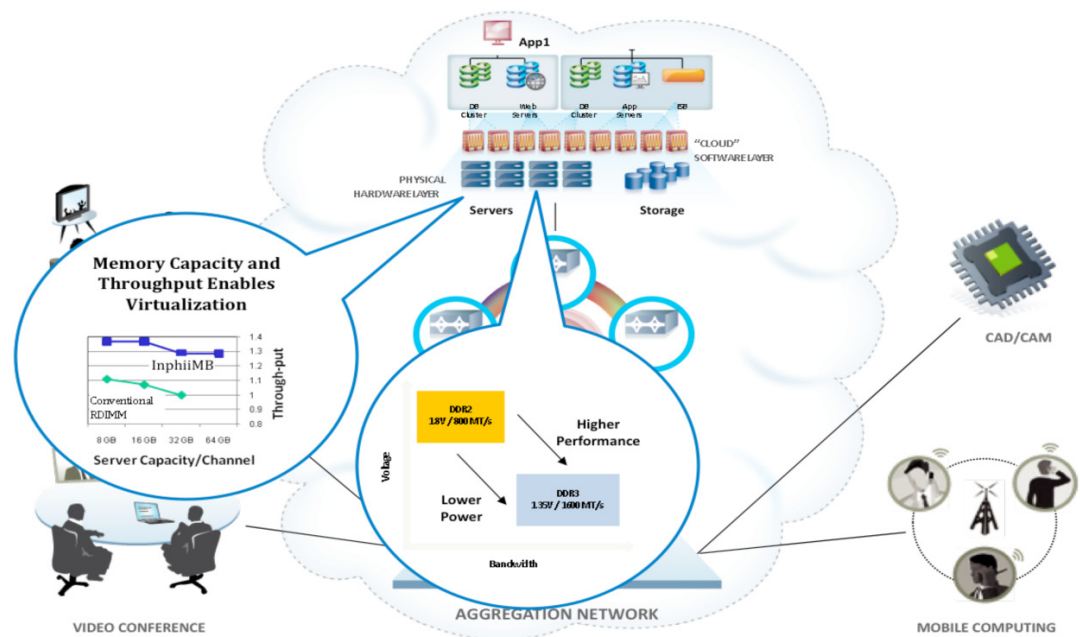


## Enabling Cloud Computing and Server Virtualization with Improved Power Efficiency

### I. Enabling cloud computing and server virtualization without power penalties

Enterprise servers are the workhorses of today's data centers, running enterprise applications such as DBMS, CRM, ERP and mail services; executing database queries; and powering search engine operations. In addition, virtualized servers are key to cloud computing. All these functions, of both physical and virtual servers, require massive amounts of memory, as well as bandwidth.



However, enterprise servers are challenged today, more than ever, to meet the growing demands for memory capacity and available bandwidth. Scaling server memory technology to higher capacity and bandwidth requires pushing beyond existing technological capabilities, forcing IT and data-center administrators to make trade-offs that prevent optimization.

As server systems push to higher speeds, the ability to drive more memory becomes a challenge due to increased memory loading.

To minimize or eliminate these issues, Inphi Corp. developed a new memory-buffer architecture – called Isolation Memory Buffer (iMB™) technology – that isolates the memory from the CPU. iMB technology can scale the server memory modules’ capacities 2x to 4x while operating at much higher frequencies. This memory scaling is accomplished by using iMB technology in a memory module called as Load-Reduced DIMM (LR-DIMM).

Data centers using iMB technology to scale memory capacity and bandwidth in their servers can increase the utilization of future server base. Accomplishing existing capacities with smaller number of DIMMs or Servers not only reduces the procurement cost of the data-center infrastructure, but also helps to lower the overall powering and cooling cost –thereby reducing the total cost of ownership (TCO).

