

Implementation Challenges and Solutions of Low-Power, High-Performance Memory Systems

Implementation Challenges and Solutions of Low-Power, High-Performance Memory Systems

Mobile devices and their demand for rapid innovation have fundamentally and forever changed the semiconductor industry. These devices have fueled tremendous innovation in the last few years to bring about drastic improvements in performance, power and cost efficiency. They also demand condensed product development cycles which accelerate the rate and need for innovation. The only thing that has remained the same is our industry's ability to drive innovation, meet new technical challenges and develop new processes and technologies.

This is the first of a two-part whitepaper that examines the key factors that are driving new requirements in mobile memory system design and analysis. Trends include:

- Memory speed rising rapidly to keep pace with processor performance
- Energy consumed per bit of transmission decreasing proportionally
- Limited footprint forcing adoption of package-on-package and other package-level integration technology
- Commoditization pushing growth from the high to the lower end of the market

When translated into electrical requirements, the trends in mobile memory pose tremendous challenges in signal and power integrity for SoC and system developers, including:

- Decreasing timing and voltage margin
- Escalating circuit jitter sensitivity to supply noise
- Rising supply noise
- Increasing signal and supply noise coupling
- Growing signal integrity degradations due to manufacturing variations

Memory Solutions Trends

Mobile processor performance has been rapidly rising, which is evident when considering the advance of mobile processors from single core at <500MHz (the original iPhone) to quad cores at close to 2GHz (Galaxy S4, US edition) in just six years. This makes it challenging to develop memory solutions that can keep pace and fully utilize the power of the processor.

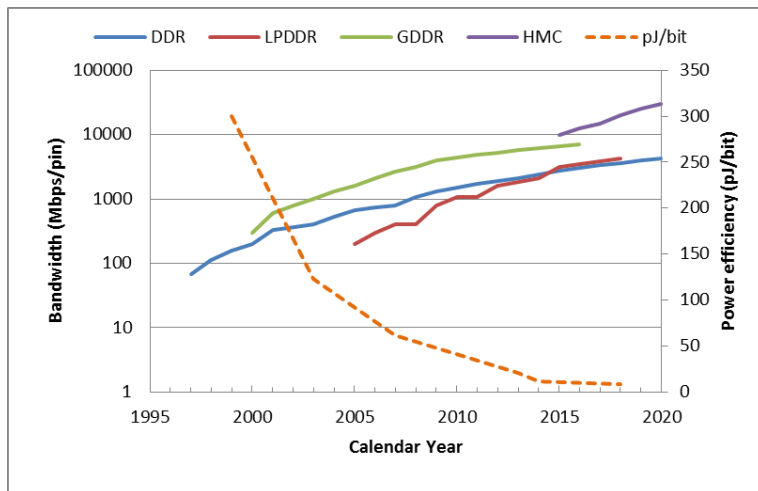


Figure 1 shows the historical bandwidth trends of three mainstream memory solutions – DDR, LPDDR and GDDR – and one emerging solution, HMC (Hybrid Memory Cube).

While it took 13 years (2000-2013) for DDR bandwidth to increase an order of magnitude from 200 to 2133Mbps, LPDDR bandwidth, driven by demand of mobile processors, experienced the same increase in only nine years (2005-2014).

Figure 1. Historical Memory Bandwidth Trends. Source: Internal Analysis

Designers look for solutions that enable memory bandwidth to rapidly trend up. At the same time, they contend with power requirements that remain relatively unchanged, including;

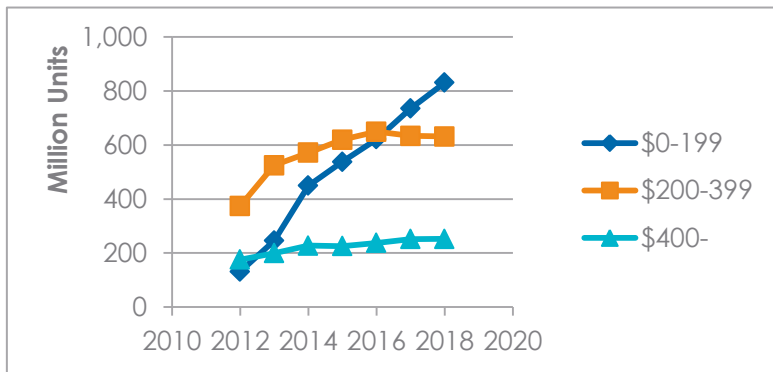
- Thermal envelope — tolerance of the human body does not change
- Power density of the IC
- Battery size

