8.1 A 47×10Gb/s 1.4mW/(Gb/s) Parallel Interface in 45nm CMOS

Frank O'Mahony, Joseph Kennedy, James E. Jaussi,
Ganesh Balamurugan, Mozhgan Mansuri, Clark Roberts,
Sudip Shekhar, Randy Mooney, Bryan Casper

Intel, Hillsboro, OR

The recent emphasis on power efficiency in serial I/O [1-4] reflects the growing need for lower-power chip-to-chip interfaces for computing systems. Board-level transceivers using a variety of low-power circuit techniques have demonstrated power efficiencies as low as 2.2mW/(Gb/s) across four data lanes [1]. Because power efficiency generally degrades as the per-lane data rate increases [2], low-power interfaces with high aggregate bandwidths must combine many parallel data lanes within the silicon, package and board area constraints. Parallel links can also reduce average power by disabling some or all lanes during periods of sub-peak bandwidth demand, but the efficiency and latency of this scheme is limited by wake-up time [4].

This paper describes a 470Gb/s binary NRZ parallel interface in 45nm CMOS that consumes 1.4mW/Gb/s. The circuitry and interconnect were co-designed to minimize power and area. Power is reduced by sharing clocking within "bundles" of data lanes, minimizing the span of clock signals and pairing a low-swing TX driver with a sensitive RX sampler. Silicon area is minimized by using on-chip transmission lines (TLs) to redistribute clock and data signals, while a dense, top-side package connector enables intra-bundle delay matching. The interface also has fast (<5ns) RX standby wake-up and an integrated wake-up time measurement circuit to enable aggressive power management. The interface and channel topology are intended for CPU-to-CPU and CPU-to-memory communication.

Figure 8.1.1 shows the interconnect topology and link schematic. The link is asymmetric full-duplex with 19 lanes in one direction and 28 lanes in the opposite direction. The data lanes are organized into groups of 9 or 10, which are referred to as bundles. A single forwarded clock transmitter and injection-locked VCO (IL-VCO) are shared for each die. The interconnect topology consists of two packaged dies connected to a bridge board through top-side package connectors. This topology is compatible with either a high-density interconnect (HDI) or Flex cable, but the HDI implementation is the focus of this work. The data signals for each bundle are routed on a single layer of the bridge, and all lanes within a bundle are length matched to <100μm. This allows clock recovery to be done on a per-bundle basis. The package-to-bridge connector is a 500μm pitch LGA, which provides approximately 4X area density advantage over socket-to-PCB routing and facilitates length matching in the package breakout. The channel is continued on-die with length matched TLs that route the data and clock signals to centrally located TX and RX bundle circuitry (Fig. 8.1.7). Each bundle occupies the area of only eight C4 bumps. The total area for active interface circuitry is 3.2mm^2.

Figure 8.1.2 shows the schematic for the TX portion of the interface. A supply-regulated IL-VCO generates the interface clock, emulating a system wherein multiple interfaces share a single PLL and filter the clock locally. Alternately, a per-interface PLL based on this VCO would generate comparable short-term jitter. The two-stage voltage regulator uses a 1.5V supply to provide adequate headroom and has a simulated supply-noise rejection of 50dB. An open-drain buffer drives the interface clock over on-die TLs to the TX bundles [2]. The duty-cycle of the low-swing clock is corrected using AC coupling and a digital DCC similar to [3]. Inverters drive the clock 200μm through a clock tree to the TX lanes. The TX driver consists of a 2-tap equalizer, a 2:1 serializer and a low-swing CML driver. To minimize power due to high-speed data multiplexers, the driver has fixed post-cursor polarity and dedicated cursor and post-cursor segments with current-controlled tap weights.

The RX bundle block diagram is shown in Fig. 8.1.3. The phase rotator (PR) consists of a digital DCC, a 6-stage/180° DLL and a two-stage 360° phase mixer with 72 phase steps/cycle. Complementary clock phases from the PR drive a pair of offset-trimmed data samplers in each lane. A replica delay line following the PR generates an edge-sampling clock that is enabled periodically to implement a sub-rate CDR. Following initial link clock-data deskew, which is done based on BER eye margining, the voltage offset trim of the edge sampler is adjusted to sample at the metastable point of rising data edges. Whenever the edge sampler is subsequently enabled, the corresponding CDR logic provides an indication of clock-data drift, which is corrected by adjusting the PR. Note that the forwarded clock is still the primary mechanism for clock-data phase tracking, but the sub-rate CDR corrects for low-frequency drift between the forwarded clock and data.