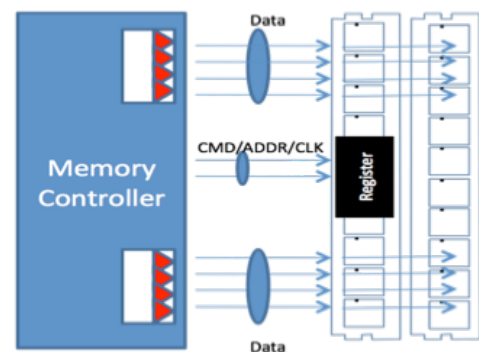
## II. Options for expanding memory capacity

Few options exist that support scaling to higher memory density and higher bandwidth. The key challenge in scaling is that the memory controller is required to drive higher numbers of modules and memory components and therefore higher loads on the Command/Address bus (C/A) and data (also called DQ) that form the memory interface. Current generations of memory controllers can drive up to three memory slots per channel. There is a trade-off between how fast the channel can run vs. the number of modules that are installed – because the memory controller must drive longer trace lengths that increase electrical loading on the memory controller drivers. These requirements make it even more difficult to achieve the timing and signal integrity needed to run high-performance applications.

Each of the following options provides a way to mitigate this challenge, resulting in varying degrees of capacity and bandwidth scalability.

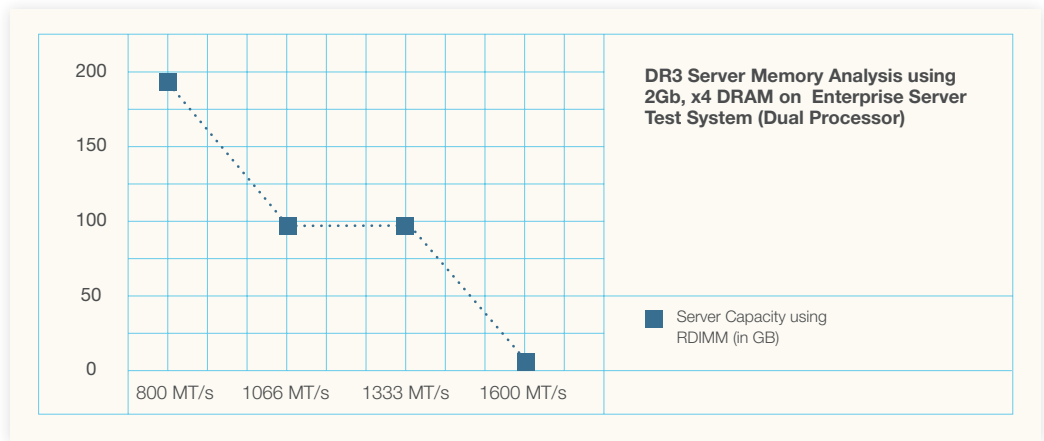
### A. Standard RDIMM speed and density baseline

RDIMM, or registered DIMM, is an architectural implementation that was introduced to facilitate higher memory capacities and bandwidths, relative to the un-buffered DIMM (UDIMM). A logic component, the register, is the core logic on an RDIMM. It “buffers” the C/A signals, or the set of signals that is routed to each DRAM component on the DIMM. Relative to an un-buffered DIMM, RDIMM provides higher capacity and faster speeds for single-, dual- and quad-ranked memory modules.



**Registered DIMM**

For the purpose of this discussion, we consider RDIMM as the baseline for speed and density. Below is an illustration of the inverse capacity and frequency scaling for an Enterprise Server test system (Dual Processor) based on DDR3 RDIMM.



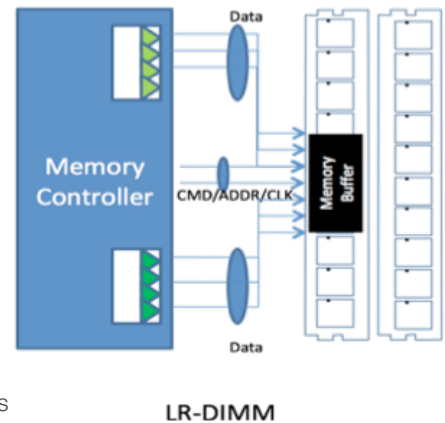
Although RDIMM improves the capacity and frequency scaling to some extent, its performance is reaching its limits. This is because at higher capacities, the data signals (DQ) driven on the memory controller and DRAM interface experience excessive loading.

### B. Load-Reduced DIMM (LR-DIMM)

LR-DIMM seeks to alleviate many of the problems faced by RDIMM by extending the concept of “buffering” to include data signals as well. LR-DIMM incorporates an iMB component that buffers DQ and command, address and control signals on its pre-buffer interface to the memory controller. The result is reduced load on the memory controller as the memory controller has to drive only a single load to the iMB component on all its interface signals. The iMB component assumes responsibility for interfacing with the memory elements on its post-buffer interface.