Figure 1 reveals that in the same period LPDDR bandwidth increased by an order of magnitude from 2005 to 2014, the general trend in energy per bit decreased roughly by an order of magnitude¹, resulting in relatively constant power.

In addition to delivering higher bandwidth at the same low power, mobile device developers must contend with low profiles and limited real estate available for adding parts. Taken together, it forces the adoption of more integration to expand functionality. This trend is directly related to the adoption of package-on-package (PoP) technology, especially by high-end smartphones. In addition to presenting some unique electrical challenges, PoP moves the memory interface out-of-sight, which significantly reduces the ability to observe one of the highest speed data transmission channels in the system.

In contrast, commoditization in the mobile market is pushing growth down from the higher end to the lower-end market. Figure 2 shows the projected smartphone unit shipment growth by price range up to 2018 [1]. The CAGR for unit price ranges of \$0-199, \$200-399 and \$400+ are 36%, 9% and 6%, respectively. Therefore, while the growth of high-end and, to a lesser extent, mid-range smartphones is tapering, low-end smartphones will experience tremendous growth in the next 5 years.

¹ The energy per bit curve in Figure 1 indicates a general trend and is not specific to LPDDR evolution.



The growth of the lower-end smartphones, together with the highly-competitive nature of the market, is putting downward pressure on costs. As a result, the same high-bandwidth memory solutions used in high-end smartphones today must be delivered by the mid-range and eventually low-end devices of tomorrow, albeit at lower implementation costs.

Figure 2. Smartphone Unit Shipment Growth by Price Range. Source: ABI Research

Signal and Power Integrity Challenges

Trends in mobile memory pose tremendous challenges in signal and power integrity to SoC and system developers. Higher data rates lead to smaller bit time and faster signal transitions. The former results in a smaller timing budget while the latter creates larger signal degradations due to loss and reflection. Together they significantly reduce system timing margin. In addition, lower energy per bit requires lower supply voltages and smaller signal swings, both leading to smaller system voltage margin.

To minimize power consumption, low-power, high-performance memory systems utilize aggressive power management as well as circuits optimized for power. Aggressive power management creates frequent power state transitions increasing both the magnitude and probability of worst case supply noise. On the other hand, circuit optimization for power is often at the expense of increased timing jitter sensitivity to supply noise. Together, they further exacerbate the problem of shrinking timing and voltage margin.

Table 1 shows the trend of shrinking timing and voltage margins as LPDDR data rate increases. Data signal (DQ) margin of read operation is used for illustration purposes. "Timing Margin for Rest of Channel" and "DRAM Output Swing" provides a measure of timing and voltage margins respectively available to the controller receiver, package and PCB after accounting for the DRAM specifications.

	Data Rate (Mbps)	Bit Time (ps)	DRAM max tDQSQ (ps)	DRAM min tQH (ps)	Timing Margin for Rest of Channel ¹ (ps)	DRAM Output Swing ² (mV)
LPDDR2	200	5000	700	2800	2100	96
LPDDR2	800	1250	240	670	430	960
LPDDR3	1600	625	135	475	340	800
LPDDR3	2133	468.8	100	356.4	256.4	690
LPDDR4	3200	312.5	TBD	TBD	TBD	350-400

Table 1: DQ Timing and Voltage Margin vs LPDDR Data Rate

¹ Timing Margin for Rest of Channel = (Bit Time) – [tDQSQ + (Bit Time – tQH)]

² DRAM Output Swing assumes ODT = 240Ω at 1600Gbps, 120Ω at 2133Gbps

Diminishing device footprint and reliance on low-cost packaging solutions to meet optimal price points for mid-range mobile devices raises several challenges that mobile device developers must contend with. More functionality is now packed into smaller form factors and demands increasingly tight integration. This results in declining pin pitch and spacing, translating into more crosstalk and supply noise coupling. Such degradations are further magnified by the use of low-cost packaging solutions driven by system cost constraints. This is especially true in the use of wire-bond packaging for the DRAM, which has much higher self and mutual inductance as compared to flip-chip packaging.

