DRAM Pricing – A White Paper

Introduction

DRAM (Dynamic Random Access Memory) is the world's best-selling electronic memory technology, commanding over half the total memory market in 2001.[28] It's not the fastest memory available, nor the densest; it uses more power than some technologies, is more susceptible to errors than others, and loses all of its data whenever it loses power. So, why is it the best seller? One simple reason: price. DRAM costs less, per bit, than any other memory technology.

DRAM was patented in 1968. It first hit the market in 1970 as a 1Kbit memory chip that sold for US$10.[20] Through the ensuing years, DRAM bits-per-chip rose dramatically, and the price-per-bit fell just as dramatically; memory speed improved, manufacturing costs fell, and the market expanded enormously. Measured in bits sold, the DRAM market grows larger every year;[40] however, measured in dollars, the history of the market is erratic. The volatility is not in the demand, but in the price. This paper examines a few of the most important factors in DRAM pricing.

Tracking DRAM Prices

Seventy-five to eighty percent of DRAM memory is sold by contract. Large customers, such as big-name PC makers, negotiate with memory suppliers for bulk contracts at favorable prices.[9] These prices are difficult to discover, as the contract parties are unlikely to reveal them to the press.

The remaining 20 – 25% of DRAM is sold on the "spot market"[7] for whatever the market will bear. The major buyers in this market are speculators and smaller PC makers who sell generic "white-boxes".[10] Spot market prices are openly advertised and very easy to track. Although the spot market is extremely volatile and does not reflect contract prices, it is a useful barometer of market conditions.[12]

Supply and Demand

DRAM's infamous boom-and-bust cycle is driven by supply, not demand.[3] Supply tends to increase in stepwise fashion, experiencing a massive surge with the opening of each new "fab" (silicon wafer fabrication facility). Memory firms have huge startup and R&D costs;[33] building and equipping a new fab takes about 3 years and costs over US$3 Billion. When a new fab comes online, producing tens of millions of chips per month,[15] the world supply of DRAM chips increases by several percent. The sudden glut drives prices down. As demand grows to match the supply, prices stabilize. Eventually, demand outstrips supply, prices rise, and manufacturers invest in fabs again. This cycle has repeated for decades,[6] and is apparently here to stay.[27]

Not all segments of the DRAM market are affected equally by this cycle. There is always one type of chip that is considered the market standard; historically, this standard has been determined by the memory requirements of PC makers. The standard chip has extremely high demand (as much as 76% of the DRAM market[4]), and is therefore made in huge quantities. Manufacturers over-produce the standard chip[16] in order to benefit from economies of scale. This makes the standard chip less expensive (per bit) than all other DRAM chips. These high-volume, low-cost chips are extremely vulnerable to the price swings of the boom/bust cycle;[39] with very little product differentiation, customers choose their DRAM sources based almost solely on price.

Other DRAM chips vary from the standard in capacity, speed, packaging, power consumption, interface, or some other specific factor; this differentiation limits the appeal of these chips to specialty markets, but it also insulates them from the brunt of the price swings. Production quantities, and hence prices, tend to be stable.

DRAM Pricing – A White Paper

Density Cycle

Technological advances allow manufacturers to craft smaller and smaller features on silicon wafers. When a two-fold capacity increase becomes possible, a new "generation" of DRAM chips is introduced. The new generation offers twice the density of the previous generation – that is, twice the memory capacity in a chip of comparable size. At first, production is limited and costly, but the high-density chips are nonetheless snapped up by customers who crave high-density memory;[38] these chips rarely appear on the spot market.

Cramming more bits onto each chip offers several advantages: systems using high-density chips can fit more memory into the same space, use fewer components, and draw less power. The designs for these systems are simpler, consuming less time and money in the design and debug cycles.

