Ace's Guide to Memory Technology: Part 3

By Johan De Gelas – August 2002

DDR333, DDR400, DDR-II, and 32-bit PC1066 RDRAM

What is the memory standard of the future? RDRAM, DDR400 or DDR-II? It is a question that continues to fuel heated discussions on many Internet hardware forums, including our own message board. In this article, we'd like to offer you some insight in how these different technologies compare. This article's difficulty level is a bit steep, but we are convinced that with a bit of technical background you will be able to make a wiser choice between the different DRAM types out there.

First, I'd like to immediately address a question that some of our readers may have: "why still include RDRAM?" Indeed, over the past months there have been many reports that RDRAM is dead. Intel has indeed stopped RDRAM chipset development efforts and it is clear that DDR-II has a much brighter future ahead. Nevertheless, RDRAM is far from dead.

Despite Intel's half-hearted launch of the i850E, the good performance of PC1066 has sparked some interest and quite a few manufactures have launched or will launch boards with official PC1066 support. The largest motherboard manufacturer, ASUS has launched two motherboards officially supporting PC1066 RDRAM. DFI, EPOX, BIOSTAR and IWill have followed in ASUS footsteps.

On top of that, Silicon Integrated Systems (SiS), the manufacturer of the first DDR400 chipset for the P4, has launched the PC1066 RDRAM- based R658. At the same time, Intel is still validating PC1066 support for the i850E. After erasing RDRAM from their roadmap a few months ago, **Elpida (NEC + Hitachi) has decided** to go ahead with the PC1066 RDRAM production anyway. So while DDR is by far the most popular DRAM in today's market, RDRAM can still be found in quite a few high-end desktops and workstations. As many of our readers are power-users, we decided to take a look at the different chipsets and DRAM technologies available for the Pentium 4 platform once more.

But before we start talking about benchmarks in our next article, this article will provide some technical background about the differences between RDRAM and DDR SDRAM. The objective is to understand which technology has the best chances in the long term. What the strong and weak point are of DDR SDRAM and RDRAM? Why RDRAM can reach so much higher clockspeeds? Can DDR-II adopt some of the techniques that make RDRAM a faster clocked technology?

This article will be pretty technical, but even if you are not so technically inclined, don't worry. Every interested reader will be able to understand the discussion. However, to understand this article well, you should understand the basics of DRAM memory, which are explained in our previous articles **here**.

High Clockspeeds for DRAM

Many articles point out that RDRAM uses a 16-bit bus for the data signals and that this narrow 16-bit path is the main reason why RDRAM is able to run at speeds up to 533 MHz (DDR, effective clockrate: 1066 MHz). Indeed, one of the problems that parallel wired systems face is skew. The longer and the faster a synchronous bus (with a clocksignal) gets, the more likely it is that some data signals will arrive "too soon" or "too late" and are not within the limits of the clocksignal. That is why we see that many I/O interfaces go from typical slow parallel busses to fast serial busses. Examples are the move from the parallel port to USB or the current IDE ATA to serial ATA.

But while a 16-bit datapath can be clocked a bit higher than a 64-bit one, it does not explain the differences in clockspeed between RDRAM and DDR SDRAM. The very best DDR SDRAM overclocks to 200 MHz or slightly more, while the best RDRAM has been able to run at 640 MHz (stable) and even more when we tested. Interestingly, overclocking from 166 MHz (PC2700) and 533 MHz (PC1066), both memory technologies achieve a 17% overclock. There'll be a full report on this later on.

Packaging is another reason why RDRAM is able to run faster than DDR SDRAM. Most DDR SDRAM uses TSOP (Thin Small Outline Package), while RDRAM uses FBGA or Fine pitch Ball Grid array. TSOP chips have rather long contact pins on each side, while FBGA chips have tiny balls at the underside to make contact with the outside world. Those very small soldered balls have a much lower capacitive load than the TSOP pins. In fact, industry sources told us that the current FBGA DDR SDRAM chips are able to run at 200-266 MHz, while the same chips with TSOP packaging are limited to 150-180 MHz. Therefore a decent DDR400 module should come FBGA and not TSOP packaged chips. Of course, this doesn't mean that FBGA modules with mediocre memory chips are superior to TSOP modules with quality memory chips.

Nevertheless, it is clear that the manufacturers who produce DDR-400 (Samsung and Micron) in TSOP packaging are somewhat stretching the possibilities of their chips. Don't expect high overclocking margins and be careful with highly loaded (= all sockets filled) memory systems. Already we have noticed that quite a few unofficial (=non JEDEC) DDR2700 CAS 2.5 won't work at 166 MHz DDR on many boards when you install two or more DIMMs of 512 MB. While we have yet to perform a comprehensive test of DDR400 modules, it is likely that we will encounter the same problems. It is no coincidence that JEDEC has decided that all DDR-II modules, which will work at 200 and 266 MHz, will be FBGA chips.

Architecture

Packaging is one thing, but the most important reason why RDRAM can clock so much higher is the special signal topology for addresses and data. Signaling and signal integrity are very important aspects of high-speed system design. Capacitive and inductive loads increase the time it takes to read a signal correctly. If we want to signal a logical "1" bit by sending a signal of +1.4V, it might take some time before voltages reaches 1.4V.