Another challenge posed by packaging is found in high-end mobile devices where small form factors and tight integration leads to the adoption of PoP packaging. Figure 3 shows the typical construction of a PoP application processor. While the upper package houses a mix of DRAM and flash dice connected to the substrate with bond wires, the lower package carries the application processor flip-chip connected to the substrate.

PoP packaging poses some unique electrical challenges. First, power distribution to the upper memory package must flow through the lower application processor package, where both packages have inferior electrical properties due to narrow inductive traces and insufficient signal reference planes. This results in both large supply noise and large supply noise coupling. Secondly, to create a cavity to accommodate the lower application processor chip, pin placement on the upper package is constrained to the periphery, limiting the number of pins available. As the application processor die grows in size to expand functionality, or the number of memory signal pins increases to support higher memory bandwidth, the package size must grow or the pin pitch must be reduced, resulting in increasing cost as well as coupling.

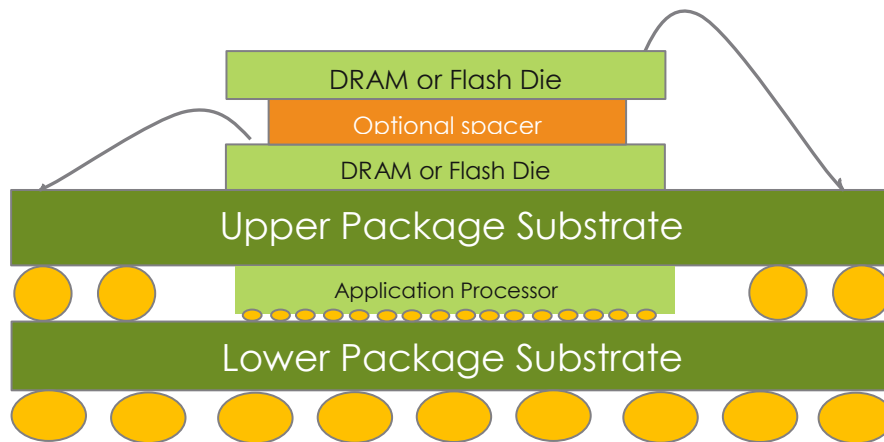
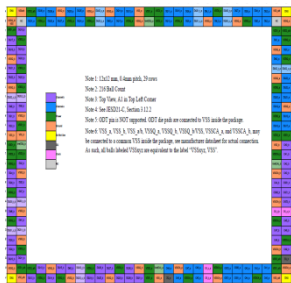


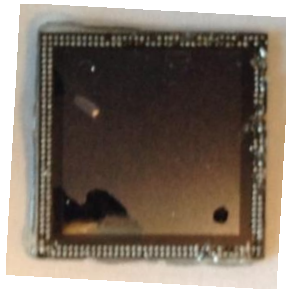
Figure 3. PoP Construction and Footprint

Figure 4 depicts a LPDDR3 footprint comparison between three implementations:

- 1) JEDEC specification
- 2) in Samsung Galaxy S4
- 3) in Apple iPhone 5S



1) JEDEC Standard Footprint
12x12mm
0.4mm pitch
216 BGA



2) Samsung Exynos 5 Octa found in Galaxy S4
14x14mm
0.35mm pitch
296 BGA

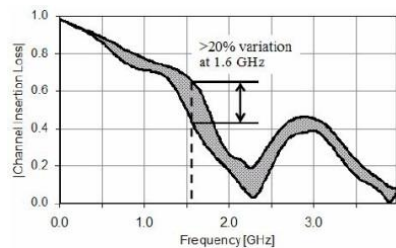


3) Apple A7 found in iPhone 5S
14x15.5mm
0.35mm pitch
456 BGA

To overcome the challenges of using PoP for application processors, designers of both Samsung and Apple devices chose to use a bigger package, a smaller pin pitch, and a higher pin count than JEDEC specification to implement 2 x32 channels of LPDDR3 rated for 1600Mbps operation,.

