Performance Model for Inter-chip Communication Considering Inductive Cross-talk and Cost

Brock J. LaMeres
Department of Electrical and Computer Engineering
University of Colorado
Boulder, CO 80309
Email: brock_lameres@agilent.com

Sunil P. Khatri
Department of Electrical Engineering
Texas A&M University
College Station, TX 77843
Email: sunil@ee.tamu.edu

Abstract— We present an analytical method to perform the design of the I/O subsystem of an IC given its throughput requirements. Our method can be used to select the IC package, along with the bus size and speed so as to minimize I/O cost. We have validated our model by conducting simulations on three industry-standard packages while varying the bus width, slew rate, and signal-to-power/ground ratio. Our experimental results track closely with the analytical model. We demonstrate for the packages considered that it is more cost effective to use faster, narrower busses rather than slower wider busses to achieve a desired system throughput.

I. Introduction

Advances in CMOS technology have led to a dramatic increase in the on-chip performance of ICs. While the computational power of on-chip circuitry continues to grow, the inter-chip interconnect significantly limits the performance of digital systems [1], [2]. The core speed for today's ICs is many times faster than the speed of inter-chip busses. As a consequence, inter-chip bus design is becoming a very important challenge in digital system design. Simply widening I/O buses to increase the total bus throughput is not practical due to the high cost of each I/O pin. In addition, the electrical parasitics of standard packaging limits not only the per channel bandwidth, but also the total number of signals that can switch simultaneously. Due to all of these factors, inter-chip bus design requires a careful analysis of the cost versus performance tradeoff.

Traditionally, inter-chip communication is performed using wide parallel busses. The standard approach to achieving the desired system bandwidth is to increase the number of pins on the package until the desired throughput is attained. There are two main problems with this approach.

- Cost of packaging. Package cost scales faster than linearly with the number of I/O pins that are needed and accounts for a large contribution to the overall chip price [3].
- Performance. Wide parallel busses experience a host of signal integrity issues associated with simultaneous switching of digital signals [1], [4], [5]. Problems such as supply bounce, glitching, and signal edge degradation occur due to dynamic currents inducing unwanted voltages across the parasitic inductance in the package. Increasing the number of signals in an off-chip bus aggravates this problem.

In order to design the most effective bus, both the cost and simultaneous switching noise associated with adding pins must be considered. This paper presents an analytical model for selecting the width and packaging of the bus, based on its throughput requirement. Our approach considers the maximum data rate that a package can accommodate as the number of channels is increased. In addition, the cost of adding I/O pins is considered for three different Signal/Power/Ground (SPG) ratios – 8:1:1, 4:1:1, and 2:1:1. SPICE simulations are performed on three industry standard packages to validate the analytical model. Experimental results matched closely with our analytical predictions.

The rest of this paper is organized as follows. Section II describes the methodology used in constructing the analytical model including the variables considered and the failure mechanism. Section III presents the analytical model. Section IV presents the experimental results and cost analysis. Sections V presents the conclusion of this paper.

II. Methodology

In order to develop the analytical model, a typical CMOS driver/receiver circuit topology was used. This circuit topology was also used in the SPICE simulations to validate the model. In this topology, the following parameters were varied:

1) Number of Channels
2) Slew Rate
3) SPG Ratio
4) Package

A. Test Circuit

The circuit used to formulate the model and for simulations is shown in the Figure 1. We used the BPTM 0.1um [6] technology using BSIM3 model cards [7]. All simulations were done using SPICE [8]. A CMOS inverter was used to model the driver and the receiver load. The driver was designed to drive a 75Ω PCB trace which was 2" long with a drive strength of 25mA. The CMOS inverter had V_{DD}=1.5v and V_{SS}=0v. A series termination resistor was placed on the PCB at the output pins of the driving IC. The resistor value was chosen so that the cumulative output impedance of the resistor in series with the R_{ON} of the inverter is 75Ω. The optimal size of the inverter that can drive 25mA into a 75Ω, 2" long PCB trace that was series terminated with an equivalent output impedance of 75Ω is W_N=80nm and W_P=260nm. The inverter is sized to have an equal drive strength on both the PMOS and NMOS transistors by using $(\frac{W_P}{W_N}) = (\frac{u_n}{u_p}) = 3.25$ [4].

