A Spike-Latency Transceiver with Tuneable Pulse Control for Low-Energy Wireless 3D Integration

Benjamin J. Fletcher, *Student Member, IEEE,* Shidhartha Das, *Member, IEEE,*
and Terrence Mak, *Senior Member, IEEE*

Abstract—Wireless 3D integration using Inductive Coupling Links (ICLs) has recently gained attention as a low-cost alternative to through silicon vias (TSVs) for interconnecting stacked silicon tiers. However, 3D integration using ICLs is often criticised for its inferior energy efficiency compared to conventional approaches. To address this challenge, in this paper, we present a low-energy ICL transceiver that combines: (1) a spike-latency encoding scheme (to reduce the number of energy-expensive analogue transmit pulses by encoding data in the time-domain), and (2) a tuneable current driver (to minimise the transmit energy depending on the given integration scenario). The proposed transceiver is modelled mathematically, simulated in 0.35um, 65nm and 28nm CMOS technologies, and experimentally validated in a 2-tier 3D stacked silicon test-chip. Silicon evaluation of the proposed modulation approach demonstrates an energy of 7.4pJ/bit, representing a reduction >13% when compared to previously reported schemes (or 7.4% when also considering the additional energy overheads of peripheral clock timing control circuits). Simulated results show even greater energy savings (up to 28%) at more advanced technology nodes. Combined with the adaptive current driver, this results in a 7.7× improvement in energy-per-bit compared to state-of-the-art implementations across the same communication distance, marking an important progression towards cost *and* energy efficient 3D integration.

Index Terms—3D-IC, Inductive, Wireless Links, Transceiver, Time-Domain Coding.

Fig. 1. Illustration of an example of a heterogeneous stacked 3D-IC for IoT sensing applications (containing technology nodes between $0.35\,\mu\mathrm{m}$ and 28nm), assembled using Inductive Coupling Links (ICLs) to communicate data vertically between tiers. The transceiver element (consisting of data encoder/decoder circuits, driver and sense-amplifier) has been highlighted, and forms the focus of this paper.

I. INTRODUCTION

THE Internet of Things (IoT) requires a new breed of low-power, technologically diverse integrated circuits (ICs) to combine analogue sensing, digital processing and novel NVM technologies in a low-power, small form-factor way [1]. To achieve this, designers are exploring 3D integration where multiple tiers, each of which may be fabricated in a different process technology, are stacked and connected vertically [2].

In order to provide vertical connectivity between stacked tiers, a range of 3D integration methodologies have been explored, the most straightforward of these being face-to-face (F2F) stacking using, for example, flip-chip bonding. F2F stacking can be performed with relatively low cost, however, limits the maximum number of stackable dies to two, making it impossible to realise the *highly heterogeneous* 3D-ICs discussed above. Other approaches to 3D integration include 3D System-in-Package (SiP) solutions, for example stacking
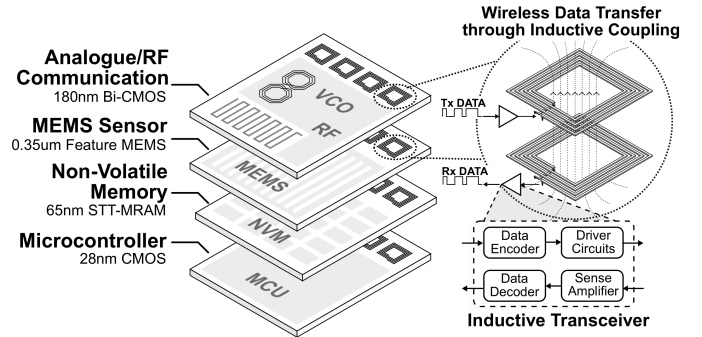
B. J. Fletcher and T. Mak are with the Department of Electronics and Computer Science University of Southampton, UK.
E-mail: {bjf1g13,t.mak}@ecs.soton.ac.uk.
S. Das is with ARM Ltd, Cambridge, UK. E-mail: shidhartha.das@arm.com. B. J. Fletcher is also a iCASE student at Arm Ltd.

multiple dies in a staggered arrangement and adding wirebonds to interconnect each die (such as used in [3]). Whilst these 3D-SiP approaches overcome the 2-tier limit associated with F2F approaches, the custom bonding patterns required in such ICs are typically expensive and difficult to scale to mass production. One final approach to realising highly heterogeneous 3D-ICs is using through silicon vias (TSVs). TSVs are conductive pathways that are etched entirely through the silicon substrate, allowing electrical communication between the front and back of the die. TSVs therefore allow face-to-back (F2B) stacking, hence providing the ability to combine several dies within the same IC, with high vertical interconnect bandwidths. This makes them a promising solution for many applications (such as 3D-stacked DRAM *etc.*), however TSVs are expensive to manufacture and presently only available at leading-edge foundry technology nodes (rather than the full diversity of technologies discussed in the opening paragraph).

When considering the context of the IoT, each of the 3D integration approaches that have been discussed (F2F stacking, 3D SiP assembly, and F2B stacking using TSVs) have associated drawbacks (stack-height, cost, and process availability respectively). Motivated by addressing these drawbacks, more recent research has looked to the use of Inductive Coupling Links (ICLs) to provide low-cost, highly reliable vertical integration at *any* technology node [4]. Fig. 1 illustrates one such heterogeneous 3D-IC (typical of an IoT application) using ICLs to interconnect layers. Here, data is encoded in a series of current pulses which are fed through planar inductors fabricated in the upper Back-End-Of-Line (BEOL)

interconnect layers of the transmitting (Tx) die. These Tx current pulses cause a magnetic field that is intersected by a similar planar inductor fabricated in the receiving (Rx) die, and hence induce a corresponding voltage signal. This signal can then be used to recover the transmitted data stream, as highlighted on Fig. 1.

In our previous work, [5], we compared TSV and ICL-based (wireless) communication approaches and found that the bandwidth-per-unit-area which can be achieved using TSVs is significantly greater than that using ICLs (approximately $25\times$ [5]). However, ICLs can be realised with substantially lower cost (in terms of design, manufacture and assembly) [6], [7] as they do not require the additional fabrication stages associated with TSV processing, nor do they require TSV-aware EDA tools. In addition to this, the near-field inductive approach to wireless data transmission can be extended to wireless clock distribution [8] and wireless power transfer [9], enabling potential for fully wireless 3D assembly. When using this approach, once manufactured, dies can be simply picked and stacked using adhesive. This makes them an attractive option for IoT applications which are driven by cost, scalability and design-time, rather than performance. Impressive 3D-ICs constructed using ICLs have been demonstrated in several publications [7], [10]–[15], however one of the reported drawbacks of using ICLs is their inferior energy efficiency [10], [16], which is of significant importance for the IoT.

In this paper we address this challenge, presenting a low-energy ICL transceiver that uses a time-domain modulation approach to encode data. Prior implementations of ICLs use coding schemes where one or two data bits are mapped to one-or-more transmit (Tx) current pulses, resulting in significant energy consumption when implemented on-chip. The approach applied in this paper uses the latency *between* pulses to encode frames of data, thereby reducing the number of Tx current pulses and overall energy. This encoding approach is also combined with a tuneable current driver to minimise the transmit current for a given integration scenario. The main contributions of this work can therefore be summarised as:

- A low-energy inductive transceiver that applies a time-domain encoding approach (spike-latency encoding) in the context of intra-chip communication for communication between tiers of a 3D-IC. The approach uses the latency *between* sequential pulses to represent data, hence reducing the required transmit energy.
- Mathematical modelling of the proposed transceiver design for evaluating best-performing algorithm parameters across a range of 3D integration scenarios.
- A tuneable current driver circuit, to precisely control the Tx energy (within 0.25pJ) depending on the channel quality (and hence compensate for up to $40\,\mu\text{m}$ of die-to-die stacking misalignment in both x and y directions by post-assembly tuning).
- Validation of the proposed transceiver using post-layout SPICE simulations in $0.35\,\mu\text{m}$, 65nm, and 28nm technologies, demonstrating an energy consumption as low as 0.26pJ/bit across a $110\,\mu\text{m}$ channel (a 28% improvement
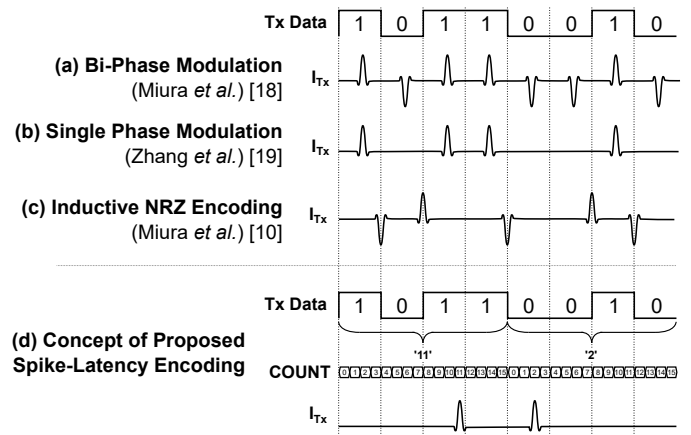

Fig. 2. Illustration of: (a) Bi-Phase Modulation (BPM) [18], (b) Single Phase Modulation (SPM) [19], [20] (c) Inductive Non-Return-to-Zero (NRZ) line code (proposed by Miura *et al.*, this is the state-of-the art approach for inductively coupled communication within a 3D-SiP/3D-IC context) [10]–[13]), and (d) The spike-latency encoding scheme proposed in this paper.

compared with the state-of-the-art)[1].
- Silicon validation of the proposed transceiver on a 2-tier 3D stacked test-chip in a $0.35\,\mu\text{m}$ CMOS technology [17], demonstrating a 13% reduction in energy-per-bit when compared with state-of-the-art transceivers.

The remainder of the paper is organised as follows: Section II presents a survey of background work related to ICLs and their modulation schemes, Section III outlines the spike-latency encoding scheme proposed in this paper (including mathematical modelling in Section III-A). Section IV outlines the hardware implementation of the transceiver, including the tuneable pulse driver circuit, before validation is performed Sections V and VI. Finally, discussion and the conclusion are presented in Sections VII and VIII respectively.

II. Background and Related Work

When using inductive coupling links (ICLs) to interconnect stacked dies, Tx data is encoded as an alternating current that is fed through a planar spiral inductor, inducing a magnetic field (corresponding to the data stream) within the die stack.

In order to minimise the power consumption of the transceiver whilst maximising dI_{Tx}/dt (and hence the magnetic flux linkage within the die stack) most ICL transceivers use *pulse-based* modulation schemes where the flow of transmit current I_{Tx} is limited to a short duration. Fig. 2(a-c) shows a range of previously published pulse-based encoding schemes for mapping the data bits to current pulses in ICL transceivers.

Bi-phase Modulation (BPM), shown in Fig. 2(a), is arguably the most straight-forward approach where '1's are mapped to I_{Tx} pulses with positive polarity, and '0's are mapped to pulses with negative polarity (or vice-versa). Whilst this is a robust solution that is often used for high bandwidth applications [18], [21], [22] (due to its favourable noise immunity), of the schemes discussed in this section, BPM suffers from the highest natural energy-per-bit with one pulse-per transmitted

---

[1]Simulated results in 28nm technology. Simulated energy per bit in 65nm is 0.66pJ and measured energy in $0.35\,\mu\text{m}$ is 7.4pJ.

bit. Works such as [10] (which demonstrates significant energy savings using digitally controlled pulse shaping), [23] (which presents a charge recycling scheme to reduce energy across a bank of many ICL channels), and [24] (which uses a dual coil transmission scheme) all use BPM in conjunction with other circuit techniques to reduce energy.

One alternative encoding scheme is Single-Phase Modulation (SPM), proposed by [19], shown in Fig. 2(b). Here, '1's are represented by the presence of an I_{Tx} pulse, and '0's are represented by the absence of an I_{Tx} pulse. This has the benefit of an intrinsic energy reduction because, assuming an equiprobable random binary bit stream, SPM requires only one pulse per two Tx bits [20]. This energy reduction, however, comes at the expense of reduced noise immunity due to the fact that there is no phase difference between '1' and '0' bits (all pulses have the same polarity).

