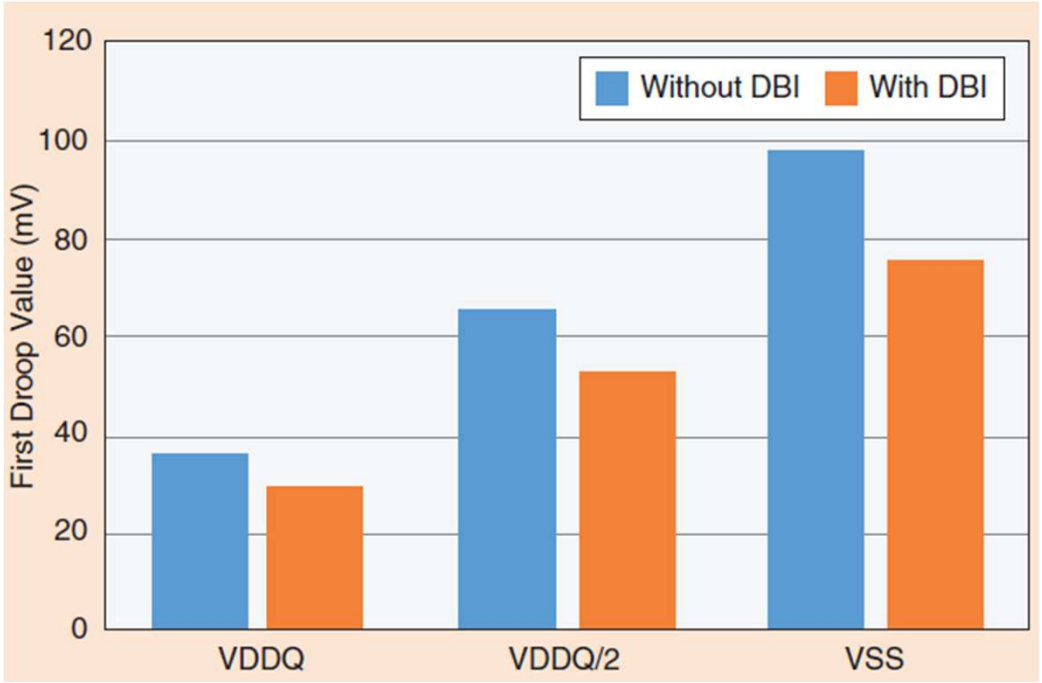
Bus Invert for Power Supply Noise Reduction



*A. Sarkar, "Challenges in IC and electronic systems verification," *Semiconductor Engineering*, May 9, 2013.

Bus Invert for DRAM Memory Bus

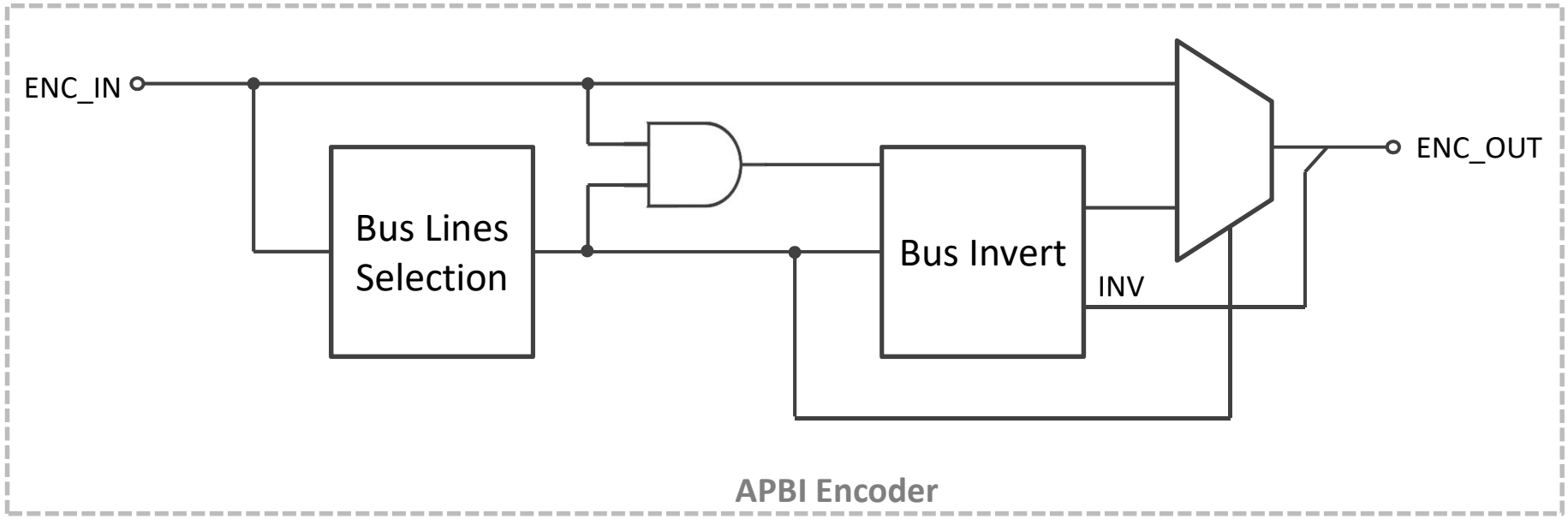




*H. Y. To, "An Analysis of Date Bus Inversion," *IEEE Solid State Circuits Magazine*, Vol. 11, No. 2, pp. 31-41, Spring 2019.

Adaptive Partial Bus Invert (APBI)

- Observe the data stream over a window of N words
- Select the bus lines with the highest probability of switching
- Apply bus inversion to these lines



[2] C. Kretzschmar, R. Siegmund, and D. Mueller. Adaptive Bus Encoding Technique for Switching Activity Reduced Data Transfer over Wide System Buses. In Workshop on PATMOS, pp. 66-75, Sep. 2000.

Adaptive Bus Encoding (ABE)

- Observe the data stream over a window of N words
- Select a bus line as basis and form a cluster of the highly correlated lines
- XOR all the clustered lines with the basis line

L_0	L_1	L_2	L_3	…	L_M
0	0	1	1	…	0
1	1	1	0	…	0
0	0	1	1	…	0
0	1	1	0	…	0
…	…	…	…	…	…
0	1	0	1	…	1

Basis Line

Clustered Lines

[3] S. Sarkar *et al*., "Adaptive Bus Encoding for Transition Reduction on Off-Chip Buses With Dynamically Varying Switching Characteristics," *IEEE Trans. on VLSI Systems*, Vol. 25, No. 11, pp. 3057–3066, Nov. 2017.

Adaptive Bus Encoding (ABE)

- Observe the data stream over a window of N words
- Select a bus line as basis and form a cluster of the highly correlated lines
- XOR all the clustered lines with the basis line

L_0	L_1	L_2	L_3	…	L_M	L_{M+1}
0	0	1	1	…	0	1
0	1	1	1	…	0	0
0	0	1	1	…	0	0
1	1	1	1	…	0	1
…	…	…	…	…	…	…
1	1	0	0	…	1	0
0	1	0	0	…	0	0

Basis Line

Clustered Lines

Redundant

[3] S. Sarkar *et al.*, "Adaptive Bus Encoding for Transition Reduction on Off-Chip Buses With Dynamically Varying Switching Characteristics," *IEEE Trans. on VLSI Systems*, Vol. 25, No. 11, pp. 3057–3066, Nov. 2017.

MANCHESTER 1824

The University of Manchester

Limitations of Encoding Schemes

- Static
  - Knowledge of data statistical properties is not always feasible
  - Statistical properties can temporally vary
- Adaptive
  - High power overhead of encoder and decoder
  - Switching reduction of adaptive schemes might not be adequate
- Coupling-based
  - Unsuitable for inter-chip interconnects, $C_g >> C_c$
  - High power overhead

MANCHESTER 1824

The University of Manchester

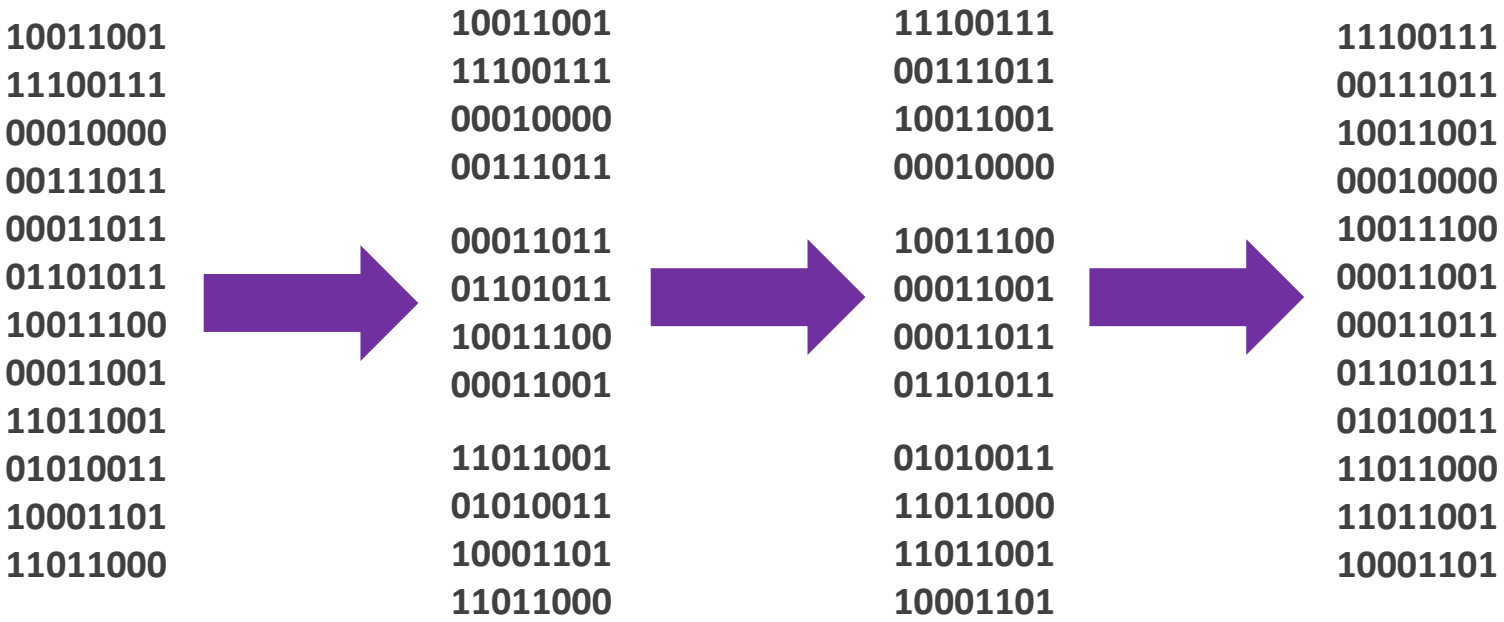
Part I – Outline

- Single-Ended Chip-to-Chip Communication
- Low-Swing Signaling for Energy Efficiency
- Data Encoding for Energy Efficiency
  - Data Encoding Approaches
  - Adaptive Word Reordering
  - Simulation Results
- Summary

Adaptive Word Reordering (AWR)

- Core idea
  - Split the data stream to blocks of N words
  - Reorder the N words in each block to minimise transitions

30% fewer transitions

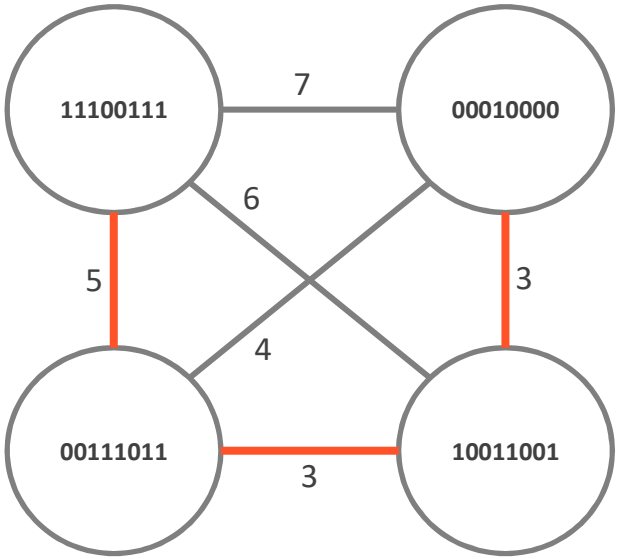


Transitions = 46

Transitions = 32

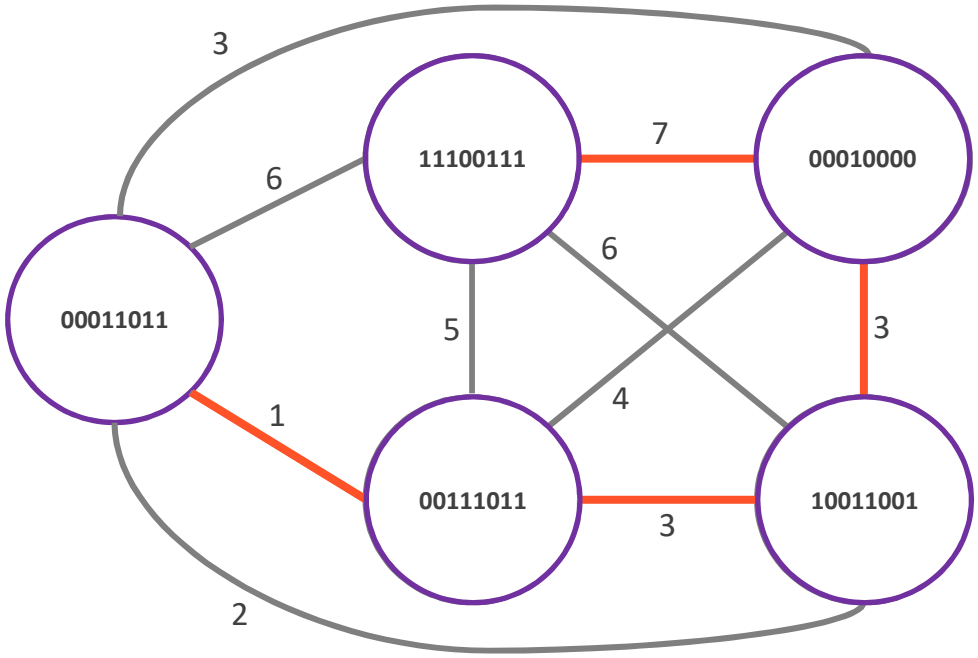
Optimal Reordering

- Word reordering is equivalent to the Travelling Salesman Problem (TSP)
- Each word is a node of a fully connected graph
- Each weight is the Hamming distance between the words
- High computational cost



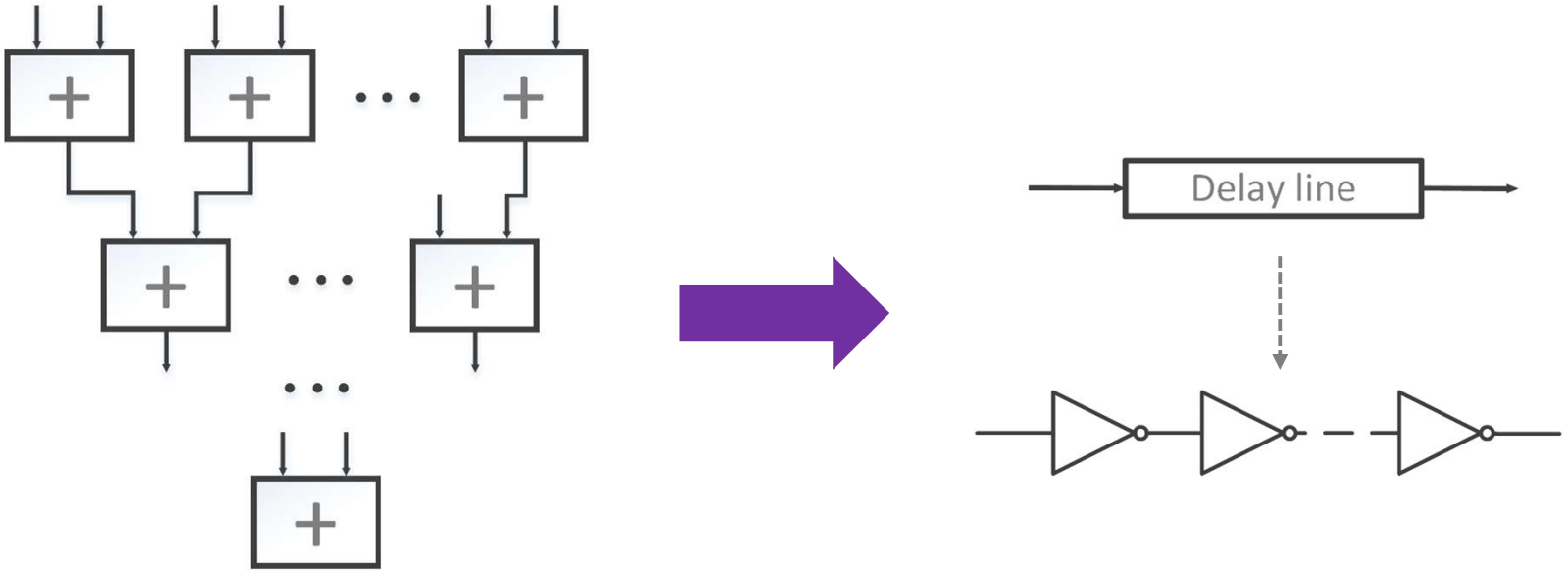
Nearest Neighbour

- In each cycle, out of the N words, select the one with the lowest Hamming distance from the previous



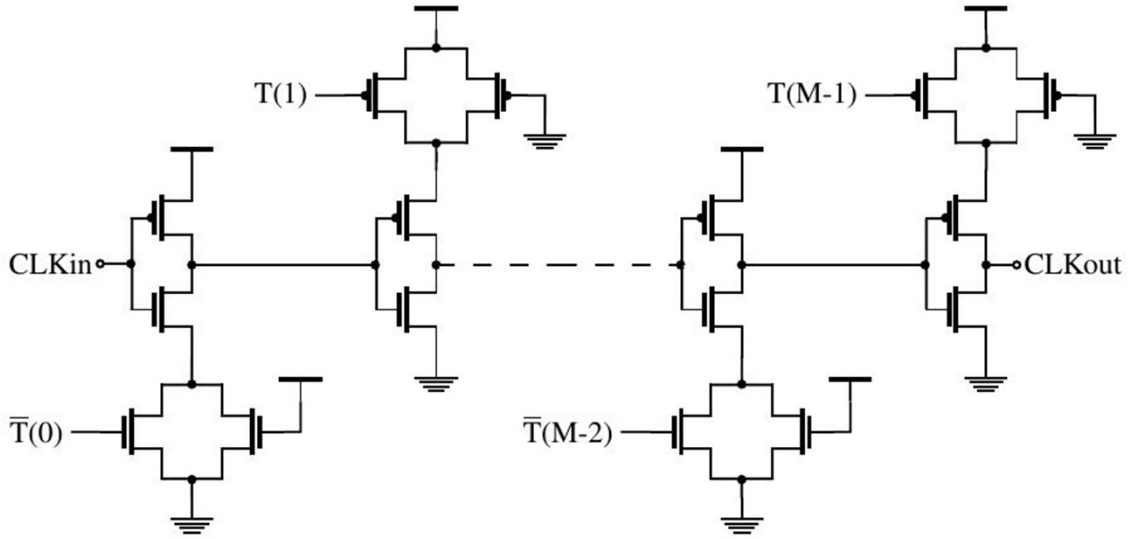
Circuit Implementation

- Challenge: power-efficient calculation of Hamming distance



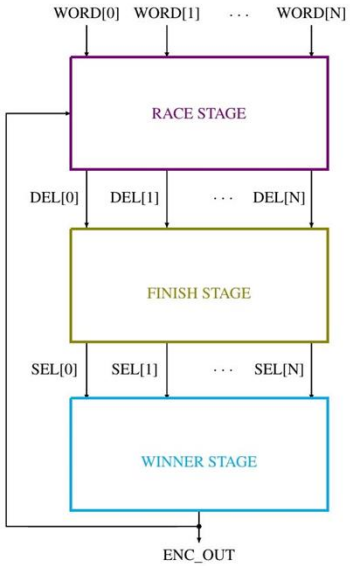
Delay Line for Hamming Distance Calculation