The package model included the self RLC of the leads and wire bonds (if used). The model included coupling capacitance out to the nearest two adjacent signals for both the package leads and the wire bonds. The mutual inductance of the leads and wire bonds was considered out to the nearest 5 signals. Coupling was not considered on the PCB since the geometries on the PCB are such that coupling can be and often is eliminated with trace spacing [5], [9].

B. Failure Condition

In our model, a failure was defined as ground bounce (or V_{DD} droop) that had a magnitude greater than 5% of the V_{DD} supply. The magnitude of the ground bounce was measured on the die of the driver. The worst case ground bounce was present when all of the CMOS inverter outputs switched from a logic 1 to a logic 0 at the same time. This failure mechanism only accounts for the magnitude of the ground bounce. Other limitations such as delay and signal shape were not considered.

0-7803-8834-8/05/$20.00 ©2005 IEEE.

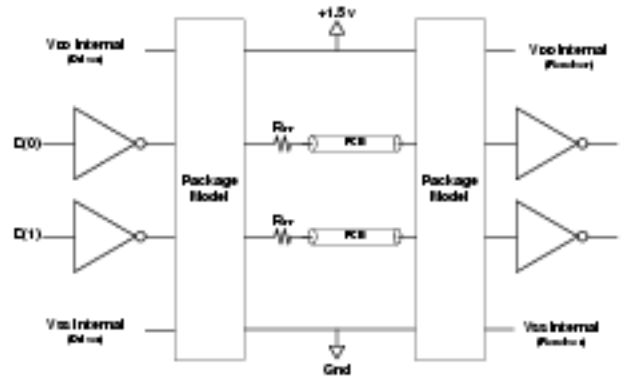

Fig. 1. Test Circuit Used to Analyzer Bus Configuration (showing a 2:1:1 SPG Ratio)

C. Ground Bounce

There are two factors that contribute to ground bounce. The first component is due to the voltage induced across the self inductance of the V_{SS} pin of the driver. This voltage follows the relationship:

$$V_1 = L_{11}\frac{di_1}{dt} \tag{1}$$

The AC current (i_1) in this expression is the cumulative drain current of the CMOS inverters as they transition in the same direction. This current is directly proportional to the number of inverters that switch. It is lowered when the SPG value has fewer signal pins. The subscripts on V_1 and i_1 represent the fact that voltage on the ground pin is caused by the current through the same ground pin. This current induces a voltage across the self inductance of the pin (L_{11}).

The second component is due to the mutual inductance from neighboring signal pins. This contribution follows the relationship:

$$V_1 = M_{1k}\frac{di_k}{dt} \tag{2}$$

For this type of contribution to ground bounce, the voltage (V_1) induced on the ground pin is caused by the mutual inductive coupling from adjacent signal pins that are transitioning. The subscript k represents an arbitrary neighboring pin that is k pins away from the V_{SS} pin. The current i_k in this k^{th} neighboring pin induces a voltage across the mutual inductance M_{1k} of the ground pin and the k^{th} neighbor.

D. Slew Rate

$\frac{di}{dt}$ is proportional to the slew rate of the bus signals. As the slew rate increases, the amount of time it takes for the charging and discharging of the load decreases which increases the data rate at which the bus can operate. This also means that as the slew rate gets faster, the more ground bounce will be present and thus limit the maximum data rate that the bus can run at.

The slew rate $\frac{dv}{dt}$ can be found as follows:

$$slewrate = \frac{dv}{dt} = \frac{di}{dt}\cdot Z_{load} \tag{3}$$

The rise time of the signal is defined as the time it takes to switch from 10% to 90% of the DC output value (80% of V_{DD}).

$$t_{rise} = \frac{0.8\cdot V_{DD}}{slewrate} \tag{4}$$

The rise time can then be used to define the minimum Unit Interval (UI) that can used in a robust digital system [2], [9], [5]:

$$UI_{\min} = (1.5)\cdot(t_{rise}) \tag{5}$$

The $UI_{\min}$ defines the minimum duration of the data valid window in order to transmit a logic symbol successfully. This corresponds to the maximum data rate of a signal as follows [2], [9], [5]:

$$DR_{max} = \frac{1}{UI_{\min}} \tag{6}$$

E. Packaging

The package selection dictates the magnitudes of the electrical parasitics present in the inter-chip bus. Packages traditionally add a large inductive component to the I/O system. This inductive component results in ground (and supply) bounce (equations 1 and 2). As package technology advances, the electrical parasitics are reduced [1]. However, these advanced packages add to the overall cost of the IC [3]. In this paper, we study three industry packages, the QFP wire bond, BGA wire bond, and the BGA flip chip [10], [11], [12], [3]. Table 1 describes the characteristics for the three packages that were studied in this paper [3].