To overcome this issue, whilst maintaining the power benefits of SPM, the majority of works exploring wireless 3D integration use the inductive non-return to zero (NRZ) signalling scheme, proposed in [13] by Miura *et al.*. This approach is illustrated by the waveforms in Fig. 2(c). Here, each rising/falling *data edge* is encoded as a current pulse with corresponding positive/negative polarity. This is a robust solution that allows data to be simply encoded using a delay buffer, and decoded using just a sense amplifier (SA) and set-reset (SR) latch [10]–[13], [15]. Assuming that the data stream is an equiprobable random bit sequence, the NRZ scheme still uses, on average, only one I_{Tx} pulse per two transmitted bits, however the 180 degree phase difference (inverted phases or pulse polarities are used to represent rising and falling edges) is maintained. This makes it particularly favourable for robust low-energy communication in prior published works.

Whilst NRZ inductive line code is the most common encoding scheme in the context of ICL communication schemes, when using NRZ, by far the largest contribution is the drive circuit power consumption (over 80% [25]). This is because each I_{Tx} pulse is very energy-expensive [13], especially when communicating over large distances (such as multiple stacked dies). As a result of this, ICL transceivers using these schemes are still power hungry.

III. PROPOSED SPIKE-LATENCY ENCODING MODULATION SCHEME

To address the high Tx power consumption of existing ICL transceivers, in this paper, we propose the use of *spike-latency* encoding to encode data frames in the time domain. Under the proposed scheme, values are not represented directly by current pulse patterns, but by the latency *between* the start of the frame, and the transmit current pulse (a form of Pulse Position Modulation). Fig. 2(d) illustrates this concept. Here, N bits (in this example N=4) are translated into a decimal value which is represented by a single I_{Tx} pulse. This pulse is transmitted with a *latency* δ, where δ is proportional to the decimal value of the N encoded bits. In other words, the *value* of the bits is represented by the transmission *latency* of the Tx pulse representing them.

In this example, 'b1011 is denoted by transceiving a pulse when the Rx/Tx counter (COUNT) is at time value 11,

TABLE I
NOMENCLATURE FOR THE EQUATIONS OUTLINED IN THIS SECTION.

Parameter	Description
N	Algorithmic parameter representing the number of binary bits mapped to a single Tx pulse
I_{Tx}	Transmitter current
I_{Rx}	Receiver current
I_{SL}	Digital supporting logic current
$E_{pb_{\text{SET}}}$, $E_{pb_{\text{BPM}}}$, $E_{pb_{\text{SPM}}}$, $E_{pb_{\text{NRZ}}}$	The energy per transmitted bit when using the proposed Spike Encoding Transceiver (SET) Modulation / Bi-Phase Modulation (BPM) / Single Phase Modulation (SPM) / Non Return to Zero (NRZ) Modulation
V	Supply voltage
f_{DAT}	Tx/Rx data frequency
f_{COUNT}	Counter frequency (when using the proposed SET scheme)
L_{Tx} or L_{Tx}	Inductance of Tx or Rx inductor
k	Electromagnetic coupling coefficient between the TX and Rx inductors
I_p	Tx Pulse current
δ	Pulse duration
I_{DFF}, I_{XOR}, I_{AND}	Current consumed by a single DFF / XOR / AND cell

and 'b0010 is denoted by transceiving a pulse when the Rx/Tx counter (COUNT) is at time value 2. This scheme is only possible provided that precise counter synchronisation[2] is available between the transmitter and receiver making it suited to 3D-IC/3D SiP applications where the channel is fixed and communication is over a short distance. As with the other encoding schemes (SPM, BPM and NRZ encoding, discussed above), the data-bit to Tx pulse ratio can be further increased by encoding one bit in terms of the *phase*, or polarity of the Tx pulse.

The advantage of using the proposed Spike-latency Encoding Transceiver (SET) is that the number of I_{Tx} pulses required to transmit a given bit stream is significantly reduced. To transmit i bits using BPM requires i, pulses. To transmit i bits using SPM or NRZ requires, on average, $i/2$ pulses, but to transmit i bits using the SET scheme requires only i/N pulses, allowing for a large Tx energy saving. However, as N increases, the COUNT frequency (and hence supporting digital logic energy required to maintain the existing data rate) increases proportionally to 2^{N-1}. This faster clock results in energy overheads in the supporting logic (in addition to extra clock distribution and synchronisation energy[3]). Therefore, when using SET, the parameter N must be carefully selected to best-exploit the trade-off between the reduction in Tx pulse energy and the corresponding increase in digital logic energy by considering the transceiver design *as a whole*. Section III-A below provides mathematical modelling to explore this trade-off in more detail.

A. Mathematical Modelling

As discussed above, when using the proposed spike-latency encoding scheme, it is important to select an appropriate value for the parameter N which trades off a reduction in the number

---

[2]Discussion of the different approaches that can be used for clock synchronisation is provided in IV-E.

[3]The contribution of these overheads is discussed in Section IV-E, and evaluated in Section V-C5
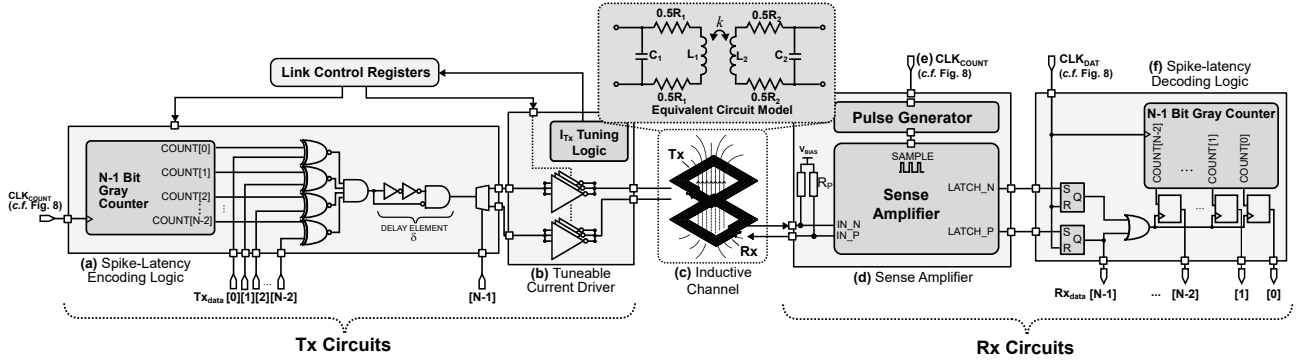
Fig. 3. Schematic diagram showing architecture of proposed low-energy inductive coupling link consisting of: (a) spike latency encoding logic, (b) an tuneable current driver, to minimise the transmit current depending on the assembly quality, (c) the inductive coupling channel, (d) a sense amplifier to sample the received voltage signal, and (e) the spike latency decoding logic.

of transmit pulses against additional digital processing. This section provides more in-depth modelling of this trade-off.

A typical ICL architecture has three main sources of power consumption:

1) The analogue transmit current (I_{Tx}) through the driver circuits and Tx inductor to form the magnetic field.
2) The analogue receive current (I_{Rx}) consumed by the Rx amplifier detecting the induced Rx voltage.
3) The current consumed by the supporting digital logic (I_{SL}), including the data encoding/decoding circuits.

For the proposed Spike-latency Encoding Transceiver (SET) scheme, the energy-per-bit ($E_{pb_{\text{SET}}}$) is therefore given by:

$$E_{pb_{\text{SET}}} = \frac{V}{N} \int_0^{\frac{1}{f_C}} I_{\text{Tx}}(t)dt + \frac{2^{N-1}V}{N} \int_0^{\frac{1}{f_C}} I_{\text{Rx}}(t) + I_{\text{SL}}(t,N)dt \tag{1}$$

where V is the supply voltage and f_C is the link counter frequency (which will be equivalent to the data frequency $f_D/2^{N-1}$). Here, the first term represents the transmit pulse current, which will decrease by $1/N$ as N increases (as more bits are encoded using a single pulse). The second term represents the current consumed by the sense amplifier; as N increases, the number of sense operations increases by 2^{N-1} (here the '-1' term corresponds to the additional bit that can be encoded using phase) and hence this term is proportional to 2^{N-1}. The final component represents the supporting logic. The number of clock edges in the supporting logic to maintain a given data-rate will also increase by 2^{N-1} and hence this term is also proportional to 2^N. Additionally, the number of gates depends on N and so I_{SL} is also a function of N (see below). These three elements (I_{Tx}, I_{Rx}, and I_{SL}) can be approximated as follows. The transmit pulse current (I_{Tx}) can be modelled mathematically by a gaussian pulse [13]:

$$I_{\text{Tx}}(t) = I_p \cdot \exp\left[-\left(\frac{t\pi}{\delta}\right)^2\right] \tag{2}$$

where I_p is the peak amplitude of the current pulse required to ensure error-free pulse detection in the receiver and δ is the minimum Tx pulse width, a technology dependent parameter. Given a wireless channel, with coupling coefficient k, using

inductors with inductance L_{Tx} and L_{Rx}, the voltage pulse amplitude induced in the Rx coil is given by:

$$V_{\text{Rx}} = k\sqrt{L_{\text{Tx}}L_{\text{Rx}}} \cdot \frac{dI_{\text{Tx}}}{dt} \tag{3}$$

For transmission to be robust, V_{Rx} must be greater than the minimum receiver sensitivity threshold V_{St}, a technology-dependent parameter indicating the minimum Rx voltage fluctuation that can be accurately distinguished by the SA. I_p can therefore be obtained using Eqn. 4 below:

$$V_{\text{St}} + V_{\text{noise}} < \max\left\{\frac{2\pi^2 I_p t}{\delta^2} \cdot \exp\left[-\left(\frac{t\pi}{\delta}\right)^2\right]\right\}_0^t \tag{4}$$

where $0 > t > 1/f_C$, $t \in \mathbb{R}+$, and V_{noise} is the maximum amplitude of transient noise in the SA supply (*e.g.* any substrate noise, supply droop etc.[4]). Once I_p has been obtained, Eqn. 2 can be used to find I_{Tx}.

The receiver current (I_{Rx}) consumed in the sense amplifier can be modelled statically, because the average current required for a single sense operation will remain constant. However, the amount of supporting digital logic (I_{SL}) in the data encoder/decoder depends on N. Approximately, $I_{\text{SL}}(N)$ can be modelled by:

$$I_{\text{SL}}(N) \approx 2NI_{\text{DFF}} + NI_{\text{XOR}} + (N+2)I_{\text{AND}} \tag{5}$$

where I_{DFF}, I_{XOR}, and I_{AND} represent the dynamic current consumption of a flip-flop, XOR and AND gate respectively [5](justification for this is provided later, in Section IV-A).

Analysing the equations presented above, the advantages of the proposed Spike-Latency scheme (in terms of reducing the I_{Tx} current) can be observed. To transmit i bits using BPM requires i, pulses. To transmit i bits using SPM or NRZ requires, on average, $i/2$ pulses, but to transmit i bits using the proposed SET scheme requires only i/N pulses. Increasing N, however comes at the cost of increasing I_{Rx} and I_{SL} and so N must be carefully selected. Section V-B evaluates this trade-off mathematically using databook logic gates parameters for

---

[4]In depth modelling of this transient noise is beyond the scope of this paper, but discussed in detail in [26].

[5]For this basic mathematical modelling, static power consumption is considered negligible and hence ignored.

TABLE II
EXAMPLE CODE-BOOK USING SET WITH PARAMETER N=3. INCORRECT
PHASE/POSITION DECISIONS RESULT IN ONLY ONE BIT ERROR.

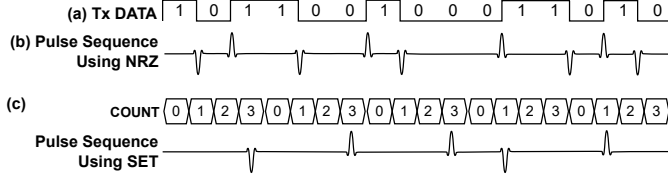| Binary | Decimal | Pulse Code | | | | Binary | Decimal | Pulse Code | | | |
		0	1	2	3			0	1	2	3
000	0					110	6				
001	1					111	7				
010	2					100	4				
011	3					101	5				

Fig. 4. Illustration of the bit-stream to I_{Tx} pulse mapping for the Tx Data shown in (a) when using (b) the existing NRZ encoding benchmark approach [10]–[13], and (c) the proposed SET scheme with the codebook shown in Table II (N=3).