In addition to reducing the load, iMB technology seeks to increase capacity by accessing more DRAM on the post-buffer interface. Using a new option called Rank Multiplication, iMB technology enables more ranks of DRAM to be populated on the memory module and seamlessly accessed by the CPU memory controller, thereby accomplishing higher capacities.

The above concept allows LR-DIMM to scale to higher capacities and frequencies using cost-effective mainstream DRAM technology. The graph below indicates that at higher operating frequencies, LR-DIMM allows for the server system capacity to increase 2x-4x compared to the capacity of RDIMM, at difference bandwidth points.

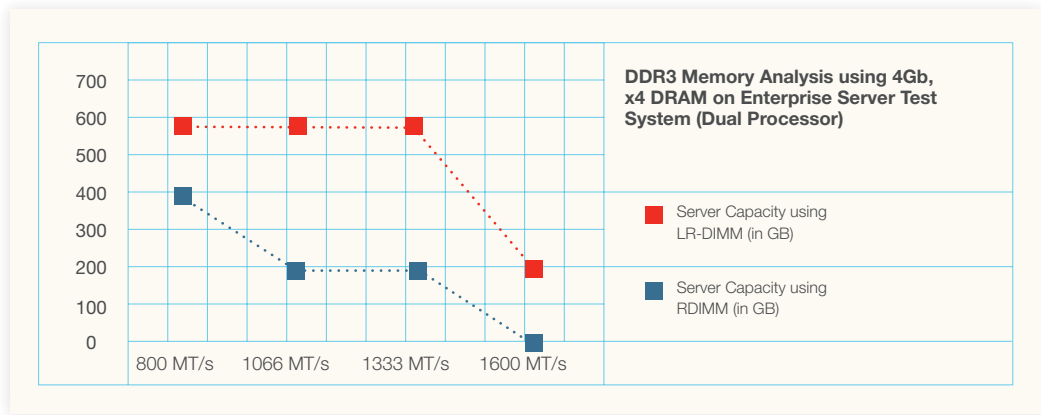
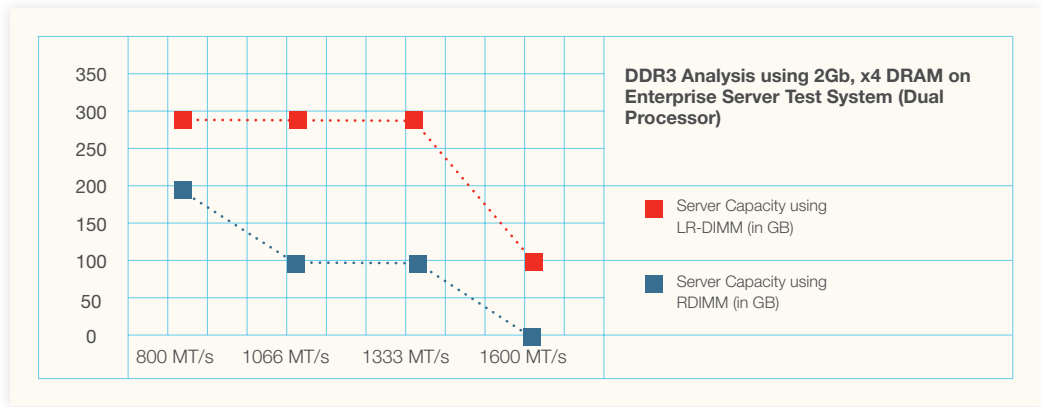


**Table 1: Enterprise Server Test System (Dual Processor) Memory and Bandwidth Scaling comparison**

2Gb, x4 DRAM Technology	Using RDIMM			Using LRDIMM		
	Max DIMM Capacity	Total DIMMs/system	System Capacity	Max DIMM Capacity	Total DIMMs/system	System Capacity
800	16GB, 4R	12	192	16GB, 4R	18	288
1066	16GB, 4R	6	96	16GB, 4R	18	288
1333	8GB, 2R	12	96	16GB, 4R	18	288
1600	0	0	0	16GB, 4R	6	96