Mobile devices geared toward the lower end of the market have another set of challenges that mobile designers must address. These challenges are mainly due to commoditization trends and downward cost pressures leading to the adoption of low-cost manufacturing with less process control and larger manufacturing variations. This, in turn, results in increasing variations in system margin.

Figure 5 depicts the variation in passive insertion loss of a memory channel due to manufacturing variations of the controller and DRAM packages, as well as PCB [2]. For 3.2Gbps operation with a Nyquist frequency at 1.6GHz, the variation in insertion loss is up to 20%, which, based on simulation of this particular channel, translates into timing jitter variations of more than 0.2UI or bit time.



The gradual introduction of sophisticated techniques to standards-based memory solutions to address these challenges is a testimony to their increasing significance. For example, optional termination and some data and command bus training have been introduced in LPDDR3 to improve signal integrity and reduce static timing offsets, while additional training and internal Vref calibration are expected to be introduced in LPDDR4 to reduce additional offsets to improve timing and voltage margin.

Figure 5. Insertion Loss Variations due to Manufacturing Variations

Design Analysis for Power and Performance – Challenges and Solutions

Signal and power integrity challenges are prominent in the implementation of low-power, high-performance memory systems. Design analysis must evolve to a more holistic approach in order to better manage shrinking timing and voltage margins. The conventional waterfall design approach where different parts of the system are designed in a sequential manner is no longer acceptable. In optimizing the cost/performance of an

● ● ●
Design analysis must evolve to a more holistic approach in order to better manage shrinking timing and voltage margins.

upstream component such as a chip, too much margin may be allocated to it to reduce cost such that little is left for the components downstream, which can drastically increase their costs. To circumvent this, the entire system consisting of chips, packages and PCBs should be designed in parallel so that the different components can be optimized concurrently to guarantee robust systems at reasonable costs. The implication is that traditionally disparate EDA tools which have been separately tailored towards chip, package and PCB design must converge to provide a concurrent design platform with feedback paths to account for mutual interactions. For example, PCB layout optimization may require re-optimization of the package pin assignments, which, in turn, will require re-optimization of the package layout.

Moreover, an efficient simulation methodology is required to enable fast yet accurate analysis of the entire system both to drive design optimization and for verification. The challenge is not only in being able to handle a large and complex model encompassing all parts of the system, but also in accurately modeling design features with dimensions spanning orders of magnitude from deep sub-micron for chips to millimeter for PCBs.

One solution is to use a divide-and-conquer approach where the same system model is optimized in different ways based on the analysis performed. For example, thorough power integrity analysis includes simulation of static IR drop as well as medium and high frequency AC noise. Since the chip design dominates both static IR drop and high frequency AC noise, the package and PCB models can be simplified for these simulations. On the other hand, since medium frequency AC noise is dominated by the package and PCB, the chip model can be simplified in its simulation.

● ● ●
To avoid this pitfall, a statistical approach needs to be introduced into the analysis methodology to minimize pessimism while accounting for worst-case variations.

As a corollary to the co-design requirement, signal and power integrity optimization must occur at every stage of the design cycle. Otherwise, later stage optimization can become costly and it may be necessary to iterate the design, significantly increasing time-to-market. To address this challenge, intelligence must be built into design tools to enable SI/PI analysis even before the design takes shape. For example, the assignments of flip-chip bumps can have a large impact on current distribution among bumps and thus both power supply IR drop and electro-migration. However, most off-the-shelf PI analysis tools require a layout before analysis can be performed. Therefore, after initial bump assignments are made, the designer must create a preliminary layout before IR drop can be estimated. If IR drop is then determined to be excessive, correction may require bump assignment re-optimization and hence iteration of layout, leading to increased design time.