Asserting the *standby* signal in Fig. 8.1.3 puts the RX clock and data paths into a low-power state by gating the clocks and most analog biases. To facilitate wake-up in a few nanoseconds, bias nodes and loops with relatively long time constants – such as the DLL control voltage (*pctl*) – are maintained during standby. During normal DLL operation, a tracking ADC based on a replica delay cell voltage DAC (VDAC) periodically samples *pctl* and stores its digital value. Then during standby, the charge pump and *nctl* are disabled and the VDAC maintains the value of *pctl* on the loop capacitor. When *standby* is de-asserted, the DLL takes back control of the loop filter.

Figure 8.1.4 shows the circuitry included in each RX lane to measure wake-up time. A shift register and MUX generate a digitally adjustable, delayed version of *standby*. When *standby_del* de-asserts, the LFSR pattern checker transitions from open-loop (seeding) to closed-loop operation. If the delay from *standby* to *standby_del* is long enough that the RX circuits have recovered, then the checker seeds correctly and no bit errors are detected. Otherwise, many errors are detected. The *standby_del* timer has 170ps to 22ns delay resolution, 4b range, and consumes only leakage power except during a power state transition.

The link is tested in the channel configuration shown in Fig. 8.1.1 with all 47 data lanes operating simultaneously. Initialization is entirely automated and controlled through software. Most calibration is done by shared FSMs within each bundle. Figure 8.1.5 shows a bathtub plot for one bundle at 10Gb/s/lane with a peak TX swing of 150mV$_{ppd}$ across a 2 inch HDI channel. A reasonable metric for the cost of sharing one PR per bundle is the difference between the minimum lane timing margin and the aggregate bundle timing margin. Based on the data in Fig. 8.1.5, the timing margin overhead of sharing PRs is ≤0.08UI.

The measured power consumption for the link is 660mW under nominal conditions, corresponding to a power efficiency of 1.4mW/(Gb/s). This excludes the on-chip pattern generators, checkers, counters and periodically enabled calibration circuitry. Figure 8.1.6 summarizes the power breakdown. RX active power is reduced by 93% during standby, and the measured wake-up time ranges from 3.3 to 4.8ns. Furthermore, Figure 8.1.6 shows how this interface compares with recent low-power and high-bandwidth links. The power efficiency is 36% lower than [1] and >10 lower than the best reported link with comparable aggregate bandwidth [5].

Acknowledgments:
The authors thank B. Provost, B. Horine, M. Trobough, H. Braunisch, H. Do, L. Armenta, A. Sanchez, D. Searles, S. Lotz, A. Ganesan, N. Tiwari, Sowmya S., V. Baca, Kumar K., K. Truong, D. Finan, D. Jenkins and H. Wilson.

References:
[1] J. Poulton, et al., "A 14-mW 6.25-Gb/s Transceiver in 90-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 42, no. 12, pp. 2745-2757, Dec., 2007.
[2] G. Balamurugan, et al., "A Scalable 5–15 Gbps, 14–75 mW Low-Power I/O Transceiver in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 43, no. 4, pp. 1010-1019, Apr., 2008.
[3] S. Joshi, et al., "A 12-Gb/s Transceiver in 32-nm Bulk CMOS," *Symp. VLSI Circuits Dig. Tech. Papers*, pp. 52-53, Jun., 2009.
[4] R. Palmer, et al., "A 4.3GB/s Mobile Memory Interface with Power-Efficient Bandwidth Scaling," *Symp. VLSI Circuits Dig. Tech. Papers*, pp. 136-137, Jun., 2009.
[5] H. Lee, et al., "A 16Gb/s/Link, 64 GB/s Bidirectional Asymmetric Memory Interface," *IEEE J. Solid-State Circuits*, vol. 44, no. 4, pp. 1235-1247, Apr., 2009.

