

## 20.8 A 2×25Gb/s Deserializer with 2:5 DMUX for 100Gb/s Ethernet Applications

Ke-Chung Wu, Jri Lee

National Taiwan University, Taipei, Taiwan

The ever growing bandwidth requirement for novel server technologies including multi-core processing, virtualization, and networked storage leads to multi-channel Internet connectivity such as 100GbE. Among the proposed standards [1], those with 4 channels (e.g., 100GBASE-ER4) are selected to arrive at a reasonable component count in discrete and photonic integration. This paper presents a 2-channel receiver prototype that amplifies, retimes, and deserializes the data in the same manner as a full 4×25Gb/s receiver, providing CMOS design examples and references for fully-integrated 100GbE transceivers.

The receiver architecture is shown in Fig. 20.8.1. Two channels process the input data independently, presenting an aggregate data rate of 50Gb/s. Each channel consists of a limiting amplifier (LA) with constant gain biasing, and a full-rate clock and data recovery (CDR) circuit. The two retimed data streams are further demultiplexed into five 10Gb/s lanes in parallel. The two 25GHz clocks distilled from the data streams are sent to a clock generator, which creates 2.5, 5, 10, and 12.5GHz clocks for the subsequent deserializer. At 25Gb/s, it is difficult to conduct 1:5 demuxing directly with reasonable power and skew, because one CDR needs to drive at least 5 flipflops simultaneously. The clock generation suffers as well since a ÷5 circuit must be made at 25GHz. Here, we perform an additional 1:2 demuxing right after the CDR to relax the stringent speed requirement. The 1:5 demuxing can therefore be realized in a relaxed way, and finally five 4:1 MUXes are incorporated to produce five 10Gb/s outputs. Note that the two channels may be subject to finite phase difference due to imbalance. A deskew circuit is placed in channel 2 to line up its 2.5Gb/s data streams with those in channel 1. This adjustment is mandatory because the middle 4:1 MUX handles inputs from both channels. A full 4×25Gb/s receiver can easily be implemented by using 2 identical chipsets proposed here.

The limiting amplifier design is illustrated in Fig. 20.8.2(a), where 5 identical gain stages are cascaded to achieve high gain ( $\approx 26$ dB) while maintaining broad bandwidth ( $f_{-3dB} > 42$ GHz). Offset is canceled by means of an RC feedback loop, which reveals a corner frequency of 2.5kHz. Each gain stage is realized as a simple differential pair with triple-resonance peaking [2], where the loading ( $150\Omega$  in parallel with a pMOS resistor) and the tail current are regulated by the constant-gain bias circuit, i.e., a constant loading resistance ( $\approx 125\Omega$ ) and a constant bias current ( $\approx 4$ mA). In other words, the small-signal gain as well as the maximum output swing ( $\approx 500$ mV) are fixed over temperature and process variations. The control signals  $V_{b,R}$  and  $V_{b,I}$  are generated as illustrated in Fig. 20.8.2(b). First, a sub-1V bandgap reference [3] is adopted to create a voltage of 0.7V ( $=1.2V - 0.5V$ ) and a current (4mA) which are immune from PVT variations. In nano-scale CMOS, a single-deck current source is prone to channel-length modulation when mirroring. Thus, we use the loop of  $R_1$ , Opamp 1, and  $M_1$ - $M_2$  to "refine" the tail current bias ( $V_{b,I}$ ), which connects to all current sources in the gain stages. It is because  $R_1$  ( $\approx 40\Omega$ ) suppresses the excessive  $V_{DS}$  of  $I_{SS}$  and  $M_2$  mimics the operation of the switching pair. Note that if we were to bias the gain cells with the bandgap current directly, the mirrored current would vary by 8%. The created  $V_{b,I}$  also biases  $M_3$ , which together with  $M_4$  and Opamp 2 forms another loop to produce  $V_{b,R}$ . The equivalent loading resistance is determined by the desired output swing and is nominally equal to  $125\Omega$  ( $\approx 500$ mV/4mA). Simulation suggests that both feedback loops are stable without any confliction. Using this compensation biasing, the gain and -3dB bandwidth variations for different processes and temperatures (0°C~100°C) are reduced from 9dB to 5dB and from 25GHz to 2GHz, respectively. Supply variation can be suppressed by adopting a voltage regulator. The CDR circuit follows the design in [4]. Taking advantage of 65nm CMOS, we increase the speed by 25% while reducing the power by 33%.