0.35 µm, 65nm, and 28nm technologies, in conjunction with the equations above, to find the optimal value of N for a given channel quality.

IV. ARCHITECTURE DESIGN AND HARDWARE IMPLEMENTATION

Following the theoretical modelling of the proposed spike-latency encoding scheme, this section explores how it can be implemented in hardware. Fig. 3 shows the architecture of the low-energy inter-tier link proposed in this paper, consisting of six key components: (a) the spike-latency encoding logic, to implement the modulation scheme discussed in the previous section, (b) a tuneable current driver, to adaptively control the transmit current such that it is absolutely minimised depending on the integration scenario, (c) the inductive channel itself, consisting of two coupled planar inductors, (d) a sense amplifier, to amplify the received voltage signal, (e) the Tx/Rx clock synchronisation infrastructure, and (f) the demodulation logic to recover the transmitted data stream. The following sub-sections outline the design of each of these six components in more detail.

A. Encoding/Decoding Logic

The most important element of the proposed transceiver design is the encoding/decoding logic. Fig. 3(a) illustrates a practical implementation of the en/decoding logic consisting of an $N - 1$ bit counter (that generates the COUNT signal) and XOR-based match logic which compares the parallel Tx data bits with the incrementing COUNT signal. This generated signal is then fed through a final multiplexer stage, controlled by the MSB of the data which selects the *phase*. Here, the impact of increasing N on the logic size can be observed. Not only will a higher N result in a faster clock frequency, as N increases, one additional flip-flop will be required in the

counter (in addition to extra match logic). To minimise the power consumption of the system, the width of the I_{Tx} pulse is limited by a delay element with length δ, as shown on Fig. 3. This is analogous to the δ delay used for modulation in the benchmark NRZ scheme.

To improve the BER of the system, the COUNT signal is implemented using a *Gray-coded* counter, as shown. The use of the Gray-coded counter means that if a pulse is detected in the wrong sub-window, the effect of the incorrect detection on the data frame is minimised (*e.g.* incorrect detection of the Rx pulse at the $N \pm 1^{th}$ COUNT value only results in 1 bit of error in the whole frame). Additionally, the multiplexer stage means that an incorrect detection of *phase* results in only a single bit error in the output. An example code-book for these bit-to-code mappings for $N = 3$ is shown in Table II. Here the first two binary bits are the Gray-coded counter value, and the final bit is the phase-based decision bit. Fig. 4 illustrates how this works in practice when compared with the benchmark NRZ scheme. Here, using the benchmark NRZ scheme to transmit the data stream 0x591A results in 9 I_{Tx} pulses (Fig. 4(b)), whilst using the proposed SET scheme, with the bit-to-code mappings from Table II, requires only 5 I_{Tx} pulses (Fig. 4(c)). To minimise the power consumption of the en-decoding logic, each of the functional blocks (counter, match logic etc.) are implemented in separate supply domains where near-threshold voltage scaling is applied.

B. Tuneable Current Driver

The second element of the proposed low-energy transceiver is the tuneable current driver circuit. One of the benefits of using *wireless* 3D integration as opposed to traditional approaches, such as TSVs, is the relaxed assembly requirements when stacking each of the individual dies. As such, ICL-based 3D integration is ideally suited to low-cost IoT devices. However, low-cost assembly means that variation from chip-to-chip is typically significant. Fig. 5 illustrates different variation mechanisms, introduced at assembly time, that will affect the channel coupling quality: Fig. 5(a) shows variation in quality between channels ① and ② due to adhesive thickness, Fig. 5(b) shows variation in quality due to lateral die-to-die stacking misalignment, Fig. 5(c) shows variation in quality between channels ① and ② due to substrate thickness, and
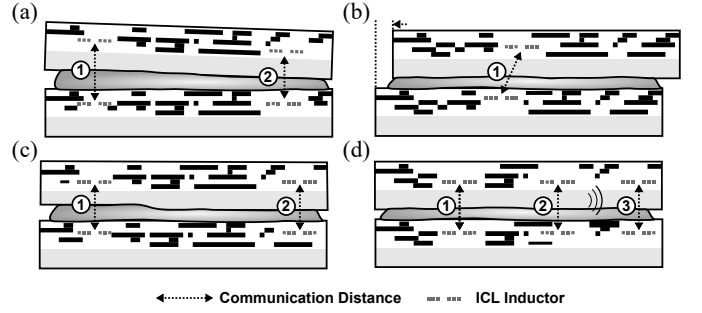

Fig. 5. Illustration of channel quality variation (across ICL channels ①, ②, and ③) in end devices due to (a) uneven adhesive thickness, (b) laterally misaligned die attach, (c) uneven substrate thinning, and (d) communication over different numbers of dies.

(a) Tx DATA 1 0 1 1 0 0 1 0 0 0 1 1 0 1 0
(b) Pulse Sequence Using NRZ
(c) COUNT 0 1 2 3 0 1 2 3 0 1 2 3 0 1 2 3 0 1 2 3
Pulse Sequence Using SET

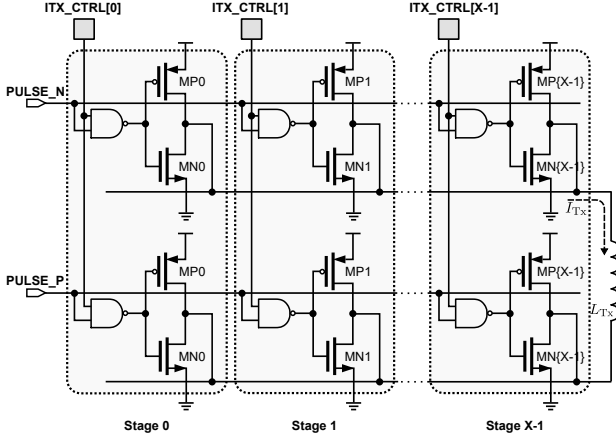Communication Distance — ICL Inductor

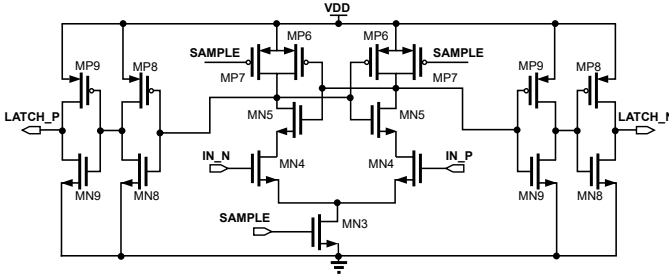Fig. 6. Schematic diagram showing structure of the proposed differential adaptive current driver circuit.

Fig. 7. Buffered sense amplifier used in the proposed transceiver. The SAMPLE signal is a short pulsed signal (with duration δ, generated at each rising edge of the COUNT signal. The buffered outputs LATCH_N and LATCH_P are passed to an SR latch (as shown in Fig. 3) and then used as inputs to the SET decoder.

Fig. 8. Illustration of the clock generation and synchronisation infrastructure in the presented test-chip. The data clock (with frequency f_{DAT}) is generated in the lower (Tx) die and delivered to the upper (Rx) die through a wire-bonded link. In each die, a multiplying delay locked loop (MDLL) is then used to generate the higher frequency f_{COUNT} clock.

C. Inductive Channel

One other important element of the ICL is the inductive channel itself, consisting of two coupled planar metal inductors. To maximise the performance of the system, it is desirable to maximise the EM coupling, k (*c.f.* Fig. 3) between the Tx and Rx inductors, as discussed previously, such that the minimum I_{Tx} pulse has maximum effect, as observed by the receiver. The coupling coefficient depends on a range of factors, however most notably the physical layout parameters of the inductor [28]. These are the inductor diameter (D), track width (w), track spacing (s), and number of turns (n). In order to determine best-performing parameters for these physical values and map them to an electrical link model, the optimisation flow outlined in [29] was used. The results from these simulations are presented in Section V-A.

D. Sense Amplifier

Fig. 7 shows the sense-amplifier adopted in the proposed transceiver. The design is similar to that used to implement the NRZ scheme and operates on the basis that, whilst SAMPLE is high, the Rx signal is amplified by the NMOS pair MN4. This causes a negative pulse based on the differential potential which is amplified in buffers MP8 and MP9, latched to avoid glitching, and then used to copy the Rx N-bit counter to the output, as shown in Fig. 3. In this implementation, the SAMPLE signal is generated by a synthesised programmable pulse generator block (incorporated in the SET logic[6]), as shown in Fig. 3.

E. Clock Synchronisation

Although external to the transceiver circuits themselves, one other important consideration is the Tx/Rx clock synchronisation infrastructure. The majority of ICLs published in existing works use *coherent* transceivers, which assume the presence of a synchronous Tx/Rx clock. To provide this Tx/Rx clock

Fig. 5(d) shows variation in quality between channels ① and ③ due to interference from other neighbouring links (②). Transient noise (*e.g.* from on-chip radios, or other devices through substrate coupling) may also cause variations in the inductive channel quality, affecting the coupling coefficient, k [27].

Because of this, ICLs must typically be designed to meet the *worst-case* assembly specification ($\mathrm{Min}(k)$) meaning that often, the Tx pulse current I_{Tx} is much larger than needed for robust operation. To address the need for this over-provisioning, in this work, we propose the tuneable current driver architecture, shown in Fig. 6. The proposed design uses a multi-stage, differential driver shown in the figure. Each stage in the driver circuit (0 to $X-1$) can be individually en/disabled according to the appropriate bit of the register ITX_CTRL. Each stage (0 to $X-1$) is also comprised of inverters with differing transistor widths $w_{\mathrm{MN0}} > w_{\mathrm{MN1}} > w_{\mathrm{MN2}}...$ etc. to allow precise control of the Tx current I_{Tx} using the control register. Using this approach, the dies can be stacked and then I_{Tx} can be tuned (post-stacking) to compensate for the assembly defects shown in Fig. 5, without using an unnecessarily large transmit current.
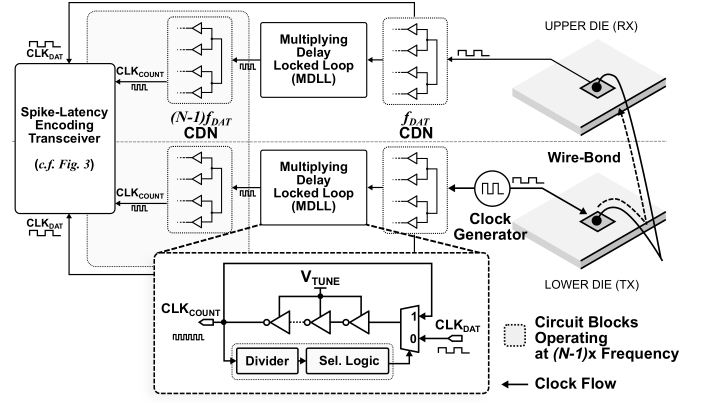
---

[6]For this reason its energy contribution is accounted for within the SET logic for the results presented in Sections V and VI.
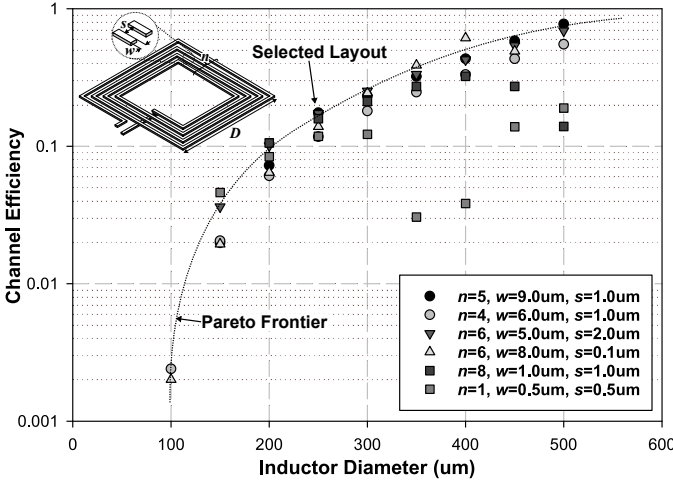
Fig. 9. Scatter plot showing the simulated efficiency vs. area trade-off, including pareto-optimal frontier (only a small sample of trialled layouts are presented for clarity). The $250\,\mu\text{m} \times 250\,\mu\text{m}$ square geometry used for silicon measurement results in this section is highlighted.

