7.6 Capacitively Coupled Non-Contact Probing Circuits for Membrane-Based Wafer-Level Simultaneous Testing

Mutsuo Daito[1], Yoshiro Nakata[1], Satoshi Sasaki[1], Hiroyuki Gomyo[1], Hideki Kusamitsu[1], Yoshio Komoto[1], Kunihiko Iizuka[1], Katsuyuki Ikeuchi[2], Gil Su Kim[2], Makoto Takamiya[2], Takayasu Sakurai[2]

[1]Association of Super-Advanced Electronics Technologies, Yokohama, Japan
[2]University of Tokyo, Tokyo, Japan

Wafer-level simultaneous testing (WLST) where all chips on a wafer are tested and burned in at the same time is preferable in reducing the cost of obtaining Known Good Dies (KGD's). At present, however, it is difficult to realize the WLST because it requires a probe card with some hundred thousand needles, leading to more than a ton of force needed for stable contact of all needles. Non-contact probing has been proposed based on a chip-to-chip inductively-coupled interface [1] which can reduce the force but it needs a probing chip built specific to a certain product, which is costly. Recently, a low-cost membrane-based probing technique has been disclosed which makes use of the atmospheric pressure and 700kg of force can be uniformly distributed over a 300mm wafer [2]. Yet, since a contacting bump is used and each bump requires 4g of force, the number of pins is limited to about 150K, which is still the world biggest pin counts ever reported.

Figure 7.6.1 shows the proposed WLST system. The contacting bumps can be used not only for power supply but also for testing ultra-high speed interfaces and analog and RF functions if needed. An expensive probe card with probe chips which implement the capacitively-coupled interfaces is common to all products and inexpensive organic interposer and membrane are to be redesigned and built when a product under test is changed. Normal atmospheric pressure is applied on top of the probe card but the space between the membrane and the wafer is depressurized to 0.1bar and the space between the membrane and the probe card is depressurized to 0.2bar. Thus, contacting bumps are pressed by 0.8bar of pressure to the wafer and non-contact pads can use 0.1bar. This two-stage depressurization approach is essential to the non-contact probing ensuring stable probing. The tester chips are placed on top of the probe card, which implement simple tester functions including test results compression, reducing the power and cost of communication between the main tester equipment and the probing system.

For this application, a capacitively-coupled interface is advantageous over an inductively-coupled interface in two aspects. One is that the same pad for wire bonding can be used for the capacitively-coupled probing. The other advantage is that capacitive coupling needs only one pad on a probe chip and one wire resource on probe card per channel, while a coil needs two pads for a driver and two wires. Considering the interconnect congestion in the probe card and the pad-limited nature of the probe chip, the inductive coupling [3] is prohibitive. This situation also hinders the use of differential signaling [4].

Figure 7.6.2 shows the system block diagram of the proposed WLST system. The signals travel between the DUT and the probe chip through the complex wiring structure. The total length of the wiring ranges around 20mm and diverse reflections occur at interfaces among layers. Consequently, de-skewing function should be implemented for each digital channel. A low-power and small-area de-skewing circuit, namely, Feed-forward All-digital De-skewer (FAD) in Fig.7.6.3 is proposed. The delay element is constantly adjusted to show one eighth of a clock cycle. For a certain channel, if the rising edges of delayed D_{IN} and CLK match at the location "A", D_{IN} is considered to rise about a half clock cycle before the CLK rising edge. Then, D_{IN} should be delayed about a half clock cycle to maximize the margin in synchronization which occurs at Flip-Flop "F" by using CLK bar. Thus, after the bubble error corrector and the 3rd order moving average digital filter, the LUT-based adaptive switch selector selects the switch "a". On the other hand, if D_{IN} rising edge comes a bit earlier (later), the matching edge is captured at location "B"("C") and the switch "b"("c") is selected. The data edge is adjusted within 8 data edges at most. Since all functions are realized by digital and the feed-forward control is employed, the area is as small as 0.04mm^2/Channel and the measured power is 1.2mW/GHz/Channel, which should be compared with 0.3mm^2/Channel of area and 6.4mW/GHz/Channel of estimated power achieved by a previously-reported DLL-based clock de-skewer [5].