Let us first take a look at DDR SDRAM signaling. Most DDR SDRAM modules today consist of 8 or 16 SDRAM chips. Let us assume that there are 8 chips on our DIMM (Dual Inline Memory Module). A 64-bit datapath comes from the memory controller to the memory subsystem. Those 8 chips on each DIMM are connected in parallel to the 64-bit datapath. So each chip connects to 8 of the 64 datalines.

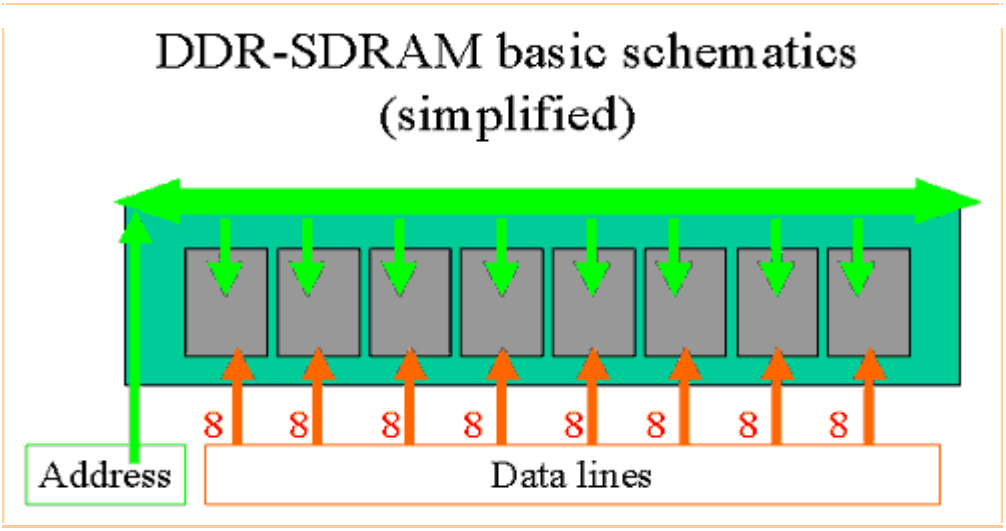


Fig 2. DDR SDRAM address and data routing

Notice that the signal topology for addresses and data is different in DDR (as well as SDRAM) systems. Take a look at figure 2, but please note that this is a simplified scheme, the exact wire routing is beyond the scope of this article. What you should grasp from the figure above is that 64 data bits of the data bus are cut up in 8 parallel data wires, each connecting to a (S)DRAM chip.

But you cannot cut up the bits of the address or commands in 8 pieces of course, as it would destroy the original address or command. So the complete address and command bus must be connected to all chips. Addresses and commands must be presented to all DRAM chips on a DDR DIMM at the same time. In order to accomplish this, addresses must be "fanned out" to all the chips on the DIMM. From the point that the address traces enter the DIMM module, all DRAMs must be equally far away. You can imagine that routing equally long traces from the address pins to each chip is much more complex than routing the data pins to the chips. Each chip has its **own exclusive eight** data pins it should be connected too (1 on 1 relation), while each chip should have access to **all** address pins (1 on N relation).

Let us take a look at the topology of RDRAM, so we can compare them both.

The figure below comes originally from a **RDRAM whitepaper** and I have colored the datapath in orange just like we did in the DDR SDRAM figure.