Package:	L_{11}	K_{12}	K_{13}	K_{14}	K_{15}	K_{16}	Cost per pin
QFP-wb	4.350n	0.744	0.477	0.352	0.283	0.263	$0.22
BGA-wb	3.766n	0.537	0.169	0.123	0.097	0.078	$0.34
BGA-fc	1.344n	0.630	0.287	0.230	0.200	0.175	$0.63

TABLE 1
CHARACTERISTICS FOR STUDIED PACKAGES

III. Analytical Model

A. Performance of the Bus

This section presents an analytical model that describes the maximum data rate per unit cost for an inter-chip bus considering the magnitude of ground bounce on the IC as the failure condition.

Using equations 1 and 2, the net ground bounce of a bus can be expressed as:

$$V_{gnd-bnc} = (\frac{W_{bus}\cdot L_{11}}{N_g})(\frac{di}{dt}) + \sum_{k=2}^{W_{bus}}(M_{1k})(\frac{di}{dt}) \tag{7}$$

In this expression, W_{bus} is the number of signals in the bus. For this model, it is assumed that all of the signal in the bus are transitioning in the same direction to represent the worst case ground bounce situation. N_g is the number of ground pins in the bus and is dictated by the SPG ratio that is selected. Increasing the number of grounds effectively reduces the inductance of the ground path. i is the current in any pin.

$V_{gnd-bnc}$ is set to an acceptable magnitude ($p\cdot V_{DD}$, where $p<1$) depending on the desired noise margin for the bus. Therefore the maximum slew rate achievable for an inter-chip bus is:

$$(\frac{dv}{dt})_{max} = \frac{p\cdot V_{DD}\cdot Z_{load}}{(\frac{W_{bus}\cdot L_{11}}{N_g}) + \sum_{k=2}^{W_{bus}} M_{1k}} \tag{8}$$

From equation 4, we get the minimum tolerable rise time as:

$$t_{rise-\min} = \frac{(0.8)\cdot[(\frac{W_{bus}\cdot L_{11}}{N_g}) + \sum_{k=2}^{W_{bus}} M_{1k}]}{p\cdot Z_{load}} \tag{9}$$

The minimum Unit Interval can be computed by combining equations 5 and 9. We can then use equation 6 to get an expression for the maximum per-pin data rate DR_{max}that can be achieved:

$$DR_{max} = \frac{p\cdot Z_{load}}{(1.5)\cdot(0.8)\cdot[(\frac{W_{bus}\cdot L_{11}}{N_g}) + \sum_{k=2}^{W_{bus}} M_{1k}]} \tag{10}$$

The total system throughput TP of the bus can now be expressed as $TP = DR_{max}\cdot W_{bus}$.

B. Cost-Effectiveness of Bus

We now formulate a method to analyze the cost-effectiveness of the bus design. The following expression defines the number of I/O pins needed to implement an inter-chip bus of width W_{bus} with an equal number of V_{SS} and V_{DD} pins set by the SPG ratio:

$$N_{I/O} = W_{bus} + 2 \cdot \lceil (\frac{W_{bus}}{SPR}) \rceil \quad (11)$$

In this expression, SPR refers to the Signal/Power/Ground ratio. For example, if SPG = k:1:1, then $SPR = k$. The cost of the inter-chip bus is given by:

$$Cost_{bus} = (N_{I/O}) \cdot (Cost_{per-pin}) \quad (12)$$

where the $Cost_{per-pin}$ will vary depending on which package is selected (see Table I).