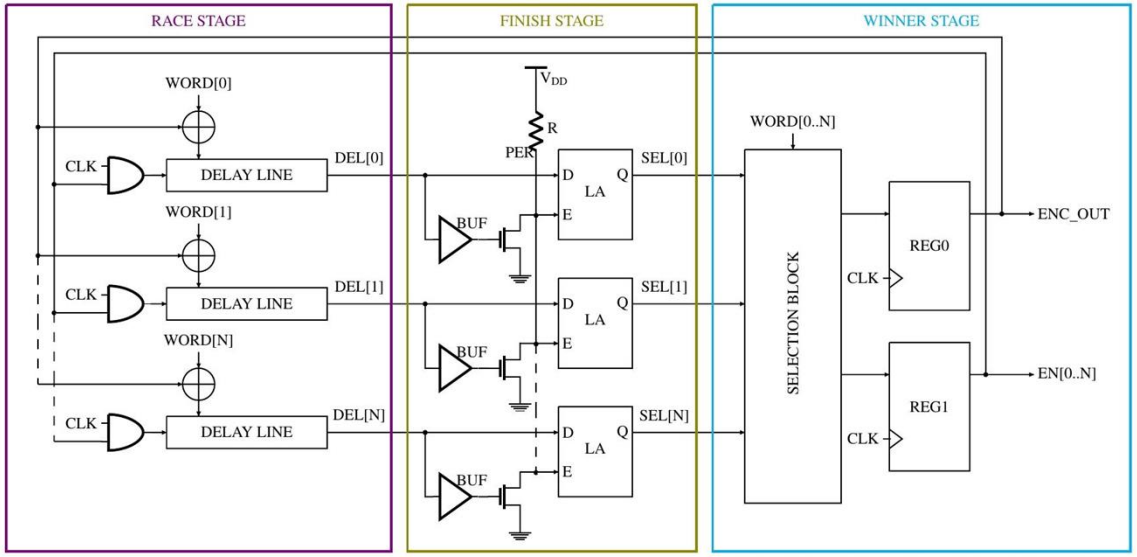
- Inverter chain with short adjustable delays
- Inverters are connected to ground or V_{DD} through a pair of transistors, one is always on
- Higher delay when a transition occurs



*M. Fujino and V. G. Moshnyaga, "An Efficient Hamming Distance Comparator for Low-Power Applications," *Proc. of the IEEE Int. Conference on Electronics, Circuits and Systems*, pp. 641–644, Sept. 2002.

Encoder Circuit

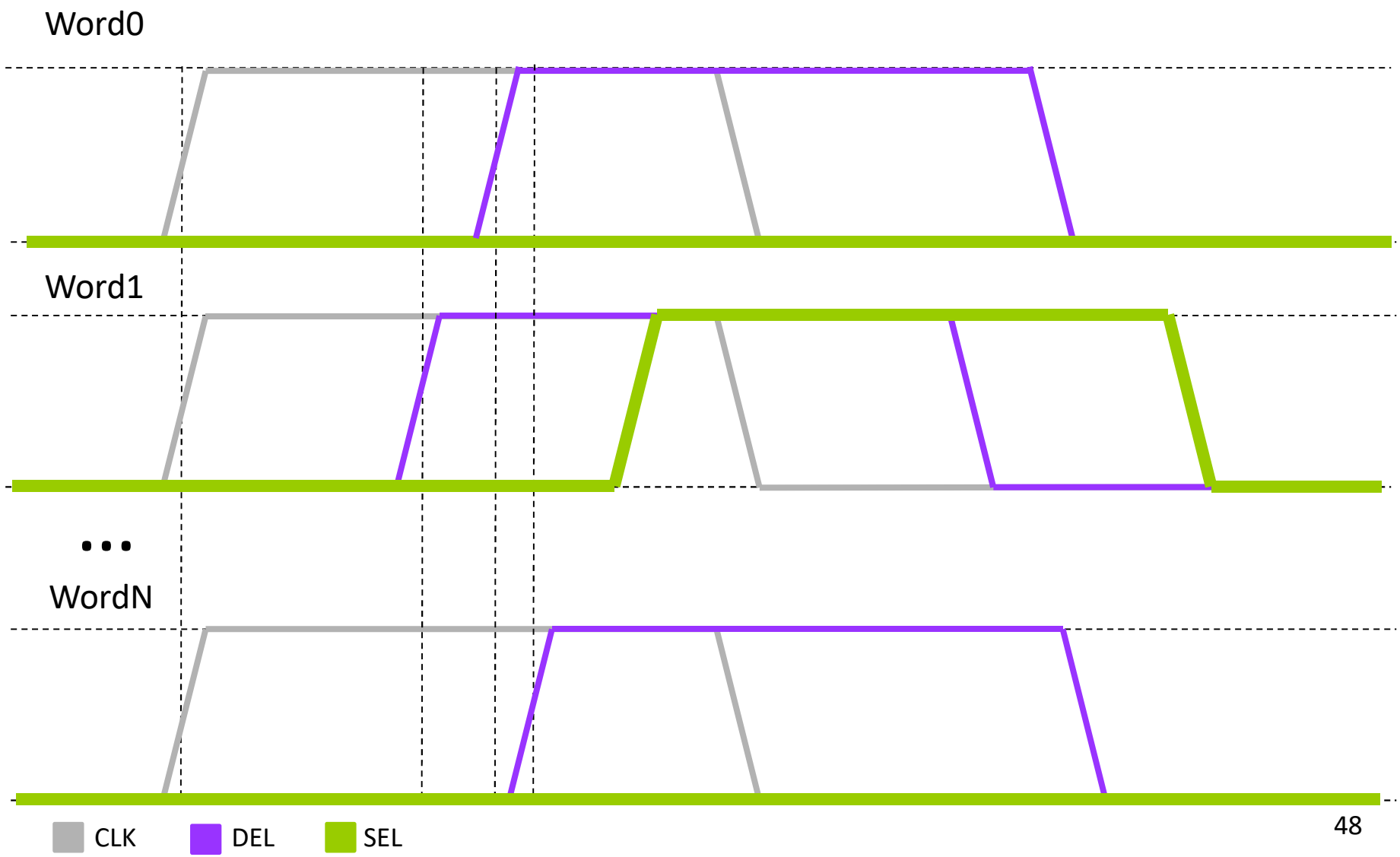




- Race stage
  - Clock is delayed according to the number of transitions
- Finish stage
  - The faster signal prevents the others from propagating
- Winner stage
  - The word that won the race is selected

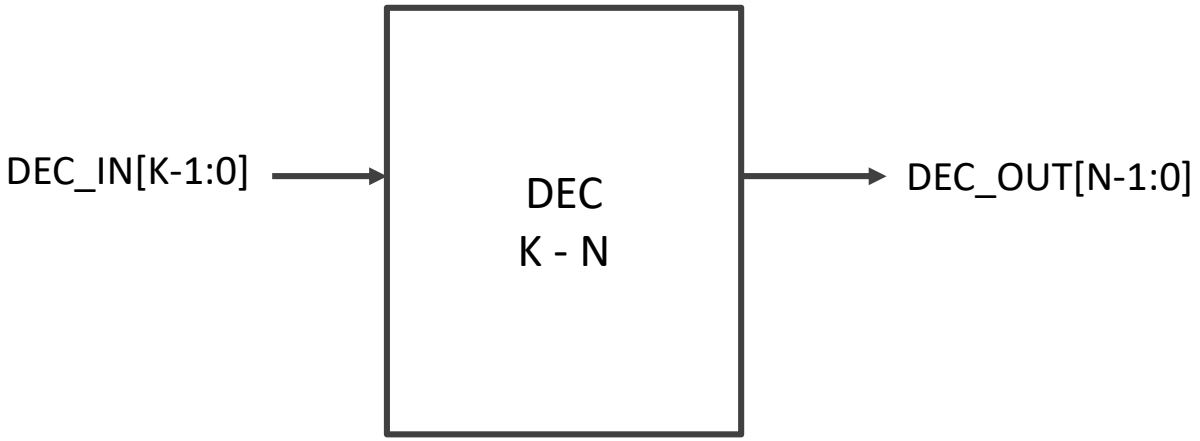
*E. Maragkoudaki, P. Mroszczyk, and V. F. Pavlidis, "Adaptive Word Reordering for Low-Power Inter-Chip Communication," *Design Automation, and Test Conference in Europe*, pp. 974-977, March 2019 .

Encoder Function



Decoder Circuit

- Low spatial redundancy is used to indicate the order
- $K = log_2 N$ bits are added to the word
- Decoder stores the words to registers in the initial order



MANCHESTER 1824

The University of Manchester

Part I – Outline

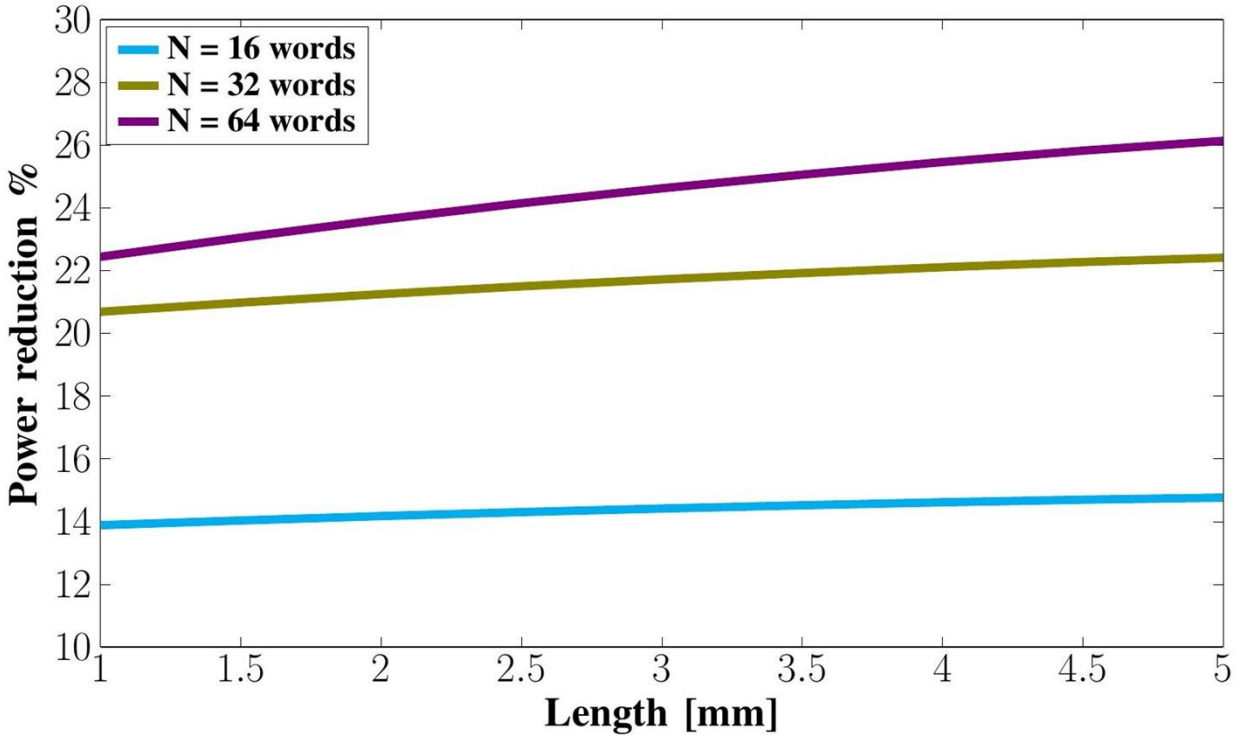
- Single-Ended Chip-to-Chip Communication
- Low-Swing Signaling for Energy Efficiency
- Data Encoding for Energy Efficiency
  - Data Encoding Approaches
  - Adaptive Word Reordering
  - Simulation Results
- Summary

Simulation Set-up

- 65 nm technology, 400 MHz frequency
- Interposer-based interconnect
- Wire parameters according to [4]
- Interconnect model consists of
  - Distributed wire model
  - $C_{ESD} = 115\ fF$
  - $C_{\mu bump} = 30\ fF$

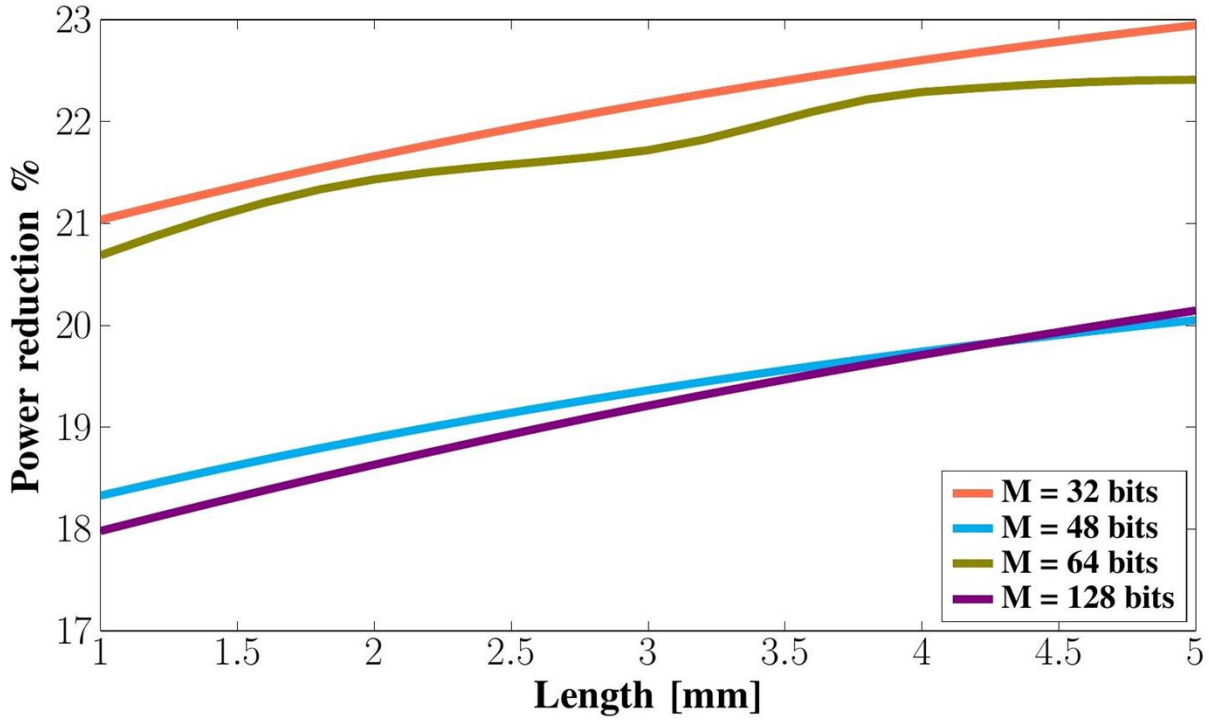
[4] [Online]: Predictive Technology Model (PTM), http://ptm.asu.edu/

Power Efficiency of AWR



- 64 bits bus, LFRic benchmark
- 23% reduction at just 1 mm

Power Decrease *vs* Bus Width

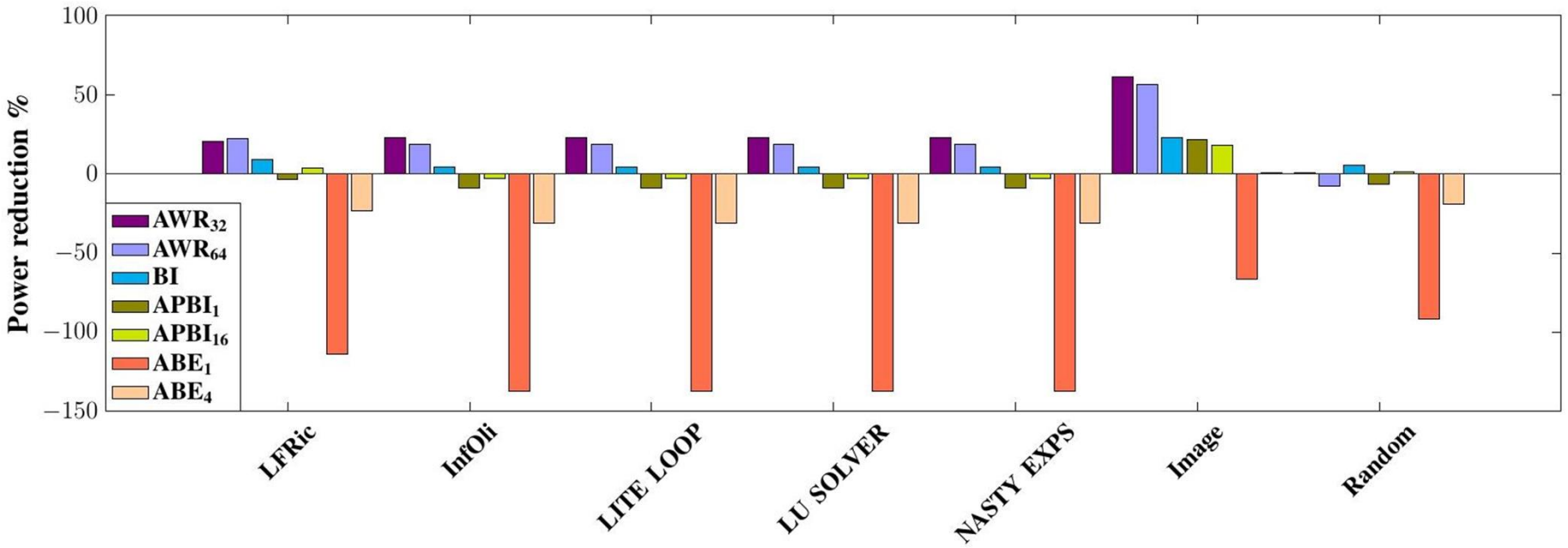


- $N = 32$, LFRic benchmark
- 200 MHz for $M = 128$ bits
- High power gains for all buses