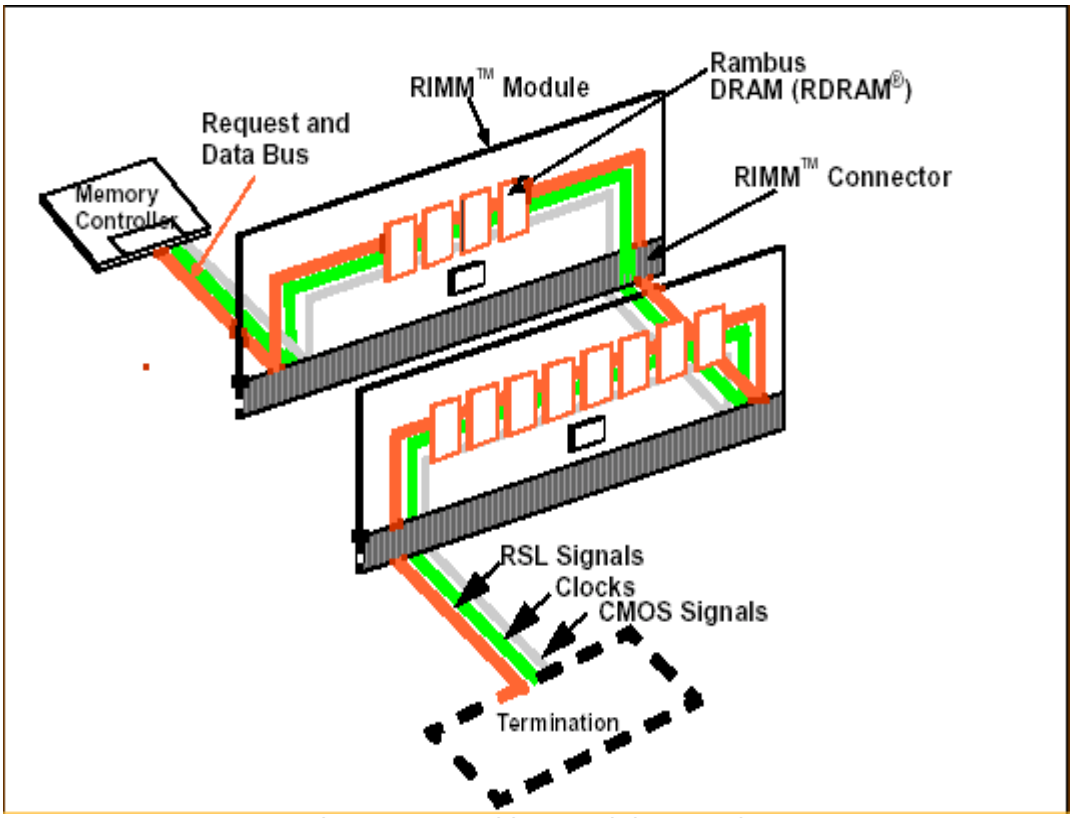


Fig 3. RDRAM address and data routing

The most important thing to notice is that an Rambus memory system consist of a transmission line (orange + green + gray line) which contains the **16-bit datapath** (orange) but also all other signals (clocks, adresses, commands). The address bus works exactly the same way as the databus and therefore the address and data buses are loaded in the same fashion. This Rambus transmission line zigzags like a snake through the two (or four) RIMMs and has to be terminated at each end. So while the dataline still consists of a 16 parallel datalines, all the Rambus chips are connected one after another or serially to the transmission line. Each chip is 16 bit wide and one chip is enough to make the RDRAM channel work.

In contrast, the DDR SDRAM datapath (in our example) consists of 8 different parallel paths of 8 bit wide. In the first 8 bit path the first chip of one DIMM is connected to the first chip of the second DIMM, in the second the second chip of the first DIMM is connected to the second chip of the second DIMM and so on. Four to 16 SDRAM (8 in our example) chips must work together to produce meaningful output.

In case of RDRAM there is only one 16 bit datapath, and the first chip of one RIMM is followed by the second chip of the same RIMM on this datapath.

RDRAM Compared to DDR SDRAM

Let us now compare these two different topologies and try to find what the exact advantages and disadvantages are. And more importantly, let us also evaluate how important an advantage is for a certain technology. Certain advantages may only exist on paper...

1. Advantage of RDRAM: Address and data are loaded equally

With RDRAM, Data and Addresses are loaded in the same fashion - they are part of the same "snake like transmission line" - as we could see in figure 3. Addresses and Data both use the zigzag snake route across the RIMMs. Therefor, each data and address wire has the same number of loads, with the number of loads depending on how many modules and how many ranks of devices are in the system.

In case of DDR SDRAM, we have a totally different situation: The address bus is loaded differently from the data wires. Let us take our example, a single module with only one side with 8 chips. In such a system, the address bus has 8 loads, while the data bus has only 1 load. The fact that the address bus and the data bus are both loaded differently, explains partially why DDR SDRAM even today does not attain more than 166-200 MHz. More importantly, the loading of the data bus changes as the number of rows (single sided module = 1 row, double sided = 2 rows) in the system increases.

With RDRAM, Data and Addresses are loaded in the same fashion - they are part of the same "snake like transmission line" - as we could see in figure 3. Addresses and Data both use the zigzag snake route across the RIMMs. Therefor, each data and address wire has the same number of loads, with the number of loads depending on how many modules and how many ranks of devices are in the system.

In case of DDR SDRAM, we have a totally different situation: The address bus is loaded differently from the data wires. Let us take our example, a single module with only one side with 8 chips. In such a system, the address bus has 8 loads, while the data bus has only 1 load. The fact that the address bus and the data bus are both loaded differently, explains partially why DDR SDRAM even today does not attain more than 166-200 MHz. More importantly, the loading of the data bus changes as the number of rows (single sided module = 1 row, double sided = 2 rows) in the system increases.

This explains **the issues about high speed DDR that we have reported back in March 2002** but also what we have experienced in recent testing. We have tested i845G boards, which support DDR333, first with one PC2700 of 512 MB at 333 MHz. Once we inserted a second 512 MB PC2700 DDR DIMM, boards get unstable in many benchmark situations if we force the two PC2700 DIMMs to run at 333 MHz.

Secondly, we have also found a reason **why the DDR SDRAM address bus transfers addresses at "single data rate"** (i.e. only on one edge of the clock) while the data bus is capable of transmitting data on both edges of the clock (double data rate): the address bus is much more heavily loaded. In case of RDRAM, addresses are sent on both edges of the clock.

From a performance point of view, this is only a very small advantage for RDRAM. For each row and column address we send, we get back 32 (burst length = 4) cycles) or 64 byte (burst length = 8 cycles) of data. Iin most cases, data bandwidth is much more important than address bandwidth.