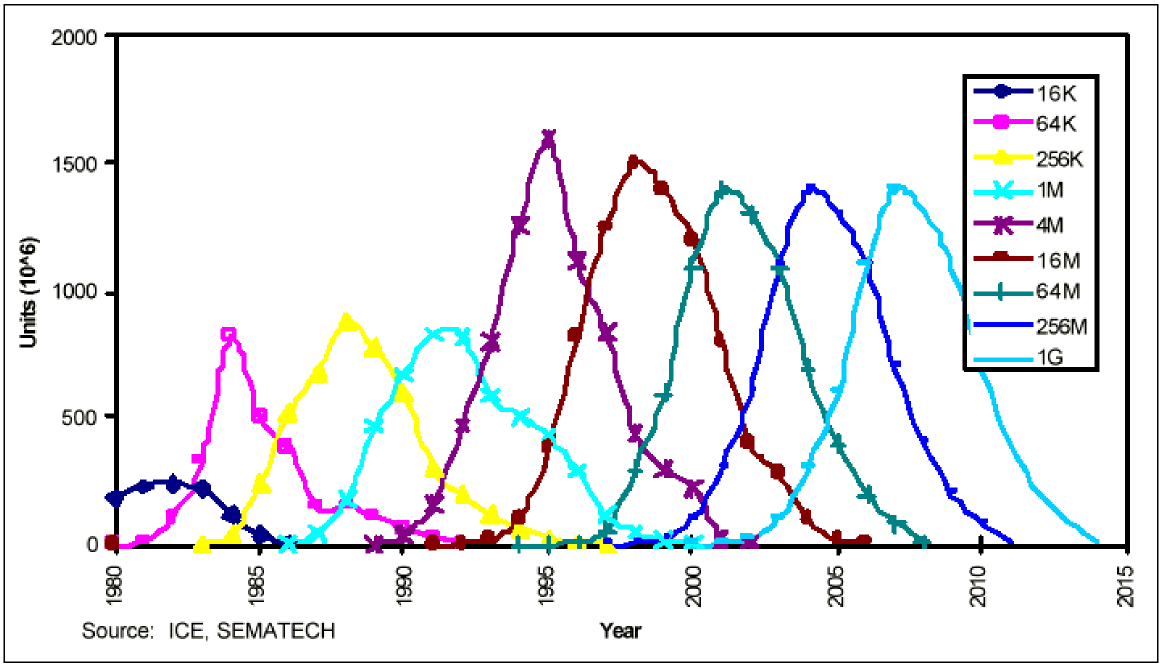
Manufacturers continually fine-tune the production of the new generation, increasing yield and trimming costs; prices decline and production rises. Gradually, the new generation begins to appeal to a wider market. As these trends continue, the denser chip's price-per-bit begins to challenge the market share of the current standard chip. Finally, a price-per-bit crossover is reached, and the higher density becomes the new standard.

When a new generation is established as the standard, two things happen: the new standard's costs continue to fall, and the previous standard's costs begin to rise. Economies of scale cease to apply to the old standard, which settles into a niche market. On the other hand, the new density standard has a built-in cost advantage: it uses less silicon (per bit) than the old chips, and can therefore drive the cost-per-bit to a new all-time low.