MANCHESTER 1824

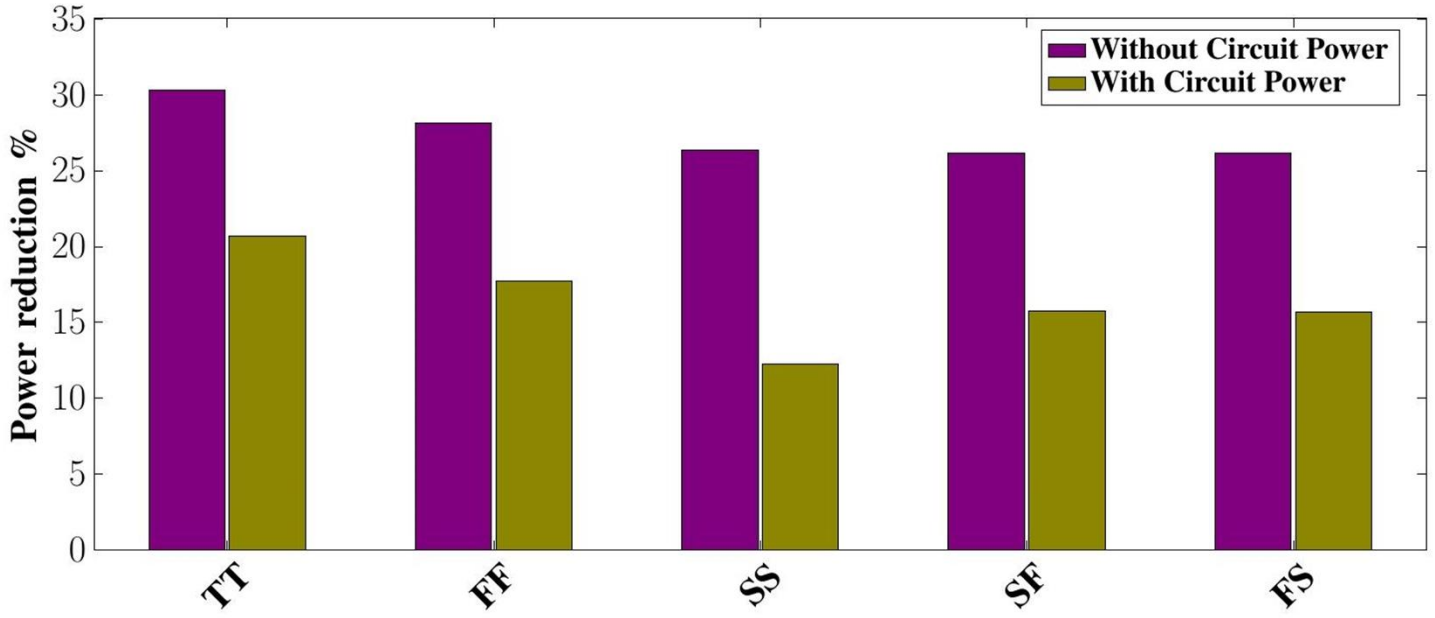
The University of Manchester

Comparison of Encoding Schemes



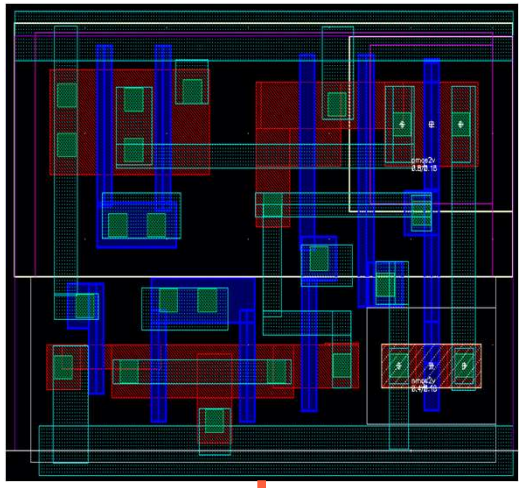
- AWR provides the highest power savings
- Up to **23%** for multiplexed address-data benchmarks and **61%** for image
- Benefits of data encoding diminish for random data

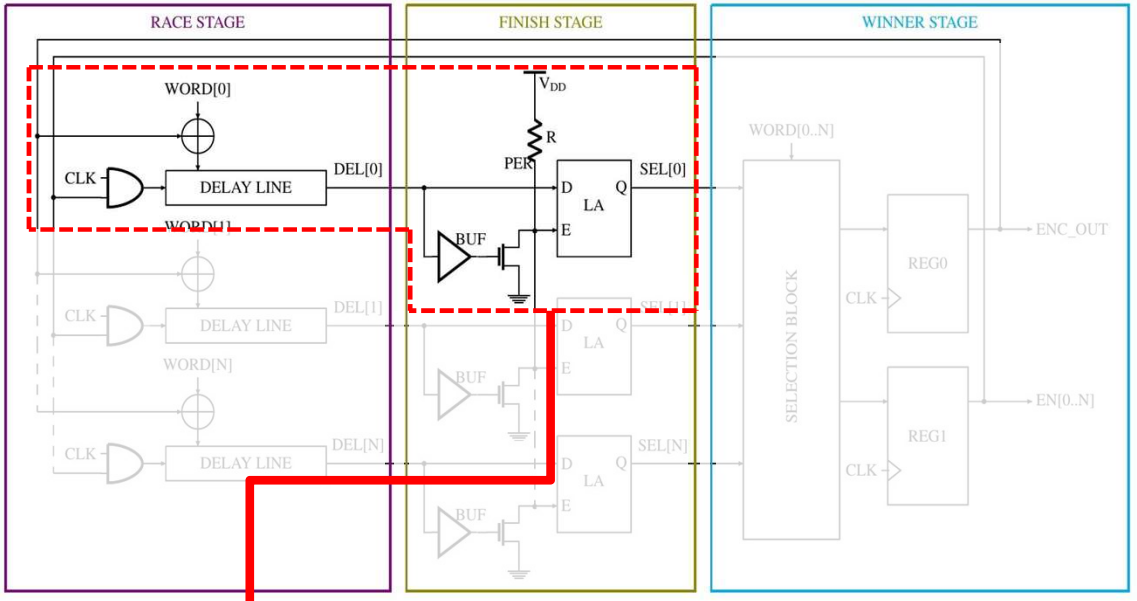
Resilience to Process Variations

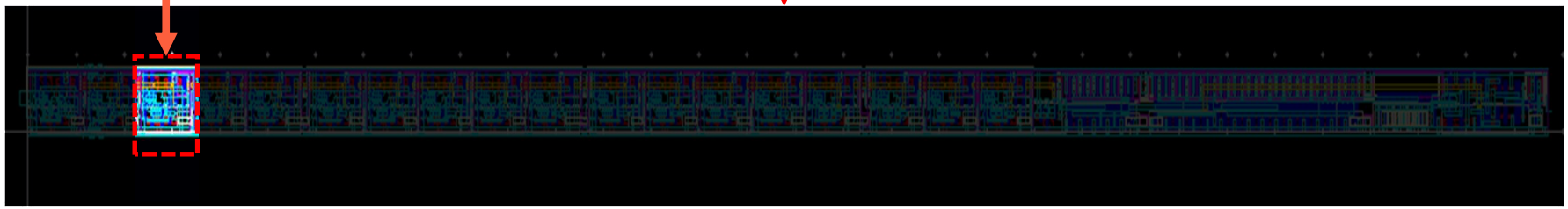


- Low drop of power savings due to variations of delays
- Size up of devices for the SS corner

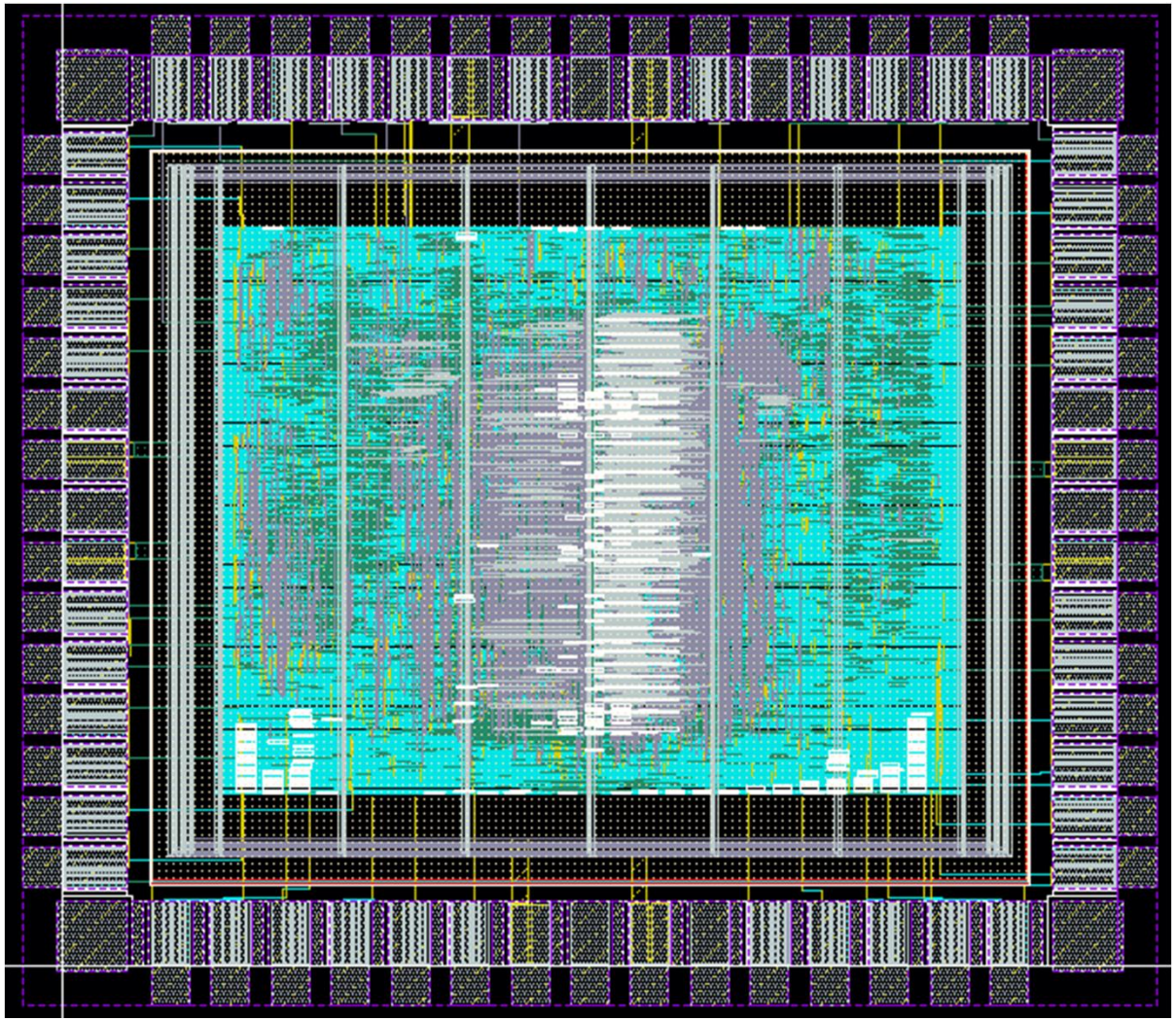
Custom Cell Design







AWR Test Chip

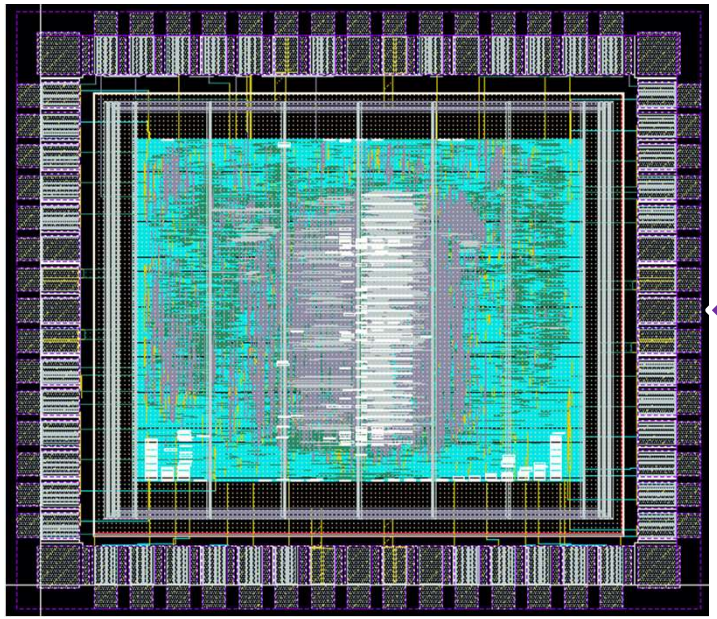


- TSMC 180 nm

Operating modes

- **Receive**
  - Store in Memory
- **Transmit**
  - Encoded
- **Transmit**
  - Unencoded

Test Plan



- Use of FPGA to exchange data
- Measure power of encoded and unencoded transmission

MANCHESTER 1824

The University of Manchester

Part I – Outline

- Single-Ended Chip-to-Chip Communication
- Low-Swing Signaling for Energy Efficiency
- Data Encoding for Energy Efficiency
- Summary

MANCHESTER 1824

The University of Manchester

Part I – Summary

- The challenge is to reduce energy below 0.1 pJ/bit across an SoC
- Wireline communication offers high efficiency, high speed, reliability, and security
- Looking for improvements on the physical layer (ExaNoDe)
  - Low swing signalling
  - Hardware trimming and training
- Looking for improvements on the data link layer (EuroEXA)
  - Reordering encoding
  - Error correction
- AWR outperforms existing techniques in terms of switching reduction
  - Transition reduction without *a-priori* knowledge of data statistics
  - Power efficiency of AWR increases for wider buses
  - The right number of reordered words depends on the capacitive load
  - Significant power reduction when $V_{DD\_IO} > V_{DD\_CORE}$

PART II – WIRELESS COMMUNICATION

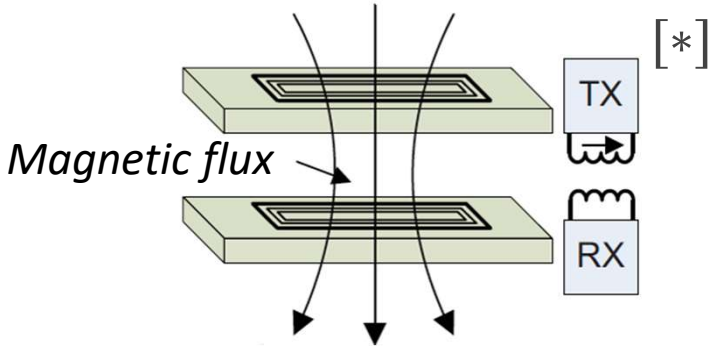
Part II – Outline

- What is Contactless Communication?
- Why Contactless Communication?
- Fundamentals of Contactless Communication
- Energy-Efficient Design of Contactless ICs
- Summary

Wireless Inter-Tier Interfaces

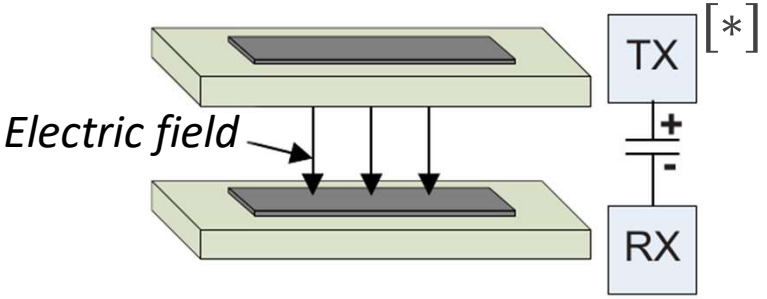
Inductive links

- Manipulates magnetic flux between on-chip inductors
- Current driven
- Long communication distances
- Support multiple integration styles



Capacitive links

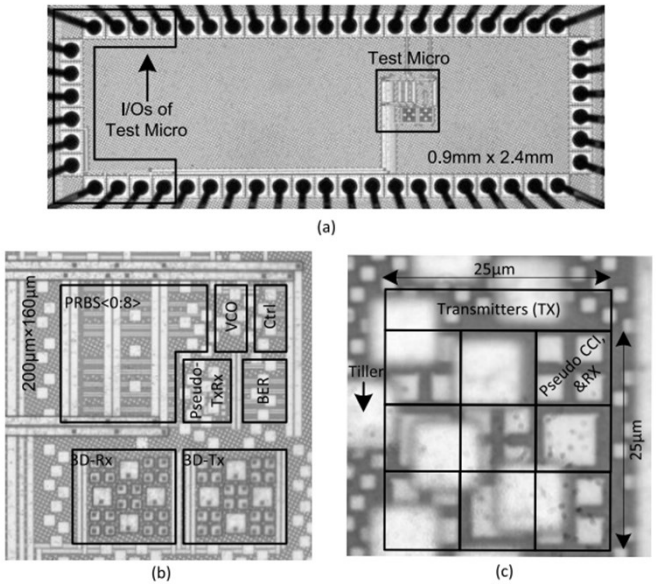
- Manipulates electric field between capacitor plates
- Voltage driven
- Short communication distances
- Limited to face-to-face integration



*J. Ouyang *et al.,* "Evaluation of Using Inductive/Capacitive Coupling Vertical Interconnects in 3-D Network-on-Chip," *Proceedings of the International Conference on Computer-Aided Design,* pp. 477-482, November 2010.

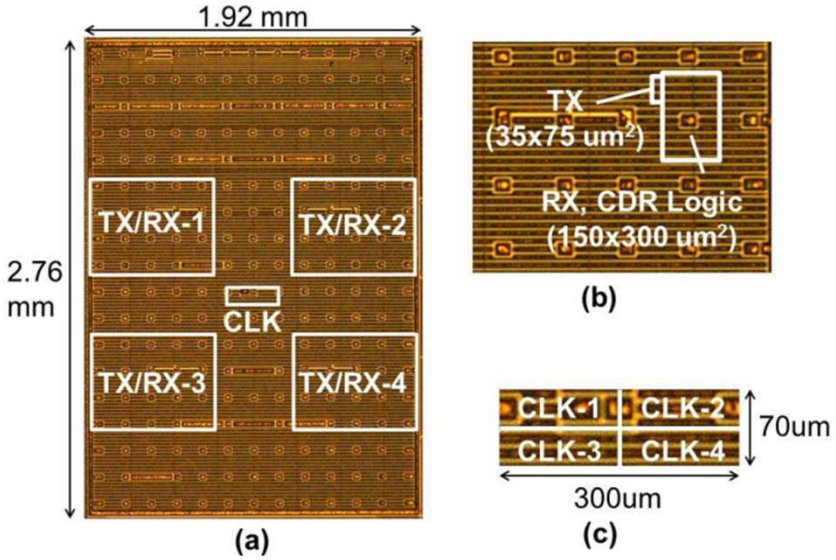
State-of-the-Art Capacitive Links

- **Crosstalk cancelled capacitive coupling**



- 65 nm process
- 2.31 Gb/s/ch
- 53 µW/Gb/s

- **Bi-directional 4 channel capacitive link**

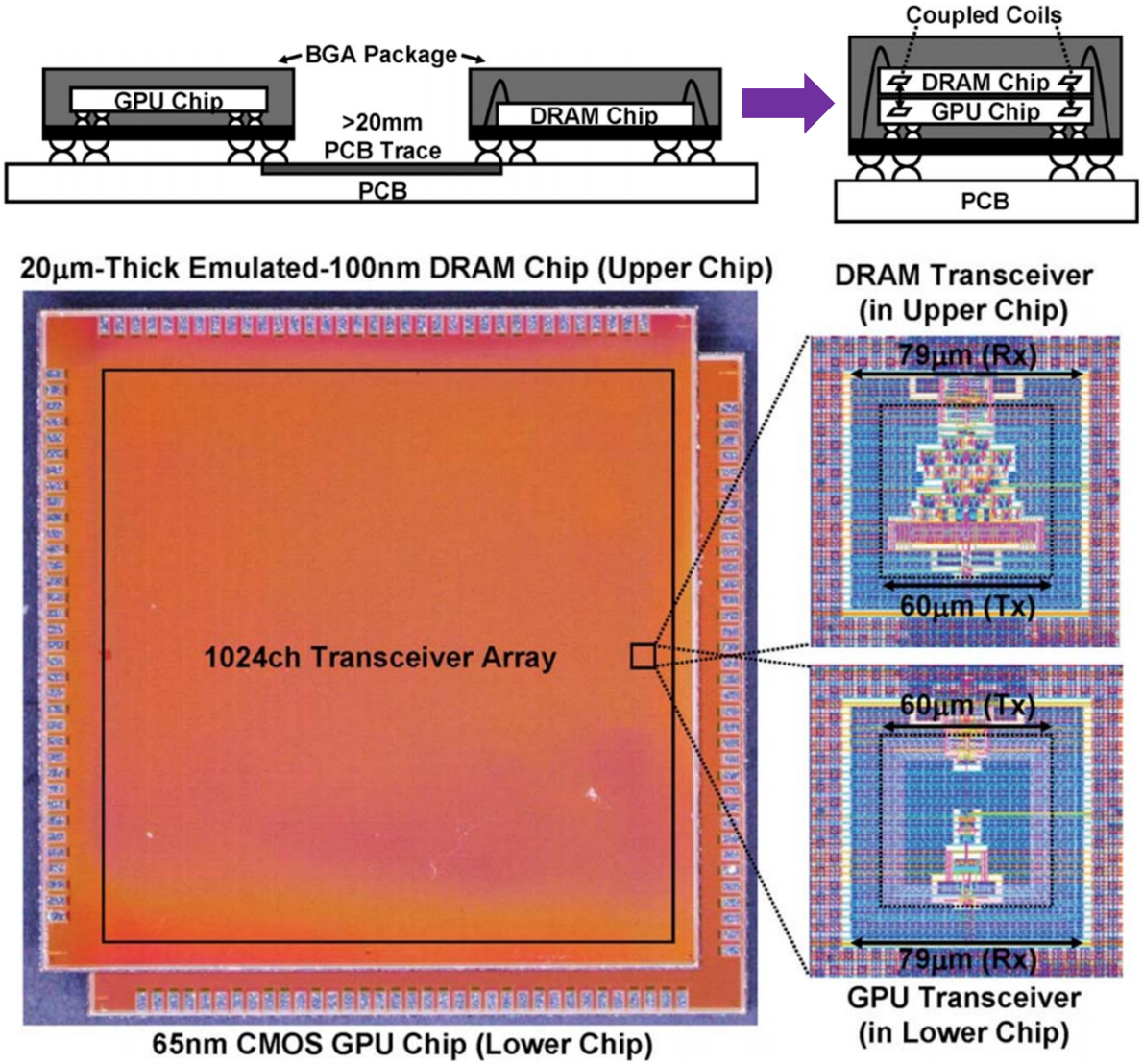


- 14 nm process
- 32 Gb/s
- 4 pJ/bit

*M.-T.-L. Aung *et al.,* "2.31-Gb/s/ch Area-Efficient Crosstalk Cancelled Hybrid Capacitive Coupling Interconnect for 3-D Integration," *IEEE Transactions of Very Large Scale Integration*, Vol. 24, No. 8, pp. 2703-2711, August 2016.