978-1-4244-6034-2/10/$26.00 ©2010 IEEE

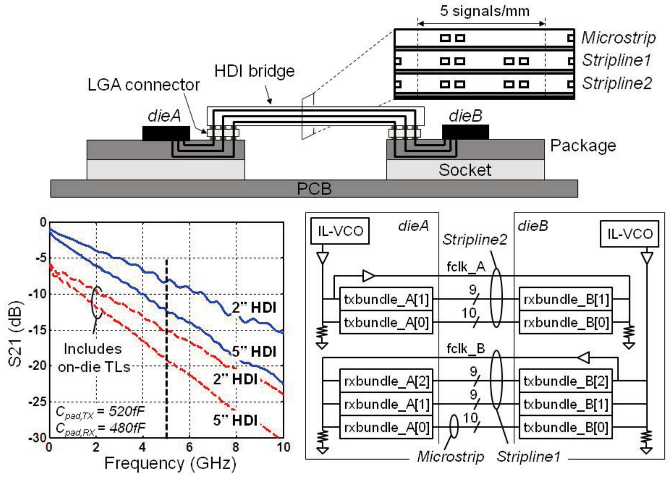


Figure 8.1.1: Interconnect topology and interface block diagram.

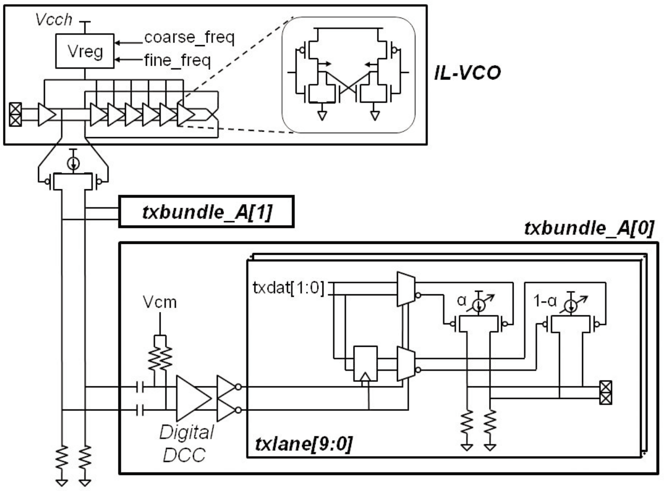


Figure 8.1.2: TX IL-VCO and bundle schematic (die A configuration).

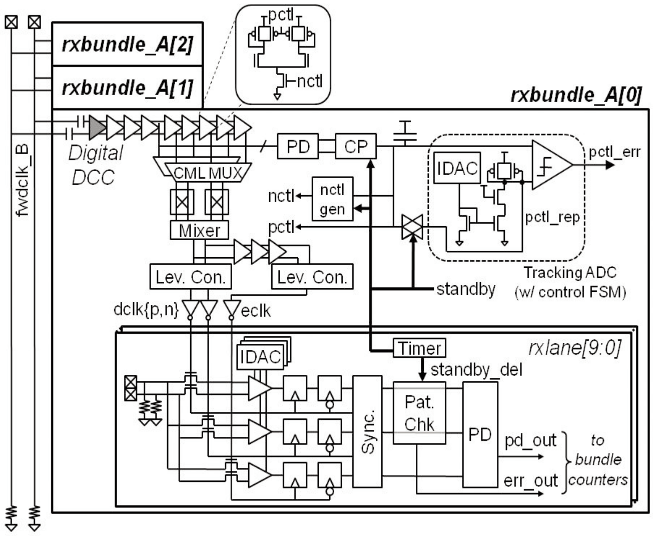


Figure 8.1.3: RX bundle clocking and lane schematics (die A configuration).

