17.7 *Sea of Leads*: A Disruptive Paradigm for a System-on-a-Chip (SoC)

Azad Naeemi, Chirag S. Patel, Muhannad S. Bakir, Payman Zarkesh-Ha,
Kevin P. Martin and James D. Meindl
Georgia Institute of Technology, Atlanta, GA

Sea of leads (SoL) is disruptive because it intends to use wafer level batch fabrication of ultra high density ($>10^4/\text{cm}^2$) x-y-z compliant input/output leads and packages as well as wafer level DC/AC testing and burn-in to enhance performance, cost, size, weight, and reliability of a mixed signal SoC. The SoL is accomplished by building flexible embedded leads in a compliant overcoat that facilitate electrical and mechanical connection of the chip to the system module without using underfill as illustrated in Figure 17.7.1. All compliant leads in SoL are concurrently fabricated over the entire wafer and all of the dice within a wafer are packaged simultaneously, which allows the package to be exactly the same size as the chip [1]. The short and highly reliable compliant leads exhibit minimal parasitics and excellent electrical performance in the GHz range. The wafer level fabrication of SoL and the ability to perform on wafer DC/AC testing and burn-in significantly reduce the manufacturing cost of the package and the time and complexity involved in testing protocols [2]. The monolithic fabrication of compliant leads makes the cost of a package independent of the number of leads. This feature of SoL allows the SoC designer to significantly increase the number of I/O connections and utilize therefor more effective signal, power, and clock distribution as well as testing purposes.

Ultra high density, 12,000 leads per cm^2, SoL technology has been demonstrated as shown in Figure 17.7.2. This array of x-y-z compliant leads is termed sea of leads (SoL). In addition, it is observed in Figure 17.7.2 that the leads are oriented along the contours of expansion to provide a higher degree of compliance moving radially outward from the center to the edge of the die.

High density SoL technology enhances the performance, especially I/O bandwidth, of a mixed signal SoC. Delay of long global interconnects limits performance. Increasing the cross-sectional area of global interconnects and using repeaters are key solutions for reducing interconnect delay. A common RC model for delay of an interconnect using an optimal number of repeaters is

$$\tau=2.5[R_{int}C_{int}R_0C_0]^{1/2}=2.5(\ /\omega)[\rho k\varepsilon_r\varepsilon_0R_0C_0]^{1/2}, \qquad (1)$$

where   and ω are wire length and width respectively, ρ is metal resistivity , R_0 and C_0 are output resistance and input capacitance of a minimum size repeater, $\varepsilon_r\varepsilon_0$ is the dielectric constant and k is a factor determined by the geometry.

Equation (1) is valid for small ω. For large ω, the impact of inductance cannot be neglected. Using optimal repeaters and Hspice simulations, wire delay versus width for the geometry of Figure 17.7.3 is shown in Figure 17.7.4. Skin effect is also considered when wire thickness is comparable with skin depth.

Figure 17.7.4 shows that as long as the RC model is valid, increasing ω decreases delay by the same ratio. However, when ω gets large enough so that the RC model deviates from the RLC model, increasing ω does not decrease the delay proportionately. Wire width, which results in a roughly 10% difference between RC and RLC models is an “optimal” value for ω that offers both small delay and high wiring density. Figure 17.7.4 shows that optimal wire width (ω_{opt}) does not depend on wire length. ω_{opt} occurs when RC delay is 1.33 times larger than time-of-flight (ToF) delay. Substituting this condition into (1) gives

$$\omega_{opt}=1.88c_0[k\rho\varepsilon_0R_0C_0]^{1/2}, \qquad (2)$$

where c_0 is the speed of light in vacuum.

Global clock frequency (f_c) *of* a SoC is usually determined by the largest interconnect delay. Therefore, the ideal value for f_c is the reciprocal of ToF delay of the longest interconnect. However, as Figure 17.7.4 illustrates, ToF delay for on-chip wires requires prohibitively wide (e.g. ω=5-10μm) wires and thick dielectrics.