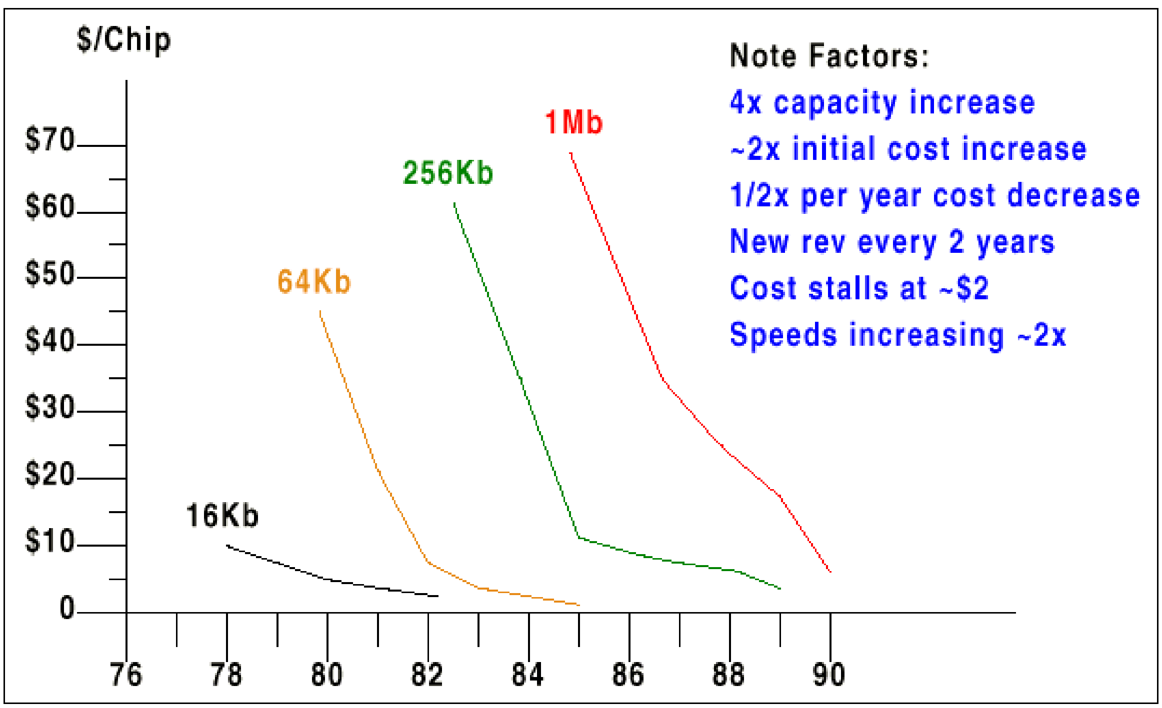
This "life cycle" has been repeated often enough to exhibit some predictable patterns:

- A new generation appears every two or three years.[45][37]
- Higher-density chips will sell for more than lower-density chips.[3]
- The first supplier to provide a new generation can enjoy high profits until competitors appear.[27]
- Each new generation starts at a record high price-per-chip, and declines to a record low price-per-bit.[45]
- A new generation needs 18 months to reach cost-effective production.[24]
- A new generation takes about 6 years to "peak."[37]
- Lowest costs occur during highest production, or soon thereafter.[3]
- Old generations never die; they just settle into niche markets.[3]

The graphs on the next page illustrate the effects of the DRAM density cycle.



DRAM Unit Volume by Generation [37]



DRAM Costs [45]

The Fragmenting DRAM Market

For decades, the driving force in the DRAM business has been the PC, the single largest market for DRAM chips.[34] The average amount of main memory in PCs grows year by year – in 2001 alone, the average DRAM in PCs increased by 30%.[34] Even when memory prices rise and/or PC sales fall, the amount of memory per PC does not retreat.[40][11] DRAM manufacturers naturally develop new chips with an eye to the needs and wants of the massive PC market. The primary demands of this market are higher density, lower price, and greater speed.

Recently, a small but rapidly growing market segment has drawn increased attention: telecommunications ("telecom") is the hot new DRAM market. In 2002, telecom may consume 10 – 20% of DRAM production;[34][19] within 4 or 5 years it may equal the PC share.[28] Selling to the telecom market has attractive advantages: prices are higher than in the PC market[23] and margins are fatter.[19][8] Because telecom products (routers and switches) have longer design cycles[8] and lifetimes[23][14] than PCs, memory needs are more predictable. Designing special DRAM chips to meet the needs and wants of the telecom market makes good business sense. The primary demands of this market are higher density, lower power consumption, and greater speed.

The definition of "greater speed" differs in these two markets. PCs want higher *data rate* while telecom wants lower *latency*. "Data rate" refers to the speed at which blocks of contiguous data can be read or written; "latency" refers to the time it takes to access each specific data location. PC applications tend to use large, sequential memory areas; telecom applications use small, widely separated memory areas. These differing needs are beginning to fragment the market, suggesting that future "standard" DRAMs will find it harder to reach super-low commodity prices.[28]

The fragmenting market introduces new niche segments with real economic value. One of those promising segments is specialty high-speed memory tailored for each market, as described in the following sections.

