Embedding Memory in Next Generation Handset Designs

Next generation handsets will have tremendous demands placed upon them to handle dozens of applications. To successfully meet that challenge, engineers will have to grapple with the types and architectures of memory and how to embed it so it provides the maximum benefit with the lowest power and minimum real estate.

By David Janas and Greg Komoto

As today's cellular networks transition to 3G, and stage for even more robust environments such as 4G, a new set of challenges is emerging for handset designers. Color screens, embedded cameras, Internet access, streaming media, e-mail, instant messaging, downloadable Java and MP3 music applications must all be considered in next generation systems. In delivering these new features, platform architects must determine the right mix of memory types to balance escalating performance and storage needs against unit costs.

This task is further complicated by a tough set of fixed requirements. Customers are excited by new features, but they are not willing to sacrifice voice quality, small size and long battery life. Power and space requirements will therefore remain stringent as new features are added.

How Much is Enough?

Designers need to estimate next generation memory requirements as accurately as possible, because the memory architecture strongly affects the customer experience. Too little memory can impair performance and limit functionality, while too much can unnecessarily increase product cost. Accurate and timely estimates are necessary to avoid design missteps that can increase development costs and delay new product rollouts.

Fortunately, relatively clear patterns can be seen in the global adoption of handset capabilities. Japan typically has led the world in feature adoption. Although Korea had lagged behind Japan by about 12 months in new feature introductions, Korea now has the highest penetration of 3G phones in the world. The United States,

Western Europe and China have lagged Japan by about two to three years. The most highly visible features, such as color screens and embedded cameras, have spread most quickly.

By looking closely at adoption patterns, and factoring in the memory required for specific features, it is possible to estimate near-term memory requirements. A forecast model indicates the phone's average flash density will reach about 210 Mb by 2005, which is about 170% more than the average density from 2003.

Voice vs. Data

In defining memory requirements for converged devices, it is common to differentiate only between voice and data traffic. However, there are many different types of data, including application code and data; highly volatile "scratchpad" data; and large, relatively static media files. Each of these data types has different performance, persistence and capacity requirements.

In handling these various workloads, the memory system must also satisfy general handset requirements, including:

Low power for long battery life&#151 The move to 1.8 VDC devices has helped, but increasing functionality will require substantial, ongoing efforts to contain power consumption. Battery life has been the Number one cited consumer and operator issue with the adoption of 3G and other advanced handsets.

High density for compact form factors&#151 Advanced silicon process technologies yielding greater bits per unit area and improved packaging strategies will be critical. Designers should also consider offering external, add-on memory modules, which customers can purchase separately. This provides the option for virtually unlimited file storage with minimal effect on handset size or cost.

Fast memory access for call management and converged functions&#151 Consumers gauge performance not only by the speed and quality of demanding applications, such as video, but also through the touch response, the speed of application launches, and the time required to start up the handset. Performance will remain an important differentiator.

Low parts count to contain costs&#151 The benefits of multiple memory types must be balanced against total memory costs. In the final analysis, the total cost per handset is always more important than the cost per bit of any individual component.

Memory: RAM, NOR and NAND

Most next generation handsets will use a combination of flash or RAM for code execution and flash for data (see Table 1).

RAM provides fast and flexible access to granular data, and supports high performance reads and writes over the expected product life of a handset. These characteristics make it the only viable current option for the fast read and write requirements of certain handset design elements, such as buffers, stacks and scratchpad memory. On the downside, RAM is a volatile storage medium that consumes substantial power to retain its contents. It is therefore not appropriate for storing large files or for any data that must be preserved when the handset is powered off.

SRAM is currently the most commonly used form of RAM in handset designs. SRAM is faster and requires less power than DRAM, but is also more expensive. Each SRAM memory cell is comprised of six elements: four transistors and two load devices. Newer RAM technologies, such as Pseudo-SRAM and Mobile SDRAM, are faster than SRAM, but also can consume orders of magnitude more power.

Pseudo-SRAM (known by a variety of proprietary names) is based on a DRAM type memory cell that employs a transistor and a capacitor, plus circuitry that mimics the interface of an SRAM device. Its power consumption is midway between the low consumption of SRAM and the much higher consumption of Mobile SDRAM. Mobile SDRAM, or LP SDRAM, is a special version of traditional SDRAM that includes power saving features, such as *partial array refresh* or *temperature compensated refresh*. Mobile or LP versions of other DRAM devices, such as DDR, are also in the works.

General recommendation: configure sufficient RAM for very dynamic data requirements (e.g., scratchpad), but limit total RAM to minimize cost and power

consumption.

NOR is a type of nonvolatile flash memory that supports fast random data access for granular data. It can be used in place of extra RAM to substantially reduce power requirements for code execution and for many types of data storage, while maintaining appropriate performance levels.

NOR is particularly well suited for code storage because it enables direct code execution from non-volatile memory. This capability is called XIP. It not only reduces RAM requirements, but also eliminates the performance overhead of copying code prior to execution. NOR supports synchronous burst reads, which improves processor utilization by feeding data at the processor's clock speed.