Finally, we define a cost effectiveness metric for any bus configuration called Bandwidth-per-Cost (BPC). This metric has units ($\frac{Mb}{s}$) and takes into account the total bus throughput for a given inductive noise margin as well as the I/O cost including the number of the power and ground pins.

$$BPC = (\frac{TP}{Cost_{bus}}) \quad (13)$$

IV. Experimental Results

Using the methodology outlined in section II, we simulated the test circuit and compared the results with the analytical model. Figures 2 through 4 show the maximum data rate per-pin (DR_{max}) for the three packages studied (QFP wirebond, BGA wirebond, and BGA flip-chip) as a function of the number of simultaneously switching channels[1]. Both the simulation and analytical model data are displayed. These results illustrate that as the number of simultaneously switching channels is increased, the per-pin data rate is decreased. They also show how advanced packaging such as flip-chip technology reduces the inductive cross-talk which in turn increases the per-pin data rate. In addition, it shows that the effect of adding more grounds can increase the per-pin data rate by reducing the self and mutual inductance in the ground path.

Figures 5 through 7 show the total throughput (TP) of the bus for the same three packages. This figure shows that the system throughput actually approaches an asymptotic limit as more channels are added to the bus. This is due to the fact that adding more channels to the bus actually degrades the speed at which each individual channel can switch. *The linear increase expected by adding additional I/O is negated due to the dramatic decrease in per-pin performance due to the package parasitics.*

All of the packages that were analyzed reached an asymptotic limit in total throughput as the width of the bus was increased. In all cases this was due to the ground bounce failure mechanism decreasing the maximum data rate per-pin at a rate that was similar to the increase in the throughput achieved by adding channels. This indicates that after the failure mechanism begins to dominate the per-pin performance, simply adding I/O to the bus does not increase system throughput. A more thorough analysis of this should include the cost of the bus.

The metric introduced in Equation 13 represents the cost-effectiveness of an inter-chip bus. This metric considers the SPR in the cost of the I/O, providing insight into the most cost-effective bus configuration.

Table II shows the BPC for the three different packages. This table illustrates that it is more cost-effective to use busses that are narrower and faster rather than expanding the bus which actually decreases the data rate per-pin. Also, it indicates that the flip-chip package is more cost-effective in general, especially with lower values of SPR.

Suppose we are given a value of total throughput desired from the bus. We would refer to Figures 5 through 7, and find the package and SPR configurations which meet this throughput. Then, we would use Table II to select the most cost-effective solution from the candidate configurations.

[1] A channel is simply a signal pin.

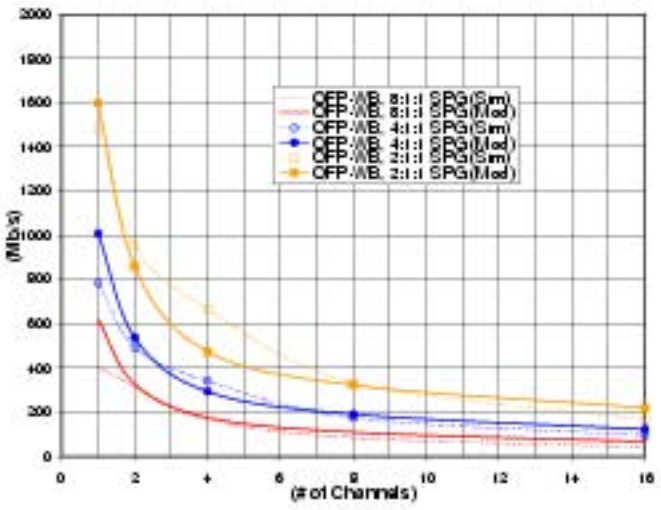

Fig. 2. Maximum Data Rate Per-Pin for a QFP Wire Bonded Package.

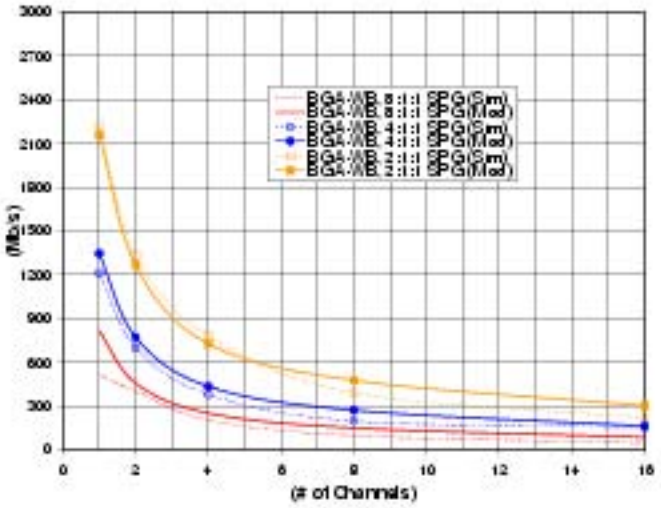

Fig. 3. Maximum Data Rate Per-Pin for a BGA Wire Bonded Package.