*C. Thakkar *et al.,* "A 32 Gb/s Bidirectional 4-channel 4pJ/b Capacitively Coupled Link in 14 nm CMOS for proximity Communication," *IEEE Journal of Solid-State Circuits,* Vol. 51, No. 12, pp. 3231-3245, December 2016.

State-of-the-Art Inductive Links

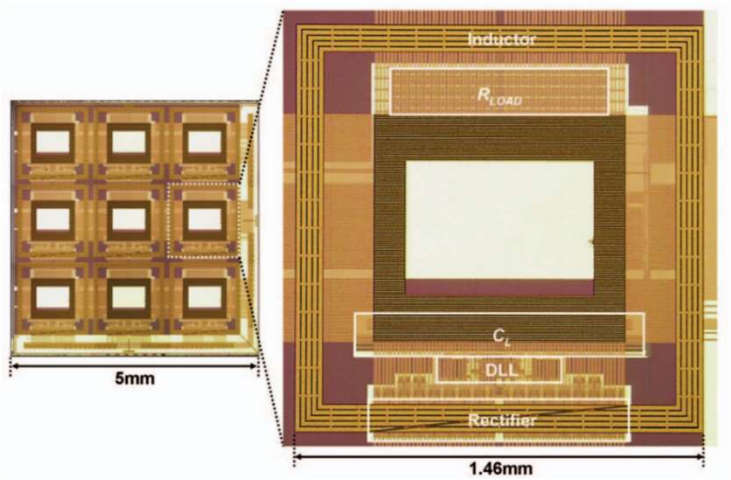


- 1 TB/s from 1,024 transceivers
- 1 pJ/bit
- 20 µm separation distance
- BER < 10^{-16}
- 65 nm process

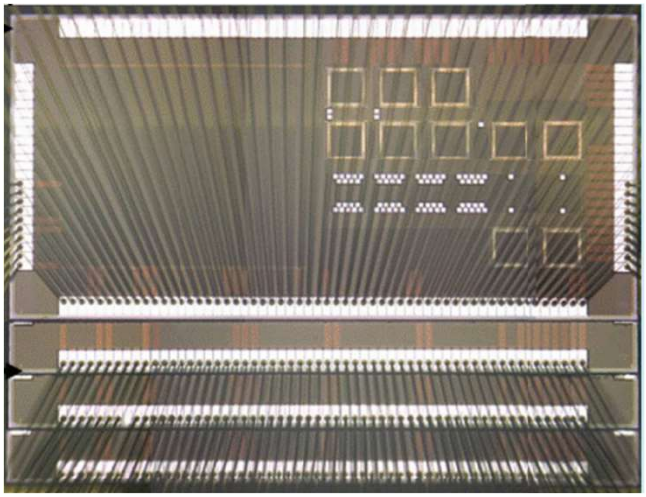
*N. Miura *et al.*, "A 1 TB/s 1 pJ/b 6.4 mm2/TB/s QDR Inductive Coupling Interface Between 65-nm CMOS Logic and Emulated 100-nm DRAM," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, Vol. 2, No. 2, pp. 249-256, June 2012.

Applications of Inductive Links

- **Non-contact wafer level testing**



- **3-D multicore CPU**



- Potential platforms for novel inductive links include
  - Internet of things edge devices
  - Biomedical circuits and micro-fluidic sensors

*A. Radecki et al., "6W/25mm^2 Inductive Power Transfer for Non-Contact Wafer-Level Testing," *Proceedings of the International Solid-State Circuits Conference,* pp. 230-233, February 2011.

*N. Miura *et al.,* "A Scalable 3D Heterogeneous Multi-Core Processor with Inductive-Coupling ThruChip Interface," *Proceedings of IEEE Cool Chips XVI*, pp. 1-3, April 2013.

Part II – Outline

- What is Contactless Communication?
- Why Contactless Communication?
- Fundamentals of Contactless Communication
- Energy-Efficient Design of Contactless ICs
- Summary

MANCHESTER 1824

The University of Manchester

Advantages of Contactless 3-D ICs

- Compatible with standard CMOS lithography
  - Exotic geometries for inductors can affect specific steps
- No need for level shifters
- Reduced ESD protection
- Comparable performance with TSV-based inter-chip communication
- Stacking at affordable cost
  - No need for TSV or micro bump processing

TSVs and contactless 3-D integration are not competitive technologies!

TSV *versus* Inductive Links

- Performance for area unit metric

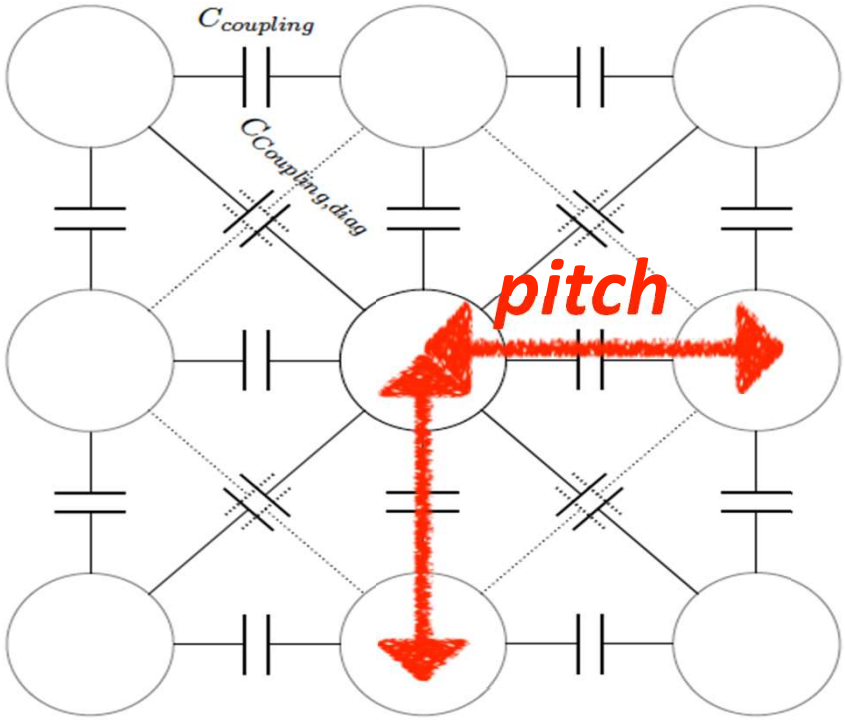
$$eff_x = \frac{BW_x}{area_x} \; [Mbps / \mu m^2] \qquad (1)$$

- Different area consumption
  - TSV vertically wiring and silicon area
  - Inductive link wiring area only

Modeling a TSV Array

- TSV to TSV coupling
- The area occupied by a TSV array is

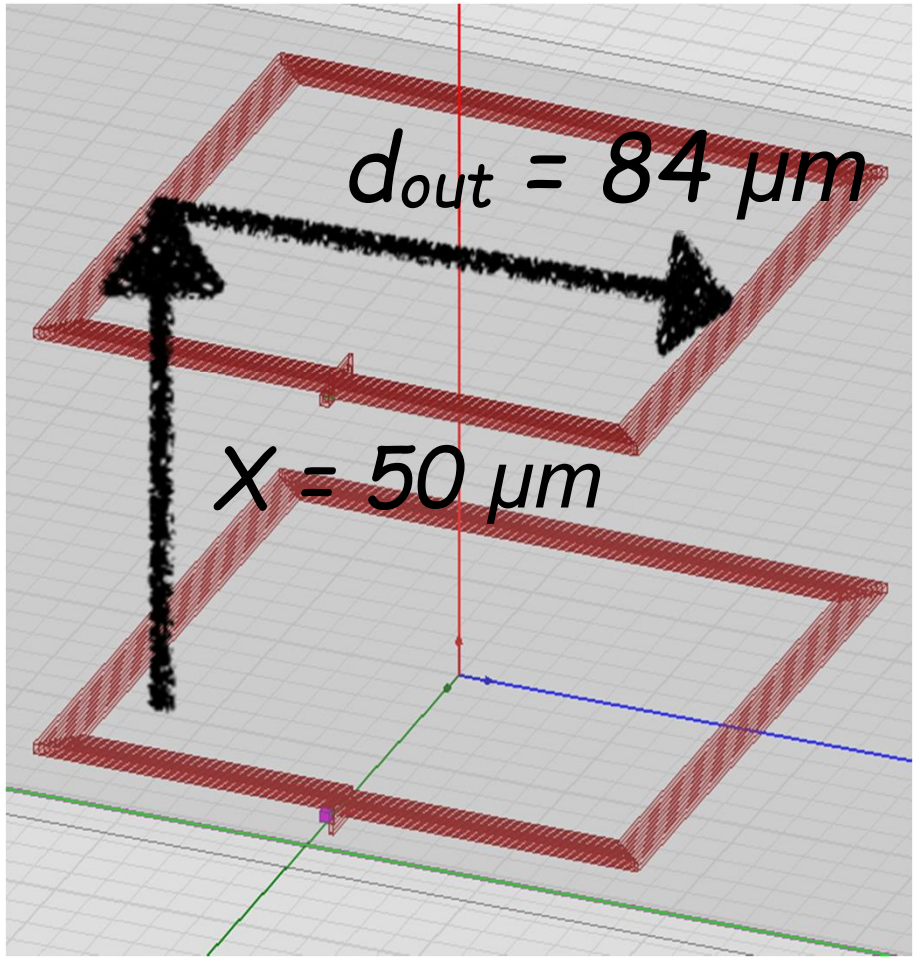
$$area_{TSV} = (N \times M) pitch^2 \quad (2)$$



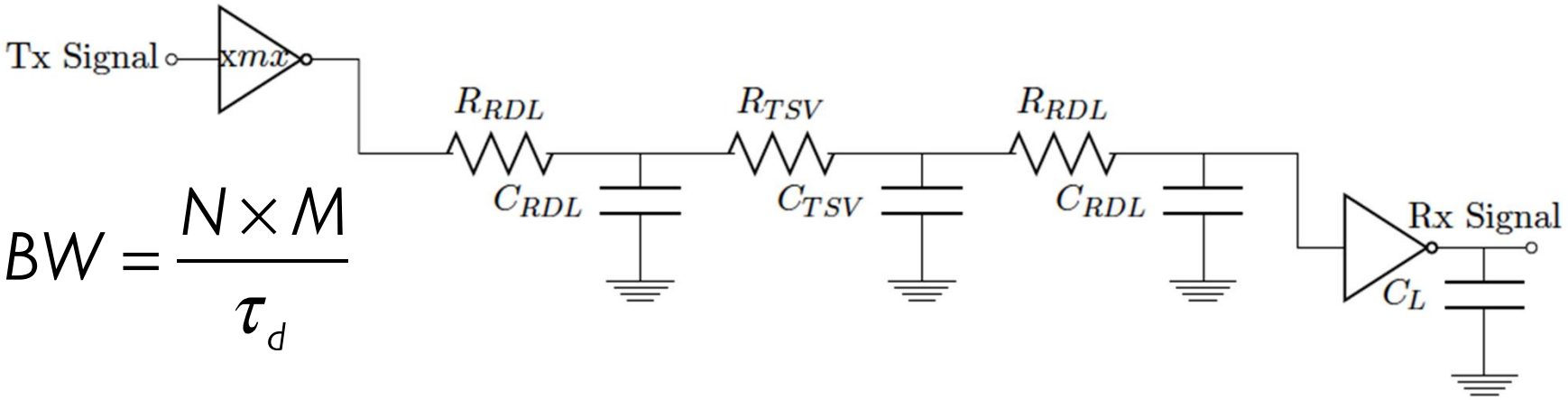
Modeling of On-Chip Inductors

- Core element
- Simple *RLC* model
- Coupling efficiency, k = 0.3
- Transceiver circuit
  - H-Bridge transmitter
  - Hysteresis comparator receiver
  - 20 Gbps
- Inductive link area

$$area_{IL} = (N \times M) d_{out}^2 \quad (3)$$



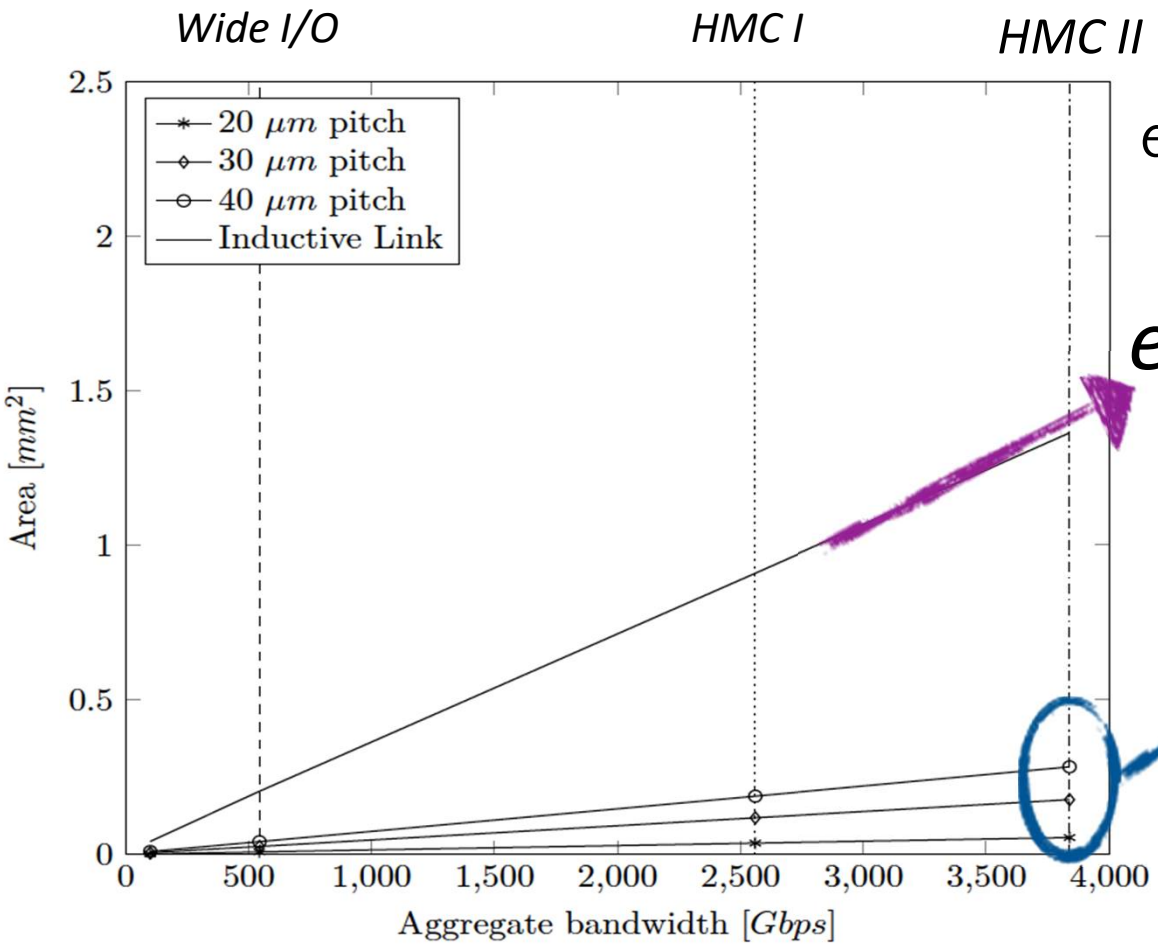
TSV Delay Extraction



$$BW = \frac{N \times M}{\tau_d}$$

- Elmore delay calculation
- Impact of TSV size, coupling, and RDL on delay
- Different TSV pitches simulated (20, 30, 40 µm)
- Increasing TSV array (4x4, 8x8, 16x16)

