Rambus DRAM (RDRAM)

- Goal
  - High Density, Low Cost, High Bandwith DRAM
- To achieve high bandwidth to memory interface can either:
  - make interface to memory faster
  - make interface to memory wider
- Wider => More Chips or More Pins => More Cost
  - e.g., “wider is NOT necessarily better”
  - more chips also decreases reliability

7/10/00 1

Speeding up the interface

- Many benefits to speeding up the interface instead of widening the datapath
  - Fewer pins, fewer chips => less cost
  - higher reliability
- Rambus DRAMS or SyncLink DRAMs uses 400 Mhz bus based on Gunning Transceiver Logic (GTL)
  - Basically same approach as used with Pentium II local bus

7/10/00 2

Pentium II GTL Bus (Host Bus)

- Gunning Transceiver Logic (GTL) used for Pentium II local bus (66Mhz now, 100Mhz later)
  - GTL bus is open drain bus where all runs are terminated
  - Termination voltage (Vtt) is 1.5 v.
- GTL bus is a differential bus with only wire!
  - Vref used by all receivers, drivers
    - Vref (1.0v) is 2/3 of Vtt .
  - Voltage swing about Vref is +/- 200 mv.
    - Less voltage swing => higher speed, less noise margin

7/10/00 3

GTL Bus (continued)

- Interconnections on a GTL bus are transmission lines so interconnect topology, termination very important.
- Interconnection is point to point to avoid stubs (stubs generate reflections)

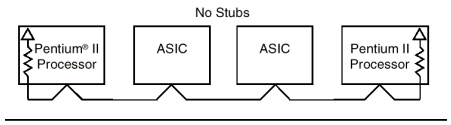
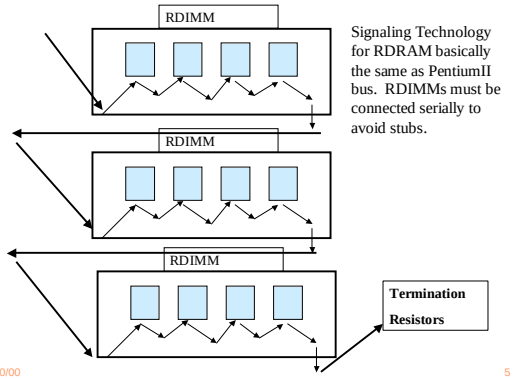


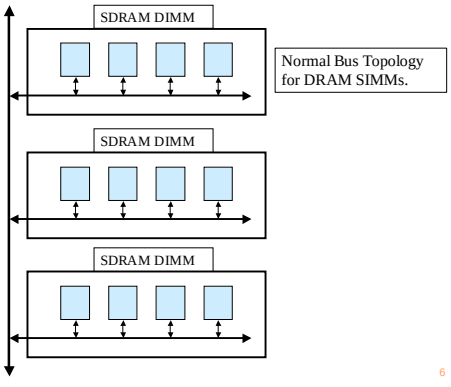
Figure 2. GTL+ Bus Topology

7/10/00 4



Signaling Technology for RDRAM basically the same as PentiumII bus. RDIMMs must be connected serially to avoid stubs.

7/10/00 5



Normal Bus Topology for DRAM SIMMs.

7/10/00 6

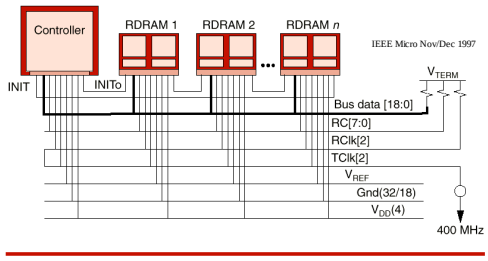


Figure 3. Direct RDRAM system.