● ● ●

A second implication of the shrinking timing and voltage margins is that the conventional budgeting and analysis approach where the worst case values of different components are combined with no probabilistic consideration to determine final system margins is also inadequate. Such an approach can either result in a large negative budget, or unrealistically tight component specifications, resulting in over-designed and expensive systems. To avoid this pitfall, a statistical approach needs to be introduced into the analysis methodology to minimize pessimism while accounting for worst-case variations. In other words, a statistical

model should be created for each parameter, which is then analyzed in conjunction with the models for the other parameters to determine the statistical distribution of timing and voltage margins of the system. The robustness of the system can then be quantified by computing the probability of the system having non-zero margins.

Figure 6 shows a comparison in simulated timing jitter of a 3.2Gbps memory system [3]. In worst-case simulation, each parameter in the system model can take on its low, typical and high value with equal probability. A full factorial simulation is run using every possible combination of parametric values and the resultant timing jitter distribution extracted. The statistical simulation, on the other hand, uses a linear regression model based on the Taguchi Design of Experiment method and uses more realistic probability distributions for the model parameters in the simulation. The worst-case approach predicts a 6- σ timing jitter of >173ps, which is more than 15% larger than that from the statistical simulation.

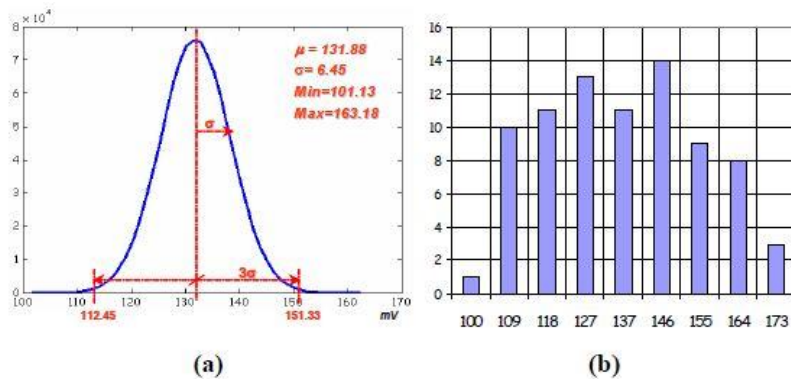


Figure 6. a) Statistical vs. b) Worst-Case Timing Jitter Simulation

Conclusion

As mobile devices continue their rapid trajectory towards increased performance and power efficiencies that are packed into shrinking form factors and manufactured at low costs, memory system analysis methodologies must evolve to address signal and power integrity challenges in terms of:

- Decreasing timing and voltage margins
- Escalating circuit jitter sensitivity to supply noise
- Rising supply noise
- Increasing signal and supply noise coupling
- Growing signal integrity degradations due to manufacturing variations

To combat shrinking timing and voltage margins, more holistic and statistical approaches to design and analysis must be adopted to optimize design for system costs, minimize design iterations and shorten time-to-market. As a result, signal and power integrity analysis must start earlier in the design cycle and chips, and packages and PCBs should be designed in parallel and analyzed statistically.

Part 2 of this whitepaper will examine the challenges and solutions of validating low-power, high-performance memory systems, including:

- Growing manufacturing variations due to cost constraints
- Increasing significance of long term impact of circuit random timing jitter
- Difficulty in extrapolating measurement results from probing to account for manufacturing variations and long term jitter
- Inability of pass or fail functional testing to quantify system robustness

References

[1] ABI Research, *ABI(MD-MDMT-159)_Mobile Device Shipments_08Jan14.xlsm*

[2] Yip, Wai-Yeung, Scott Best, Wendemagegnehu Beyene, Ralf Schmitt, "System Co-design and Co-Analysis Approach to Implementing the XDR Memory System of the Cell Broadband Engine Processor," ASP-DAC 2007

[3] Beyene, Wendemagegnehu, Newton Cheng, June Feng, Chuck Yuan, "Statistical and Sensitivity Analysis of Voltage and Timing Budgets of Multi-Gigabit Interconnect Systems," DesignCon 2004