Interface Performance Density

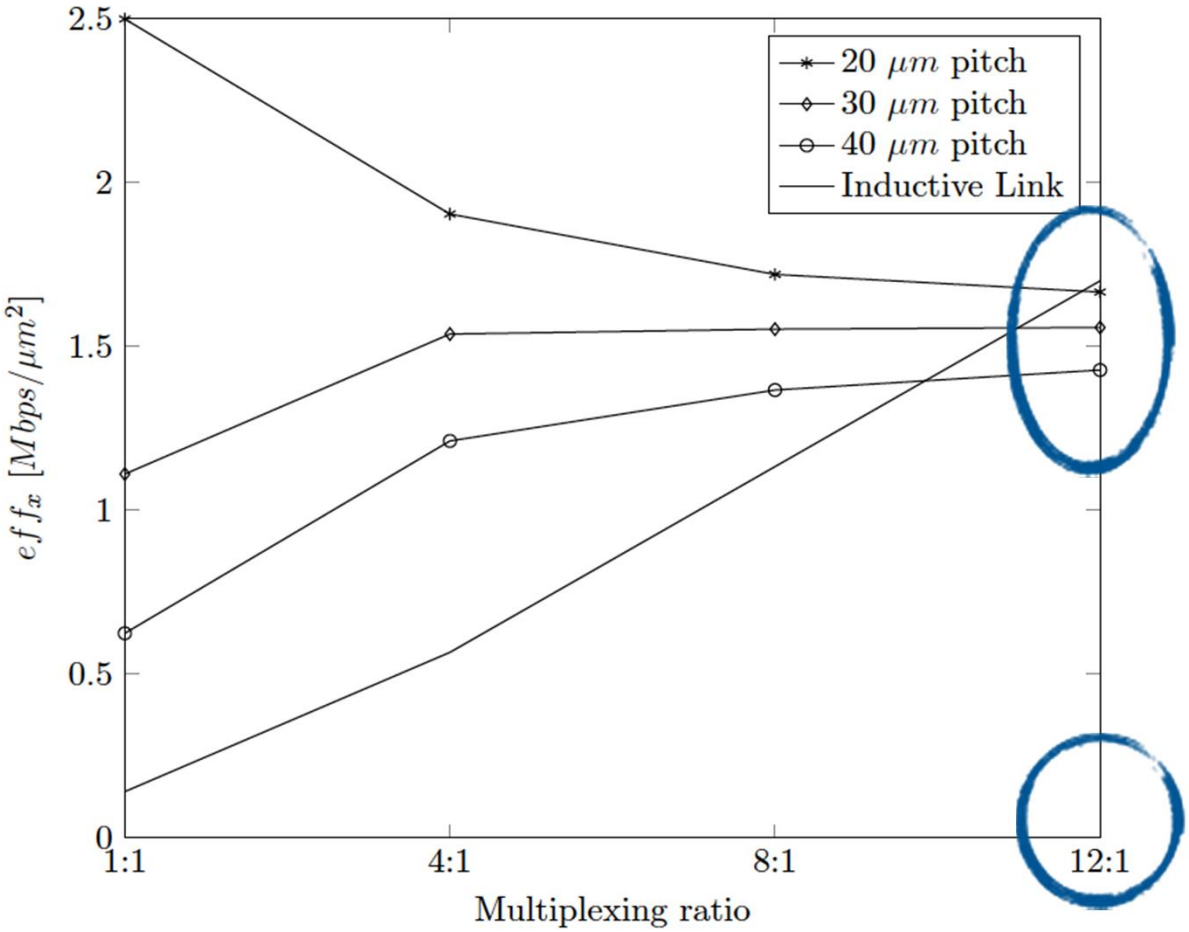


$$eff_x = \frac{BW_x}{area_x} \; [Mbps / \mu m^2]$$

eff_{IL}=2.8

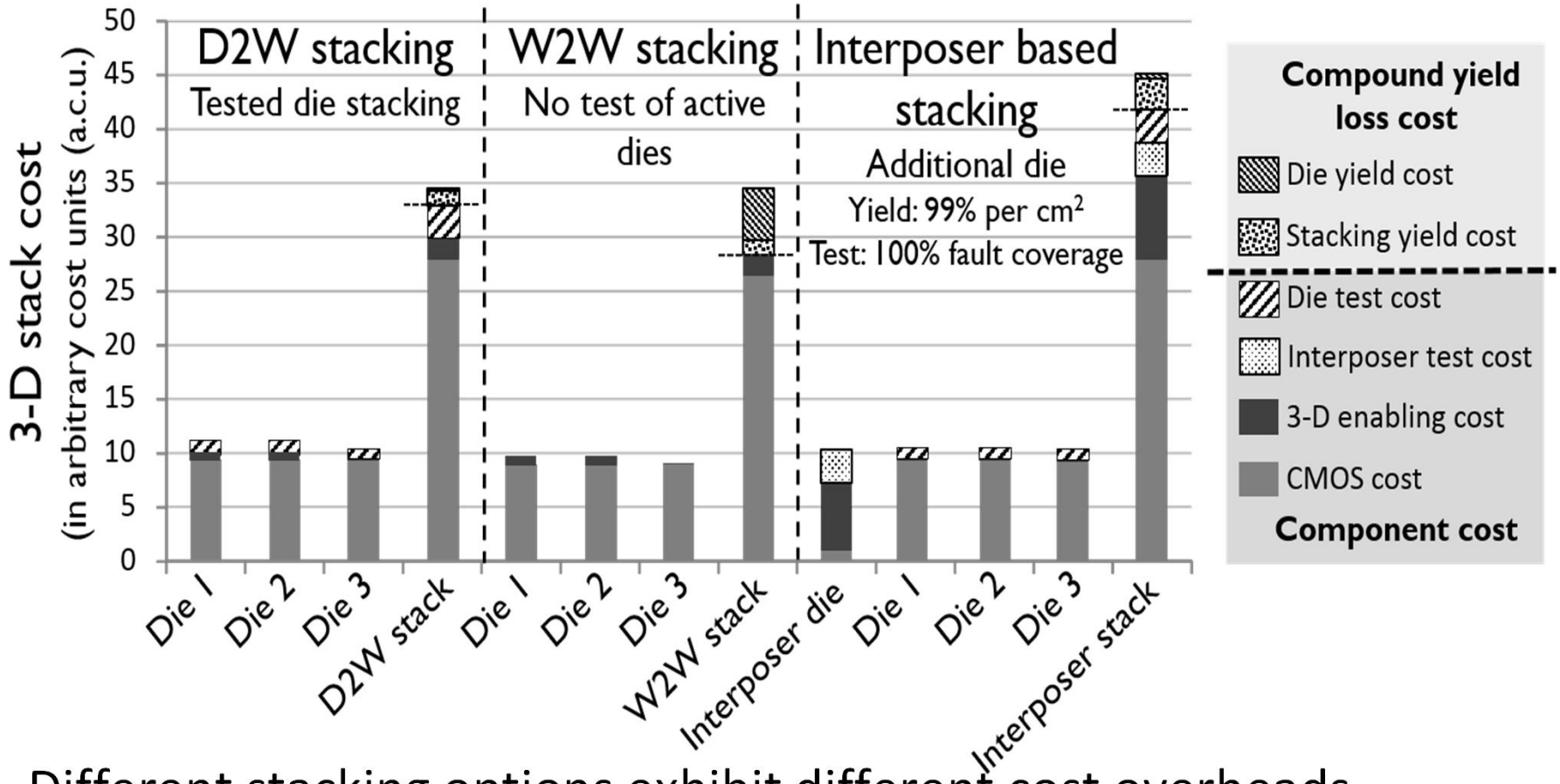
eff_{TSV}=13-72

Signal Multiplexing Efficiency



*I. Papistas and V. F. Pavlidis, "Bandwidth-to-Area Comparison of Through Silicon Vias and Inductive Links for 3-D ICs," *Proceedings of the IEEE European Conference on Circuit Theory and Design*, August 2015.

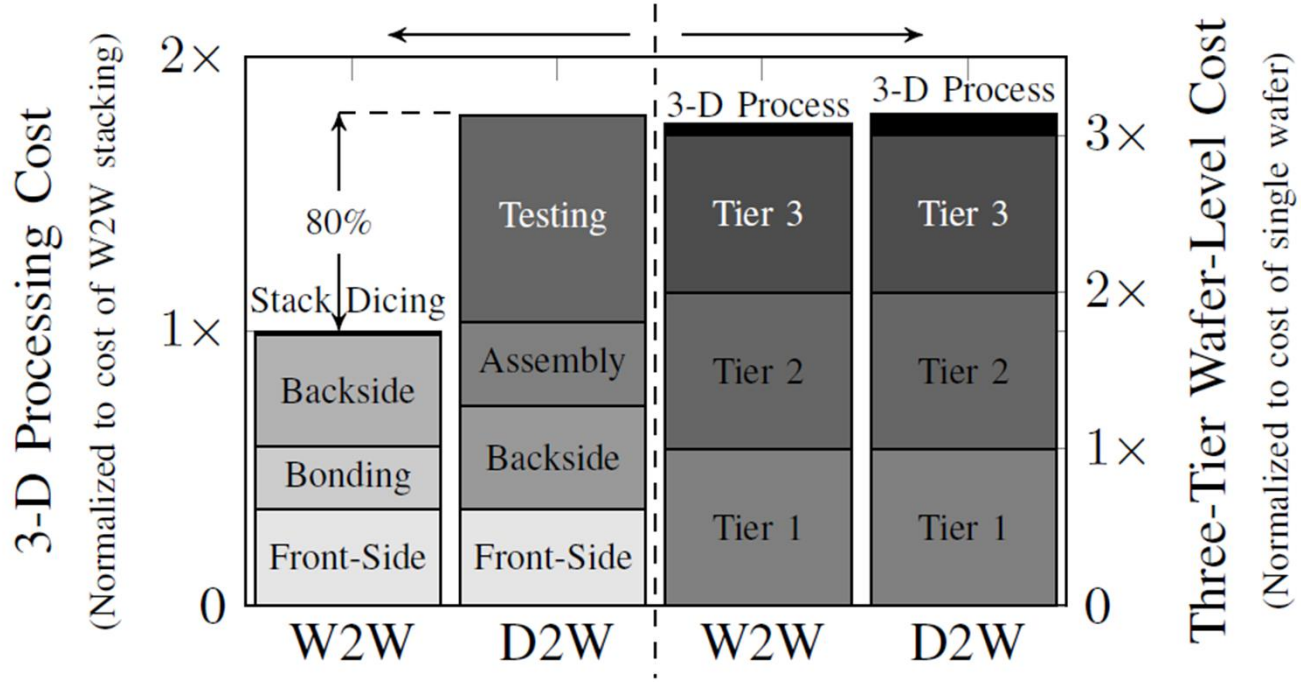
TSV Processing Impact on Fabrication Cost



- Different stacking options exhibit different cost overheads
  - TSV processing and stacking add between 15%-35% cost overhead
  - For 2.5-D systems the increase in cost reaches 66%

*V. F. Pavlidis, I. Savidis, and E. G. Friedman, *Three-Dimensional Integrated Circuit Design, 2nd Edition, Morgan Kaufmann Publishers, Elsevier,* July 2017.

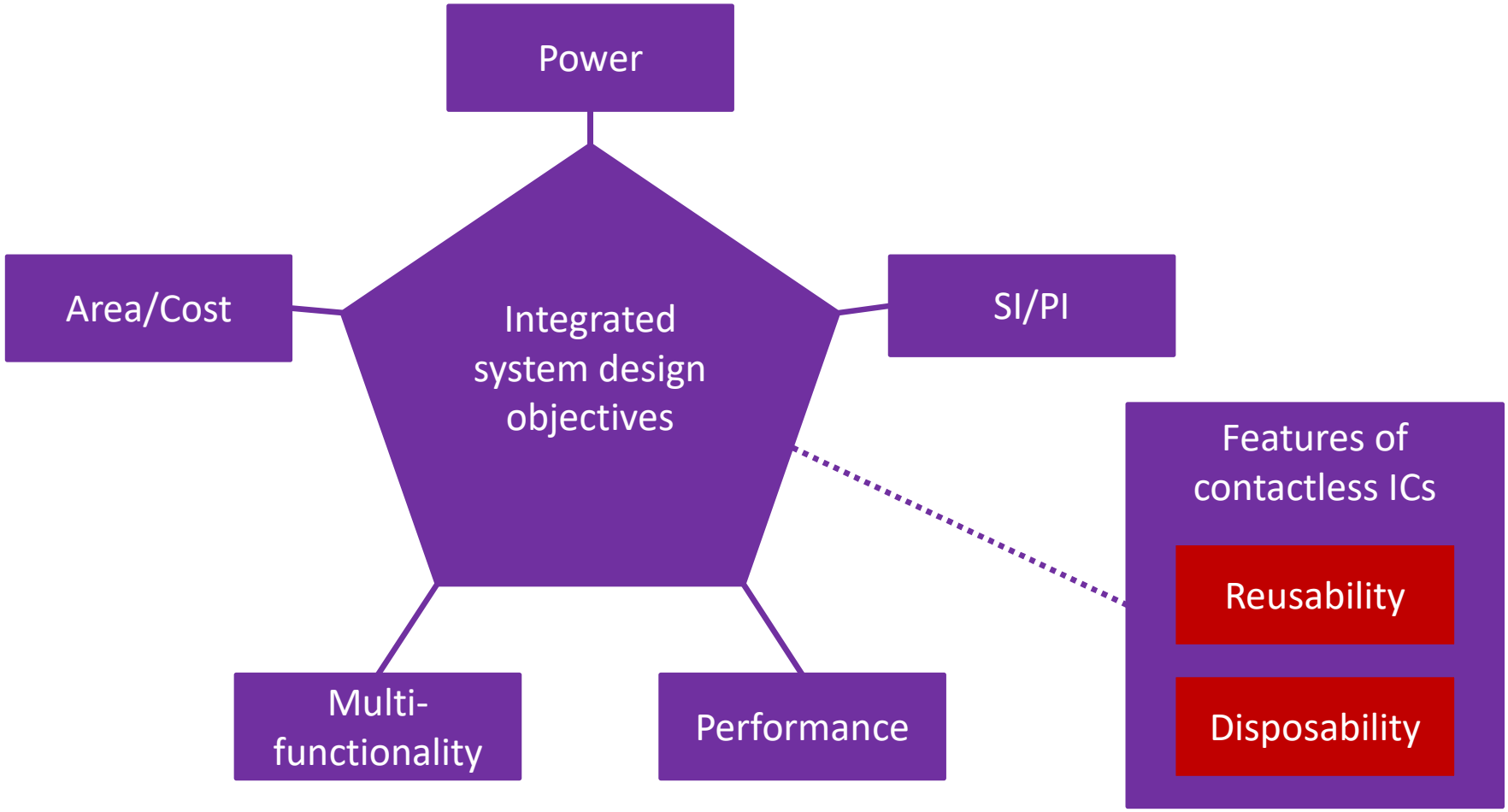
Cost Benefits from Contactless 3-D Integration



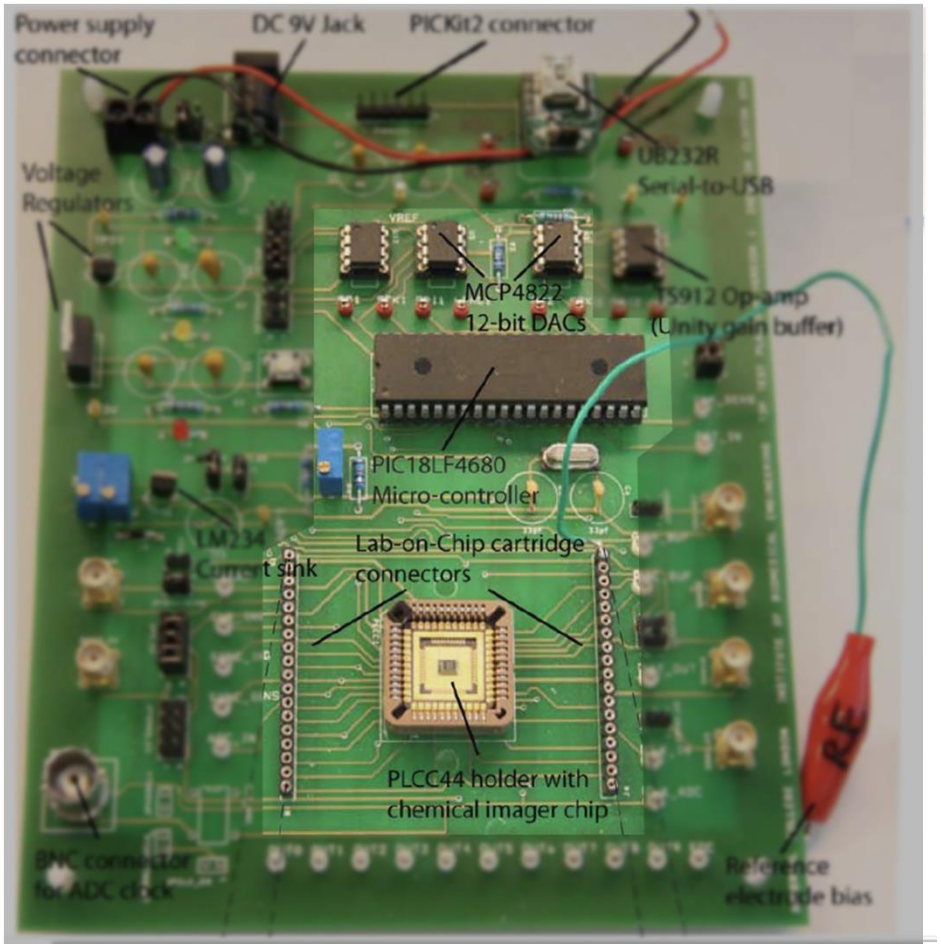
- The increase in cost from vanilla CMOS processes does not exceed a merely 5%
  - A significant cost advantage over TSV-based stacking
- This cost includes the additional test for KGD detection

*I. Papistas, V. F. Pavlidis, and D. Velenis, "Fabrication Cost Analysis for Contactless 3-D ICs," *IEEE Transactions on Circuits and Systems II: Express Briefs*, Vol. 66, No. 5, pp. 758-762, May 2019.

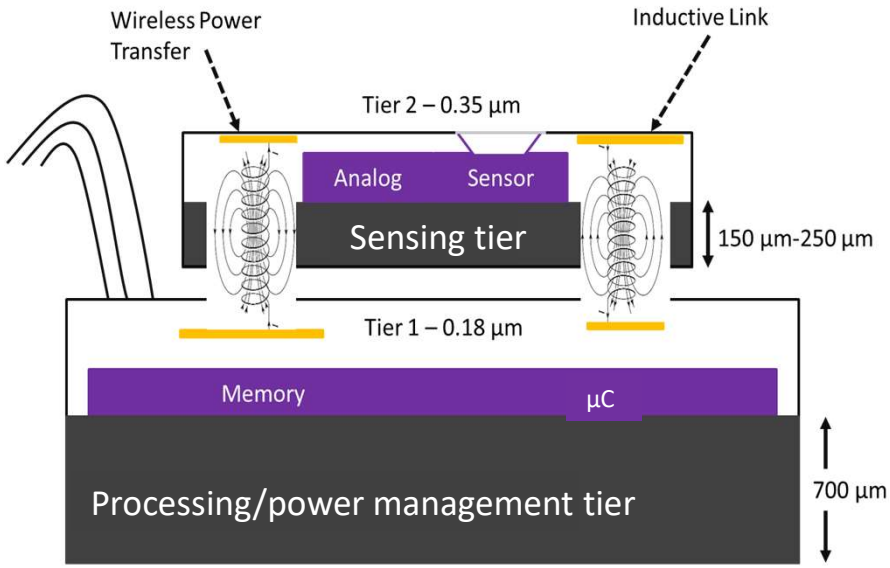
Design Objectives for Contactless ICs



Cost and Form Factor Driven Systems

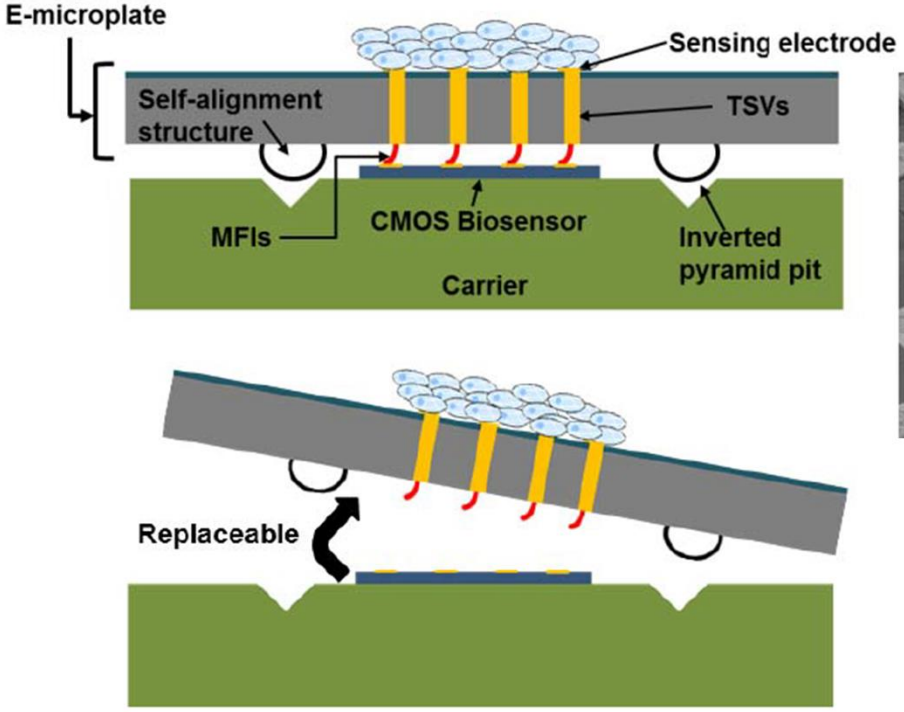


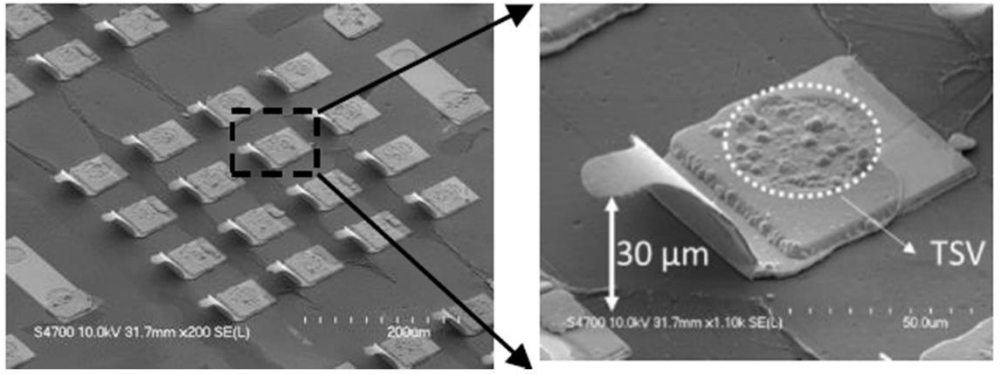
- Lab-on-chip applications
- Conventional SoC/SiP approaches do not support integration of fluidics
  - Required for chemical sensing



[*] P. Georgiou et al, Institute of Biomedical Engineering, Imperial College London.