Another points of difficulty in the membrane-based capacitively-coupled interface over a chip-to-chip interface lies in the transmission nature of wiring and the smaller received signal due to the large wiring capacitance of up to 2pF. To overcome the issue, a low-pass filter is inserted before a preamplifier, which suppresses the undesirable ringing as shown in Fig.7.6.4. A DC-cutting capacitor M1 is inserted to eliminate interference between precharger 1 and precharger 2 even if the precharge levels of the two are different due to process variation. A high-sensitivity receiver circuit is introduced, denoted as weak-feedback receiver. The previously reported receiver in [6] is robust but needs intermittent resetting of circuits which is difficult to implement in the test environment. Compared with the previously reported receiver [7], the proposed receiver shows the higher sensitivity as shown in Fig.7.6.5 over large PVT variations. The proposed circuit is less sensitive to parameters because the feedback is rather moderate through series-connected transistors. The whole receiver occupies 36x7.5μm^2 of area.

The probe chip and DUT-wafer are fabricated with a 90nm 1.2V CMOS technology. Two types of pad are implemented: one is an uncovered pad and the other is an insulator covered pad both of which worked successfully. Two versions of pad size, 80x80μm^2 and 160x160μm^2, are tried but turned out that 80x80μm^2 is sufficient even with the maximum alignment error which is less than ±5μm over the whole 300mm wafer. Figure 7.6.6 shows the measurement results. The circuit can operate up to 1Gbps, which is sufficient to cover the at-speed test for most of the next-generation SoC's for mobile and/or consumer digital applications. When the de-skew function (FAD) is turned off, the bit error rate (BER) is measured to be larger than 10^{-6} but the BER is improved below 10^{-12} when the FAD is turned on. The power consumption is 0.5mW/Channel for a RX core at 1.2V 1Gbps operation. Figure 7.6.7 shows a chip micrograph and a performance summary. RX for DUT is smaller in size than RX for a probe chip because low-pass filter is not needed for the latter since there is no long interconnect after the coupling capacitor.

Acknowledgement:
This work is supported by NEDO (New Energy and Industrial Technology Development Organization) entrustment project. Advantest Corp., Sharp Corp., Panasonic Corp. and Yamaichi Electronics kindly offered existing technologies and helped defining the specifications. Valuable discussions by Shigeru Matsumura and Shigeru Kannari from Advantest are appreciated.

References:
[1] Y. Yoshida, et al., "Wireless DC Voltage Transmission Using Inductive-Coupling Channel for Highly-Parallel Wafer-Level Testing," *ISSCC Dig. of Tech. Papers*, pp. 470-471, Feb. 2009.
[2] S. Sasaki and Y. Nakata, "A Challenge of 150k Probes on 300mm," IEEE SW Test Workshop, June. 2009.
[3] H. Ishikuro, T. Sugahara and T. Kuroda, "An Attachable Wireless Chip Access Interface for Arbitrary Data Rate Using Pulse-Based Inductive-Coupling through LSI Package," *ISSCC Dig. of Tech. Papers*, pp.360-361, Feb. 2007.
[4] L. Luo, et al., "A 3 Gb/s AC coupled chip-to-chip communication using a low swing pulse receiver," *ISSCC Dig. of Tech. Papers*, pp.522-523, Feb. 2005.
[5] A. Agrawal, P. K. Hanumolu and G.Y. Wei, "An 8×3.2Gb/s Parallel Receiver with Collaborative Timing Recovery," *ISSCC Dig. of Tech. Papers*, pp. 468-469, Feb. 2008.
[6] A. Fazzi, et al., "3D Capacitive Interconnections with Mono- and Bi-Directional Capabilities," *ISSCC Dig. of Tech. Papers*, pp. 356-357, Feb. 2007.
[7] S. A. Kühn, et al., "Vertical Signal Transmission in Three-Dimensional Integrated Circuits by Capacitive Coupling," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 1, pp.37-40, Apr. 1995.