4Gb, x4 DRAM Technology	Using RDIMM			Using LRDIMM		
	Max DIMM Capacity	Total DIMMs/system	System Capacity	Max DIMM Capacity	Total DIMMs/system	System Capacity
800	16GB, 4R	12	384	16GB, 4R	18	576
1066	16GB, 4R	6	192	16GB, 4R	18	576
1333	8GB, 2R	12	192	16GB, 4R	18	576
1600	0	0	0	16GB, 4R	6	192



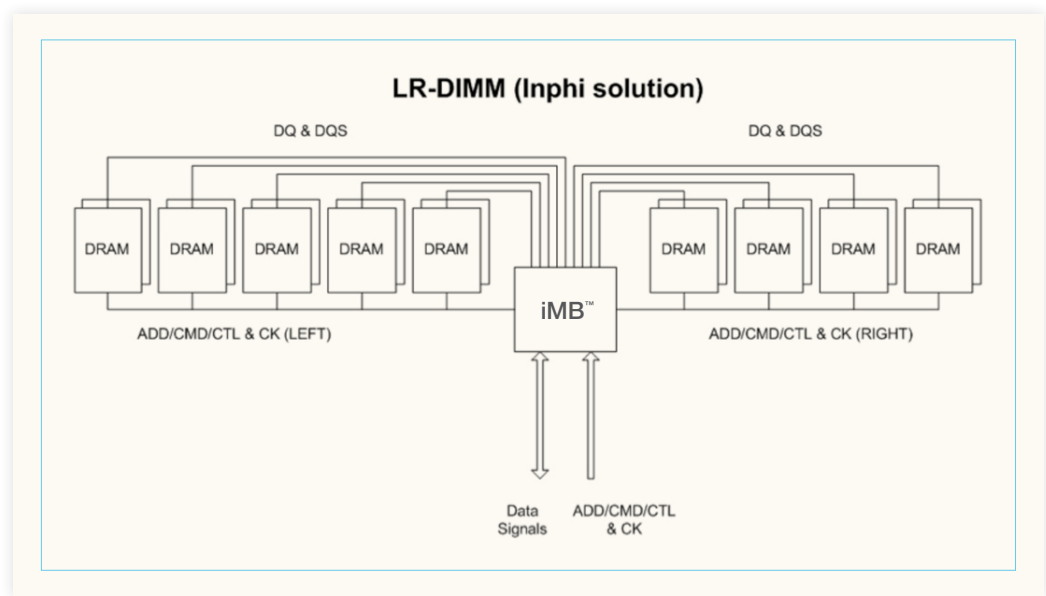
### C. Proprietary ASICs on the motherboard

On some enterprise servers, a concept similar to iMB technology is incorporated on the system motherboard itself. A logic memory element is implemented on the system board, also sometimes known as “buffer on board” (BoB). BoB essentially performs the buffering of the DQ, C/A and control signals on the interface with the memory controller on the motherboard itself. On the DRAM interface, BoB provides increased memory channels that incorporate memory slots.

BoB can be a compelling solution for even higher memory capacity. Yet, this comes at the cost of a fixed cost overhead and reduced flexibility for end customers.

### III. Architecture of the iMB technology

iMB technology is central to the load-reduced DIMM architecture. It is the logic to which all the memory controller interface signals get routed; from there, they are fanned out on the other side of the DRAM interface.



Besides the load-reduction feature, an iMB component also contains logic that allows the memory controller and BIOS to program the drive strengths and the termination to optimal values. iMB component also assumes responsibility for ensuring a proper initialization of the DRAM interface during the initial system power-up. Specifically, iMB component performs particular functions such as write-leveling and read-training to ensure that the iMB and the DRAM devices operate at high speeds within the margins of timing and signal integrity.

Following are some of the high-level features of iMB technology.

## A. Key iMB features and benefits

### 1. Reduced memory bus load

iMB technology allows DDR3 server systems to scale to higher memory capacities and operating frequencies by reducing the loading on the memory controller's DQ, C/A and control signals.

This distributes the capacitive loading on the memory sub-system uniformly, which enables the memory controller and the rest of the memory sub-system to operate at higher frequencies, even with higher loading.