The 25Gb/s 1:2 DMUX is realized as a tree structure of CML flipflops with two outputs aligned in phase [5]. The subsequent 1:5 DMUX necessitates proper

phase arrangement to assemble the 20×2.5Gb/s data. As shown in Fig. 20.8.3(a), a 5-phase 2.5GHz clock is used to sample the 12.5Gb/s incoming data sequentially. Here, the outputs need to be separated by an angle as close as  $180^\circ$ . Retiming flipflops driven by  $\phi_3$  and  $\phi_5$  are therefore placed to launch  $\{D_{out1}, D_{out2}, D_{out3}\}$  and  $\{D_{out4}, D_{out5}\}$  at the rising edges of  $\phi_3$  and  $\phi_5$ , respectively. A deskew circuit must be incorporated here. For example, if channel 2 data need to be aligned with phase 3 in channel 1 ( $\phi_{3,CH1}$ ), one retiming flipflop (FF<sub>1</sub>) is sufficient theoretically [Fig. 20.8.3(b)]. However, in practice an error may occur if channel 2 data transition is too close to the rising edge of  $\phi_{3,CH1}$ . To ensure proper sampling, we assign  $\pm 72^\circ$  around  $\phi_{3,CH1}$  as forbidden area (gray). If channel 2 data transition locates in this region, a pre-sampling is made (by FF<sub>2</sub>) so that the data is shifted before the main sampling of  $\phi_{3,CH1}$ . Depending on the skew, the control logic picks  $Q_A$  or  $Q_B$  from the 2-to-1 selector. Such a design covers skew range of one complete UI (400ps).

The clock generator is shown in Fig. 20.8.4. Two 25GHz clocks from CDRs are first divided by 2 and then divided by 5. The 25GHz ÷2 circuit is realized as static topology with an inductively-peaked flipflop. The 5-phase ÷5 circuit is implemented as that in [6]. All blocks here are made of CML structures including the NAND gate. One 2.5GHz clock is sent to the 10GHz CMU (made of LC-tank oscillator, type IV PFD, and third-order on-chip loop filter) to create clocks for the 4:1 MUXes. The 10GHz clock synchronizes all five 10Gb/s data outputs.

The circuit has been fabricated in 65nm CMOS technology. It consumes a total power of 510mW for 2 channels, of which 32×2mW dissipates in LAs, 99×2mW in CDRs, 128mW in the 2:5 DMUX, and 120mW in the clock generator. A 1.2V supply voltage is used throughout the chip except the 2:5 DMUX, which requires a 1.4V supply to accommodate larger data swing. The input matching presents an  $S_{11}$  less than -13dB from dc to 25GHz. The CDR presents an operation range of 640MHz, across which no performance degradation is observed. Figure 20.8.5(a) and (b) depict one CDR recovered data (25Gb/s) and one final output data (10Gb/s) in response to a  $2^{31}-1$  PRBS, revealing jitter of 1.02ps,rms/6.00ps,pp and 1.45ps,rms/8.89ps,pp, respectively. The CDR bandwidth is about 10MHz. The phase noise plots of the recovered clocks at 25GHz and 10GHz are also recorded in Fig. 20.8.5(c), suggesting -108 and -111dBc/Hz at 1MHz offset. The integrated rms jitters from 100Hz to 1GHz are 254fs and 340fs, respectively. It achieves BER< $10^{-12}$  for input greater than 20mV<sub>pp</sub>. Figure 20.8.6(a) shows the jitter tolerance of one 10Gb/s output in response to a 100mV<sub>pp</sub>  $2^7-1$  PRBS input, which exceeds the extrapolated IEEE 802.3ae mask by at least 0.27UI<sub>pp</sub> for all the measurable jitter frequencies (the highest allowable modulation magnitude and rate of our BERT are 160UI<sub>pp</sub> and 10MHz, respectively). The on-chip inter-channel crosstalk is also measured. Figure 20.8.6(b) depicts the channel 1 output BER as a function of input power with channel 2 turned on and off, implying a power penalty of 1.3dB. Figure 20.8.7(a) shows the die photograph, which occupies 1.9×1.3mm<sup>2</sup>. A table summarizing the receiver performance is shown in Fig. 20.8.7(b).

### Acknowledgment:

This work is supported in part by NSC, MOEA, and MediaTek. The authors thank TSMC University Shuttle Program and Dr. Sheu for chip fabrication and C. Cole from Finisar for discussion.

### References:

- [1] IEEE P802.3ba 40Gb/s and 100Gb/s Ethernet Task Force [Online]. Available: <http://www.ieee802.org/3/ba/index.html>
- [2] S. Galal and B. Razavi, "40Gb/s Amplifier and ESD Protection Circuit in 0.18μm CMOS Technology," *ISSCC Dig. Tech. Papers*, pp. 480-481, Feb. 2004.
- [3] H. Banba *et al.*, "A CMOS Bandgap Reference Circuit with Sub-1-V Operation," *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 670-674, May 1999.
- [4] J. Lee and K.-C. Wu, "A 20Gb/s Full-Rate Linear CDR Circuit with Automatic Frequency Acquisition," *ISSCC Dig. Tech. Papers*, pp. 366-367, Feb. 2009.
- [5] K. Kanda *et al.*, "40Gb/s 4:1 MUX/1:4 DEMUX in 90nm standard CMOS," *ISSCC Dig. Tech. Papers*, pp. 152-153, Feb. 2005.
- [6] E. Tournier *et al.*, "High-speed dual-modulus prescaler architecture for programmable digital frequency dividers," *IEE Electron. Lett.*, pp. 1433-1434, Nov. 2001.