Only when you get page miss on a open page (the wrong information is still there in the sense amp and must be unloaded) and you have to send a RAS, CAS, and a Precharge command to get your data, the extra address bandwidth can help. It depends on the software, but on average this situation happens less than page hits and page misses without precharges. It is estimated that pagemisses, which require precharging, happen between 20 and 5% of the time.

Aside from the small performance advantage, sending addresses at double rate allows you to have less address traces, saving pins on the memory controller.

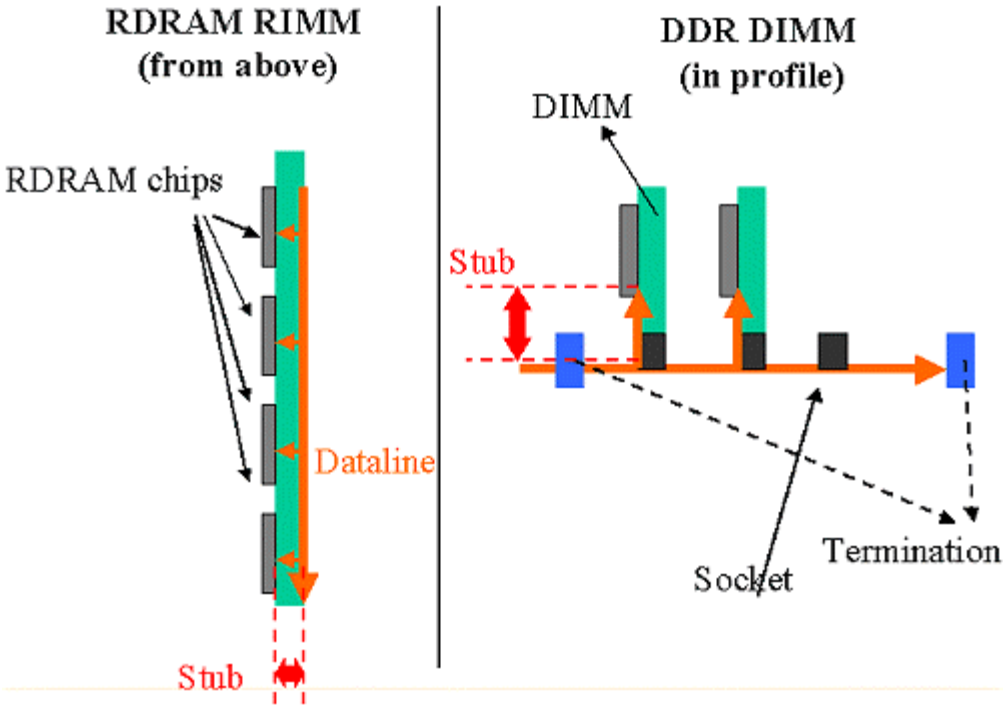
Evaluation of this advantage:

- ***Very important for maximum clockspeed. Allows RDRAM to attain much higher clockspeed.***
- ***Allows RDRAM to send Data and Addresses at double rate. Address at single data rate is however only a small performance disadvantage for DDR.***
- ***High speed DDR encounters issues when you use more than one double sided DIMM. Stability will depend on the motherboard design, and will not be consistent across many motherboards.***

2. Advantage of RDRAM: "Stubs" are Much Smaller

Another advantage of RDRAM's "serial snake architecture" becomes clear when we look at how signals travel between the memory controller and the DRAM chips.

With DDR SDRAM, the bus routes onto the module and into the DRAM chip but not back off the module again. When an electrical signal passes from the memory controller down the SDRAM/DDR channel, it hits a "fork" at the connector. If you look at the figure below, you can see that "the fork" consists of two paths: one path leads onto the module and to the DRAM chip (the **vertical** orange arrow on the DDR DIMM figure, the "**stub**"), and the other path continues along the motherboard towards the second connector (The **horizontal** arrow on the DDR DIMM figure).



At the end of the first path there is a capacitive load, (part of) the DRAM chip (the DRAM's input receiver and ESD structure). The issue here is that the "fork" in the signal path ends up introducing a discontinuity in the bus. The problem is that this discontinuity introduces signal reflections on the bus. To understand this, you must know that when a signal hits a discontinuity such as this "fork", some of the energy continues in the same direction as the original signal, but some is reflected backwards towards the direction that the signal came from. You can imagine that if there are a number of discontinuities on the channel, there are lots of signals reflecting back and forth. These reflections cause to big problems:

- The original signal gets weaker.
- As signals get reflected back and forth, up and down the channel, these reflections propagate to the DRAMs and the memory controller, and if they reach these devices at the same time they are trying to read or write data, these energy waves can interfere with each other. In the worst case, these reflections can even cause the wrong voltage to be received by a DRAM, leading to data errors.

Thus, reflections can wreak havoc on a bus. This is one of the biggest problems today when trying to clock the DRAM higher. To keep signal integrity, it is important that the the original signal is still strong enough when it arrives at the intended DRAM, and that the number of reflected signals is low enough. The faster the signal, the more troublesome a discontinuity becomes.

