



The Sweet Spot for Embedded DRAM

Introduction

Handheld devices – cell phones, PDAs, games, MP3 players, etc – require long battery life, small size, light weight, and demand ever-growing processing power. Historically, these requirements have been in conflict, but embedded DRAM (eDRAM) can improve in all these areas simultaneously. eDRAM has emerged as a technology that brings a unique combination of features and capabilities to many design challenges, including the challenges of handheld devices.

eDRAM is based on a simple concept: use an existing DRAM design, an existing DRAM process, and an existing DRAM fabline to produce a single chip that is part DRAM and part microprocessor. These "system on chip" (SOC) designs combine the logic circuits needed to process information with the DRAM circuits needed to store it. Such SOCs offer four important advantages over designs where processor and memory are in separate packages:

- Power: dramatically reduced active and standby power consumption.
- Performance: very high performance capability, because of the high memory bandwidth available.
- Size: smaller designs; only one chip is required, instead of two or more.
- Cost: fewer components, smaller circuit boards.

In order to understand these improvements, this paper will compare an eDRAM-based SOC design with a conventional CPU plus DRAM design in each of the four areas: power, performance, size, and cost. We will examine a typical 3-chip design, consisting of a 32-bit processor and 2 DRAM chips versus the equivalent eDRAM SOC design. The advantages apply in two-chip, 16 bit systems as well. (see sidebar)

Power Advantage

eDRAM saves power primarily because it eliminates many power-consuming off-chip signals. In order to compare the relative power consumption of an eDRAM-based SOC versus a three-chip design consisting of a processor and two standard DRAMs, one must consider the three variables which determine power dissipation: clock speed, voltage levels, and I/O capacitance. Of these, capacitance is often the variable most subject to change based on the board design chosen. The capacitance encountered when driving signals from one chip to another is typically 20 to 30 pF. It is often the case that over half of the total power dissipation is in the I/O buffers that must drive this capacitive load. In a conventional CPU plus DRAM design, there are typically 54 I/Os specifically required to drive the

memory bus. By eliminating the off-chip interconnects between processing elements and DRAM, eDRAM-based SOC designs save substantial power. Furthermore, this savings occurs during operation, not just during standby mode, because power is saved on every cycle to memory.

eDRAM SOC designs reduce standby power as well. All DRAMs require on-chip charge pumps and analog bias circuits to generate certain internal voltages needed for operation, and these circuits are active all the time. During standby periods, these circuits account for a substantial portion of the power consumption. Discrete designs require this overhead circuitry on every chip and this needless duplication increases system-level standby power substantially.

The trend in all IC design is toward lower voltage operation. Because power dissipation is a function of the square of voltage, a small reduction can have a noticeable effect. However, one must remember to base calculations on the actual voltage at which various portions of the IC run. A processor with a 1.2 volt core still uses 3.3 volt I/O buffers. It saves a little power in its core, but most power dissipation is in the large, 3.3 volt buffers as they swing the 20 to 30 pf loads. An eDRAM design at 1.8 volts uses less power because all of its internal connections are at 1.8 volts instead of 3.3 volts.

The high-capacity memory bus bandwidth provided by eDRAM allows the memory clock speed to be lowered, thus saving power.

A Typical Design: Two chips or Three?

Most handheld devices have either 16 bit or 32 bit data buses; thus the core chip count is either 2 (one CPU plus a 4 M by 16 DRAM); or three (one CPU and two DRAMs). When you want to compare a SOC solution to a discrete one, which do you compare to?

The answer is that it matters less than you might think. Tight size and power constraints drive most 16-bit designs; SOC solutions are smaller and use less power than 2-chip discrete solutions. A need for processing power drives designers to 32-bit solutions; again SOC solutions offer greater performance than 3-chip solutions, as well as smaller size and lower power.

Ultimately, every design has unique trade-offs, and must be evaluated on its own merits. But whether an eDRAM-based SOC is compared against a two-chip or 3-chip solution, eDRAM still wins.

While a two-chip design typically has a 16-bit data bus, an eDRAM design can be arbitrarily wide. Many eDRAM designs have a 256-bit wide data bus. Thus, fewer cycles are needed to fetch the same amount of data. Solutions which are properly architected for eDRAM designs can leverage this increased memory bandwidth with clever processing element design. We will look more closely at this in the next section.