Fig. 10. Mathematical modelling results showing how the transceiver energy (and optimal N value) varies as a function of N across three different technology nodes (28nm, 65nm and $0.35\,\mu\text{m}$ and three different channel coupling strengths (k=0.05, k=0.10, and k=0.15).

synchronisation in this work, we use the clock architecture shown in Fig. 8. Here, the data clock (with frequency f_{DAT}) is generated in the lower (Tx) die and delivered through a wire-bonded link to the upper (Rx) die. To minimise jitter, this low-frequency (f_{DAT}) clock is then passed to a Multiplying Delay Locked Loop (MDLL) in each die which also generates the higher frequency COUNT clock ($f_{\text{COUNT}} = (N-1)f_{\text{DAT}}$). Compared with the existing NRZ benchmark scheme, the areas that operate at higher frequency when using the SET scheme (and hence incur additional energy overheads) are: the pulse generator, the high frequency CDN ($(N-1)f_{\text{DAT}}$), and the MDLL control logic. These elements are highlighted in grey on Fig. 8, and their energy overheads are evaluated in Section V-C5.

In general, however, it is often more convenient to transmit the clock wirelessly using a separate ICL channel (such as used in [22] and [10]). The SET approach proposed in this paper could be combined with such a scheme (which would result in a different set of energy trade-offs), however wireless clock synchronization is beyond the scope of this paper (which is focussed on the *data* transceiver design).

V. EXPERIMENTAL VALIDATION AND RESULTS

This section presents experimental validation of the proposed low-energy inductive transceiver outlined previously. Initially, Section V-A evaluates the geometric layout parameters of the Tx and Rx inductors for forming the inductive channel using the COIL-3D software tool [29]. Following this, Section V-B evaluates the spike-latency encoding concept using the mathematical modelling presented in Section III-A, and Section V-C performs post-layout simulation of the transceiver using SPICE.

A. ICL Layout Parameter Selection

As outlined in Section IV-C, the geometric parameters of the inductive channel (which largely determine the EM

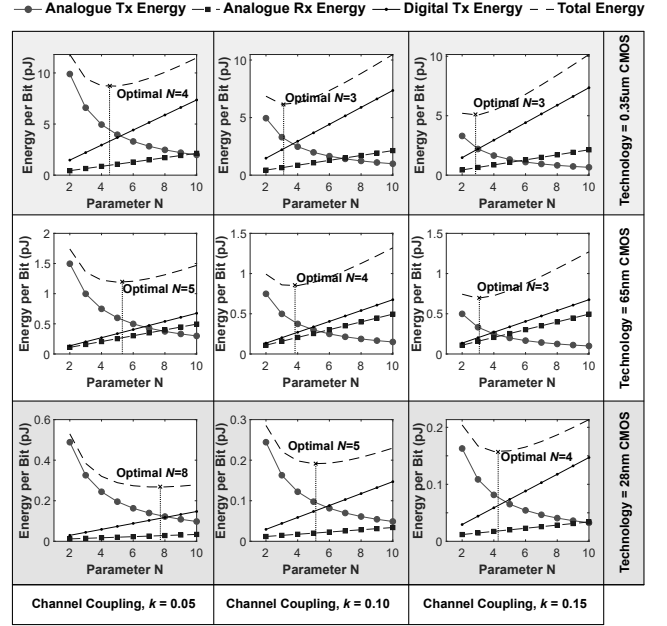coupling coefficient, k) were selected using the optimisation flow in [29]. Fig. 9 shows a scatter plot of channel efficiency ($V_{\text{Rx}}/V_{\text{Tx}}$) vs. diameter for a selection of optimal geometries. As can be observed from the figure, a strong trade-off exists, and therefore the $250\,\mu\text{m} \times 250\,\mu\text{m}$ layout on the 'knee' of the pareto curve was selected for use. The selected design has physical parameters $D = 250\,\mu\text{m}$, $w = 9\,\mu\text{m}$, $w = 1\,\mu\text{m}$ and $n = 5$, corresponding to a channel efficiency ≈ 0.13. For validation in this paper (through simulation in Section V, and silicon measurement in Section VI) it is assumed that no circuits are placed within the ICL channel. However, prior research by Niitsu *et al.* [30] has demonstrated that SRAM cells can be placed within the channel area without significant performance degradation, and that standard logic cells (automatic place and route) can be placed within the channel area with only a minimal performance impact (which can be overcome by increasing the Tx power by around 9%) [30]. This implies that for certain applications (digital logic/memory) the area overhead of the ICL inductors is limited to the coil tracks themselves (typically in a high metal layer), and the interposed silicon can still be utilized.

B. Validation using Mathematical Models

Having established the approximate coupling coefficient k that can be achieved within the $250\,\mu\text{m} \times 250\,\mu\text{m}$ area, this section evaluates the energy breakdown of the proposed scheme using the equations from Section III-A in conjunction with databook logic gates parameters for $0.35\,\mu\text{m}$, 65nm, and 28nm technologies across a range of values for parameter N. As predicted, a trade-off between E_{pb} and N can be observed. In each case, the energy of the transceiver is projected for every value of N between 2 and 10, and an optimal point

(typically around N=4) exists where a good balance between I_{Tx} and I_{SL} is established. At the less advanced process technology nodes ($0.35\,\mu m$) the predicted optimal N value is lower (between 3-4) because the digital logic is expensive in terms of energy. As the process technology scales down to 28nm, the digital logic energy decreases and hence the optimal N shifts to the right, increasing up-to a maximum of 8.

The results presented in Fig. 10 also illustrate how the optimum value of N varies as the coupling strength k between dies changes. As the EM coupling deteriorates, k reduces, the Tx current required for robust operation increases, and hence the best-performing value of N increases. For the inductor layout determined in the previous section, k is in the order of 0.13 and hence the modelling predicts that the optimal value of N will be around N=3-4, depending on the technology.

C. Validation using SPICE

Following the theoretical modelling of the proposed spike-latency encoding scheme, the presented transceiver was compared to the existing inductive NRZ design using commercial EM and circuit simulators in $0.35\,\mu m$, 65nm and 28nm CMOS technologies (to represent the full spectrum heterogeneity that would likely be found in IoT devices, the context of this work). For each chip, a total die thickness of $100\,\mu m$ was assumed (in line with presently available low-cost wafer lapping technologies) and an adhesive thickness of $10\,\mu m$ was assumed for die attach.

Ansys HFSS was used for EM modelling of the inductive coupling channel, using the EM simulation setups shown in Fig. 11. This figure shows the technology stackups for each process node (Fig. 11(b)-(d)), and the 3D view of the chip design assumed in simulation (Fig. 11(a)) (only measurements from the central channel (port S(0) → port S(1)) are used, with the neighbouring channels (N(0) and N(1)) simulating noise effects for BER analysis. The analogue circuit blocks (discussed above) were each sized for their respective technologies with the circuit architecture remaining the same between simulations. The only notable difference was that, in the 28nm node a level shifter was inserted between the encoding logic and the driver circuits, allowing the driver to be implemented using thick-oxide transistors to meet the $Min(I_{Tx})$ requirements. A number of different comparisons were performed and the results are documented in the following subsections.

1) Area Overhead: Fig. 12(a) shows the layout of the proposed low-energy transceiver in 65nm CMOS technology consisting of the Tx/Rx inductor ($250\,\mu m \times 250\,\mu m$), the sense amplifier ($15.4\,\mu m \times 43.7\,\mu m$), and the tuneable Tx driver circuit ($36.0\,\mu m \times 22.0\,\mu m$). When compared to the existing state-of-the art transceivers using BPM, SPM or NRZ-encoding, the only additional area overhead is derived from the supporting digital logic which is highlighted on the figure ($13.6\,\mu m \times 17.8\,\mu m$ at the pictured 65nm technology node). These area overheads are itemised in Table III for all three considered technology nodes compared to existing schemes. For the BPM/SPM/NRZ approaches, the *digital* area includes the SAMPLE pulse generation logic and the RX-side latch. For the proposed SET scheme, the digital logic area also includes
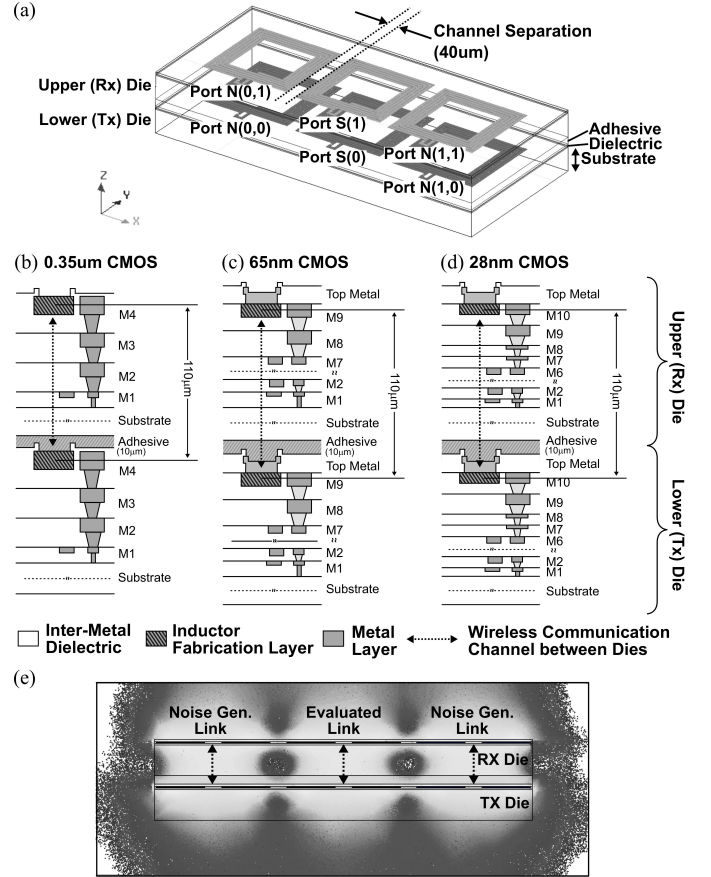

Fig. 11. (a) Illustration of EM simulation setup in Ansys HFSS including stacking cross-sections for (b) $0.35\,\mu m$, (c) 65nm and (d) 28nm CMOS technology. Channel S is used for analysis, with channels N(0) and N(1) transmitting random bit streams to simulate noise. (e) Simulated magnetic field strength within the die stack showing the directional intra-link near-field coupling, and inter-link fringe coupling for evaluating cross-talk/BER.
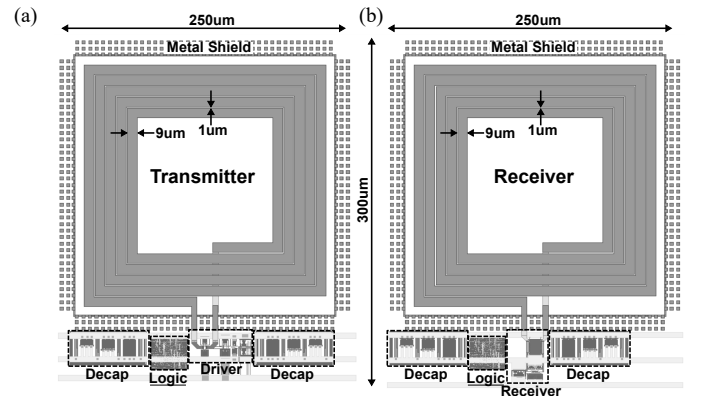

Fig. 12. Layout of proposed low-energy transceiver in 65nm CMOS technology including (a) transmitter, and (b) receiver highlighting the additional digital coding logic area [underlined] (the only additional silicon area overhead when compared with the state-of-the-art scheme).

the input/output registers, counters and match logic (shown in Fig. 3). As can be observed from the table, the additional SET control logic does not add significant overhead to the footprint of the transceiver, in fact only contributing between 0.1% (in the case of 28nm technology) and 14% (in the case of the $0.35\,\mu m$ technology).