The length of the stub also plays a very important role. The stub is the trace from the parallel wires on your motherboard to DRAM chip on the module (the vertical orange arrow on the DDR DIMM figure). This stub is large compared to the wavelength of the signal, and the result is the "reflections issue" gets worse. To counter this effect, termination resistors are used to try and dampen out reflections somewhat, but it is very hard to eliminate the reflections totally.

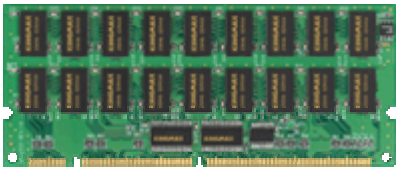
The faster the clockspeed, the shorter wavelengths become. Hence, the effect of stubs gets more and more pronounced, as they look larger and larger compared to the wavelength of the data signals. To sum it all up, **the bigger the stubs and the higher the clockspeed of the signal, the worse the signal integrity**. Stubs should be kept as small as possible if you want to achieve decent clockspeed and maintain signal integrity at the same time.

The "serial snake" topology of RDRAM solves the "stub causes reflections at high speed" issue. The stubs are much shorter than those in a DDR SDRAM system. Why? Because the signal travels on the RDRAM channel which runs onto and off of the RIMM module the RIMM, very close to RDRAM chips. You might recall seeing a "snake-like signal highway" on the RIMM (fig 3).

Therefore, the stub does not consist of a trace from the motherboard wires (through the connectors) to the chip, but from the "snake signal highway" to the RDRAM chip. More specifically, the stub is the distance from the solder ball (which originate in the "snake signal highway") to the actual RDRAM die itself. The result is that the stub is **only 3 mm big**, where the stub on the DDR SDRAM can be **as big as 25 mm**. Since the RDRAM stub is so short, it means that the data rate can be much higher.

But that is not all. Take a look at the figures 2 and 3 again. As the DDR SDRAM topology uses 8 datapaths in parallel, each datapath of 8 bit will hit only two or three discontinuities, one for each DIMM that you insert in your motherboard. In case of RDRAM however, as every chip is connected to the same datapath, the data signal will encounter as much discontinuities (stubs) as there are chips. For a signal traveling away from the memory controller, even though the stubs are small (relative to the wavelength of the signal), they look like a "lumps" of capacitance on the memory bus itself. To make sure that a lot of chips can be connected to the snake, the memory bus traces compensate for this. The traces are made narrower to increase the impedance and compensate for the reduced impedance of the capacitive load of the stubs. So, they have higher impedance right where the DRAM is attached to the memory bus. At the end of the RDRAM channel we find termination that matches the characteristic impedance of the channel, thus absorbing the signals that reach the end of the bus and eliminating signal reflections.

Nevertheless, the number of capacitive loads on the bus ("drops") must be restricted somewhat. The more chips, the more stubs, and thus the more capacitive loads. Therefore each RDRAM channel is limited to 32 chips, though memory repeaters can be used to enable more memory on a single Rambus channel. The dual channel i850 support 64 devices and the i860 can be outfitted with a memory repeater to support 128.



DDR SDRAM chipsets are also often limited to 32 chips too in practice - "registered stacked" DIMMs need to be used to extend this limit. For example, the Intel i845E can support 4 double sided DIMMs. Each double sided DIMM can contain at most contain 16 chips, the i845E does not support the stacked format with 32 chips. VIA and SIS support up to 3 double sided dimms, or 48 chips.

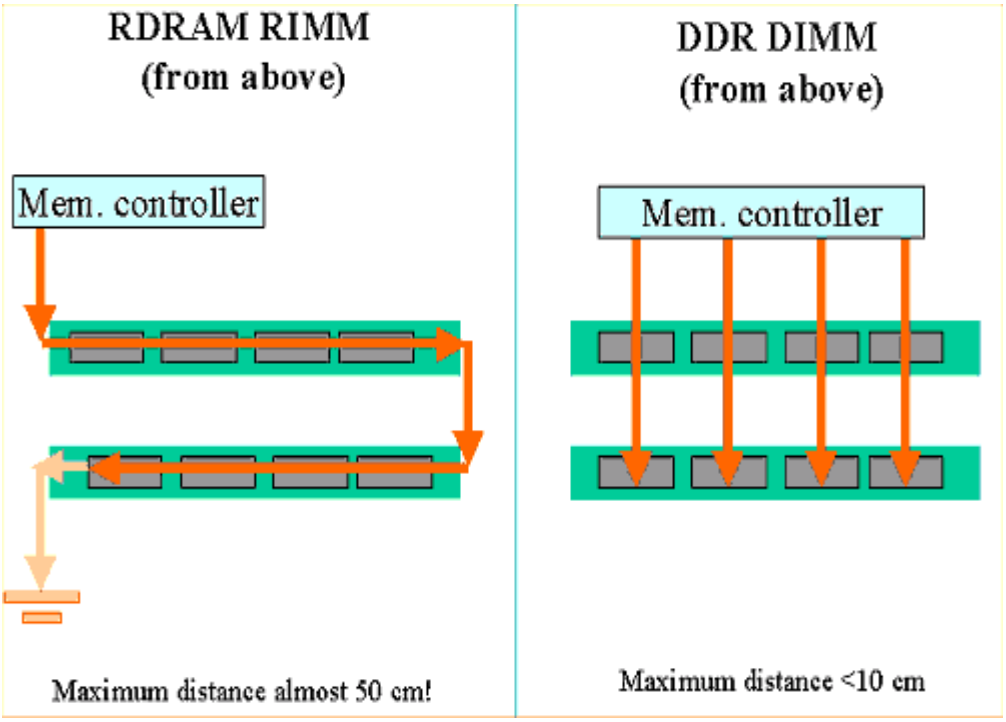
Stacked 32 Device SDRAM

Evaluation of this advantage:

