Serial vs. Parallel interface SRAMs—A Comparison

Anirban Sengupta, Reuben George, Cypress Semiconductor - July 29, 2015

Traditionally, external SRAMs feature a parallel interface. Given the memory requirements for most of SRAM-based applications, it’s no surprise that parallel is a better option. For the high performance (primarily caching) applications where SRAMs are used, a parallel interface has considerable advantages over serial interface options. However the status quo is about to change.

Parallel interfaces, while providing superior performance compared to serial interfaces, have their own tradeoffs. The most significant is that a parallel interface has a significantly larger footprint in terms of both board space and pin count requirements. For example, a simple 4Mb SRAM may require up to 43 pins to interface with a controller. Considering that when using a 4Mb SRAM our requirements could be:

A. Storing up to 256K, 16bit long words
B. Storing up to 512K, 8bit long words

In the case of “A”, we need 18 pins for selecting an address (since there can be $2^{18}$ possible addresses) and an additional 16 more pins for the actual data input/output. On top of these 34 pins, we need more connections for functions such as Chip enable, Output enable, write enable, etc. In the case of “B”, we need relatively fewer pins: 19 pins for address selection and 8 bits for Input/output. But the overheads (chip enable, write enable, etc.) remain unchanged. For a package that accommodates these many pins, even the footprint in terms of area will be larger.

Once an address is selected, a word (or its multiples) is read or written at one go, which makes these SRAMs a great choice when a high access speed is required. In most the popular systems where SRAMs have been used, this performance advantage made the “too-many-pins” limitation negligible. These are systems where the controller is performing extremely complicated functions and thus requires a large cache. Historically these controllers have typically been large and had sufficient pins to interface. Applications with smaller controllers and limited pin-out have had to make do with embedded RAM.

In a memory chip with a serial interface, the bits are accessed serially (from between 1 bit to 4 bits at a time). This automatically makes serial interface simpler and smaller than parallel, but many times slower in terms of throughput. This limitation practically eliminates the use of a serial interface in most systems where SRAMs are used. However, the memory requirements of next-generation applications may soon lead to a major change in balance of pins vs. speed.

Trends in the industry
Processors continue to become more powerful while shrinking in size. And, with more power, processors require commensurate improvements in cache memory. At the same time, however, increasing the embedded cache memory has become an equally increasing challenge with every new process node. SRAM has a six-transistor architecture (logic area is typically four transistor/cell). This means that with smaller process nodes, the number of transistors per square centimeter will be extremely high. Such high transistor density can cause many problems, including:

**Increased susceptibility to soft errors:** Soft error rates are expected to increase seven-fold as process technology scales from 130nm to 22nm.

**Lower yields:** Due to the shrinking of the bit-cells coupled with higher transistor density, the SRAM area is more susceptible to defects due to process variation. Such defects reduce the overall yield of the processor chip.

**Increased power consumption:** If the SRAM bit cells have to be of same size as the logic bit cells, then the SRAM transistors will have to be scaled smaller than the logic transistors. The small size of the transistors causes an increase in leakage current which in turn increases the standby power consumption.

Another trend in technology is the advent of wearable electronics. For wearables such as smart watches and health bands, size and power are critical factors. Due to limited board size, the MCU has to be very small and able to run on the frugal power provided by portable batteries.

To fulfill the above requirements, on-chip cache is limited. In future generations, we can expect more functionality to be associated with wearables. In such a case, the on-chip cache will fall short and the need will arise for an external cache. Of all the memory options available, SRAMs would be the most fitting option to act as an external cache. This arises from their lower standby current.
consumption compared to DRAM, and lower access time than both DRAM and Flash.

The Rise of Serial

When we observe how a particular electronic product has progressed over the years – we note one key trend: device footprint continues to shrink in size while maintaining/improving performance with each successive generation. This shrinking phenomenon can be attributed to the fact that each component in a board has contributed to the overall effect by shrinking itself. The downscaling of electronic circuits was predicted way back in 1965 by Gordon Moore in his famous Moore’s law. However, downscaling has not happened uniformly in all types of circuits. For example, a logic circuit has shrunk in size many times more than an SRAM circuit. This leads to a major problem where the embedded SRAM begins to occupy up to 90% of the controller area. The lack of shrinking of the embedded SRAM also prevents the shrinking of the controller at a level commensurate to that of the logic area. Thus the cost (which is proportional to the area of the die) doesn’t lower as much as it should. Since the logic area performs the core functions of the processor/controller, moving the embedded SRAM out of the chip and replacing it with an external SRAM begins to make sense.