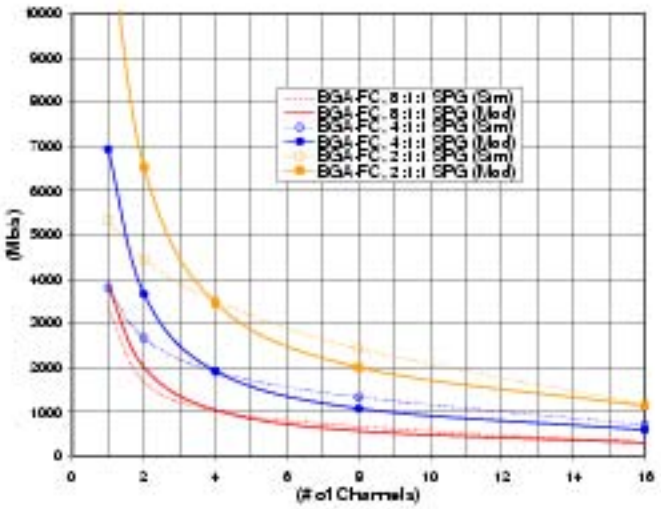

Fig. 4. Maximum Data Rate Per-Pin for a BGA Flip-Chip Package.

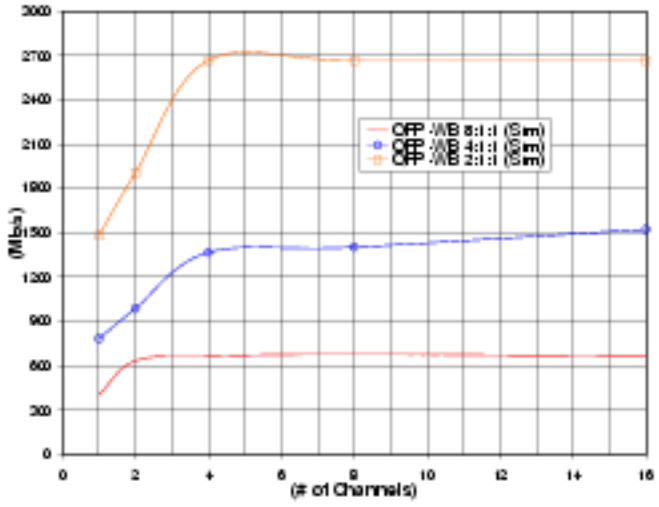

Fig. 5. Total System Throughput for a QFP Wire Bonded Package.

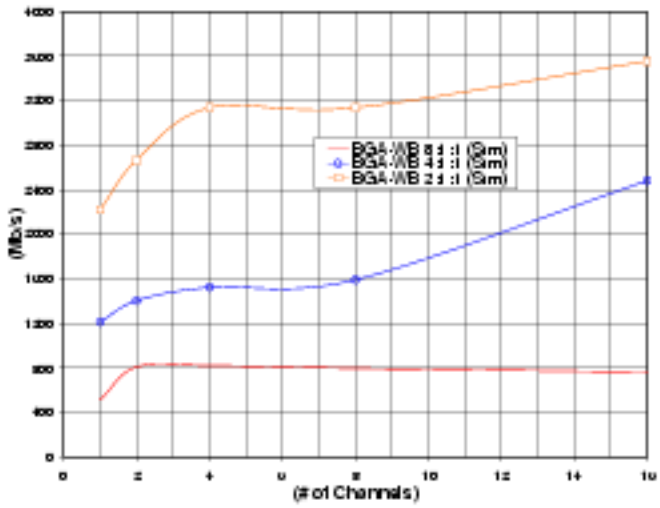

Fig. 6. Total System Throughput for a BGA Wire Bonded Package.