TABLE III
SIMULATED PERFORMANCE OF THE PROPOSED LOW-ENERGY TRANSCEIVER (WITH OPTIMAL PARAMETER N), COMPARED TO BI-PHASE MODULATION
(BPM) [18], SINGLE PHASE MODULATION (SPM) [19], AND NON-RETURN TO ZERO (NRZ) [5], [10]–[13] ACROSS THREE TECHNOLOGY NODES.

	Performance Metric	Bi-Phase Modulation (BPM) [18]	Single Phase Modulation (SPM) [19]	Existing State-of-The-Art NRZ [10]–[13]	Proposed Approach
28nm	Total Footprint	0.064mm²	0.064mm²	0.064mm²	0.064mm²
	Inductor Area (Tx, Rx)	0.0625mm², 0.0625mm²	0.0625mm², 0.0625mm²	0.0625mm², 0.0625mm²	0.0625mm², 0.0625mm²
	Digital Logic Area (Tx, Rx)	33 μm², 33 μm²	33 μm², 33 μm²	33 μm², 33 μm²	72 μm², 72 μm²
	Analogue Circuits Area (Tx, Rx)	586 μm², 500 μm²	586 μm², 500 μm²	586 μm², 500 μm²	586 μm², 500 μm²
	Max. Bandwidth (N=5)	2.4Gbps	2.4Gbps	2.4Gbps	800Mbps
	BER	9.8E-7	9.1E-5	2.2E-6	2.8E-6
	Tx → Rx Transmission Latency	1 cycle	1 cycle	1 cycle	5 cycles
	Energy-per-bit	**0.70pJ**	**0.36pJ**	**0.36pJ**	**0.26pJ (28.1% Reduction)**
65nm	Total Footprint	0.066mm²	0.066mm²	0.066mm²	0.066mm²
	Inductor Area (Tx, Rx)	0.0625mm², 0.0625mm²	0.0625mm², 0.0625mm²	0.0625mm², 0.0625mm²	0.0625mm², 0.0625mm²
	Digital Logic Area (Tx, Rx)	110 μm², 110 μm²	110 μm², 110 μm²	110 μm², 110 μm²	242 μm², 242 μm²
	Analogue Circuits Area (Tx, Rx)	792 μm², 673 μm²	792 μm², 673 μm²	792 μm², 673 μm²	792 μm², 673 μm²
	Max. Bandwidth (N=4)	1.6Gbps	1.6Gbps	1.6Gbps	1.05Gbps
	BER	9.2E-7	8E-5	1.15E-6	2.0E-6
	Tx → Rx Transmission Latency	1 cycle	1 cycle	1 cycle	4 cycles
	Energy-per-bit	**1.60pJ**	**0.93pJ**	**0.84pJ**	**0.66pJ (21.4% Reduction)**
0.35um	Total Footprint	0.075mm²	0.075mm²	0.075mm²	0.0855mm²
	Inductor Area (Tx, Rx)	0.0625mm², 0.0625mm²	0.0625mm², 0.0625mm²	0.0625mm², 0.0625mm²	0.0625mm², 0.0625mm²
	Digital Logic Area (Tx, Rx)	1196 μm², 1196 μm²	1196 μm², 1196 μm²	1196 μm², 1196 μm²	4986 μm², 4986 μm²
	Analogue Circuits Area (Tx, Rx)	17 060 μm², 5465 μm²	17 060 μm², 5465 μm²	17 060 μm², 5465 μm²	17 060 μm², 5465 μm²
	Max. Bandwidth (N=3)	450Mbps	450Mbps	450Mbps	300Mbps
	BER	8.0E-7	6.3E-5	2.3E-6	1.2E-6
	Tx → Rx Transmission Latency	1 cycle	1 cycle	1 cycle	3 cycles
	Energy-per-bit	**14.96pJ**	**8.80pJ**	**8.50pJ**	**7.56pJ (11.1% Reduction)**

2) Bit Error Rate and Latency: The BER of the proposed scheme was then evaluated in each technology using the channel model generated by Ansys HFSS. As shown in Fig. 11, the EM setup includes 3 channels, each of which transmits an equiprobable random bit stream. This generates noise in the channel of interest, facilitating estimation of BER through simulation as shown in Fig. 11(c). The results are presented in Table III with comparisons to simulations implementing the BPM, SPM and NRZ benchmark schemes. Across these simulations, the measured BER when using the proposed transceiver is approximately equal to the BER achieved when using the BPM or NRZ approaches (and better than that achieved using SPM). This is due to the combination of using Gray-coded pulse mappings and phase-coding (in which 180 degrees of phase shift exist between MSB '1' and '0' values). The latency when using the SET approach is, however, greater than that when using the existing NRZ approach as the full data frame must be present before transmission. When using SET, the latency is N clock cycles, rather just a single cycle.

3) Sample Timing Margin Sensitivity: The sensitivity of the proposed approach to Tx/Rx clock jitter was also evaluated. Fig. 13 shows the results of these simulations across three technology nodes, (a) 0.35 μm, (b) 65nm and (c) 28nm CMOS. For each node, the timing sensitivity was evaluated by inducing jitter in the Rx clock signal and simulating the BER. The grey bathtub curves show the timing sensitivity of the proposed SET scheme and the black bathtub curves show the timing sensitivity of the benchmark NRZ scheme for the same f_{DAT} frequency. As can be observed, the proposed scheme is more sensitive to Rx clock jitter (by between 3.6× and 8.3×, depending on the technology node), due to the increased

SA sample frequency. Whilst this does not limit the *BER performance* for the presented data rates (as the timing margin is greater than the maximum expected COUNT clock jitter, shown by the shaded area), it does have the effect of limiting the maximum transceiver *bandwidth*, as shown in Table III. This bandwidth reduction represents the most significant trade-off for the additional energy gains achievable using SET.

4) Energy-per-Bit Evaluation: The effectiveness of the proposed transceiver in reducing energy consumption (the primary motivation for this study) was then evaluated. The energy-per-bit of the proposed approach was measured for a range of N values and compared with the BPM, SPM and NRZ transceivers. Fig. 14 shows the energy required to transmit a single bit for the case of the benchmark transceivers, and the proposed SET design for varying values of N across each of the three technology nodes. As can be observed from the figure, the proposed transceiver is successful in reducing the energy consumption by up to 62.7% when compared with previously published BPM transceivers, and 28.1% compared to the existing state-of-the-art in low-energy modulation, NRZ encoding. Fig. 14 also validates the mathematical modelling in Section III-A, demonstrating that N=3-5 performs optimally across the range of technologies considered.

5) Additional Clock Requirements: Although the additional dynamic energy associated with the $(N-1)\times$ faster clock and SAMPLE pulse generation is accounted for in the simulation of the SET transceiver block, consideration should also be given to the additional energy overheads associated with *implementation* of a faster clock (as discussed in Section IV-E). These additional energy overheads are derived from two main sources: (1) The additional energy consumed by

(a) 0.35um CMOS Technology

$f_{DAT} = 300\text{Mbps}$ $f_{COUNT} = 400\text{Mbps}$

(b) 65nm CMOS Technology

$f_{DAT} = 800\text{Mbps}$ $f_{COUNT} = 1.6\text{Gbps}$

(c) 28nm CMOS Technology

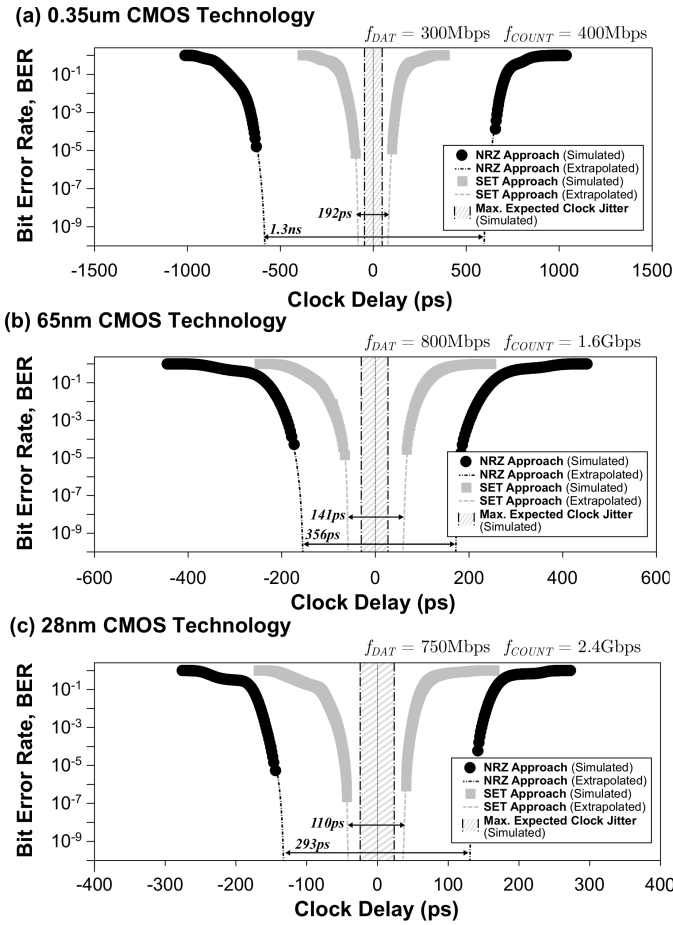$f_{DAT} = 750\text{Mbps}$ $f_{COUNT} = 2.4\text{Gbps}$

Fig. 13. Bathtub curves showing the SAMPLE signal timing margin when using the proposed SET approach when compared with the inductive NRZ benchmark approach across 3 technology nodes: (a) 0.35 μm CMOS, (b) 65nm CMOS, and (c) 28nm CMOS. Silicon measurement of this timing margin in the 0.35 μm technology is presented in Section VI-B.

TABLE IV
TABLE SHOWING THE ENERGY OVERHEAD ASSOCIATED WITH THE HIGHER FREQUENCY CLOCK INFRASTRUCTURE, ΔE_{CI}.

Source	Additional Energy Contribution (per Bit)		
	0.35 μm	65nm	28nm
(1) ΔE_{CDN} (per bit)	0.11pJ	0.019pJ	0.007pJ
(2) ΔE_{DLL} (per bit)	0.36pJ	0.034pJ	0.032pJ
Total Clock Infrastructure Overhead, $\Delta \mathbf{E_{\text{CI}}}$ (per bit)	**0.47pJ**	**0.053pJ**	**0.039pJ**

the Clock Distribution Network (CDN) from the output of the MDLL to the sink node in the SET modulator, and (2) the additional energy consumed by the MDLL control logic to maintain a higher output frequency.

To evaluate the energy overhead of (1) and (2) in our design, the relevant modules of the clock distribution architecture were simulated in each of the three technology nodes. Simulations were performed at the data frequency (f_{DAT}) which is representative of the benchmark NRZ/SPM approaches, and at $(N - 1) \times f_{\text{DAT}}$, representative of the SET approach. Table IV shows the energy difference (per bit) $\Delta_{E_{pb}}$ resulting from using the SET scheme for the CDN (ΔE_{CDN}) and the MDLL (ΔE_{DLL}). As can be observed, these energy overheads are

small in comparison to the overall energy per bit (between 6-15% depending on the technology node), but still important to consider when designing such a system. It should also be noted, however, that these additional energy contributions are implementation dependent (*e.g.* would change if a wireless clock link was used) and can often be amortised when multiple parallel data links are present on the same chip.

The summary table (Table VI) on page 14 calculates the energy benefits of the proposed scheme (compared with the NRZ benchmark approach) taking into account this additional penalty, assuming 1 clock link (wire-bonded) per data link. Even considering these additional energy overheads, the proposed SET link still offers competitive energy reductions between 7.4% and 16.9% depending on the technology.