- ***Extremely Important for maximum clockspeed. Allows RDRAM to attain much higher clockspeed without reflections.***

3. Advantage of DDR SDRAM: Datapath from memory controller to farthest memory chip is always much shorter

So far, the "chips serially connected to snake" topology has given us nothing but advantages. But this topology comes also with a very important disadvantage. With RDRAM, the signal has to travel up to 25 cm to the farthest DRAM chip. With DDR SDRAM, the farthest "chip" is simply all the chips on the farthest DIMM. The differences in "reaction time" between nearest and farthest chips in the DDR SDRAM topology is very small, while the difference in distance between the first and the last chip in the RDRAM snake channel is very high. As electrical signal travel with about 14 cm per nanosecond, this big difference in distance can cause quite a bit of problems. First the Address signal has to travel 25 cm to the farthest chip, and when the chip answers the right data, it has to travel 25 cm back. At 533 MHz (PC1066), the clocksignal is only 1.88 ns long, so if each chip would respond as fast as it can, data would arrive "out of order".



These enormous differences in "reaction" time must be compensated. To do this, the memory controller assigns each memory chip a different extra delay (latency) called TPARM. The closer the chip to the memory controller, the higher the extra latency. The result is that all DRDRAM chips appear to be as slow or as fast as the farthest chip. This also explains why the typical latency of RDRAM is (a bit) higher than DDR SDRAM.

Evaluation of this advantage: Important for latency. Long RDRAM channels have a higher latency than DDR SDRAM memory subsystems.

4. Advantage of RDRAM: Each Chip is Independent

Where a DDR SDRAM DIMM needs 8 or 16 chips to function and deliver 64 bits to the datapath, each single DRDRAM chip can function at his own and is able to offer the full bandwidth of a RDRAM channel. That is probably the reason why RDRAM chips found their way in for example Sony's Playstation 2 and some network devices. With only two chips, each on a different channel, you can get 3.2 GB/s (PC800) and even 4.2 GB/s (PC1066). Samsung might even produce chips that are speced to run at 666 MHz in these short channel implementations, delivering up to 5.3 GB/s in a dual channel configuration.

But this chip independency comes also with it's disadvantages. While a typical RDRAM **RIMM** does not dissipate (much) more power than a typical DDR SDRAM DIMM, a RDRAM **chip** might absorb 3 times as much current than a DDR SDRAM chip. Therefore a heat-spreader is a bare necessity on a RIMM as it spreads the heat of one chip over the total RIMM. It also means that it is necessary to make sure that chips are not always in "Active" mode, but also in "Standby" mode where it consumes much less power. In "standby" mode a chip will answer with additional latency however.

Evaluation of this advantage:
big advantage for "compact" devices such as game consoles, network devices
disadvantage for thermal management and latency.

5. Advantage of RDRAM: 32 Banks Per Device

Note: To understand the following, you should read **Ace's Guide to Memory Technology: Part 1**. The basic operation of getting data out of DRAM goes as follows. To get a certain piece of data, a row and column address is send. First you search the right row in the memory matrix, you copy this row into an amplifier (sense amp) and you chose the right column out of the sense amp. If the chipset keeps this sense amp charged, it is said that the "page" (row) is kept open. If the next piece of data comes out of the same row, you have a page hit. From our first part about memory technology you should remember that open pages allow you to eliminate the RAS latency, and thus we can reduce the total latency that the CPU sees significantly. Open pages have the same effect as a very small cache: if your next piece of data is available in the open page, you reduce latency. If not, the latency is increased.

Both DDR and RDRAM make use of the same DRAM technology. In case of RDRAM, each RDRAM chip has a 2x16d bank architecture, with "d" standing for dependent. Each chip has thus 32 banks but there is only one sense amp for each two banks. So in theory, up to 16 pages can be kept open per chip. As a typical RDRAM solution contains about 16 to 32 chips, you could have up to 256 to 512 open pages, but only in theory. First of all, most chips are not in the active state. Typically, only 4 chips are kept active, so at most 64 pages could be active at the time. This is no big deal as at a certain point, more open pages give you diminishing returns. The drawback of keeping a page open, is that if the next piece of requested data is not in the open page (a page miss), the sense amp has to close (write back) the old row before it can charge the new one. So an additional latency will occur on top of the RAS and CAS latency.

Nevertheless, on average, more open pages gives you slightly better performance. The i850(E) supports only 8 open pages, and therefore it does not take advantage of this special feature of RDRAM. This huge amount of banks per chip is only interesting in systems like the Playstation 2, which contains only 2 chips.