Printed wiring board (PWB) wires usually have a large cross-sectional area and therefore, ToF delay. If the longest interconnects are routed off-chip through PWB "*exterconnects*", the maximum on-chip interconnect length is then reduced and as (1) shows, the maximum delay is reduced. If maximum on-chip interconnect length decreases so that maximum on-chip delay equals ToF delay of the longest interconnect, the maximum possible f_c is achieved. Hence, the ideal maximum on-chip interconnect length is calculated as

$$l_{max}=0.75(2D_{chip}), \qquad (3)$$

where D_{chip} is the chip edge dimension. All "interconnects" for $l_{max} < l < 2D_{chip}$ should be on-PWB *exterconnects*.

The total number of PWB layers and chip I/O leads required for maximum f_c of a SoC implemented in year 2011 are projected using a recently derived stochastic global net length distribution [3]. Figure 17.7.5 shows key parameters of the projected chip and net length distribution. Results are summarized in Figure 17.7.6. Four PWB layers and 1800 I/O leads for exterconnects increase f_c by 45%, while on-chip wiring density does not change.

Although more on-PWB layers and I/O pads may be available and more *exterconnects* could be used, f_c can not be increased due to ToF delay of corner-to-corner *exterconnects*. However, as (1) shows, by reducing the maximum on-chip interconnect length, ω can be reduced by the same ratio, while delay is kept constant. In other words, by using more PWB wires, on-chip wiring density can be improved without any performance loss.

Dealing with inductance of on-chip interconnects is a treacherous problem mainly because of unknown return paths and interference of far wires. However, by keeping on-chip wire width small, the impact of inductance is negligible, and the design process is much simpler. Remaining in the RC regime of interconnect delay on-chip results in better signal integrity. Also for PWB wires, because of existing ground planes, effects of inductance are well known and local. Hence, in this way, the problem of on-chip inductance is avoided without clock frequency reduction and SoC “designability” is greatly enhanced.

The area occupied by the on-chip power distribution network reduces significantly by using an adequate number of area array bonding pads [4]. Using low-loss on-PWB transmission lines for clock distribution enables reduced skew and jitter and permits recycling the energy of reflected pulses to reduce clock power dissipation significantly [5]. SoL also facilitates using high quality electrical and optical passive components embedded in PWB to enhance mixed signal SoC performance.

In short, SoL is disruptive because it intends to use ultra high density I/O leads to improve performance, cost, size, and reliability of an SoC.

References

1. C. S. Patel, et al., "Low Cost High Density Compliant Wafer Level Package," International Conference on High-Density Interconnect and Systems Packaging, Denver, Colorado, April 26-28, 2000, pp. 262-268.
2. C. S. Patel, et al., "Cost Analysis of Compliant Wafer Level Packages," 50th Electronic Components and Technology Conference (ECTC), Las Vegas, Nevada, May 22-24, 2000, pp. 1634-1639.
3. P. Zarkesh-Ha et al, "Stochastic Net length Distribution for Global Interconnects in a Heterogeneous System-on-a-Chip", Symposium on VLSI Technology Digest of Technical papers, June 1998, pp44-45.
4. P. Zarkesh-Ha et al, "An Integrated Architecture for Global Interconnects in a Gigascale System-on-a-Chip", Symposium on VLSI Technology Digest of Technical papers, June 2000, pp194-195.
5. P. Zarkesh-Ha et al, "Asymptotically Zero Power Dissipation Gigahertz Clock Distribution Networks" Conference on Electrical Performance of Electronic Packaging, October 23-25, 1999, pp. 57-59.