NOR is also well suited for storing application data, especially for on-board functions that require relatively frequent and granular access to memory. Though NOR is slower than NAND for accessing large block data, it typically delivers better performance for frequent, byte-level access. NOR also supports RWW capabilities, which further accelerates performance by enabling simultaneous reads and writes to different memory sectors.

Single-bit-per-cell NOR is more expensive per megabit than NAND, but its *price per device* is typically lower than NAND at the densities that are appropriate for most on-board requirements. In general, the price per bit of NOR is also less than SRAM and PSRAM, though it may be somewhat higher than that of LP SDRAM. New MLC NOR solutions enable two data bits to be stored per cell, and substantially reduce the total cost per megabit of NOR memory. With this development, NOR density and cost are tracking with the &quotsweet spot&quot for on-board handset memory requirements.

General recommendation: use NOR for code execution (XIP); for on-board, non-volatile memory requirements; and for on-board data storage where RAM-like speeds are not needed. As a general rule, a 4-to-1 ratio (NOR-to-RAM) will continue to be appropriate in next generation designs.

NAND is another derivative of non-volatile flash memory that has been targeted toward mass storage applications. It supports serial readout of data from a buffer, rather than random byte-level access. It is therefore characterized by slow initial

data access, followed by moderately fast delivery of the rest of the accessed page, which is contained in a buffer. In general, NAND comes in higher-density devices and may be priced less per megabit than NOR, but can be more expensive per device in the densities required for many on-board functions.

NAND is more complex to implement than either NOR or RAM. It cannot be used out of the box, but requires relatively complex software drivers for block management, non-linear address maps and error correction. The error correction software is necessary, since bit-errors and bad memory blocks are relatively common in NAND devices.

It is possible to store code in NAND and download it into faster RAM for execution. However, this provides no real performance gain in comparison with XIP, and the extra RAM increases handset costs and power consumption. Because of the initial delay in accessing NAND data, this approach also results in slower boot and launch times.

General recommendation: use NAND for high volume external storage (plug-in memory cards), rather than for on-board memory functions that require fast access to granular data. This strategy takes advantage of the low cost per megabit for very high-density NAND storage devices, without burdening handset performance or cost.

Of course, designers can expect ongoing advances in density, performance and cost reduction for all three memory types. In general, flash memory solutions will advance faster than RAM, because of growing issues with scaling traditional RAM devices.

Other Design Considerations

Addressing limitations in the platform architecture will always set specific limits on supportable memory. Memory packaging parameters are also critical. Currently, as many as five memory chips of various types can be supported in a single 100 mm^2 package. Designers should be aware that some memory combinations are more widely supported in packaging and software, which can affect both cost and development complexity.

Assessing the key handset variables, including processors, core and bus frequencies, operating systems, code types and many specific application workloads, the following are critical factors with respect to memory considerations.

Write performance&#151 Although raw write speeds are faster for NAND than for NOR, there is little user-perceptible difference in actual usage. NOR write speeds are even sufficient for streaming large multimedia files. However, NAND (or data optimized NOR) provides better performance for bulk transfer of files larger than 1 MB.

Code vs. data performance&#151 Cache miss rates, a good indicator of overall performance, are low for code execution in cellular architectures. However, they are an order of magnitude higher for data manipulation. This suggests that the faster granular access rates of LP SDRAM may be important for highly volatile data, though not for code execution. Extensive use of system RAM for code shadowing or paging is both wasteful and unnecessary. Furthermore, when LP SDRAM is used for both code execution and data manipulation, cache line fills can stall access for both functions.

Data integrity&#151 NOR memory is generally preferable for storing critical information, such as code and configuration data. This is because data in both RAM and NAND is more vulnerable to accidental or malicious corruption. With NOR, it is easier to isolate critical data from user-accessible memory regions.

Power consumption&#151 In general, power consumption for the memory system is most important in call standby and power-off modes, because other handset functions dominate power usage during extended active use. Large amounts of LP SDRAM, in particular, can significantly affect the overall power profile if is not well managed. The use of XIP for code execution can reduce power usage by 14% or more compared with typical NAND and LP SDRAM shadowed solutions. This can be important in both active and standby modes.

Core Recommendations

RAM, NOR and NAND memory will all continue to play a role in next generation handset designs. In general:
&#149 RAM is best used in relatively small amounts for volatile data manipulation.
&#149 NOR is best used for code execution, configuration data and other less volatile on-board storage functions. &#149 NAND is best used for high-density,

external storage such as removable memory cards.

Though these are good rules of thumb, there is no one-size-fits-all solution, because functionality and workloads will vary between devices and users. Optimal memory architectures will continue to depend on evolving features sets, customer expectations and memory price structures. In general, handset designers should consider all memory requirements, including both RAM and flash, and develop balanced platform solutions that leverage the unique strengths of all three technologies.

Source URL (retrieved on *03/09/2014 - 5:33pm*):
http://www.wirelessdesignmag.com/product-releases/2005/10/embedding-memory-next-generation-handset-designs

Links:
[1] http://www.wirelessdesignmag.com/sites/wirelessdesignmag.com/files/legacyimages/0510/wd510f2_table_lrg.jpg