Wired Disposable Sensor Fabrication





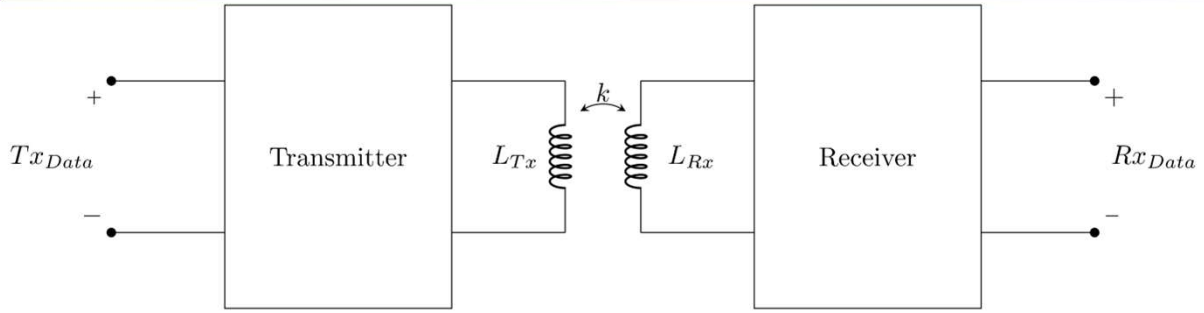
	Dimension	Value (µm)
TSVs	Diameter	50
	Height	300
	Pitch	100
MFIs	Thickness	3.5
	Vertical Height	30
	Pitch	100

*M. Zia *et al*., "3-D Integrated Electronic Microplate Platform for Low-Cost Repeatable Biosensing Applications," *IEEE Transactions on Components, Packaging, and Manufacturing Technology,* Vol. 6, No. 12, pp. 1827-1833, December 2016.

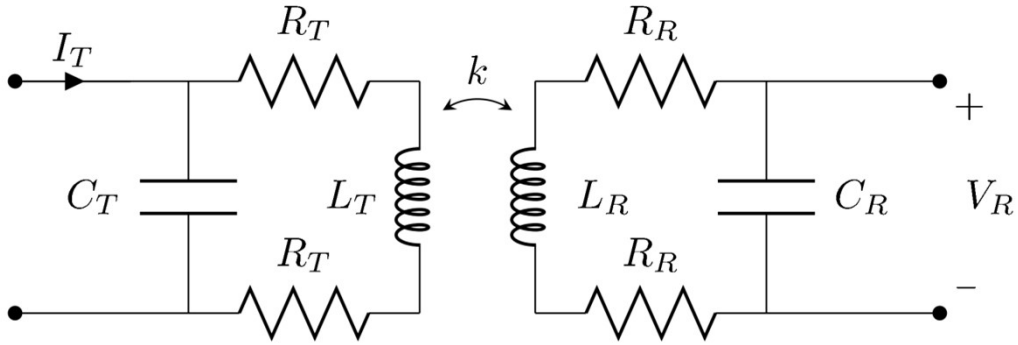
Part II – Outline

- What is Contactless Communication?
- Why Contactless Communication?
- Fundamentals of Contactless Communication
- Energy-Efficient Design of Contactless ICs
- Summary

Fundamentals of Contactless 3-D IC



$$f_{SR} = \frac{1}{2\pi\sqrt{LC}} \qquad (4)$$



$$V_R = \frac{1}{(1-\omega^2 L_R C_R)+j\omega C_R R_R} \cdot j\omega M \cdot \frac{1}{(1-\omega^2 L_T C_T)+j\omega C_T R_T} \cdot I_T \qquad (5)$$

- The short communication distance between the two inductors makes pulse modulation preferable over carrier modulation as the communication scheme

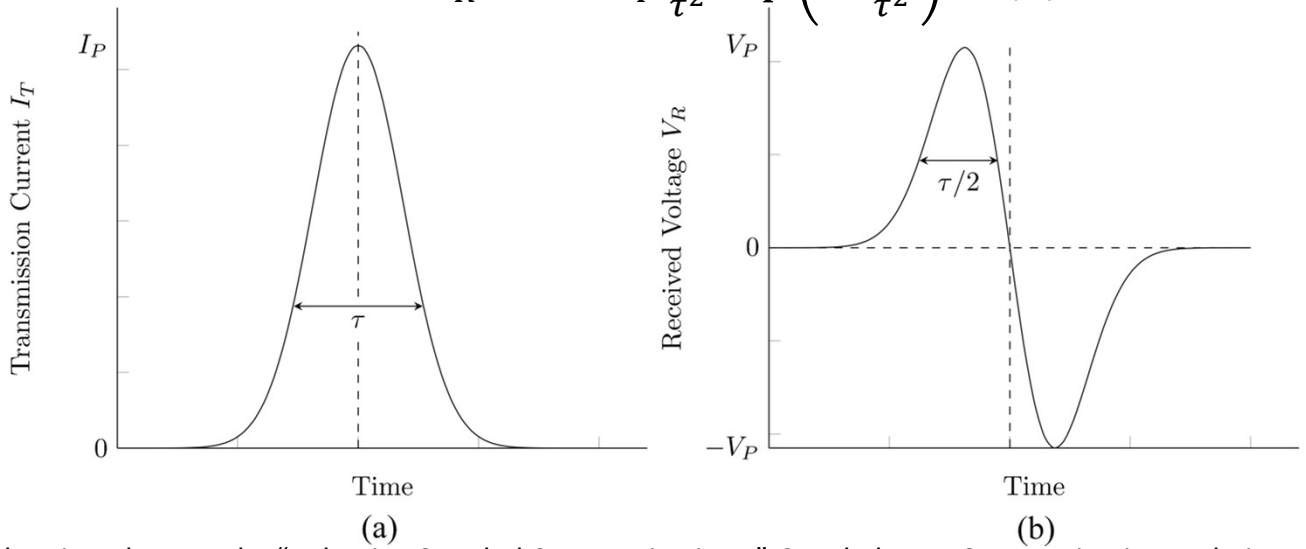
Link Modeling

- The current I_T of the transmitter produced by the digital pulse Tx_{data} is modeled as a Gaussian pulse

$$I_T = I_P \exp\left(-\frac{4t^2}{\tau^2}\right) \qquad (6)$$

- Assuming an ideal inductive link, the voltage induced on the receiver is the derivative of the transmitted pulse

$$V_R = -MI_P \frac{8t}{\tau^2} \exp\left(-\frac{4t^2}{\tau^2}\right) \qquad (7)$$



*N. Miura, T. Sakurai, and T. Kuroda, "Inductive Coupled Communications," *Coupled Data Communication Techniques for High Performance and Low-Power Computing*, pp. 79–125, Springer, 2010.

N. Miura *et al.,* "A 1 Tb/s 3 W Inductive-Coupling Transceiver for 3D-Stacked Inter-Chip Clock and Data Link," *IEEE Journal of Solid-State Circuits*, Vol. 42, No. 1, pp. 111-122, January 2007.

Bound of Resonant Frequency

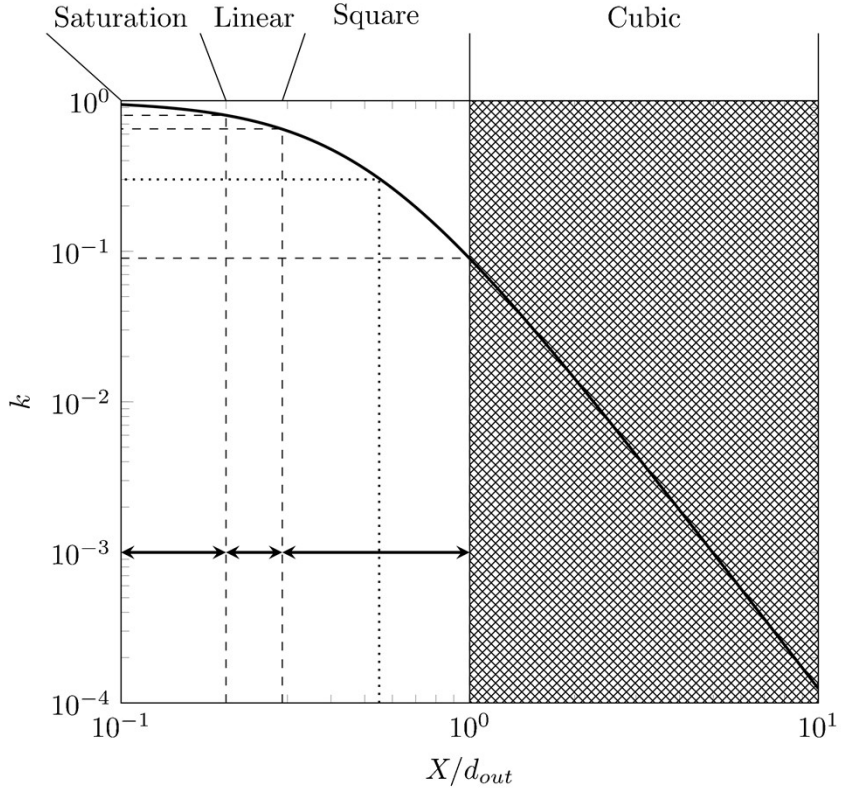
- The width τ of the current pulse is one of the primary parameters characterizing an inductive link
  - The sensitivity margin of the receiver is directly related to this width
- To avoid aliasing (or intersymbol interference), the operating frequency of the link should be greater than $2\,f_p$

$$f_{SR} > 2f_p = \frac{2\sqrt{2}}{\pi\tau} \approx \frac{0.9}{\tau} \qquad (8)$$

*N. Miura *et al.*, "A 195-Gb/s 1.2-W Inductive Inter-Chip Wireless Superconnect with Transmit Power Control Scheme for 3-D-Stacked System in a Package," *IEEE Journal of Solid-State Circuits*, Vol. 41, No. 1, pp. 23-34, January 2006.

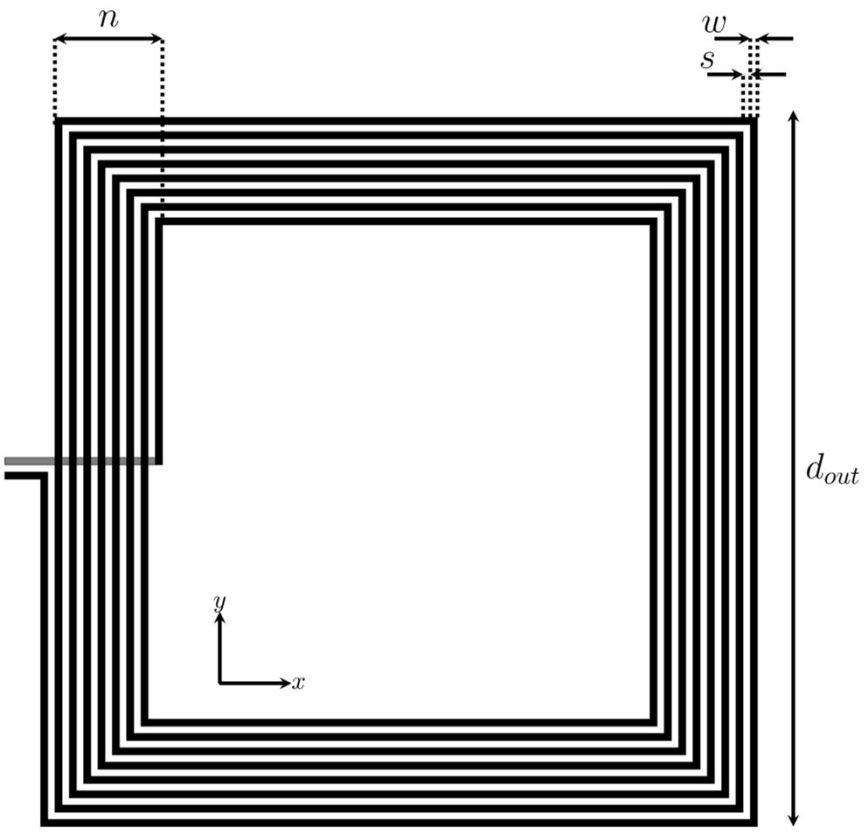
Coupling Efficiency *versus* Communication Distance

- The exponent of the dependence of coupling efficiency on distance is not constant



*N. Miura, T. Sakurai, and T. Kuroda, "Inductive Coupled Communications," *Coupled Data Communication Techniques for High Performance and Low-Power Computing*, pp. 79–125, Springer, 2010.

Inductor Diameter



- d_{out} is the outer diameter of the inductor
- n is the number of turns of the inductor
- w is the width of the turns
- s is the space between turns

$$L \propto d_{out} n^2 \qquad (9)$$

$$C \propto d_{out} n \qquad (10)$$

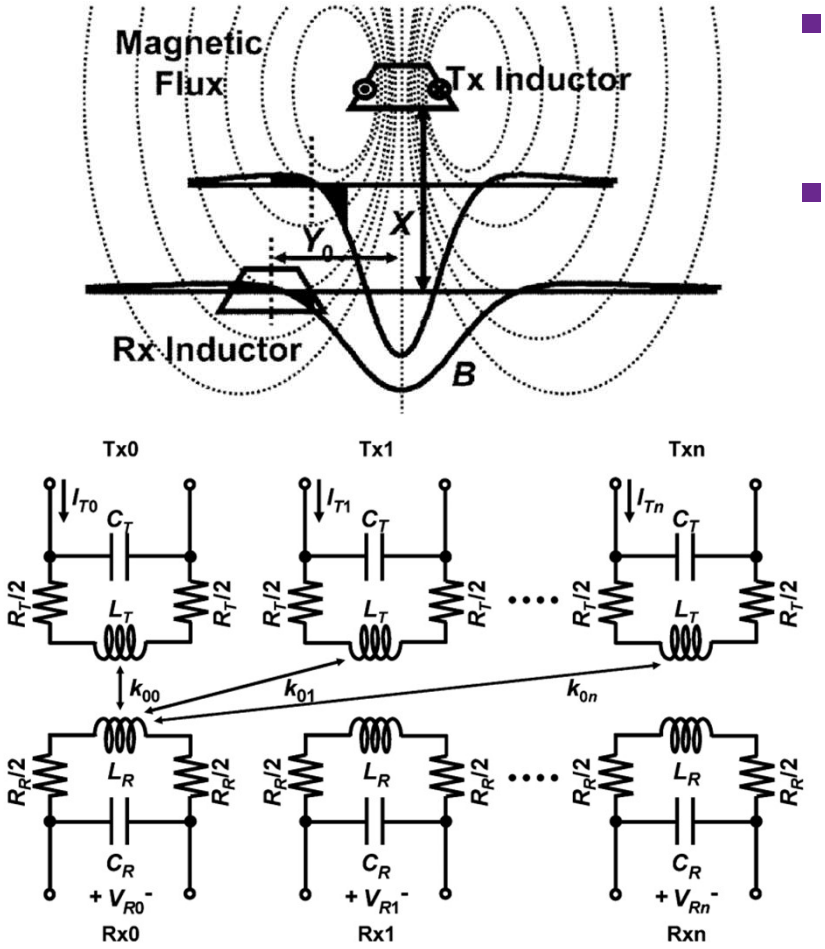
$$f_{SR} \propto \frac{1}{2\pi d_{out}\sqrt{n^3}} \qquad (11)$$

- By substituting to (4)

Inductive Link Interference

- Wireless signals couple with nearby circuits & interconnects
- Interference between inductive links is not negligible
- Effect varies depending upon the nature of “victim” circuit
- “Victim” circuits can be categorized as
  - Nearby inductive links
  - Digital circuits
  - Analog and sensing circuits
  - Signal and power on-chip interconnects

Interference on Neighboring Inductive Links



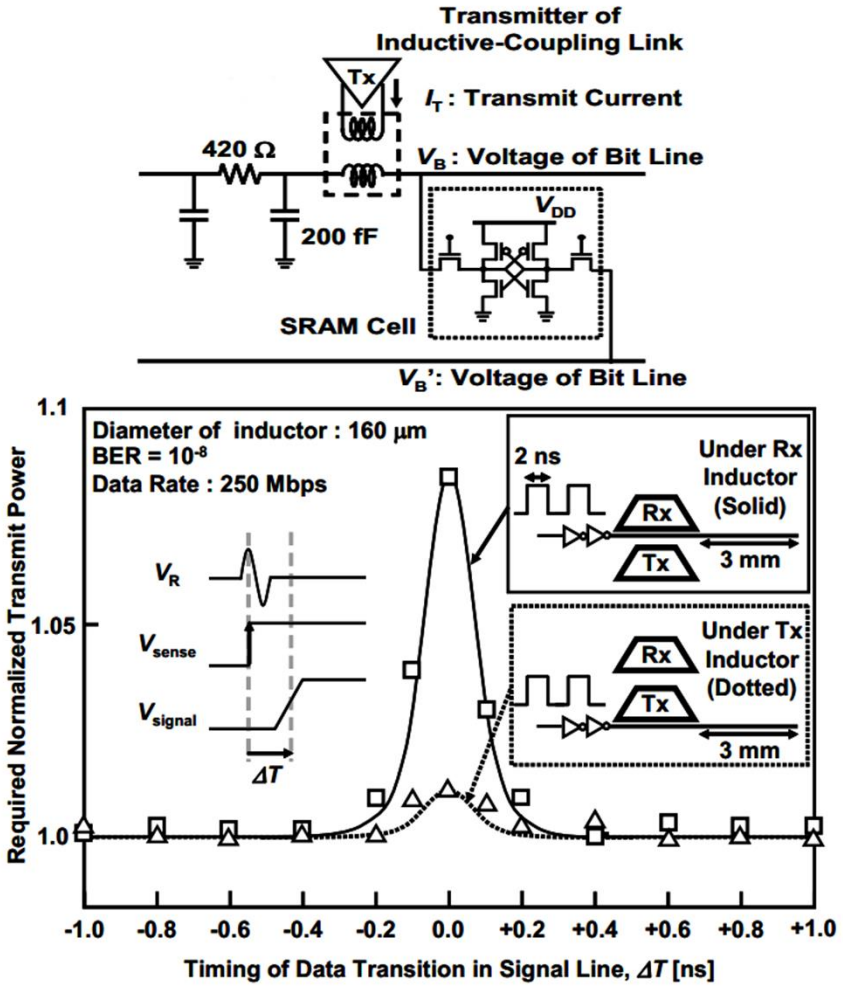
- Crosstalk to adjacent links similar to received signal (50 *mV*)
- Solutions to reduce crosstalk
  - Increase distance between links
    - Not suitable for high density applications
  - Time division interleaving technique
    - Maximum division depends upon performance constraints
    - 4-phase division sufficiently mitigates crosstalk

For 1 *Gbps* datarate time division

- 2-phase ⇒ crosstalk of 25 *mV*
- 4-phase ⇒ crosstalk of 10 *mV*

*N. Miura *et al.,* "Crosstalk Countermeasures for High-Density Inductive-Coupling Channel Array," *IEEE Journal of Solid-State Circuits*, Vol. 42, No. 2, pp. 410-421, February 2007.