6) Tolerance to misalignment: Finally, the tolerance of the proposed transceiver to lateral die-to-die stacking misalignment was also explored. As discussed in Section IV-B, one of the benefits of using wireless 3D integration is that it avoids the need for precise (and hence expensive) pick-and-place accuracy when performing the die stacking. To evaluate the tolerance of the channel to lateral placement misalignment, the channel coupling coefficient, k, was evaluated for various levels of offset. Fig. 15 presents simulation results illustrating the effect of alignment accuracy on k. As shown, the channel will tolerate up-to 40 μm of die-to-die misalignment in both x and y directions (a total diagonal offset of 56 μm) whilst maintaining performance within 10% of the optimum (representative of that which can be tolerated by tuning the ITX_CTRL register). When compared to 3D assembly using TSVs, which typically mandates sub-micron placement accuracy [31], this represents an approximately 100× improvement.

VI. CASE STUDY: TEST-CHIP DEMONSTRATION

Following the success of the proposed low-energy transceiver in SPICE modelling, the design was implemented on a 2-tier 3D stacked silicon test-chip in 0.35 μm CMOS technology for silicon performance evaluation. Fig. 16(a) shows a photograph of the assembled 2-tier test-chip with the upper (Rx) and lower (Tx) dies highlighted. Before stacking, each die was thinned to a height of 100 μm and attached using epoxy adhesive with 10 μm thickness as shown in Fig. 16(b). The dies were stacked in a face-to-back (F2B) arrangement resulting in a total communication distance of 110 μm through the silicon substrate, BEOL, and adhesive layers.

A. Tuneable Current Driver Evaluation

Initially, the transmit pulse amplitude I_{Tx} was selected using the tuneable current driver circuit. To find the optimal value of the ITX_CRL register (and hence I_{Tx} amplitude), the BER of the link (missed pulses vs. total pulses, without the spike-latency modulation scheme) was measured whilst gradually sweeping the ITX_CRL register from 0 to 32. Fig. 17 shows the results of this sweep for two separate test-chips: Chip **A** (which is assembled with perfect alignment in the inductive channel), and Chip **B** (which is assembled with an offset of 20 μm in the inductive channel, to demonstrate the effects of stacking misalignment during assembly). At the smallest
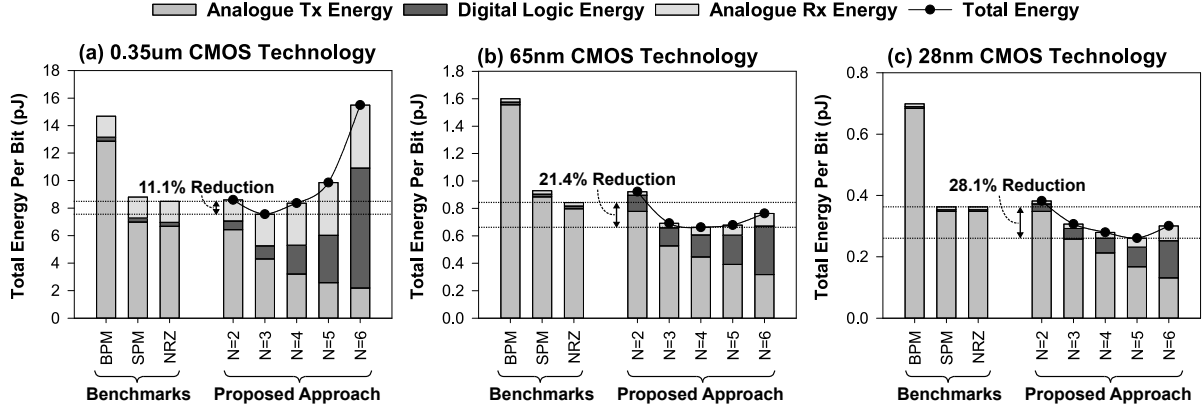
Fig. 14. Simulated performance of the proposed transceiver (for various values of N) when compared to the compared to Bi-Phase Modulation (BPM) [18], Single Phase Modulation (SPM) [19], and Non-Return to Zero (NRZ) modulation [10]–[13] benchmark designs at three different technology nodes. Results show improvements between 11.1% and 28.1% using the proposed scheme.
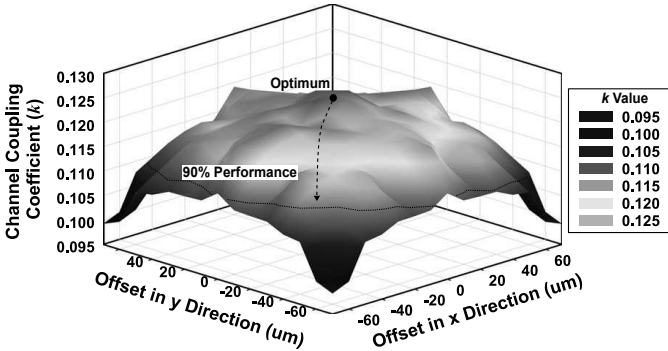

Fig. 15. Simulated channel performance with respect to x and y die-to-die stacking misalignment (in terms of coupling coefficient, k).
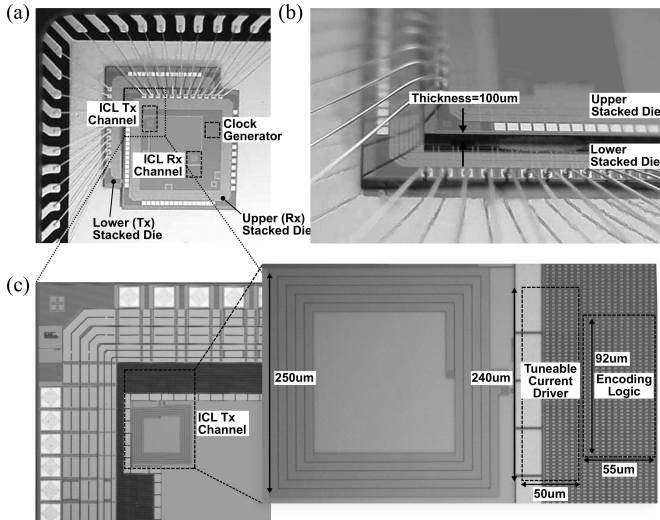
Fig. 16. Micrograph of (a) the 2-tier stacked IC with wire-bonded power, reset and debug pins. (b) Side elevation showing vertical die stacking arrangement and communication distance. (c) A single die layout, showing the dimensions of the proposed transceiver and the 250 μm square channel used for evaluation.

Fig. 17. Measured link BER and energy-per-pulse as I_{Tx} control register (ITX_CTRL) varies for two test chips: Chip **A** assembled with perfect die-to-die stacking alignment, and Chip **B** assembled with a significant 20 μm stacking offset (equating to almost 10% of the channel size) to explore the effects of die-to-die misalignment in the stacking process.

values (ITX_CTRL = 16 in Chip **A**, and ITX_CTRL = 26 in Chip **B**, due to the assembly offset). This demonstrates that the proposed tuneable driver circuit can be used to overcome significant packaging variations whilst maintaining performance within the specification. At its tuned value, Chip **A** achieves a BER in the order of 10^{-5} with a pulse energy of 12.6pJ.

B. Bias Tolerance and Timing Margin Evaluation

Following this, the transceiver's tolerance to variations in Tx/Rx clock delay (evaluated through simulation in Section V-C3) was measured. Fig. 18 revisits the bathtub curves presented in Section V-C3, this time comparing the *simulated* bathtub timing curve with the *measured* curve (varied by adjusting V_{TUNE} (*c.f.* Fig. 8)). As shown on the figure, the

settings (1,2,3) the Tx current is low, and hence the pulses are not detected. As the ITX_CTRL register is incremented further, the link begins to operate. Eventually, both chips reach the target threshold BER (1E-5) at different tuning register
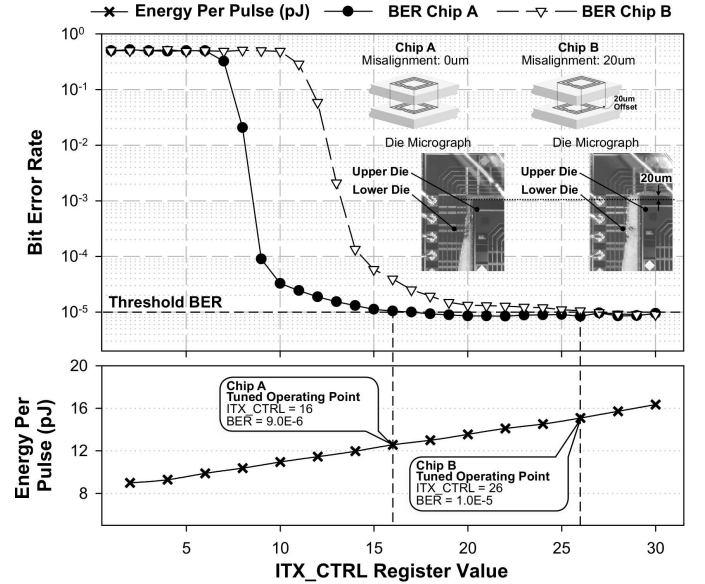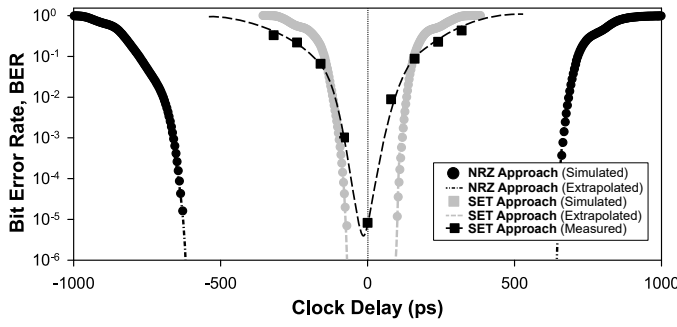
Fig. 18. Bathtub curves showing the measured `SAMPLE` signal timing margin when compared to the simulated margin for the proposed SET approach and NRZ benchmark approach in $0.35\,\mu$m CMOS. Silicon measurement results are taken from Chip **A** with f_{DAT}=300MHz and f_{COUNT}=400MHz.
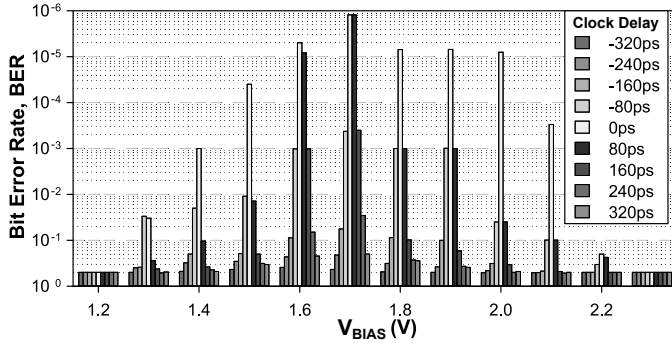

Fig. 19. Chart showing measured tolerance of the presented transceiver to variations in bias voltage (V_{BIAS}). The x-axis represents the bias voltage and the y-axis shows the Bit Error Rate (BER) at this bias voltage, for a range of clock delays (-320ps, -240ps, -160ps, -80ps, 0ps, 80ps, 160ps, 240ps, 320ps).

measured timing margin is very close to the margin predicted by SPICE with the small variation likely attributed to on-chip noise in the V_{TUNE} supply. These silicon measurements also show that, whilst the sample margin is reduced when using the proposed scheme compared with the benchmark NRZ scheme (by approximately 90%), the transceiver can still operate within this margin, providing a low BER $< 10^{-5}$.

The tolerance of the proposed transceiver to variations in bias voltage was also evaluated in the test-chip. For this, the BER of the transceiver was measured for a pseudo-random binary bit stream whilst varying V_{BIAS} between 1.2V and 2.3V. For each value of V_{BIAS}, the Rx clock delay was also swept (from -320ps to +320ps). The results of these measurements are presented in Fig. 19. Here it can be observed that the proposed design is can tolerate up to 400mV of bias variation, without significant performance degradation (whilst maintaining the BER below the 10^{-5} target). This agrees well with SPICE results which predict an acceptable V_{BIAS} range of exactly 1.8V $\pm$ 200mV.