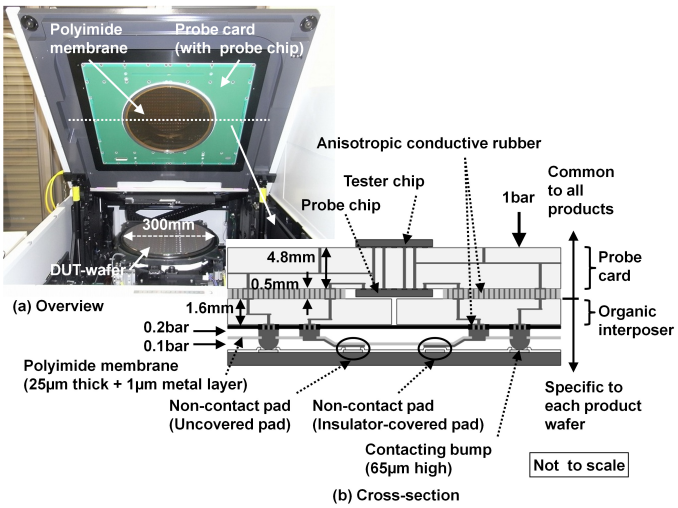


Figure 7.6.1: Wafer-level simultaneous testing scheme.

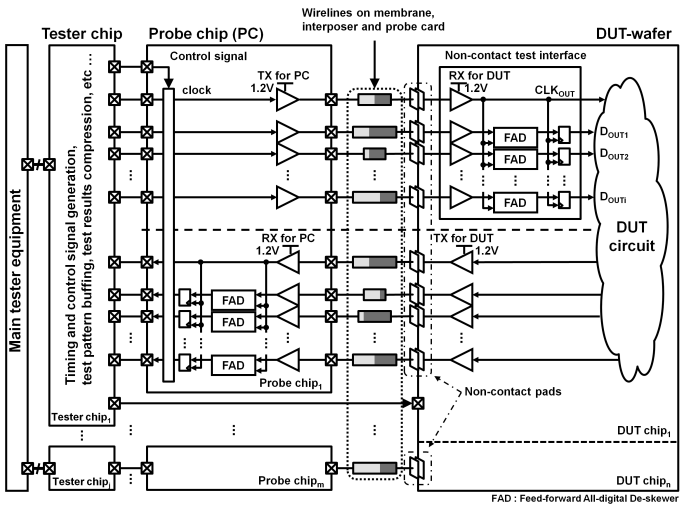


Figure 7.6.2: System block diagram of the proposed WLST system.

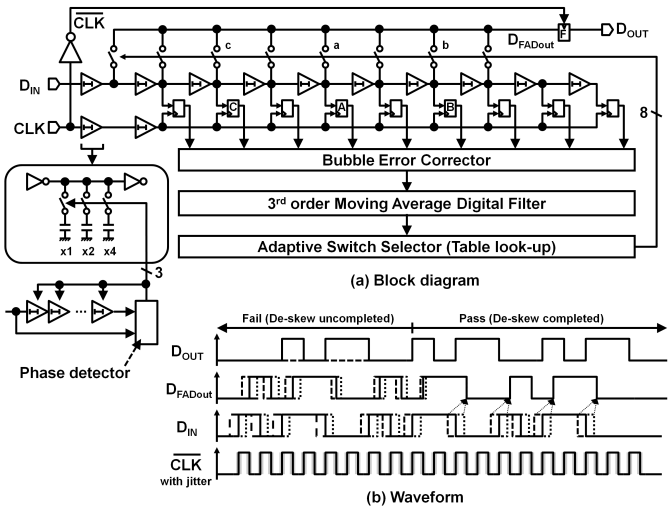


Figure 7.6.3: Feed-forward All-digital De-skewer.

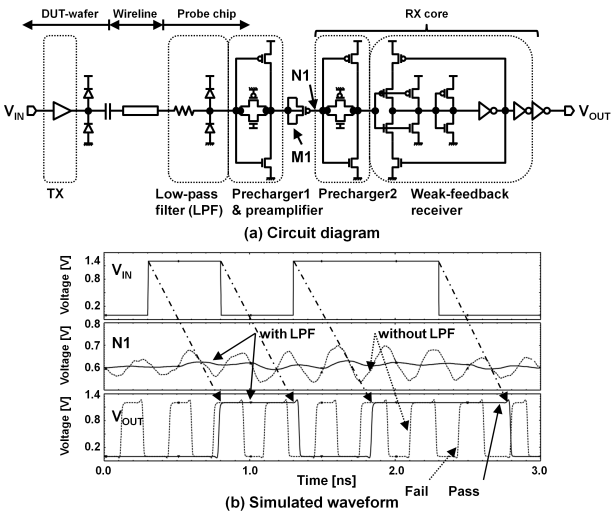


Figure 7.6.4: Proposed RX for a Probe chip.