#### Low-cost multi-sourced DIMMs

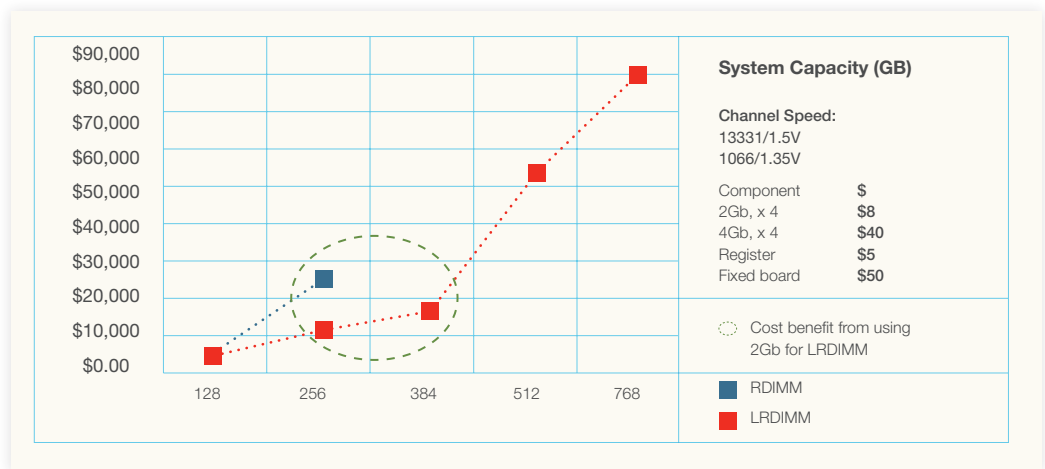
In conjunction with its iMB development, Inphi has been working with many ecosystem players, in particular DRAM vendors, to ensure that higher-capacity DRAM components such as dual-die and quad-die packages are available to the market at the same time as iMB components. This coordination will allow 16GB and 32GB LR-DIMM to be available in the first half of 2010. Inphi also has championed the iMB technology at JEDEC for more than a year in an effort to standardize this architecture for the benefit of the enterprise server market. In addition to making progress within the ecosystem, this standardization effort will ensure multiple sources for the LR-DIMM and make higher capacities cost-effective while establishing a robust supply chain.

### 2. Support for extending the use of a given memory technology

iMB technology increases the ranks of memory components on the post-buffer interface, using a concept commonly known as "rank multiplication." This means that the controller addresses x number of logical ranks on the pre-buffer interface, and the iMB can increase the number of physical ranks on the post-buffer interface to 2x or 4x. In order to fully exploit this capability, DIMM and DRAM vendors are developing a wide range of memory densities using 1Gb (2Gb dual-die packages and 4Gb quad-die packages), 2Gb (4Gb dual-die packages and 8Gb quad-die packages) and 4Gb (8Gb dual-die packages) DRAM densities. This allows increased memory capacity using mainstream x4 memory components, or similar capacities using cheaper x8 memory components – both of which help reduce procurement costs by as much as 60%, depending on the lifecycle and cross-over of the current-generation and next-generation memory technologies.