C. Energy-per-Bit Evaluation

The energy of the proposed transceiver (the primary motivation for this work) was then evaluated for a range of values of parameter N between 2 and 6 at the tuned `ITX_CTRL` register value. Energy was measured using knowledge of the transmit frequency combined with power measurements, taken with a
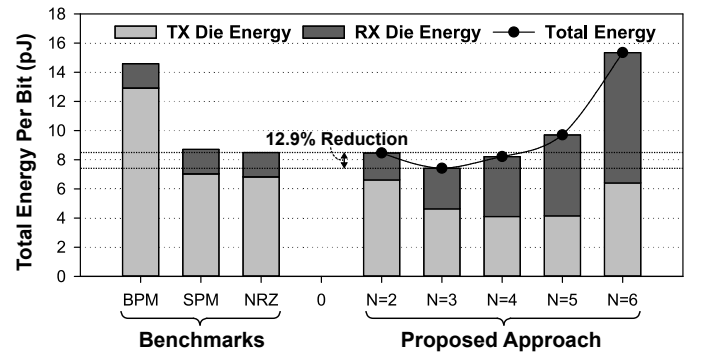

Fig. 20. ICL energy consumption in Tx and Rx dies, as N varies, compared with Bi-Phase Modulation (BPM) [18], Single Phase Modulation (SPM) [19], and Non-Return to Zero (NRZ) modulation [10]–[13] benchmarks [measured silicon results in $0.35\,\mu$m technology].

Keysight B2900A source meter unit. Fig. 20 shows the results of these experiments highlighting the energy split between the Tx and Rx dies[7] when compared to the benchmark approaches. Here it can be observed that the optimal parameter of N=3 yields a 13% energy reduction compared to the state-of-the-art NRZ encoding benchmark *with* I_{Tx} tuning, representing a significant overall energy reduction when using the SET scheme. It can also be observed that the results closely match the simulation predictions with the measured energy-per-bit being 7.4pJ, and the simulated energy-per-bit being 7.6pJ, indicating high confidence in the SPICE-based *energy* results presented in Section V-C.

Fig. 21 shows an eye diagram of the least-significant-bit (LSB) of the Rx data output when using the proposed transceiver with parameters `ITX_CTRL`=16 and N=5 at the maximum operation frequency, f_{DAT}=47.6MHz (for N=5). Although the eye opening in `RX_DATA` signal at this frequency is still wide, in order to meet the data-rate target of f_{DAT}=47.6MHz with parameter N=5 requires a `COUNT` frequency, f_{COUNT}=0.762GHz which represents the upper-bound when considering the sense-amplifier timing margin (discussed in Section V-C3). This has the effect of limiting the maximum frequency of the transceiver. For the algorithmic parameter N=3 (corresponding to the optimal *energy* efficiency), the maximum data rate was measured to be 266Mbps. Although this is a reduction when compared to the NRZ scheme, the 266Mbps data-rate is ample for most IoT applications (which form the motivation for this paper). Table V summarises the measured performance of the transceiver from the test-chip presented in this section.

To demonstrate the benefits achieved by combining this approach with the tuneable pulse driver circuit, Fig. 22 compares this proposed design with leading published research. Works [32] and [33] implement near-field *capacitive* communication, and [11], [13], [34], [35] use *inductive* communication (as adopted in this paper). Fig. 22 plots the *energy-per-bit* against *communication distance* for each approach. When compared to prior-art, results indicate a 7.7× reduction in energy consumption for wireless 3D communication across the

---

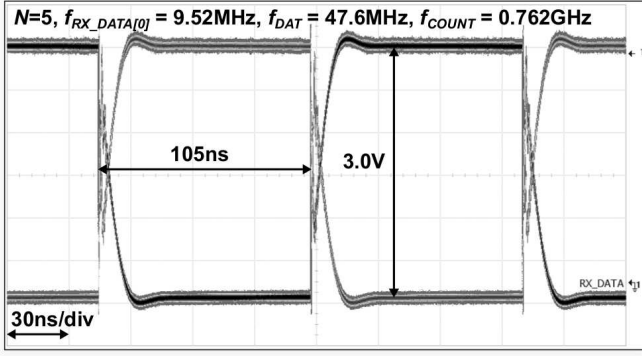[7]As the digital Tx logic and drivers are implemented using a shared supply rail.

Fig. 21. Measured eye diagram showing `RX_DATA[0]` (the LSB of the data output) from the proposed transceiver implemented on the $0.35\,\mu m$ 2-tier test-chip. $f_{COUNT} = 0.76$GHz, $N = 5$, $f_{DAT} = 47.6$MHz, $f_{RX\_DATA[0]} = 9.52$MHz.

TABLE V
MEASURED PERFORMANCE OF THE PROPOSED INDUCTIVE TRANSCEIVER
(COMPARED TO SIMULATED RESULTS FROM SECTION V-C).

| Evaluation Metric	Simulated Performance	Measured Performance
Technology	2-tier stacked $0.35\,\mu m$ CMOS	
Communication Distance	$110\,\mu m$ ($100\,\mu m$ chip + $10\,\mu m$ adhesive)	
Average Energy Per Bit	7.6pJ/bit	7.4pJ/bit
Average Bit Error Rate	1.2E-6	9.0E-6
Channel Area	$250\,\mu m \times 250\,\mu m$ (0.063mm^2)	
Transceiver Circuits Area	Tx:0.0225mm^2, Rx:0.0264mm^2	
Maximum Data Rate	300Mbps/channel	266Mbps/channel
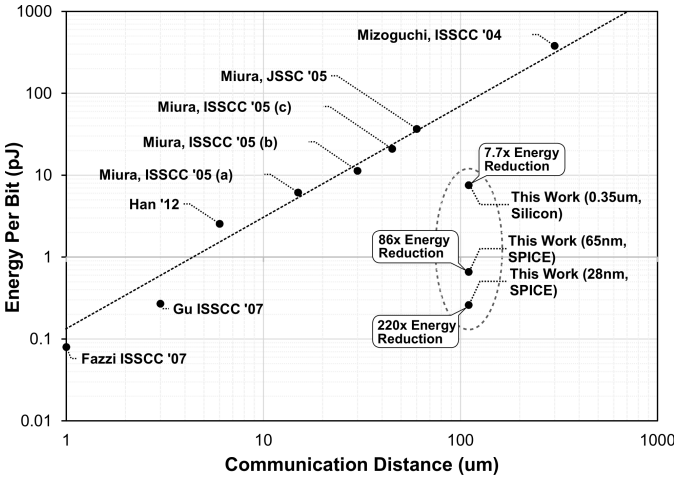
Fig. 22. Comparison of proposed transceiver design with other state-of-the art published works (*Han '12* [34], *Gu '07* [33], *Fazzi '07* [32], *Miura* [13], [35], and *Mizoguchi '04* [11]).

$110\,\mu m$ channel, based on silicon measurements in $0.35\,\mu m$ technology. This improvement is even more significant when considering the simulated performance results in 65nm and 28nm technologies which are representative of improvements of $86\times$ and $220\times$ respectively.

VII. DISCUSSION

Having validated the proposed transceiver through simulation and physical test-chip measurements, this paper has

demonstrated that significant energy savings ($>28\%$) can be achieved through using the proposed Spike-latency Encoding Transceiver (SET). Table VI shows an overall comparison of SET, and the existing state-of-the-art in terms of *energy* efficiency, NRZ encoding, combining physical test-chip results from Section VI and SPICE results from Section V-C. As can be observed from the table, the proposed approach outperforms prior-art across all test-cases (in terms of energy) by between 11% and 28%, depending on the technology node.

Whilst the proposed approach minimises *energy* (which was the goal of this work, motivated by the requirements of IoT devices), this paper also highlights the importance of *tailoring* the modulation approach to suit the target application/integration scenario. Applications requiring high-bandwidths with low error-rates may favour Bi-Phase Modulation, (BPM), however this is energy-expensive as one Tx pulse is required per transmitted bit. Conversely, the proposed SET scheme is ideally suited for low-energy applications where latency and bandwidth are less important.
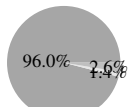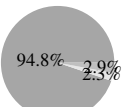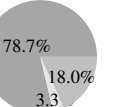
The modelling presented in this paper also shows that even more pronounced energy savings can be achieved using the proposed SET approach (compared with the NRZ/SPM benchmarks) when the channel coupling is weaker. This may be, for example, in systems that communicate across greater distances, or where smaller Tx and Rx inductors are used. This will result in worse coupling, and hence require a higher transmit energy per pulse.

By the same reasoning, in systems where the communication distance is reduced (for example if *face-to-face* die stacking is performed) the NRZ benchmark approach may provide superior energy efficiency. Transient noise will also influence this trade-off. One advantage of using the proposed scheme in favour of existing approaches is that the algorithmic parameter N (and the tuneable current driver strength) can be dynamically tuned at runtime to compensate for channel noise. For example, dynamically increasing the drive current to counteract noise caused by an on-chip radio, and simultaneously increasing N to compensate and maintain a constant energy consumption. This runtime adaptation with respect to on-chip noise is an ongoing area of our research.

Finally, as IoT devices are becoming increasingly heterogeneous, another important factor is evaluating how the proposed approach will perform at more advanced process nodes. As SET trades-off expensive analogue transmit pulses (which map to the magnetic field strength, and hence will not scale with process technology) in favour of additional digital processing (which will reduce in energy as process technology scales), the results presented in Section V-C indicate that the energy of the proposed approach will scale at a faster rate than existing schemes with process technology size. To illustrate this, Fig. 23 shows a plot of technology node vs. projected transceiver energy consumption on a logarithmic scale. The marked points show the three technology nodes explored in this paper (28nm, 65nm and $0.35\,\mu m$) and the dashed-line illustrates the expected trend with technology scaling[8]. Whilst the maximum energy savings compared to the state-of-the-art demonstrated in this

---

[8]This trend is extrapolated based upon the results presented in this paper.

TABLE VI
OVERALL COMPARISON OF PROPOSED TRANSCEIVER WITH THE EXISTING STATE-OF-THE-ART (INDUCTIVE NRZ ENCODING [5], [10]–[13]).

| Evaluation Metric	Technology = 28nm CMOS		Technology = 65nm CMOS		Technology = 0.35um CMOS		
	State-of-the-Art (NRZ) [5], [10]–[13]	Proposed Approach	State-of-the-Art (NRZ) [5], [10]–[13]	Proposed Approach	State-of-the-Art (NRZ) [5], [10]–[13]	Proposed Approach	Proposed Approach
	(SPICE)		(SPICE)		(SPICE)		(Silicon)
Transceiver Circuits Area	1152um^2	1230um^2	1685um^2	1949um^2	24917um^2	32497um^2	
Total Area	0.064mm^2	0.064mm^2	0.066mm^2	0.066mm^2	0.075mm^2	0.0855mm^2	0.0855mm^2
Tx Die → Rx Die Transmission Latency	1 cycle	5 cycles	1 cycle	4 cycles	1 cycle	3 cycles	3 cycles
Energy Per Bit	0.36pJ	0.26pJ (28.1% Reduction)	0.84pJ	0.66pJ (21.4% Reduction)	8.5pJ	7.6pJ (11.1% Reduction)	7.4pJ (13% Reduction)
ΔE_{CI}	0pJ	0.039pJ	0pJ	0.053pJ	0pJ	0.47pJ	
Energy Per Bit inc. ΔE_{CI}	0.36pJ	0.30pJ (16.9% Reduction)	0.84pJ	0.71pJ (15.1% Reduction)	0.85pJ	0.79pJ (7.4% Reduction)	
Digital Logic Power (Total [static contribution])	12.3uW [389nW]	50.8uW [514nW]	30.8uW [1.4uW]	167.7uW [4.1uW]	127.5uW [9.0uW]	320.6uW [92uW]	
Energy Breakdown	96.0% 2.6% 1.4%	76.2% 6.8% 17.0%	94.8% 2.3%	67.5% 8.4% 24.1%	78.7% 18.0% 3.3%	57% 12.6% 30.4%	Analogue Tx Energy / Digital Logic Energy / Analogue Rx Energy / Area Represents Total Energy
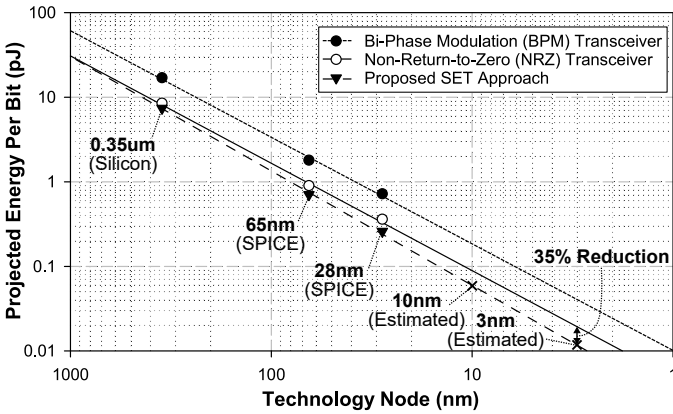
Fig. 23. Projected energy savings when using the proposed spike-latency encoding scheme, when compared with the BPM [18] and NRZ Benchmark designs [10]–[13], as process technology scales.

paper are in the order of 28%, following the trend to the 3nm node and beyond suggests that the proposed approach has potential to offer improvements of over 80% when compared to BPM transceivers and 35% when compared to the existing state-of-the-art NRZ inductive transceiver designs.