PC Memory Speed: Data Rate

The cost of memory makes up 5 – 6% of the cost of a PC.[6] Using standard DRAM keeps the cost down, but it also strangles the speed of today's powerful processors. An early answer to this dilemma was VRAM – specialized DRAM designed to speed up the video display. VRAM commanded higher prices than conventional DRAM[3] from its introduction in the late 1980s until the speed of conventional DRAM caught up with it in the late 1990s.

In 1997, SDRAM (Synchronous DRAM) was introduced, and quickly became the mainstream technology.[38] Soon a faster version appeared: PC133, one-third faster than the original SDRAM. The price was higher than standard DRAM, yet it achieved 40 – 50% market penetration before reaching price parity with the old standard.[26] By early 2001, PC133 was the new standard.

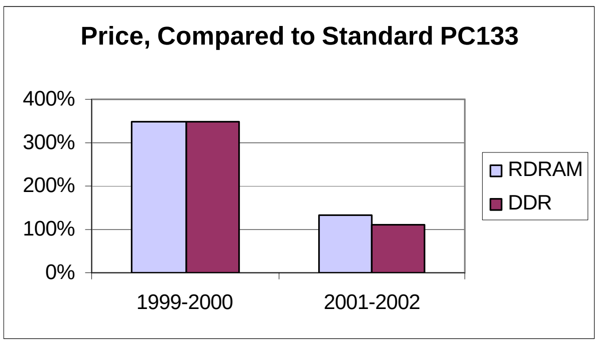
The next evolutionary advance was "Double Data Rate" SDRAM (DDR), introduced in 1998. As the name implies, its data rate (266MHz) was twice that of PC133.[25] Of course, the price was higher, too. In March of 2001, DDR was priced three or four times higher than PC133;[26] by December of that year, the price difference was still about 2-to-1, yet demand was "more than originally expected".[11] In fact, DDR captured 30% of the DRAM market in 2001.[44] By February of 2002, DDR was set to become the most commonly used PC memory, with manufacturers shifting as much as 50% of their production to the new technology.[42] As of August 2002, spot market DDR prices were still markedly higher than PC133,* and demand for DDR remained "positive".[29]

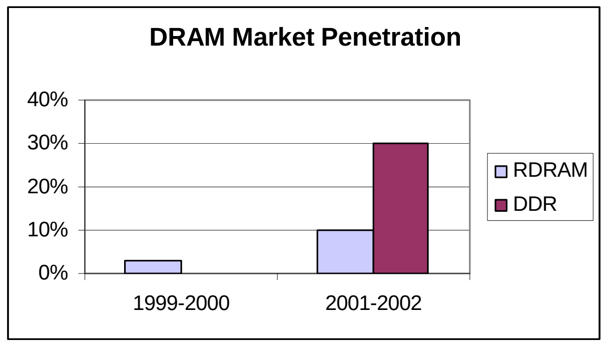
* Prices found online at www.overclockers.co.uk , www.dramexchange.com , www.fujiace.com.hk July, 2002

In the meantime, the "fast data rate" arena was challenged by RDRAM, a proprietary technology from Rambus. RDRAM was released in 1992[28] with a blazing data rate of 800MHz, but its entry to the PC market was difficult. RDRAM requires a unique interface with special controllers,[28] and can be used only in PCs explicitly designed for it. In addition, retooling a memory fab to produce RDRAM involves considerable expense.[26]

In 1997, Intel publicly endorsed RDRAM; by August of 1999, RDRAM had captured 3% of the DRAM market.[21] The price, however, remained high longer than expected; in May of 2000, it was still three to four times more[17] than standard memory, and DDR was starting to attract notice. In 2001, RDRAM market share was a mere 10%,[43] while DDR had reached 30%. By mid 2002, RDRAM was still priced 20% higher than DDR,[36] and Rambus's royalty income from RDRAM was falling.[35] RDRAM's early lead and much-touted data rate could not compensate for other factors: high prices and poor reputation,[1] resistance from designers,[28] manufacturing difficulties,[44] and an "ill-prepared infrastructure".[27] The final blow was a set of reports declaring that RDRAM was, after all, not much faster than DDR in actual PC tests.[2][32][30][41] RDRAM is not going away; even at prices too high for the larger PC market, it fits some niche markets (video games and very high-speed processors) where its data rate advantage is worth the cost.