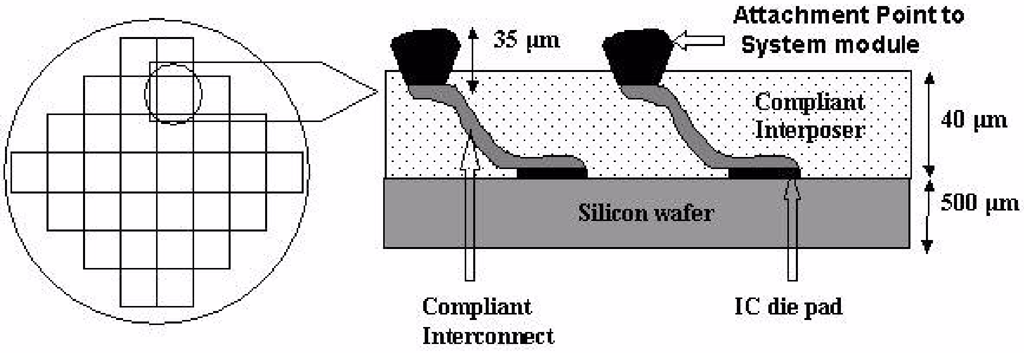


Figure 17.7.1: Sea of Leads (SoL) wafer level batch fabrication.

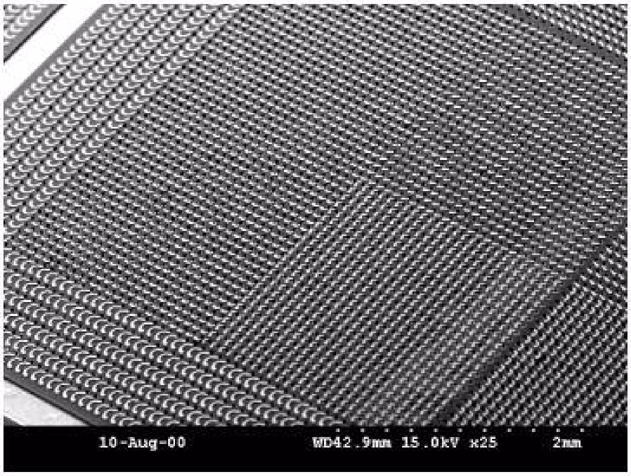


Figure 17.7.2: SEM micrograph of 12,000 compliant leads per 1 cm^2 Sea of Leads (SoL) die.

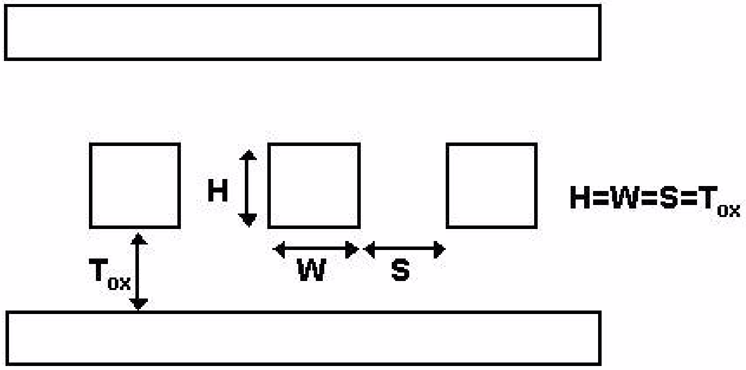


Figure 17.7.3: The cross-section of an interconnect and its neighbors with dimensions assumed for sample calculations.

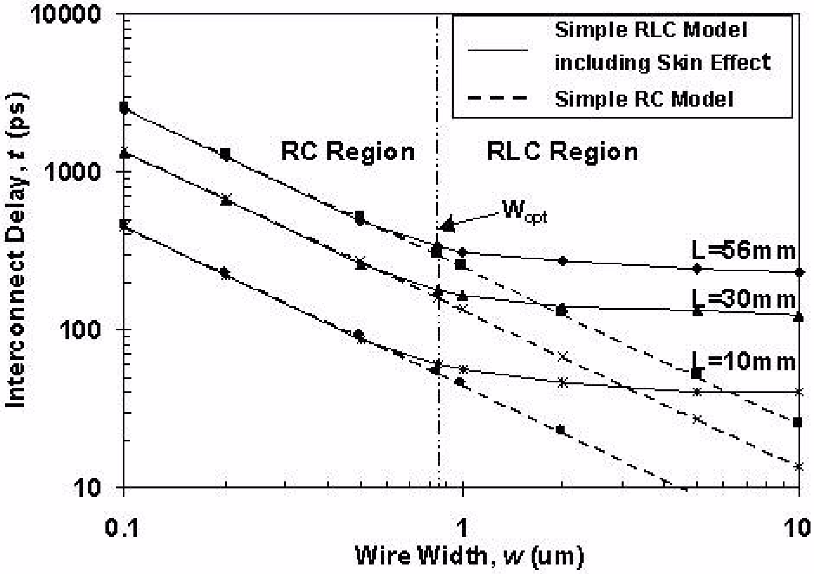


Figure 17.7.4: Interconnect delay verses width when the optimal number of repeaters is used.