In a PC, this 2x16d bank architecture is pure overkill: it makes the RDRAM chip die bigger than SDRAM and decreases yields. It is clear that 4i (4 independent banks) RDRAM architecture is much better adapted to the needs of the PC market. In similar volumes, 4i RDRAM is less than 10% more expensive to produce than DDR SDRAM, while there is no (measurable) performance hit compared to the current 2x16d RDRAM.

Unfortunately, Intel's Tulloch chipset has been cancelled, so 4i RDRAM will probably be limited to game consoles or network devices. **The SIS' R658 chipset however supports 4i RDRAM**, which still leaves a small door open to 4i RDRAM.

Evaluation of this advantage:

1. it is in fact a DISADVANTAGE for the PC market for the following reasons:
- bigger die and lower yields result in higher production cost.
- no performance increase as chipset does not support many open pages.

2. Small advantage for small systems with few chips (i.e. PlayStation 2)

6. Advantage of DDR SDRAM: Better Chipset Support

Intel has stopped developing chipset for RDRAM technology, and has focused most of its chipset engineering resources on DDR SDRAM and DDR-II. The result is that the current i845E chipset is more advanced than the i850E chipset.

First of all, the i845E supports 24 open memory pages, the i850 only 8. One double sided DDR SDRAM has at most 8 banks, and the i845E supports two double sided DIMMs. So in practice the i845E can keep up to 16 pages open, twice as much as the i850(E).

The i845E chipset has also 12 level deep in order queue (IOQ, a chipset buffer) so it is capable to send up to 12 consecutive 64 bit reads to the CPU. The i850E chipset supports only an 8 level deep IOQ.

Thirdly, the i845E refresh is also more dynamic. As you know DRAM cells have a very small capacitor. Those very small capacitors leak current so they have to be refreshed from time to time (about every 64 ms). Intel's flexible memory refresh is able to postpone the refresh of open pages if data is requested from that page. This way, the i845E is able to eliminate an extra latency that can occur from time to time.

Evaluation of this advantage: Performance wise these optimizations are small (1-2%?), but DDR SDRAM's potential is more efficiently used.

7. Advantage of DDR SDRAM: Critical Word Support

DDR SDRAM supports critical word first bursting. This reduces the latency that the CPU sees. When the CPU requests to refresh one of it's cache-lines due to a cache miss, the chipset will deliver typically a 64 byte word. But the instruction that caused the cache miss probably needs only 64 bit (8 bytes, the critical word) of those 64 bytes to continue execution, and therefore this piece of data will get priority. In other words, the critical word is immediately dispatched to the CPU, while the rest of cache-line comes later.

RDRAM works with 16 byte sub-blocks, and while not the whole cache-line (64 bytes) must be loaded into the cache to get the critical word, the whole sub-block must be read.

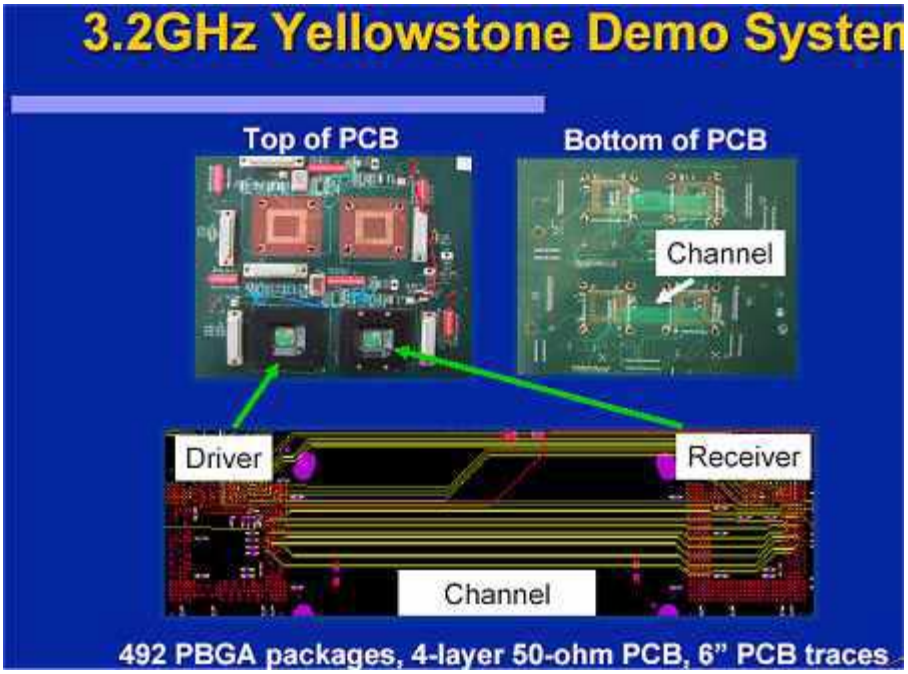
Evaluation of this advantage: A small but not negligible extra amount of latency for RDRAM.