Also contributing to this is the boom in wearable electronics and IoT (Internet of Things) applications. These devices require the overall design to be small over nearly every other design constraint. Hence, the smallest possible MCU is the best fit for such a board. For the reasons mentioned above, there is a high chance that this “smallest MCU” is likely to be devoid of an embedded cache. Similarly, it is likely to have not too many pins. A look at the future

All these trends lead to a single requirement: an external SRAM that is small in size, can perform the role of the only cache, and be connected with a minimum number of pins. A serial SRAM is tailor-made for such a requirement. In other memories (DRAM, Flash, etc.) where high-speed performance isn’t the most important requirement, serial interfaces have already replaced parallel interfaces. Serial SRAMs have been a tiny niche in the SRAM market, due to the applications where SRAMs are needed. They have been low power, small footprint alternatives in very specific applications where space has been a critical issue. Today, the highest density serial SRAM is 1Mbit with a peak clock rate of 20MHz (10MB/s bandwidth). In contrast, parallel SRAMs have bandwidth up to 250MB/s and support densities up to 64Mbit. The below table contrasts a generic 256Kbit parallel SRAM with a 256Kbit serial SRAM.

<table>
<thead>
<tr>
<th></th>
<th>Parallel 256Kbit SRAM</th>
<th>Serial 256Kbit SRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. Bandwidth</td>
<td>18.2 MBps</td>
<td>2.5 MBps</td>
</tr>
<tr>
<td>Max. Active current</td>
<td>50 mA</td>
<td>10 mA</td>
</tr>
<tr>
<td>Max. Standby current</td>
<td>10 uA</td>
<td>0.5 uA</td>
</tr>
<tr>
<td>Smallest package</td>
<td>28-pin TSOP-I</td>
<td>8-pin SOIC</td>
</tr>
</tbody>
</table>

Due to fewer pins that need to be powered and lower speeds, serial interfaced memories typically consume less power compared to parallel interfaced memories. However, the greatest gain is in footprint – both in terms of device size and pin count. While the smallest parallel SRAM package is 24-ball BGA, Serial SRAMs are available in packages as small as 8-pin SOIC. However, it must be noted that WL-CSP is the smallest package, and many parallel and serial memory manufacturers support CSP packages. Where parallel SRAMs gain over serial SRAMs in the market is in
performance – specifically in terms of access time. Due to their significantly wider bus, parallel SRAMs can support up to 200MBps while most widely available serial SRAMs peak at 40MBps.

As noted in the table above, serial interfaced memories lag behind parallel interfaced memories in terms of performance. Due to the sequential data flow, they have been unable to provide the same throughput. Hence serial memories are most useful in portable devices where size and power consumption matter more than access time, such as for handheld devices and wearable.

**What the future holds**

Until the IOT and wearable boom, serial SRAM didn’t have a market lucrative enough for major SRAM makers to take note. In fact, the key manufacturers of serial SRAM happened to be Microchip and On-semi. For both these companies SRAMs are not a part of core business and constitute a miniscule portion of overall revenue. On the other hand the market leaders in the Static RAM space (such as Cypress, ISSI, Renesas) have so far focused only on parallel SRAMs. The future may however be different. With more opportunities emerging for serial SRAMs, we may soon witness the entry of traditional SRAM makers in the serial SRAM space. This will definitely lead to roadmaps for Serial SRAMs in the years to come. The two major areas of thrust will be density and bandwidth. Cypress has already included serial SRAM in its Asynchronous SRAM Product Roadmap. In fact, the recent Cypress/Spansion merger implies that the new Hyperbus technology (pioneered by Spansion) that can provide throughputs up to 400MBps on a serial interface and thus outperform DRAMs substantially. With the entry of major SRAM manufacturers into the market, developers can expect to have access to cutting-edge Serial SRAMs in the near future.

A high throughput, small sized, serial interface SRAM brings to the table tremendous possibilities. It could thus eventually be the universal successor of the current day embedded SRAM and the current generation parallel SRAM on many boards.