Interference on Circuit Components



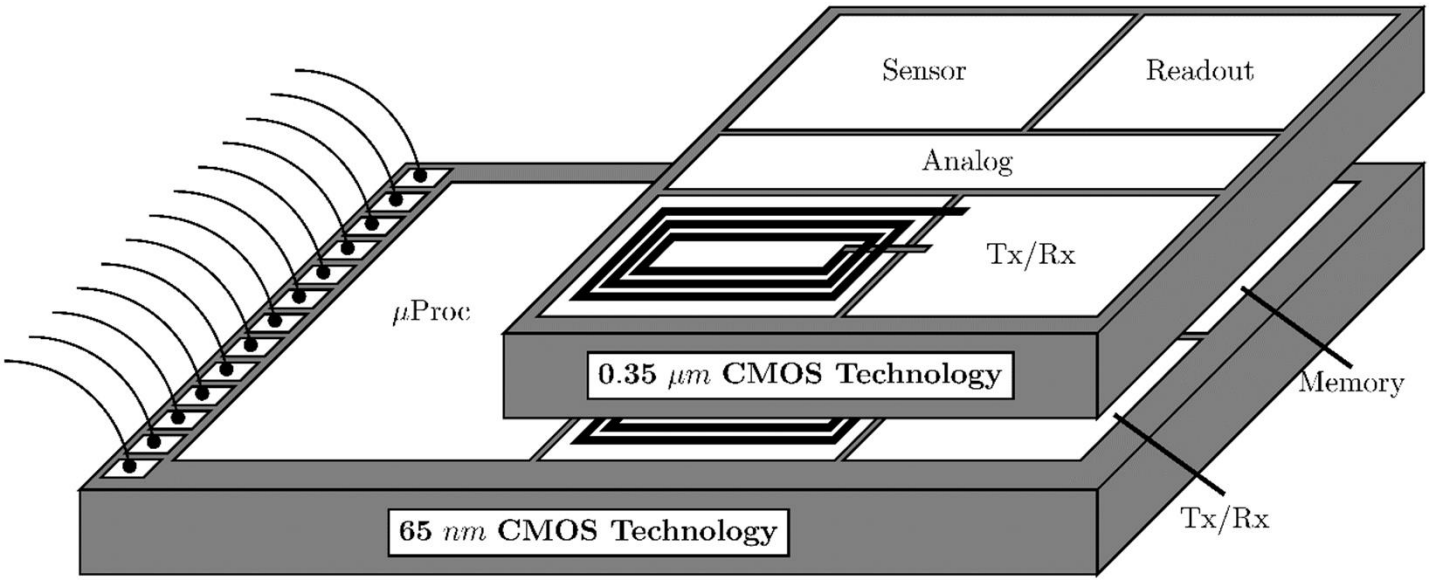
- Noise on digital circuits
  - Coupling through local interconnects
  - Crosstalk on local interconnects is negligible
    - < 1 *mV*
- Noise on global interconnects is significant
- Power integrity may be compromised in high density interfaces

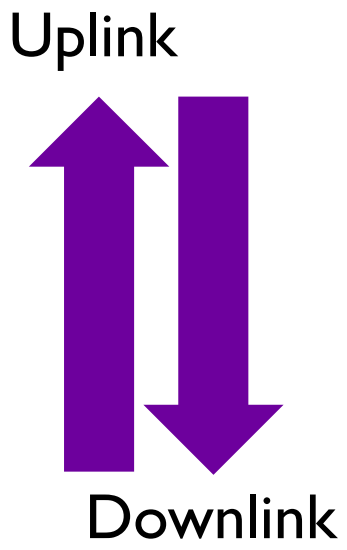
*K. Niitsu *et al.,* "Interference from Power/Signal Lines and to SRAM Circuits in 65nm CMOS Inductive-Coupling Link," *IEEE Asian Solid-State Circuits Conference*, pp.131-134, November 2007.

Part II – Outline

- What is Contactless Communication?
- Why Contactless Communication?
- Fundamentals of Contactless Communication
- Energy-Efficient Design of Contactless ICs
- Summary

Heterogeneous IoT Edge Device

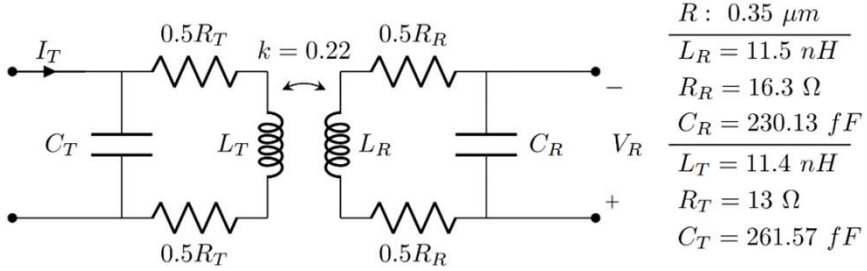




- Processing tier in 65 *nm*
- Sensing tier in 0.35 *μm*
- Stacked face-up for fluidic sensing applications
- Half duplex communication supported
- Substrate thinned to 80 *μm*

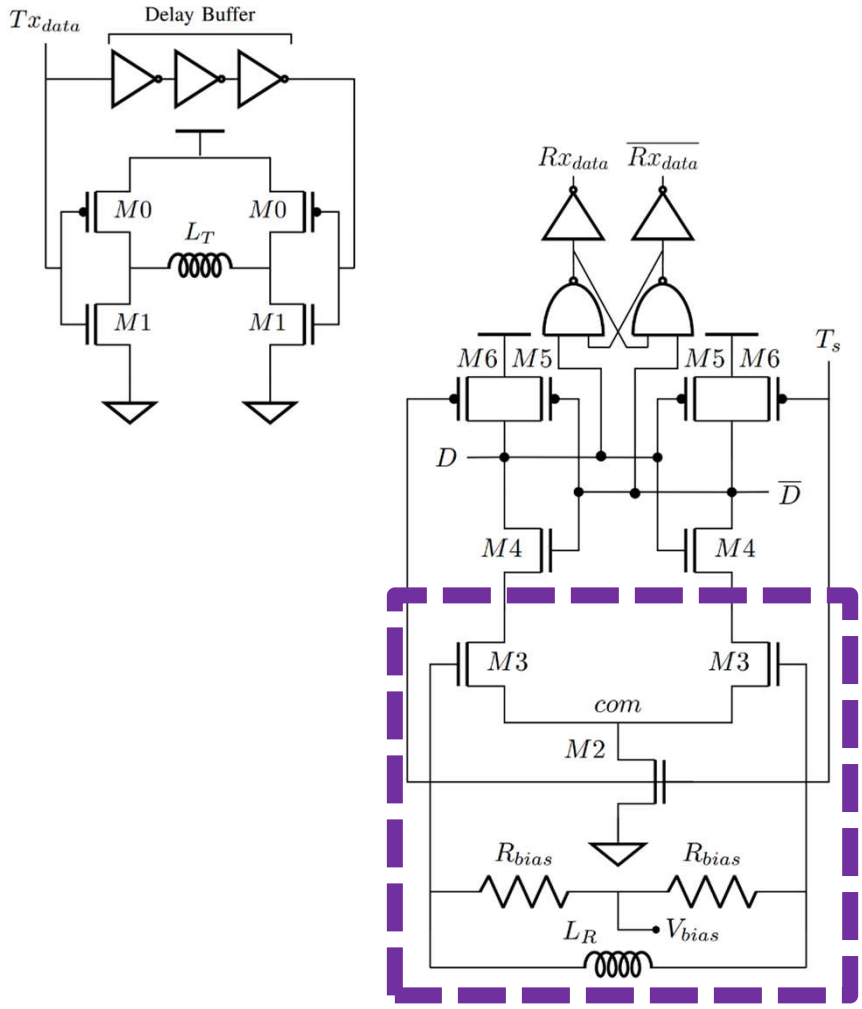
Inductive Link Area Considerations

- Coupling depends upon outer diameter and communication distance
- Minimum coupling k = 0.1
- This implementation
  - k = 0.22
  - d_{out} = 300 μm
- DRC/DFM free inductor layout using VeloceRF*
- Both inductors used as transmitters and receivers



*Helic Inc, Veloce Raptor X User Manual, November 2013, v3.

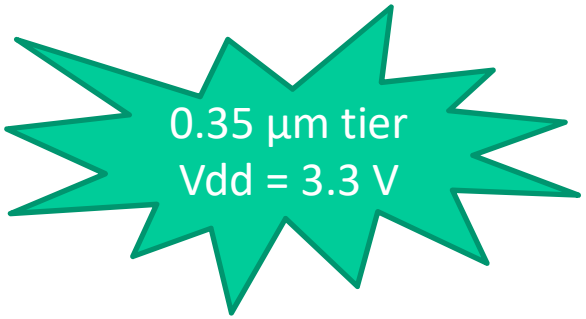
Transceiver Circuit



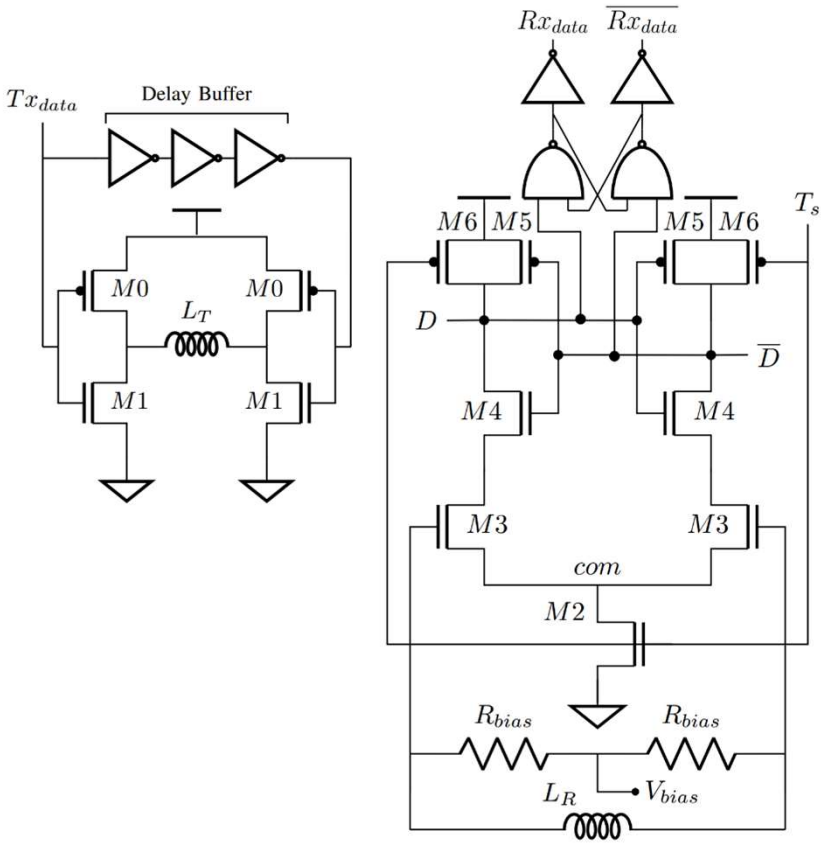
- Transceiver circuit
  - H-Bridge transmitter
  - Sense amplifier receiver
- Received pulse is sampled within a specified time interval
  - Crosstalk noise and accidental switches are reduced
- Biasing of differential pair important for circuit operation

Methodology for Power Efficiency

- Power efficiency is the primary objective
  - $P_{tot} = P_{Tx65} + P_{Tx350} + P_{Rx65} + P_{Rx350}$
- Two design approaches can be followed
  - Minimization of each power component individually ⇒ *nominal* design
  - Exploitation of core voltage in each process node ⇒ *proposed methodology*
- Tradeoff between power and sensitivity exists



Device Sizing

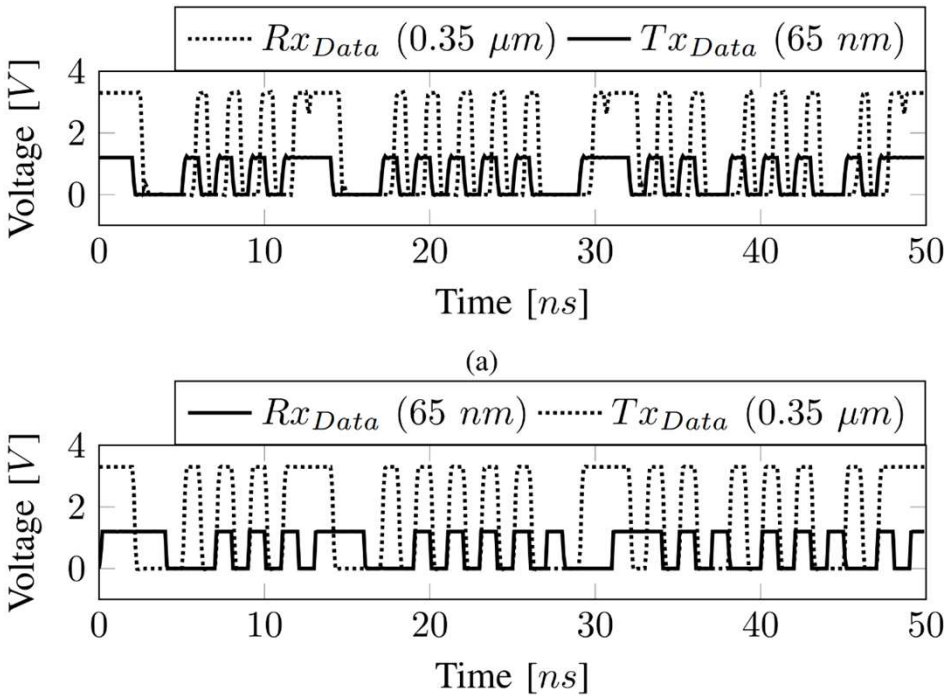


Device Name	Size [μm] 0.35 μm	65 nm
$M0$	3.74 →	6.75
$M1$	1.7 →	4.5
$M2$	1.3 →	1.5
$M3^*$	5.5 →	7.15
$M4$	0.5 →	0.6
$M5$	0.9 →	1.2
$M6$	2.6 ←	1.7

- 0.35 μm tier sized for minimum power
  - Sensitivity of 300 *mV*
- 65 nm tier sized for highest sensitivity
  - Sensitivity of 75 *mV*

- 70% decrease in 0.35 μm device width is achieved!

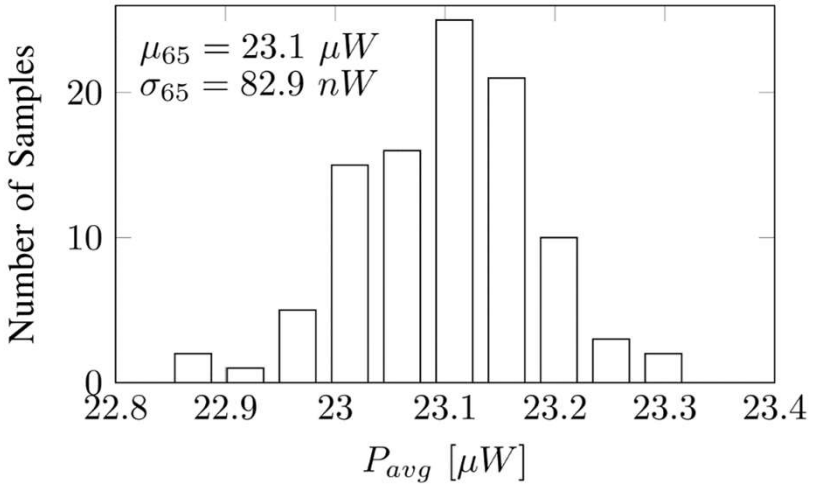
Simulation Results

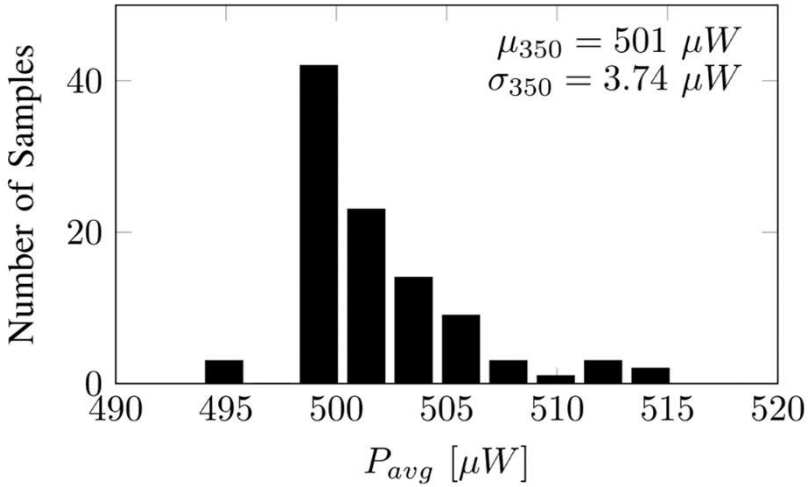


- Full swing signal at nominal voltage
- No level shifters required
- $P_{uplink} = 5.28\ mW$
  - $P_{uplink,\ avg} = 2.5\ mW$
- $P_{downlink} = 8.67\ mW$
  - $P_{downlink,\ avg} = 2.38\ mW$

*I. Papistas and V. F. Pavlidis, "Contactless Inter-Tier Communication for Heterogeneous 3-D ICs," *Proceedings of the International Symposium on Circuits and Systems*, pp. 2585-2588, May 2017.