The following two charts compare RDRAM and DDR over a rough four-year span:





This abbreviated history of the PC memory market illustrates a few generalized points:

- The extreme high end of the market will pay a hefty price for almost any increase in data rate.
- Thirty to fifty percent of the PC market is willing to pay a reasonable premium for a reasonable increase in data rate.
- The lower half of the PC market waits for commodity pricing before adopting faster memory.
- New interfaces are not easily accommodated by the PC industry.

Telecom Memory Speed: Latency

The telecom market uses many more megabytes per processor than the PC market,[5] and it wants those megabytes to display lower latency than current DRAM technologies can supply.[22] Latency is measured by the amount of time it takes to access a specific memory address. This "access time" has fallen from 300ns for the original 1103[20] to 40ns for today's DRAM[41] – still not fast enough for telecom purposes.

The lack of low-latency DRAM has caused network developers to turn to SRAM (Static RAM) for its excellent 12ns access time.[5] This tremendous speed advantage comes burdened with much lower density and a much higher price – in August of 2002, 18Mbit SRAMs sold for $50 each.* The density trade-off is a real problem for network developers; the price is hardly even mentioned. SRAM manufacturers are working to provide higher densities, but they still fall far short of DRAM density; in October of 2001, samples of 72Mbit SRAM were $180 apiece.[18]

Low-latency DRAM options have appeared recently – all proprietary and all very new, still in early (i.e. high-cost) production phases. Two of the most discussed options are FCRAM and RLDRAM:

> FCRAM (Fast Cycle RAM) appeared in January of 2000; 64Mbit samples with 20ns access time sold for $25 apiece.[13] By mid 2001, FCRAM was available in a 256Mbit chip with 25ns access time;[28] in January of 2002 these chips were priced in the $60 range.[23]
>
> RLDRAM (Reduced Latency DRAM) was announced in May of 2001 with an access time of 25ns.[28] In February of 2002, 256Mbit samples became available at a price of $54.[31]

Prices for these parts will undoubtedly fall as production volumes rise, but they are already much more affordable than equivalent SRAM. Network equipment manufacturers have concerns about standard interfaces,[23] reliability of supply,[19] and a clear upgrade path;[14] price seems to be a non-issue.

Summary

Historically, the DRAM market has been PC-driven, with the bulk of its sales in one standard chip. Manufacturers over-produce the standard chip to achieve economies of scale, driving costs and prices down in a drastic and volatile boom-and-bust cycle. Within this cycle, non-standard chips command a higher price-per-bit and maintain a more stable supply and demand system.

Recently, the market has become less uniform. The proliferation of non-PC consumer devices and the growing telecom industry may signal the end of true commodity pricing. Differentiation among DRAM products is increasing to serve the diverse needs of the market segments. With no single DRAM standard to dominate the manufacturing scene, prices are likely to reflect true costs more accurately than in the past.

New products appear at a steady rate, initially carrying very high manufacturing costs that are reflected in the price. As the manufacturing process is refined, yields improve and costs decline, allowing prices to become more attractive. If a product succeeds in capturing a large enough market share, the higher volume generates even lower costs and prices. Three factors seem to strongly hinder market acceptance:

- Significant changes to the memory interface
- Doubtful supply (few sources)
- Prices too high for perceived value

If a new technology avoids these pitfalls and offers real value (density, speed, etc.), it will support price premiums until over-supply prompts competitors to sell at a lower price.