A Look Into the Future: DDR-II and Yellowstone

What is so interesting about knowing the disadvantages and advantages of the current memory technology, is that you get a better idea what the next memory technology is all about. To some degree, DDR-II builds further on the architecture of DDR, but incorporates some of the advantages of RDRAM. DDR-II have smaller stubs, has programmable impedance (RDRAM compensates independence by making the traces narrower at certain places) and on die termination to avoid reflections. It also uses FBGA packaging to reduce the capacitive load. Interesting is also that in order to reach the higher clockspeeds of 266 MHz DDR (DDR-II 533), latencies go up significantly. Latencies were estimated from 3-4-4 to up to 5-5-5 (DDR-II 533), and DDR-II doesn't seem to support "critical word first" anymore. To compensate, DDR-II will feature 4 bit prefetch instead of 2 bit prefetch.

Also interesting is the fact that the die overhead of a DDR-II chip is about 10% higher than SDRAM. For comparison, DDR has a 5% bigger die overhead compared to SDRAM, 4i RDRAM between 5 and 10% and the current 2x16d RDRAM has a die overhead that is 15% bigger. DDR-II for PC desktops should reach volume production in the second half of 2003.

What about RDRAM? Well, RDRAM will continue to exist in 2003, but without proper chipset support, it will probably fade away around the time DDR-II really hit the market. So it seems that DDR-II will phase out both DDR and RDRAM, making both solutions "obsolete". There is still a small chance that RDRAM will continue to exist alongside DDR-II. The current PC1066 RDRAM can reach - with the proper clock generators - PC1333 (667 MHz) speeds, so Samsung and Elpida have announced that they are studying PC1333 RDRAM and even 800 MHz or PC1600! Of course, it will depend on SIS and Intel whether or not this PC1333 RDRAM will be available in the PC Market. It is most likely that PC1333 4i RDRAM will probably be used in other markets (consumer, network) than the PC Market. The same thing can be said about 64 bit (4 channel) RDRAM, about which we reported a year ago.



This doesn't mean that RDRAM is completely eliminated from the PC market. Rambus has already developed a new technology with codename "Yellowstone", with products expected in a few years. At first this technology should reach 400 MHz, and use octal data rate, and therefore could be called 3.2 GHz memory. At first, this memory technology should be able to deliver at least 12.4 GB/s. But the initial market of this technology will be the graphics markt. With a 128 bit interface (which is almost low end now in the graphics market) Rambus promises up to 100 GB/s of bandwidth. At this point this technology limits to prototype demos and marketing promises, but it must be said that no known memory technology has such an impressive roadmap.

As has been the case in the past, game boxes and graphics cards have incorporated the most advanced memory technology first, and Yellowstone will be no exception. It seems that Yellowstone technology won't be found in the main memory systems of pc's before 2005.

Conclusion

Glad you made this far. Now that we understand memory technology much better, let us see what practical knowledge we can get out of all these technical details. A few interesting points:

1. Overclockers should prefer one big high capacity DIMMs instead of smaller DIMMs.

Should you get two 256 MB or one 512 MB DIMMs? More DIMMs gets you only one advantage: more open pages are possible. But that will give only a tiny negligible performance boost (with 4 bank interleaving enabled, otherwise there is probably no performance boost at all). Each DIMM you add, results in a more (capacitively) loaded address and data bus, and therefore a less overclockable system.

You might find that your motherboard throttles back your memory speed back to assure stability.

2. DDR is not built to reach the clockspeeds of RDRAM.

Rambus is a very controversial company, but it must be said that their engineers know how to build high-speed memory interfaces. The small stub size, the snake alike data and address highway, it is quite an interesting design with still a lot of potential. It is open for discussion, but it is possible that this kind of memory topology will replace the "parallel with stubs" topology in the long term. In the long term, simply adapting SDRAM to reach slightly higher clockspeeds (DDR-II) might not be enough. Just like Fast Page RAM, which was not scalable enough and replaced with SDRAM.

3. RDRAM (should?) could have been a much better solution for the PC Market than it is today.

DDR SDRAM is the most cost effective and thus most popular solution for the PC Market. RDRAM which was build to satisfy all kinds of markets, was too complex and should have been made more cost effective before bringing it to the market in 1999. A 4i RDRAM RIMMs together with an improved i850 chipset (The dead before born Tulloch) would have been much more competitive.

4. DDR-II at 533 MHz will probably cause less trouble than some DDR at 400 MHz.

DDR-II is adapted to high clock speeds thanks to on-die termination, FBGA packaging and programmable impedance. DDR400 however is mostly a high speed-binned form of DDR333. The result is that DDR400 will probably be limited to lower capacities and to (more expensive) motherboards that use top-notch components if you want good stability. We will investigate this matter in more detail in our upcoming review, but is safe to say that DDR400 is for the overclocker and die-hard hardware freak community, but not for the general public.

5. Small stubs are the future.

RDRAM high clockspeeds is possible thanks to the small stubs. As DDR-II's stubs are smaller, this will allow higher clockspeeds as well. DDR-III might have no stubs at all.

6. Share your opinion!

We do not pretend that this article to be totally complete or perfect, so please don't hesitate to your insight or opinion. We welcome all constructive criticism and feedback on our message board...