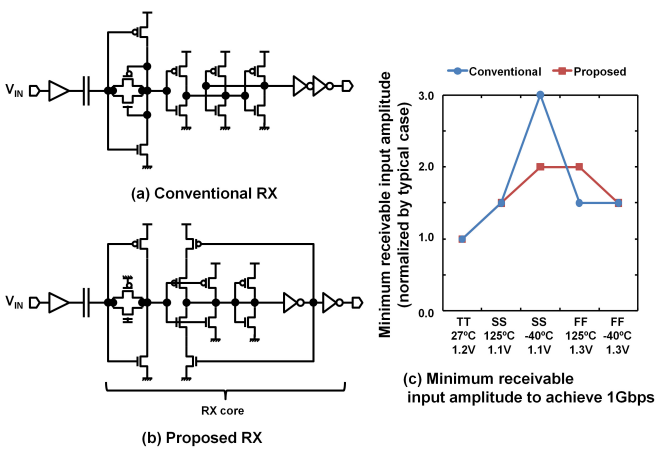


Figure 7.6.5: Comparison of Proposed RX with Conventional RX.

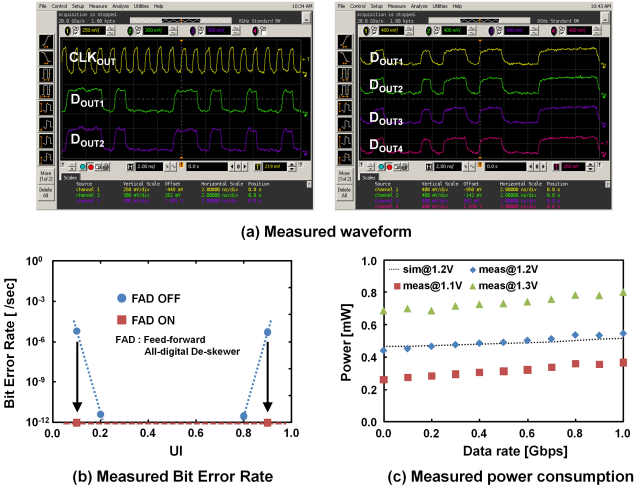
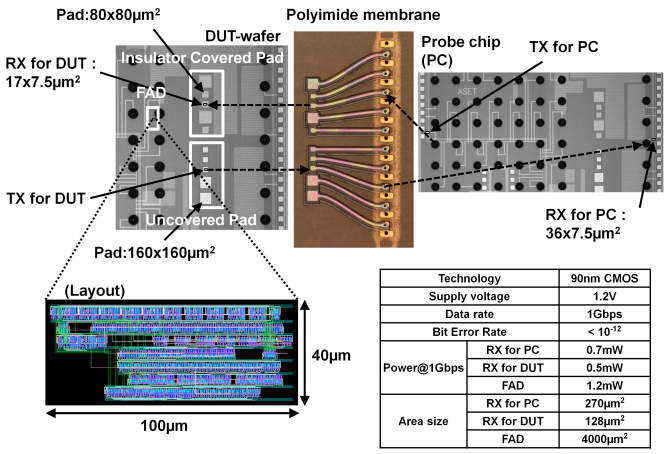


Figure 7.6.6: Measurement Results.



Technology		90nm CMOS
Supply voltage		1.2V
Data rate		1Gbps
Bit Error Rate		$< 10^{-12}$
Power@1Gbps	RX for PC	0.7mW
	RX for DUT	0.5mW
	FAD	1.2mW
Area size	RX for PC	270μm²
	RX for DUT	128μm²
	FAD	4000μm²

Figure 7.6.7: Chip micrograph and performance summary.

978-1-4244-6034-2/10/$26.00