Performance Advantage

Processing-system performance is more than just clock rate, it is the amount of useful work done per clock as well. Systems with wider data paths, or parallel processing elements, are capable of far more useful work per clock than more conventional designs which rely solely on increased clock speeds.

Discrete DRAMs have a 16-bit data bus. Therefore, designs which require wider data paths for performance must use multiple chips, which results in larger size and higher power consumption. eDRAM SOC designs, on the other hand, can easily have arbitrarily wide data paths. This dramatically enhances memory system bandwidth, especially in applications such as video processing, video compression, and 3D graphics, all of which have unusually large data transfer requirements.

All computing systems consist of processing logic to manipulate data, and storage for the data. Processing throughput is usually limited by the speed with which data can be moved to and from storage. (This is why most PCs have cache memory.) eDRAM SOC designs bypass this limitation. Clever eDRAM designs capitalize on the high memory bandwidth with more parallelism in their data processing. In addition to having a conventional central processing block, eDRAM designs make use of additional processing elements running in parallel. These elements may be general-purpose or special-purpose. The large memory bus capacity means the two processing elements do not compete with each other for data access.

Size Advantage

An eDRAM design is a single chip, so it takes up less board space than the two (or more) chips of a CPU-plus-DRAM design. By eliminating the off-chip DRAM bus, pin count and wiring complexity is reduced, allowing a denser layout. The absence of a high-speed external bus also reduces radiated RF, so regulatory compliance is easier to achieve.

Cost Advantage

The smallest DRAMs available today are 64 Mbits, with 128 Mbit devices in volume production and 256 Mbit devices beginning to ship. By using the high-volume process on which current-generation DRAMs are made, eDRAM gets the full benefit of economies of scale and the learning curve.

But eDRAM has another advantage – one only buys the amount of DRAM actually needed, rather than whatever part is in volume

production in the DRAM industry. Designers who only need a moderate amount of DRAM must buy parts bigger than they need. eDRAM designs avoid this waste. For example, if one only needs 64Mbits of DRAM with a 32 bit data bus, one would be forced to purchase 128Mbits because the smallest DRAM available is 64Mbits with a data bus width of 16 bits

eDRAM is always based on the current-technology DRAM process, but instead of building 100% DRAM on it, typically builds about half the die as DRAM and half as logic. (eDRAM designs can be as little as 30% memory, or as much as 70%.) In today's process, a typical eDRAM will offer 64 Mbits (8 Mbytes) of memory, with the remainder of the die dedicated to processing functions. Thus, eDRAM users get the lowest cost-per-bit offered by the current process, while only buying the amount of memory actually needed for the application. Over the longer term, this develops into another advantage for eDRAM: as the 256 Mbit DRAM process becomes mainstream, the eDRAM design can be ported, transparently, for cost reduction.

Because it's a commodity product, DRAM pricing and availability undergo big swings based on supply and demand. Also, DRAMs are typically made by different suppliers than microprocessors. With eDRAM one deals with a single supplier with predictable pricing and delivery.

So what's the catch?

It might sound like eDRAM is the best technology for all designs. That's not quite the case. Systems which require large amounts of DRAM – amounts which require four or more DRAM chips – are best done as multi-chip designs. Also, low-end desktop applications may not need the lower power and smaller size of eDRAM designs.

Summary

eDRAM technology reduces power consumption by eliminating many power-wasting off-chip I/Os. eDRAM performance is increased by leveraging the wide data path available with on-chip memory. eDRAM SOC designs reduce system size by substituting a single-chip solution for one requiring two or three chips. And eDRAM is cost-effective because it's built on a high-volume DRAM process, but only includes the amount of memory actually needed for the design.

Handheld and portable device designs consist of a combination of tradeoffs among power consumption, size, performance, and cost. Like a teeter-totter, pushing one factor down pushes another one up, as the system pivots around the fulcrum.

eDRAM shifts the fulcrum. It offers lower power without sacrificing performance. It offers smaller size without raising costs; indeed, it reduces costs as well. Handheld devices, with their tight requirements on power, performance, size, and cost, are the sweet spot for eDRAM technology.