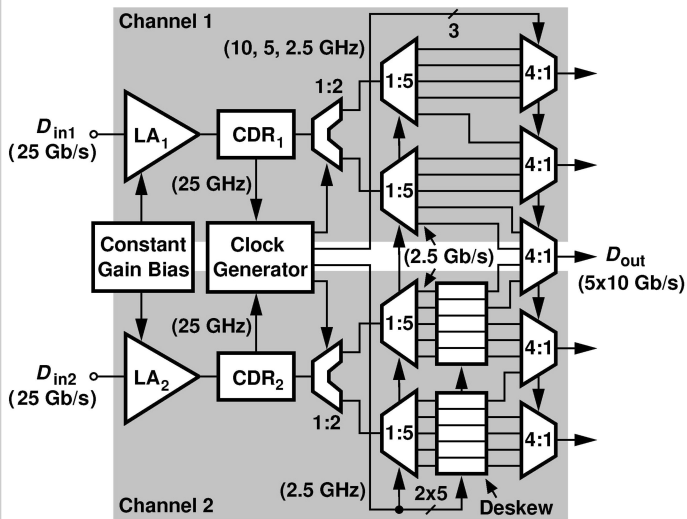


Figure 20.8.1: Architecture.

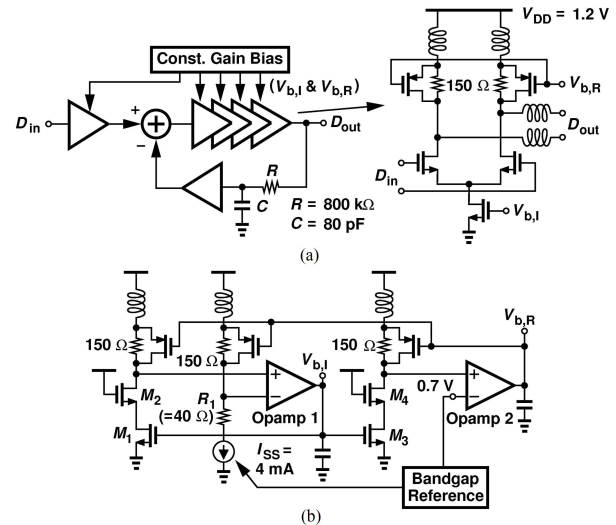
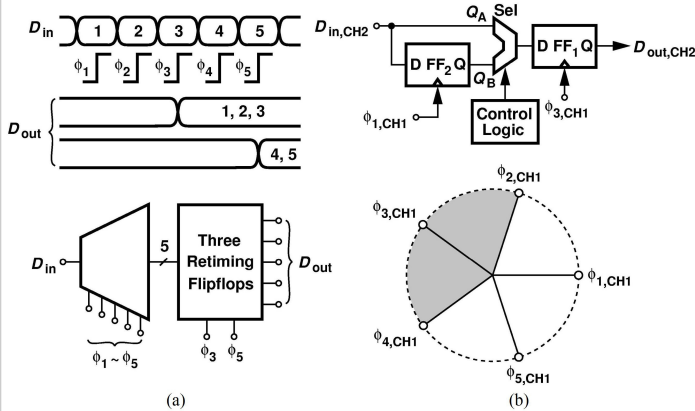
Figure 20.8.2: (a) Limiting amplifier, (b) bias circuit for  $V_{b,R}$  and  $V_{b,I}$ .

Figure 20.8.3: Example of (a) 1:5 DMUX with phase adjustment, (b) deskew circuit.

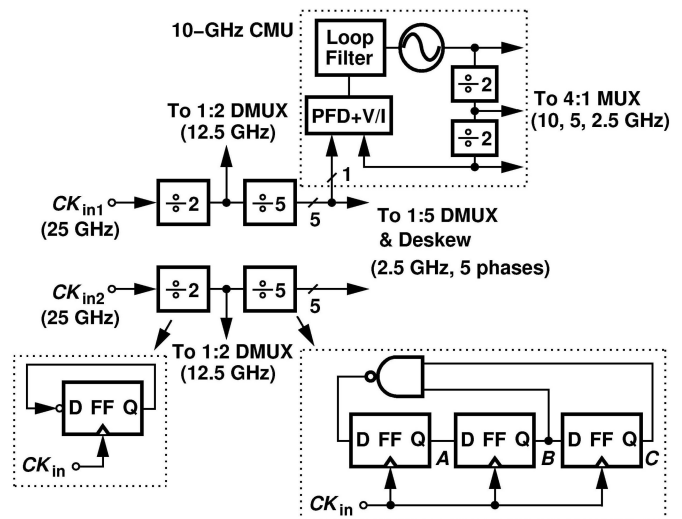


Figure 20.8.4: Clock generator.