Differential Pair Mismatch Analysis

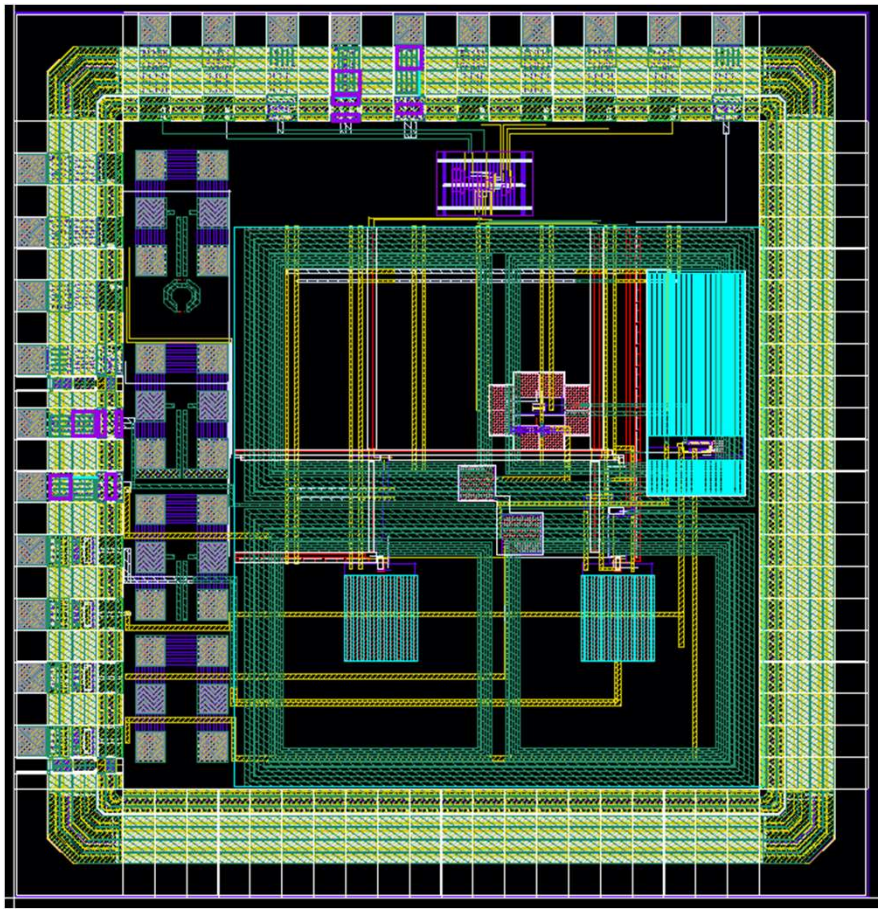




- Differential pair susceptible to device mismatch
- Length is increased to reduce the effect of random mismatch
  - L_{65} = 120 nm ⇒ overhead $\Delta Pavg$ = 7 µW
  - L_{350} = 500 nm ⇒ overhead $\Delta Pavg$ = 30 µW

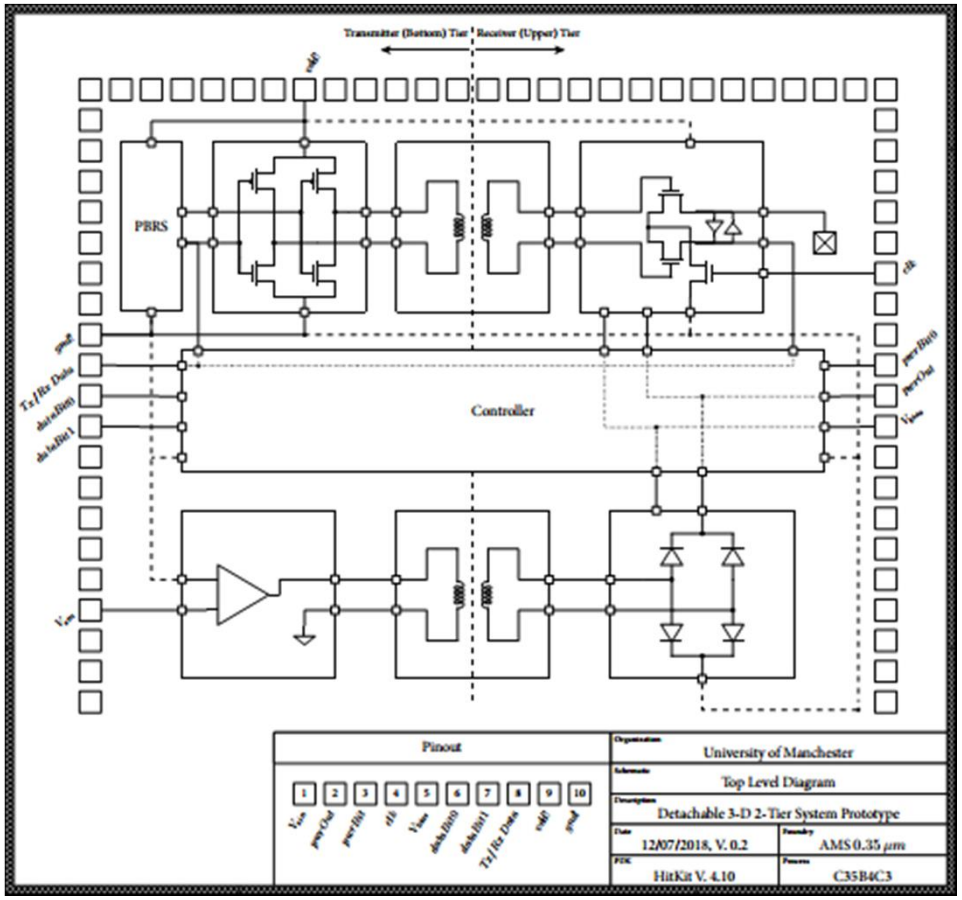
*I. Papistas and V. F. Pavlidis, "Contactless Heterogeneous 3-D ICs for Smart Sensing Systems," *Integration, the VLSI journal,* Vol. 99, No. 99, April 2018.

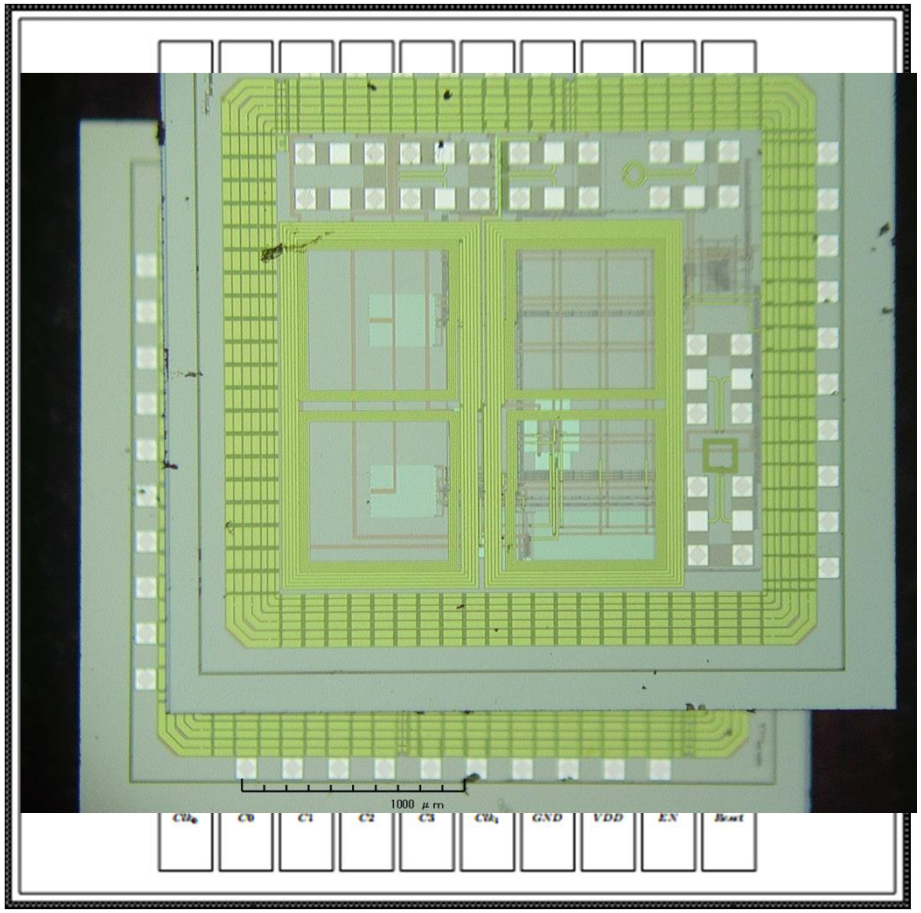
Fully Contactless 3-D Test Circuit



- Contactless power transfer
  - Two on-chip inductors
- Contactless signal transfer
  - Half/full duplex communication
- 250 µm substrate thickness
  - Overall communication distance ~270 µm
- 0.35 µm AMS technology
  - Samples delivered in June 2019

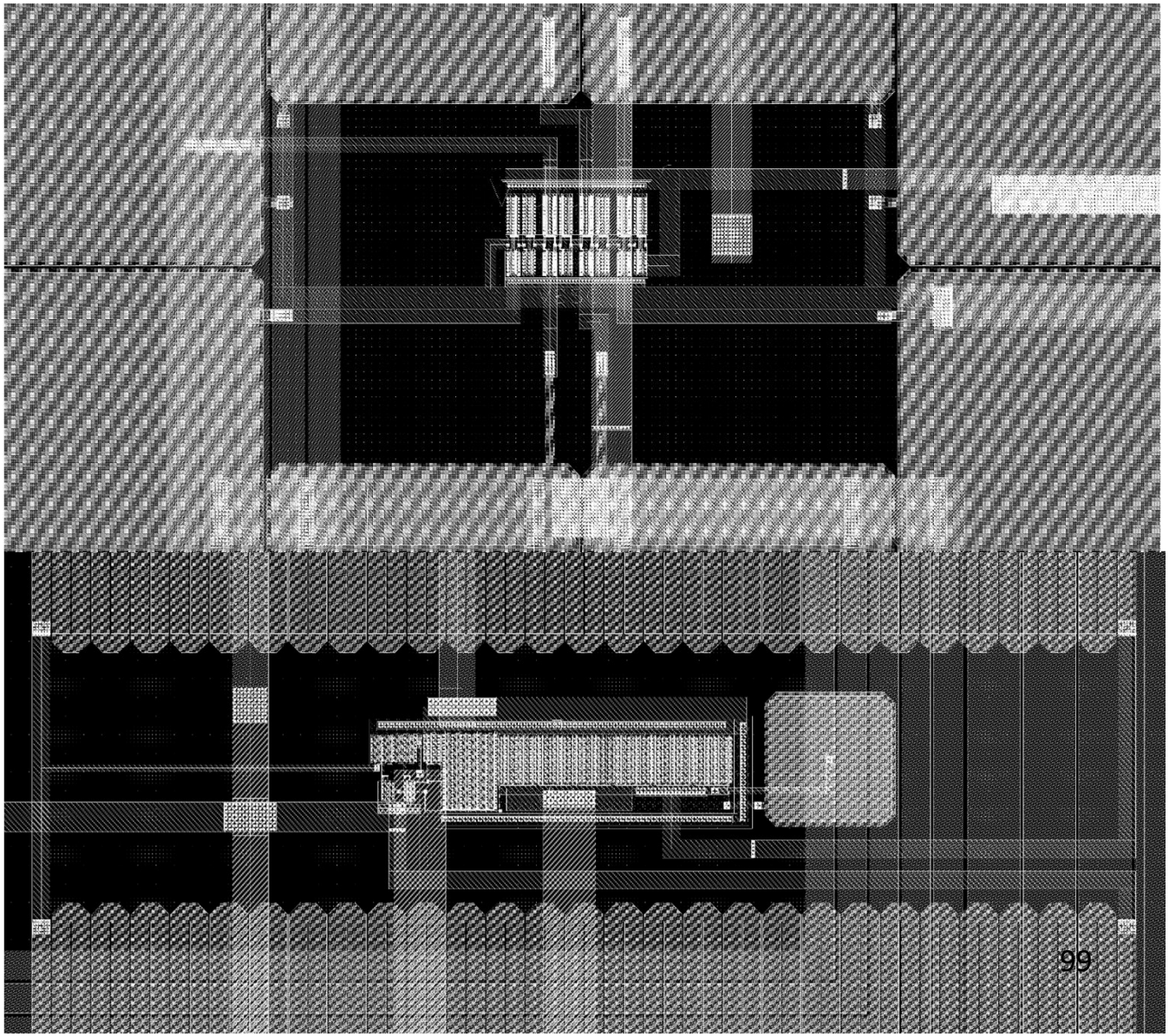
Core Components of the Test Circuit



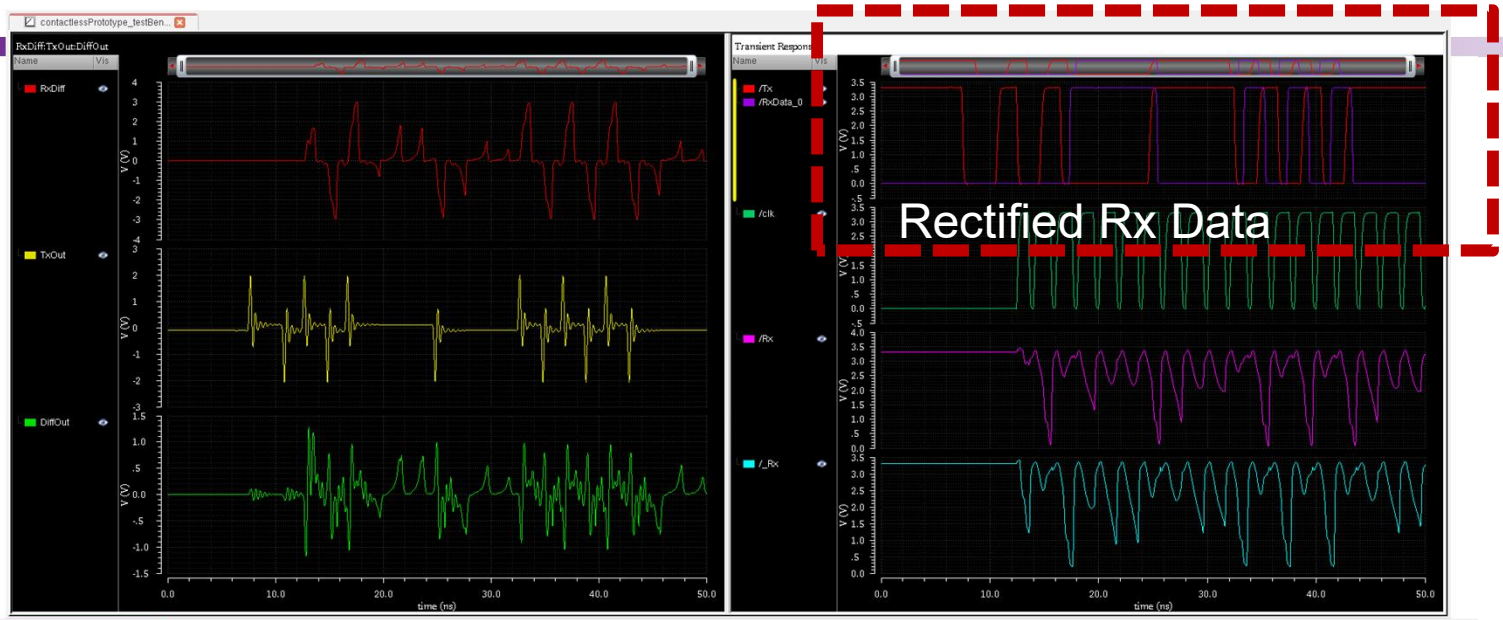


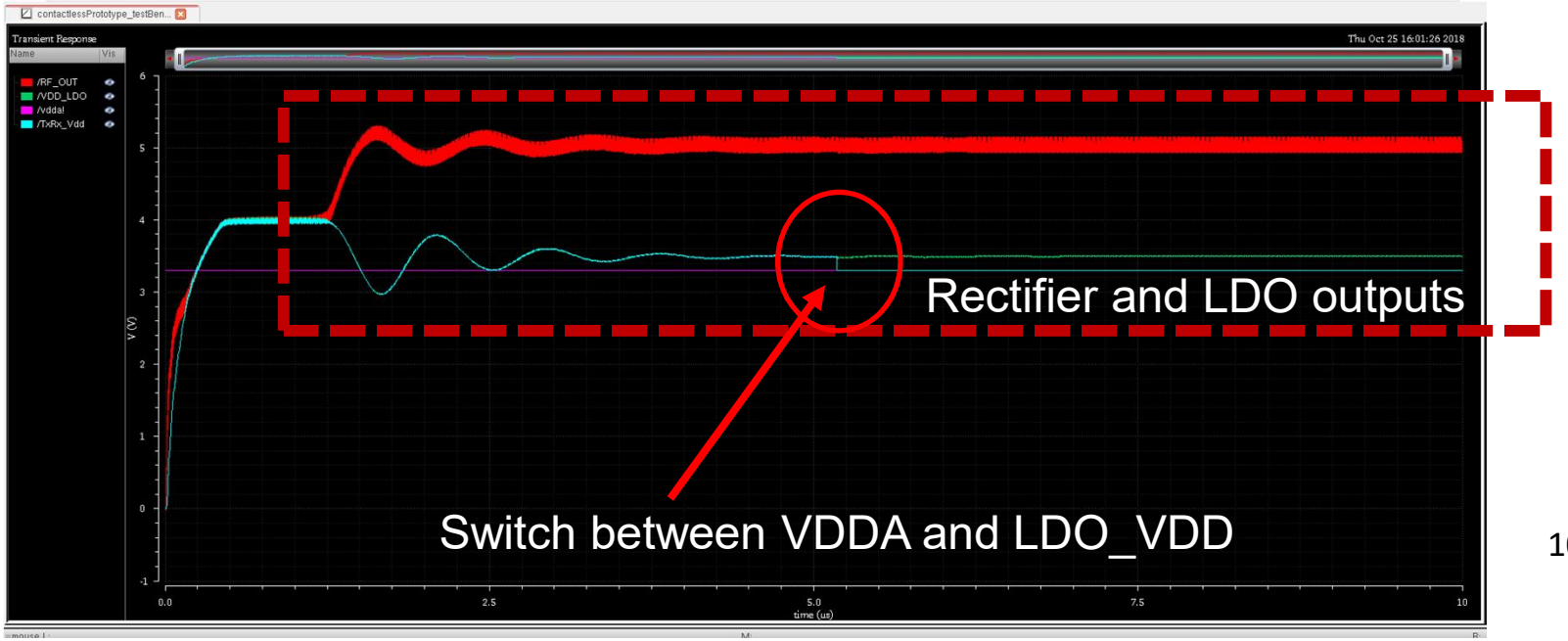
Contactless Power Transfer Block

- 4-Stage Rectifier
  - Ultra Low power diodes
  - Output is 5 *V*
- LDO
  - Output 3.4 *V*
  - Improved phase margin stability with output capacitor



Performance (Simulation) Results





MANCHESTER 1824

The University of Manchester

Part II – Outline

- What is Contactless Communication?
- Why Contactless Communication?
- Fundamentals of Contactless Communication
- Energy-Efficient Design of Contactless ICs
- Summary

Part II – Summary

- Contactless circuits are a promising 3-D integration approach enabling *disposability* and *reusability*
- Fabrication cost overhead is low (< 5%)
- Noise from on-chip inductors can be significant but mitigation techniques are available with small overhead
- Heterogeneous inductive links can lead to more economic and energy efficient solutions
- Data pulse width, communication distance, and outer diameter of the inductors are primary design parameters for inductive links

MANCHESTER 1824

The University of Manchester

Thank you for your attention!

Questions?