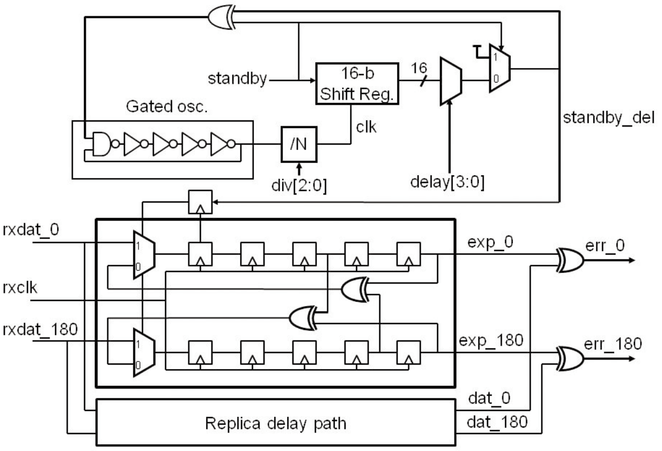
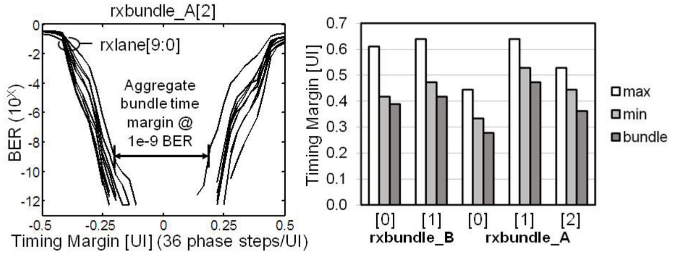


Figure 8.1.4: LFSR-based pattern checker and wakeup timer.



	Lane [mW]	Bundle [mW]	Port [mW]	Total [mW]	Pwr. Eff. [mW/Gb/s]
TX driver	2.46	—	—	116	0.25
TX serializer/equalizer	2.82	—	—	133	0.28
TX bundle clocking	—	11.1	—	56	0.12
RX samplers (switching)	1.74	—	—	82	0.17
RX amplifiers (data only)	0.68	—	—	32	0.07
RX bundle clocking	—	23.9	—	119	0.25
Forward clock driver	—	—	18.8	37	0.08
IL-VCO	—	—	20.0	40	0.09
Port clock buffer	—	—	22.6	45	0.10
			TOTAL	**660**	**1.40**

Figure 8.1.5: Measured timing margins and power with 2" HDI channel.

Supply Voltage	0.8V/1.5V (global clock)
Aggregate bandwidth	47x10Gb/s
Active power	660mW
Active area (TX / RX / Port clock)	1.36mm^2 / 1.36mm^2 / 0.47mm^2
TX swing	~150mV$_{pp\text{-}diff}$
RX input-referred noise	<1.3mV-rms
Measured BER (0 bit errors for all lanes)	<10^{-14} (each lane) < 2x10^{-16} (aggregate)
RX wake-up time (90% RX power reduced)	<5ns

	[1]	[2]	[3]	[4]	[5]	[This work]	
Total bandwidth [Gb/s]	25	10	12	43	512	**470**	
Lane data rate [Gb/s]	6.25	10	12	4.3	16	**10**	
Power eff. [mW/Gb/s]	2.2	3.6	3.15	2.7	13 / 8	**1.40**	**1.51**
Loss @ fundamental [dB]	15	7	10	—	15	**2" HDI: 8 (15 w/ TL)**	**5" HDI: 12 (19 w/ TL)**
Area [mm^2/Gb/s]	0.058	—	—	0.021*	0.027*	**0.007 (active) 0.017 (C4 bump area)**	
CMOS Technology [nm]	90	65	32	40	65	**45**	

** Includes bidirectional circuitry*

Figure 8.1.6: Performance summary (2" HDI channel) and comparison.

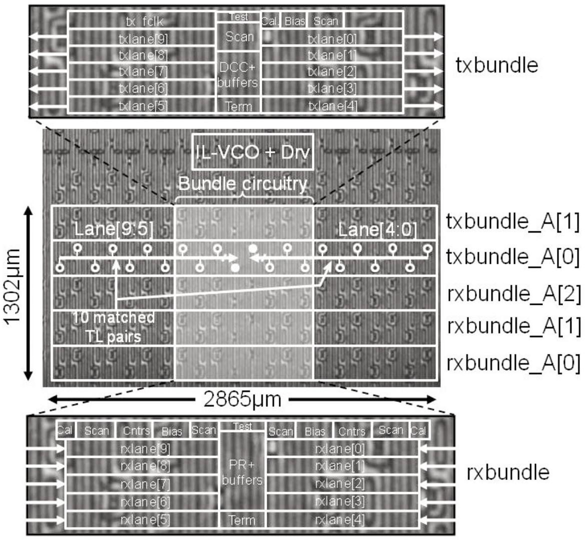


Figure 8.1.7: Die micrograph and floorplan.