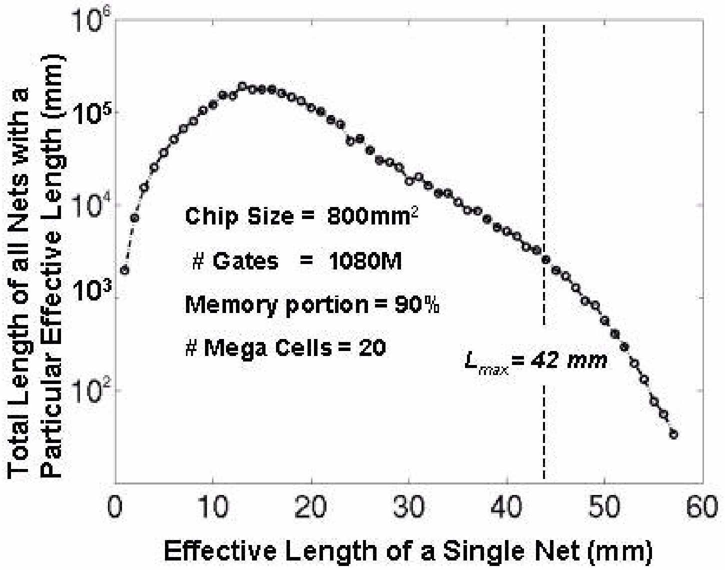
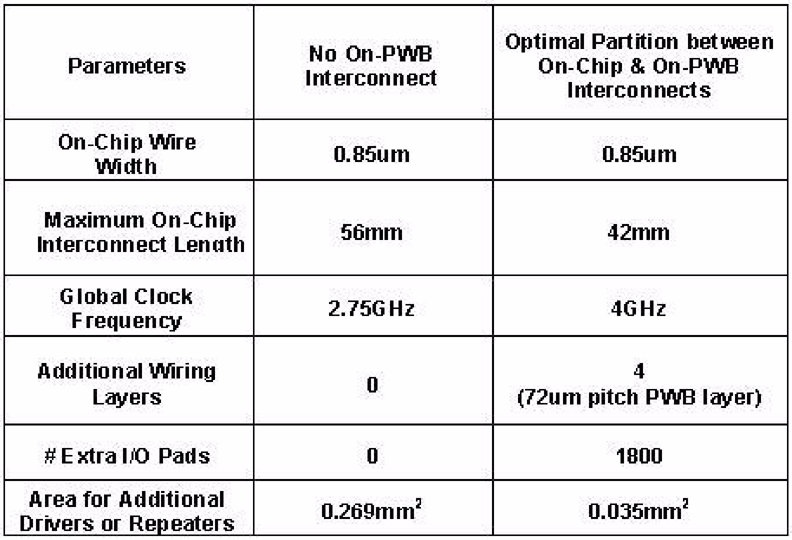


Figure 17.7.5: Effective global net-length distribution for a projected SoC in year 2011 based upon stochastic models for global net fan-out distribution, net boundary area and net length [3].



Parameters	No On-PWB Interconnect	Optimal Partition between On-Chip & On-PWB Interconnects
On-Chip Wire Width	0.85um	0.85um
Maximum On-Chip Interconnect Length	56mm	42mm
Global Clock Frequency	2.75GHz	4GHz
Additional Wiring Layers	0	4 (72um pitch PWB layer)
# Extra I/O Pads	0	1800
Area for Additional Drivers or Repeaters	0.269mm^2	0.035mm^2

Figure 17.7.6: Performance Improvement using exterconnects for the projected SoC in year 2011.