* Prices found online at Avnet Electronics Marketing, Aug 28 2002

References

[1] AsiaBizTech: "Intel Struggles to Promote Direct RDRAM of Rambus" June 5, 2001
[2] AsiaBizTech: "Will Rambus Wage Litigation Battle Against Intel?" August 14, 2000
[3] Bureau of Economic Analysis: "Price Indexes for Selected Semiconductors, 1974-96" February 1998
[4] Cahner's Instat Group: "Electronics Year Book 2002 Edition" page 244
[5] Communications System Design: "Memory Options Explode for Network Processors" May 1, 2002
[6] Computerworld: "More DRAM vendors involved in Justice Department probe" June 21, 1001
[7] ebn: "Buyers enjoy fire sale in 256Mbit DRAMS" July 5, 2002
[8] ebn: "Closer Look: Network DRAMs are a hot commodity" May 8, 2002
[9] ebn: "Closer Look: Smoking gun in the DRAM market?" June 25, 2002
[10] ebn: "Closer Look: Spot DRAM prices don't tell the whole story" July 16, 2002
[11] ebn: "DDR memory prices soar pending Intel introduction" December 27, 2001
[12] ebn: "DRAM prices tumble, but Q3 shortage looms" May 24, 2002
[13] ebn: "Fujitsu's FCRAM tries to spark application-specific DRAM market" January 18, 2000
[14] ebn: "Future unclear for specialty DRAMs" July 19, 2002
[15] ebn: "Micron discusses future fab plans amid poor Q3 results" June 26, 2002
[16] ebn: "Opinion: DRAM group may face anti-trust issues" July 17, 2001
[17] ebn: "Rambus availability, yields could dictate Intel chip performance" May 5, 2000
[18] ebn: "SRAM densities jump to satisfy network gear makers" October 12, 2001
[19] EE Times: "Elpida, Hynix entries further muddy network DRAM waters" April 19, 2002
[20] EE Times: "Ferrite cores: to remember forever" Special Issue, 2000
[21] EE Times: "IDF: Micron, Rambus address DRAM cost issues" August 30, 1999
[22] EE Times: "Juniper founder sounds off on QoS, memory blocks" November 12, 2001
[23] EE Times: "Low-latency DRAM market at risk of fragmenting" January 30, 2002
[24] EE Times: "Memory makers slow to pursue faster DDR DRAMs" January 31, 2002
[25] Electronic Business Asia: "Cover Story: Sweet Memories" April, 2001
[26] Electronic Business Asia: "DDR Chipsets: First Off the Block" March 2001
[27] Electronic Business Asia: "The Return of the Chaebol" June 1999
[28] Electronic Design: "DRAM Derby Heads for the Stretch" June 18, 2001
[29] Fuji Ace: "HK DRAM Spot Market Commentary" August 23, 2002
[30] HardwareCentral (www.hardwarecentral.com): "The Future of RAM: RDRAM vs. DDR Explained" February 20, 2000
[31] Infineon Technologies AG: "Press Release" February 11, 2002
[32] InQuest Market Research: "DDR vs. Dual Channel RDRAM" April, 2000
[33] the inquirer: "Dramurai criminal case starts over DRAMatic price" June 20, 2002
[34] the inquirer: "Memory market data shows DRAM will recover" June 20, 2002
[35] the inquirer: "Rambus profits soar on increased DDR royalties" July 12, 2002
[36] the inquirer: "Samsung 1066 RDRAM costs small fortune" June 12, 2002
[37] Intel: "300mm Conversion Timing" Q2 1997
[38] Kingston Technology: "Chip-Scale Packaging and High-Speed Memory" January 21, 2002
[39] Kingston Technology: "Product Marketing Newsflash: DRAM market price predictions abound" November 1, 1999
[40] Morgan Stanley: "Global DRAM Quarterly Supply/Demand Update" January 28, 2002
[41] Overclockers (www.overclockers.com): "RDRAM/DDR" May 21, 2000
[42] PC World: "Faster DDR Memory to Move Into Mainstream" February 22, 2002
[43] Rambus: "Financial Releases" July 12, 2001
[44] Taipei Times: "VIA's bet on DDR pays off" February 21, 2002
[45] University of Utah Department of Computer Science (course materials at www.cs.utah.edu): "DRAM Costs" 2000