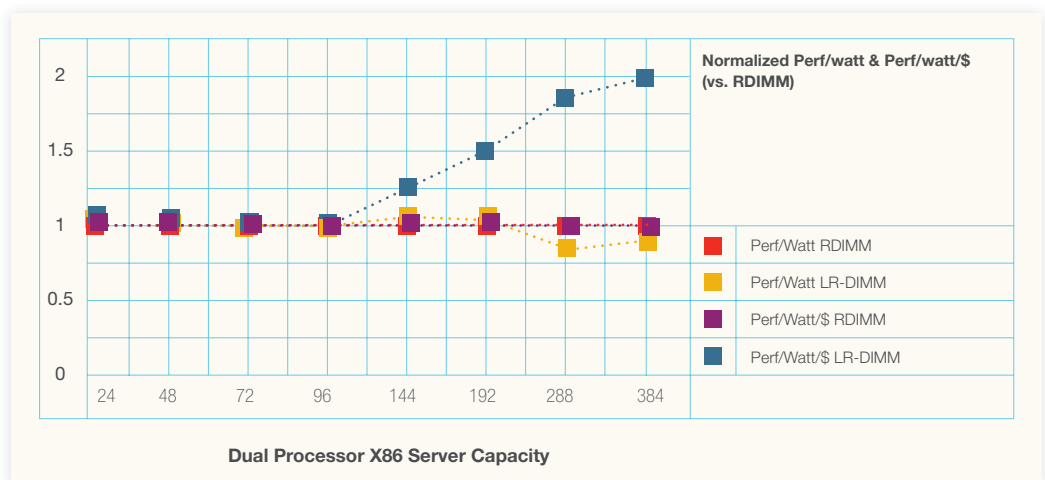
### 3. Power consumption

An obvious question is: Does an increase in capacity and frequency scaling come at a significant power penalty? It is true that relative to the register components, iMB components consume higher power. However, the iMB architecture allows thermal power optimization on the post-buffer interface. Additionally, features such as Slow-exit, CK power down, CKE power down and active CKE management provide optimal power management. The iMB architecture provides inherent performance/watt and performance/watt/\$ advantages.



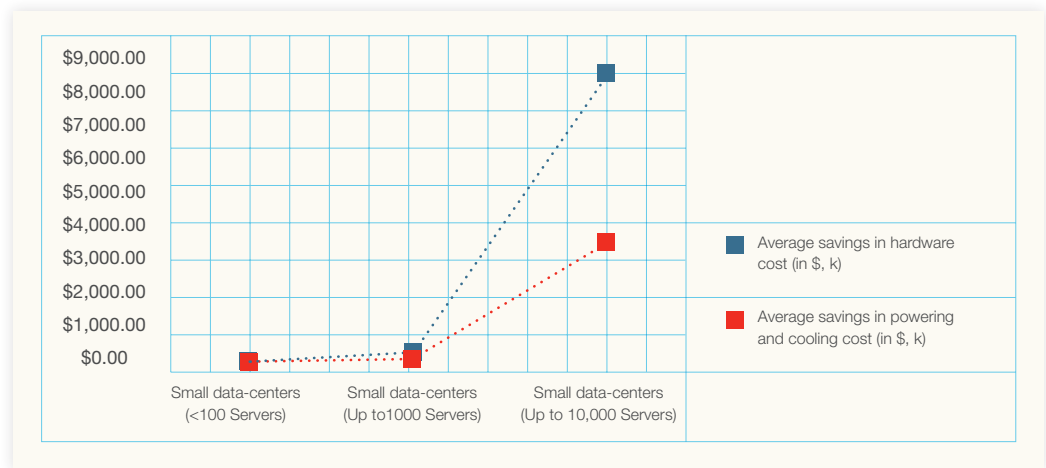
### 4. Expansion to DDR4

iMB technology allows scaling to frequencies even beyond the generation of DDR3-based server systems. Inphi is currently involved in architecting and defining revolutionary memory sub-system architecture for DDR4 based on iMB technology, while preserving the existing host interface architecture. The new architecture is targeted to enable memory interfaces operating at DDR4 speeds of 3200 MT/s.



#### IV. Greatest impact to Total Cost of Ownership

As mentioned in the introduction, data centers of any scale can use iMB-based LR-DIMMs to reduce their TCO. By maximizing the memory capacity and bandwidth on the servers, administrators can avoid purchasing additional servers that would otherwise be needed to achieve equivalent capacities and bandwidths. In addition, data centers can lower their power profile by reducing their powering and cooling requirements. The following chart illustrates how data centers of any size can reduce their TCO using iMB technology.



#### V. Conclusion

Inphi's iMB technology is a compelling approach that solves serious challenges for data center servers. iMB technology has immediate application in the current generation of DDR3-based servers, while the scalability of the technology ensures a long life and continued application for future DDR3 and DDR4 memory-based servers. The load reduction and rank multiplication features provide 2x – 4x increase in the memory capacity of the servers at different frequencies, while using mainstream memory components. Savings on procurement costs of memory components might be as high as 60%, depending on the lifecycle and cross-over of the current-generation and next-generation memory technologies. Overall, the iMB-based LR-DIMM technology is poised to significantly reduce the total cost of ownership of data centers by reducing procurement costs as well as costs related to powering and cooling. Inphi is committed to developing iMB technology-based products to serve the enterprise needs of today and tomorrow.



**Inphi Corporation**  
 2393 Townsgate, Suite 101  
 Westlake Village, CA 91361  
 Phone: (805) 446-5100

**Inphi Corporation**  
 3945 Freedom Circle, Suite 1100  
 Santa Clara, CA 95054  
 Phone: (408) 217-7300

**Web: [www.inphi.com](http://www.inphi.com)**