VIII. CONCLUSIONS

This paper presented a low-energy inductive transceiver for ICL-based 3D-ICs. The proposed transceiver combines: (1) a novel modulation scheme (spike-latency encoding) to perform time-domain data coding, and (2) a tuneable current driver circuit to adjust the transmit current to a minimum, depending on the 3D assembly quality. The proposed transceiver was modelled mathematically, simulated in 0.35 µm, 65nm and 28nm CMOS technologies, and experimentally validated in a 2-tier 3D stacked silicon testchip. Silicon evaluation of the proposed modulation approach demonstrates an energy of 7.4pJ/bit, representing a reduction >13% when compared

to previously reported schemes (or 7.4% when considering the additional energy overheads of peripheral clock timing control circuits). Simulated results show even greater energy savings (up to 28%) at more advanced technology nodes. Combined with the adaptive current driver, this equates to a 7.7× improvement in energy-per-bit compared to state-of-the-art implementations. Whilst these gains come at the cost of a slight decrease in maximum data-rate, the transceiver proposed in this paper shows strong promise for use in low-power, low-cost IoT devices which do not require gigabit operating bandwidths.

REFERENCES

[1] S. Oh et al., "IoT^2 - the internet of tiny things: Realizing mm-scale sensors through 3D die stacking," in *Design, Automation Test in Europe Conf. (DATE)*, 2019.

[2] M. Koyanagi, "Heterogeneous 3D integration for Internet of Things," in *IEEE Int. Conf. on Solid-State and Integrated Circuit Technology (ICSICT)*, 2014.

[3] G. K. et al., "A millimeter-scale wireless imaging system with continuous motion detection and energy harvesting," in *Symp. on VLSI Circ.*, 2014.

[4] I. A. Papistas and V. F. Pavlidis, "Contactless heterogeneous 3-D ICs for smart sensing systems," *Integration*, vol. 62, pp. 329–340, 2018.

[5] B. Fletcher, C.S. Poon, S. Das and T. Mak, "Low-power 3D integration using inductive coupling links for neurotechnology applications," in *Proc. Conf. Design, Automation and Test in Europe*, 2018.

[6] I. A. Papistas, V. F. Pavlidis, and D. Velenis, "Fabrication cost analysis for contactless 3-D ICs," *IEEE Trans. on Circuits and Systems II: Express Briefs*, vol. 66(5), pp. 758–62, 2019.

[7] D. Ditzel and T. Kuroda, "Low-cost 3D chip stacking with ThruChip wireless connections," in *IEEE Hot Chips 26 Symp. (HCS)*, 2014.

[8] Y. T. et al., "3d clock distribution using vertically/horizontally-coupled resonators," in *IEEE Int. Solid-State Circ. Conf.*, 2013.

[9] Y. Y. et al., "Chip-to-chip power delivery by inductive coupling with ripple canceling scheme," *Jap. J. of Applied Phys.*, vol. 47, 2008.

[10] N. Miura et al., "A 0.14pj/b inductive-coupling inter-chip data transceiver with digitally-controlled precise pulse shaping," in *IEEE Int. Solid-State Circuits Conf.*, 2007.

[11] D. Mizoguchi et al., "A 1.2Gb/s/pin wireless superconnect based on inductive inter-chip signaling (IIS)," in *IEEE Int. Solid-State Circuits Conf.*), 2004.

[12] N. Miura *et al.*, "An 11gb/s inductive-coupling link with burst transmission," in *IEEE Int. Solid-State Circuits Conf.*, 2008.

[13] N. Miura *et al.*, "Analysis and design of inductive coupling and transceiver circuit for inductive inter-chip wireless superconnect," *IEEE Jour. of Solid-State Circ.*, vol. 40(4), pp. 829–37, 2005.

[14] K. Ueyoshi *et al.*, "Quest: A 7.49TOPS multi-purpose log-quantized DNN inference engine stacked on 96MB 3D SRAM using inductive-coupling technology in 40nm CMOS," in *IEEE Int. Solid - State Circuits Conf. (ISSCC)*, 2018.

[15] S. Gopal *et al.*, "Dual-equalization-path energy-area-efficient near field inductive coupling for contactless 3D IC," in *IEEE MTT-S International Microwave Symp. (IMS)*, 2019.

[16] N. K. *et al.*, "Maximizing the data rate of an inductively coupled chip-to-chip link by resetting the channel state variables," *IEEE Trans. on Circ. and Sys. I*, vol. 66(9), 2019.

[17] Austria Microsystems AG, "AMS 0.35um CMOS Technology (C35B4)," 2019. [Online]. Available: https://ams.com/process-technology

[18] N. Miura *et al.*, "A 1Tb/s 3W inductive-coupling transceiver for inter-chip clock and data link," in *IEEE Int. Solid-State Circuits Conf.*, 2006.

[19] L. Zhang *et al.*, "A single phase modulation for pulse-based inductive-coupling connection in 3D stacked chip," *IEICE Electronics Express*, vol. 14(20), no. 20, 2017.

[20] L. Zhang, T. Li, B. Wang, and X. Zou, "A 50% power reduction in inductive-coupling transceiver for 3D-stacked inter-chip data link," in *IEEE Int. Nanoelectronics Conf. (INEC)*, 2016.

[21] X. Sun *et al.*, "Inductive links for 3D stacked chip-to-chip communication," in *IEEE Electronic Components and Tech. Conf. (ECTC)*, 2019.

[22] N. Miura *et al.*, "A high-speed inductive-coupling link with burst transmission," *IEEE J. of Solid-State Circ.*, vol. 44(3), pp. 947–55, 2009.

[23] Kiichi Niitsu *et al.*, "A 65 fJ/b inductive-coupling inter-chip transceiver using charge recycling technique for power-aware 3D system integration," in *IEEE Asian Solid-State Circ. Conf.*, 2008.

[24] N. Miura *et al.*, "A 0.55V 10 fJ/bit inductive-coupling data link and 0.7V 135 fJ/Cycle clock link with dual-coil transmission scheme," *IEEE J. of Solid-State Circ.*, vol. 46(4), pp. 965–973, 2011.

[25] B. J. Fletcher, S. Das, and T. Mak, "A low-energy inductive transceiver using spike-latency encoding for wireless 3D integration," in *IEEE Int. Symp. on Low Power Elec. and Design (ISLPED)*, 2019.

[26] I. A. Papistas and V. F. Pavlidis, "Crosstalk noise effects of on-chip inductive links on power delivery networks," in *IEEE Int. Symp. on Circuits and Systems (ISCAS)*, 2016.

[27] I. A. Papistas and V. F. Pavlidis, "Efficient modeling of crosstalk noise on power distribution networks for contactless 3-D ICs," *IEEE Trans. on Circ. and Systems I*, vol. 65(8), pp. 2547–58, 2018.

[28] U. M. Jow *et al.*, "Design and optimization of printed spiral coils for efficient transcutaneous inductive power transmission," *IEEE Trans. on Biomedical Circuits and Systems*, vol. 1(3), pp. 193–202, 2007.

[29] B. J. Fletcher, S. Das, and T. Mak, "Design and optimization of inductive-coupling links for 3-D-ICs," *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, vol. 27(3), pp. 711–723, Dec. 2018.

[30] Kiichi Niitsu *et al.*, "Interference from power/signal lines and to SRAM circuits in 65nm CMOS inductive-coupling link," in *IEEE Asian Solid-State Circ. Conf.*, 2007.

[31] K. Tu, "Reliability challenges in 3D IC packaging technology," *Micro-electronics Reliability*, vol. 51(3), pp. 517 – 523, 2011.

[32] A. Fazzi *et al.*, "3D capacitive interconnections with mono- and bi-directional capabilities," in *IEEE Int. Solid-State Circuits Conf.*, 2007.

[33] Q. Gu *et al.*, "Two 10Gb/s/pin low-power interconnect methods for 3D ICs," in *IEEE Int. Solid-State Circuits Conf.*, 2007.

[34] S. W. Han, "Wireless interconnect using inductive coupling in 3D-ICs." Ph.D. dissertation, University of Michigan, 2012.

[35] N. Miura *et al.*, "A 195Gb/s 1.2W 3D-stacked inductive inter-chip wireless superconnect with transmit power control scheme," in *IEEE Int. Solid-State Circuits Conf.*, 2005.

Benjamin J. Fletcher received the B.Eng. degree (honors) in electronic engineering from the University of Southampton, U.K., in 2016 where he is currently a PhD candidate studying as part of the ARM-ECS research centre (a joint collaboration between the University of Southampton and Arm Ltd, based in Cambridge, UK). His research interests include analogue and mixed-signal circuit design, low-power VLSI and 3D integration. In 2018 he was the recipient of the Institute of Engineering Technology Postgraduate Prize for his research on low-cost 3D integration approaches, and in 2019 won the International Symposium on Low Power Electronic Design Best Paper award. More recently, in 2020, he also received the IEEE Communications Society (ComSoc) award for outstanding contributions to future communications networks.

Shidhartha Das received the M.Sc. and Ph.D. degrees from the University of Michigan, Ann Arbor, MI, USA, in 2003 and 2009, respectively. He is currently a Senior Principal Research Engineer at Arm Research, Cambridge, U.K. His current research interests include emerging non-volatile memory technologies, microarchitectural and circuit design for variation measurement and mitigation, on-chip power delivery, and VLSI architectures for digital signal processing accelerators. Dr. Das was a recipient of the Arm Patent Cube in 2017 and the Arm Inventor of the Year Award in 2016 for his contributions to emerging nonvolatile memory technologies, multiple best paper awards (ISLPED 2019, CAL 2017, ISLPED 2015, SAME 2010, and MICRO 2003), and the Microprocessor Review Analysts Choice Award in Innovation in 2007. He served as a Guest Editor for the Journal of Solid-State Circuits and an Associate Editor for IEEE Solid-State Circuits Letters. He serves on the Technical Program Committees of of ISSCC and MICRO.

Terrence Mak is an Associate Professor at Electronics and Computer Science, University of Southampton, UK. Supported by the Royal Society, he was a Visiting Scientist at MIT during 2010, and also, affiliated with the Chinese Academy of Sciences as a Visiting Professor since 2013. His research areas includes computer architecture design, optimisation and adaptation for VLSI systems, network-on-chip, 3D-IC and, lately, wireless-on-chip. Throughout a spectrum of publications, he has awarded six Best Paper Awards, and one nominated, from prestigious conferences, at EMBS'05, DATE'11, VLSI-SoC'14, PDP'15, EUC'16, DATE'18 (nominated) and ISPLED'19. He has granted two US patents of his engineering designs, *i.e.* US16/685,090 and US13/638,330. He was also awarded the IET Premium Yearly Best Paper Award for Computer & Digital Techniques in 2013, and his newly published journal based on 3D-IC was awarded the prestigious 2015 IET Computers & Digital Techniques Premium Award. His publication at IEEE Transactions has been selected as "Top 25 Downloaded Manuscript" in 2015. He has published more than 150 papers in both conferences and journals, and jointly published 4 books.