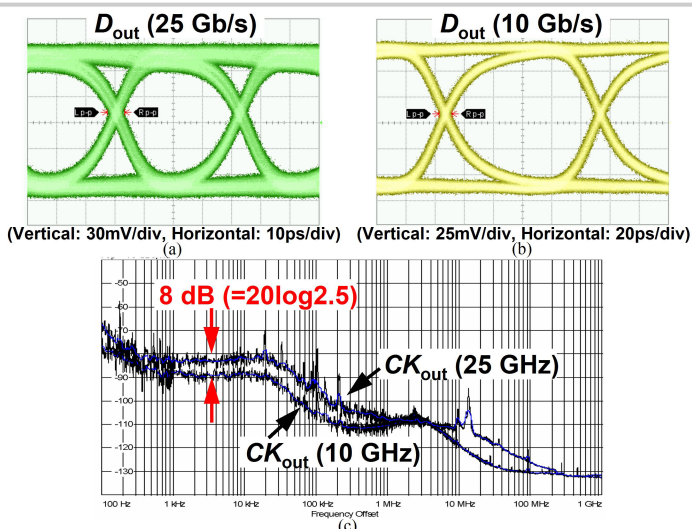
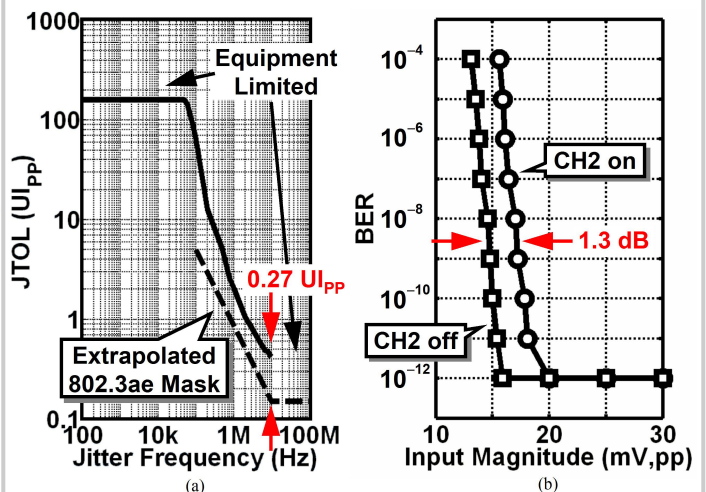
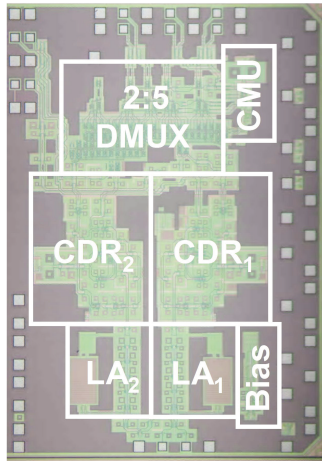


Figure 20.8.5: (a) CDR output data, (b) final output data, (c) phase noise of 25GHz and 10GHz clock outputs.

Figure 20.8.6: (a) JTOL for  $2^7-1$  PRBS with BER threshold of  $10^{-12}$ , (b) BER of channel 1.



(a)

Input Data Rate	25Gb/s x 2
Output Data Rate	10Gb/s x 5
Input Return Loss	< -13dB
Rec. Clock Jitter (with $2^{31}-1$ PRBS)	25GHz: 254fs,rms 10GHz: 340fs,rms
Rec. Data Jitter (with $2^{31}-1$ PRBS)	25Gb/s: 1.02ps,rms 6.00ps,pp 10Gb/s: 1.45ps,rms 8.89ps,pp
BER (with $2^7-1$ PRBS)	< $10^{-12}$
Jitter Tolerance	Exceeds extrapolated IEEE 802.3ae mask by 0.27UI <sub>pp</sub>
Supply	1.2V <sup>*</sup>
Power Diss.	510mW
Chip Area	1.9 x 1.3mm <sup>2</sup>
Technology	65nm CMOS

\* 1.4V used in 2:5 DMUX.

(b)

Figure 20.8.7: Die micrograph and performance summary.