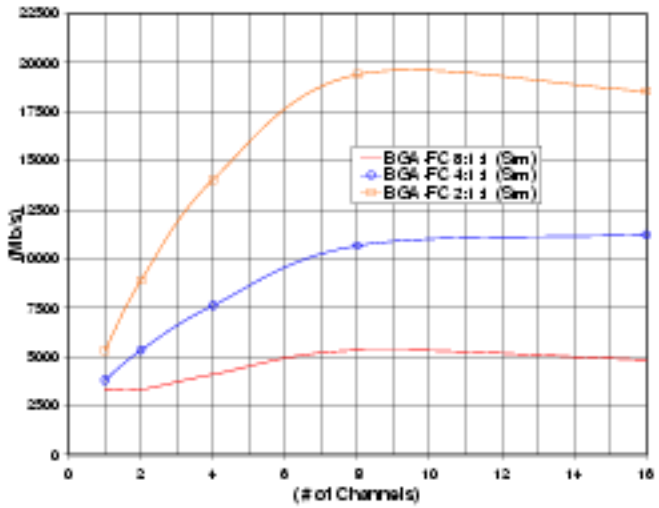

Fig. 7. Total System Throughput for a BGA Flip-Chip Package.

TABLE II

BPC OF DIFFERENT BUS CONFIGURATIONS (Mb/$)

	Number of Channels				
Bus Configuration	1	2	4	8	16
QFP-WB 8:1:1	612	722	505	309	152
QFP-WB 4:1:1	1188	1122	1036	532	289
QFP-WB 2:1:1	2345	2165	1515	758	379
BGA-WB 8:1:1	503	594	402	234	112
BGA-WB 4:1:1	1188	1032	747	390	304
BGA-WB 2:1:1	2179	1961	1153	577	327
BGA-FC 8:1:1	1764	1323	1085	847	385
BGA-FC 4:1:1	2016	2116	2016	1411	743
BGA-FC 2:1:1	2822	3527	2785	1924	920

V. Conclusion

In this paper we presented an analytical performance model for inter-chip bus design. Our model considered performance and cost as the number of channels, grounding scheme, and various packaging options were explored. We demonstrated that the maximum data rate per-pin decreased significantly as the number of simultaneously switching channels was increased. This shows that *simple expansion of an inter-chip bus does not yield a linear increase in the throughput* of the system as one would expect. It was also shown that the *total system throughput reached an asymptotic limit as the number of channels was increased*. This means that the same throughput can be achieved by using faster narrower busses rather than a traditional wider and slower bus design.

A cost analysis was also performed which considered various packaging and grounding schemes. A new Bandwidth per Unit Cost (BPC) metric was defined as a means to evaluate the most cost-effective bus configuration. It was found that the most cost-effective bus was faster and narrower rather than slower and wider. By running the individual channel near its theoretical maximum data rate (i.e., with no mutual inductive coupling), a cost advantage is achieved because additional I/O are not needed to obtain the desired system throughput. The BGA Flip-Chip package was found to be the most cost effective. Even though the cost per channel is higher for this advanced style of package, the increased bandwidth far outweighs the cost increase when considering BPC.

The technique presented in this paper to analyze the cost-effectiveness of a bus configuration (considering cost, package, and grounding schemes) can be applied to any style of packaging.

References

[1] M. Horowitz, C. Yang, and S. Sidiropoulos, "High-Speed Electrical Signaling: Overview and Limitations," *IEEE Micro.*, vol. 18, pp. 12–24, Jan 1998.

[2] C. Kimmkid, "Standards are key to optimizing high-speed data bus communications," *Planet Analog (planetanalog.com)*, Oct 2002.

[3] Agilent Packaging Group, Ft. Collings, CO. Personal Communication, 2004.

[4] W. Dally and J. Poulton, *Digital Systems Engineering*. Cambridge, U.K.: Cambridge University Press, 1998.

[5] H. Johnson and M. Graham, *High-Speed Digital Design*. Prentice Hall PTR, 2003.

[6] "BPTM Homepage." www-device.eecs.berkeley.edu/~ptm/.

[7] "BSIM3 Homepage." http://www-device.eecs.berkeley.edu/~bsim3/.

[8] L. Nagel, "SPICE: A Computer Program to Simulate Computer Circuits," in *University of California, Berkeley UCB/ERL Memo M520*, May 1995.

[9] H. Johnson and M. Graham, *High-Speed Signal Propagation*. Prentice Hall PTR, 2003.

[10] ASAT Inc., "Application Note: Peak Performance Enhanced Lead Packages." www.asat.com.

[11] ASAT Inc., "Application Note: Peak Performance Array Standard, HS fpBGA Packages." www.asat.com.

[12] ASAT Inc., "Application Note: Peak Performance Flip Chip Packages." www.asat.com.