18 bit wide external data bus which expands into 128 bit wide datapath internal to chip

7/10/00 7

Bandwidth

- External bus is 18 bits wide (2 bytes + 2 parity bits)
- External clock cycle is 400 Mhz, but data is clocked on each edge
  - Actually, external clock is a differential pair and data is sampled at each crossing
- Total Bandwidth is 1.6 GBytes/s
  - 2 bytes * 400 Mhz * 2 edges => 1.6 Gbytes
  - Initial configurations are 4 M x 18 (72 Mbits)

7/10/00 8

Comparison

- Recall that the Voodoo2 board had a 2.2 GB/s memory interface, used fast EDO DRAM
  - 12MB total, took 24 chips (two rows of 12, interleaved, used 256K x 16)
  - Would only need two RDRAM chips
    - 16 MBytes total (actually more than this, each byte is '9' bits).
    - Data Rate => 3.2 GB/s
    - Drawback is that we would need two separate RDRAM controllers, one for each chip if we want to double bandwidth.
- Some new Digital Signal Processors (DSP) already support the RDRAM interface

7/10/00 9

Uneven Net Loading in Conventional DRAM

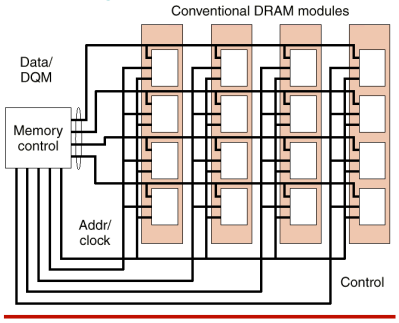


Figure 1. Nets in a conventional DRAM system.

IEEE Micro Nov/Dec 1997

7/10/00 10

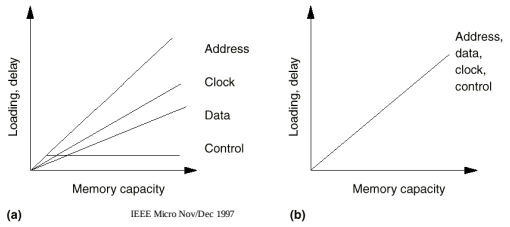


Figure 2. Signal loading versus memory capacity for SDRAM (a) and Direct RDRAM (b) systems.

Loading increases linearly as # of RDRAM chips increase. Makes for easier timing design.

7/10/00 11

Internal Architecture

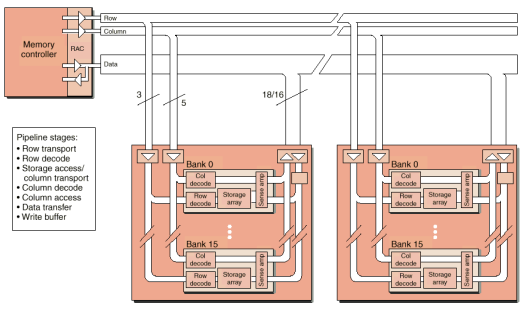
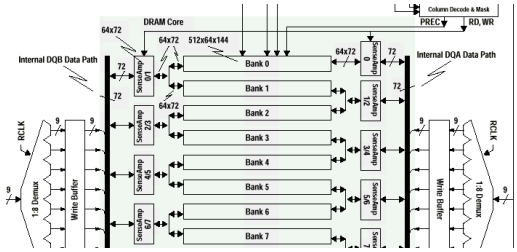


Figure 5. Direct RDRAM microarchitecture.

IEEE Micro Nov/Dec 1997

7/10/00 12



Portion of internal architecture (4M x 16 or 4M x 18)

16 banks of 512 rows of 64 dualocts (1 dualoct = 16 bytes = 128 bits)

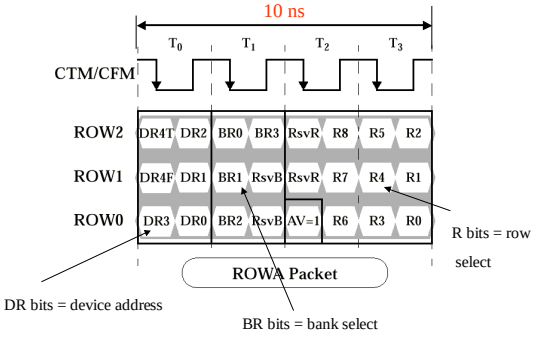
2^4 (banks) * 2^9 (rows) * 2^6 (dualocts) * 2^7 (one dualoct) = 2^{26} (64 Mbit)

A dualoct is the smallest addressable unit.

Addressing

- 3-Bit Row bus used to give commands to RDRAM
- ROW Activate command used for read
  - 4 clocks transfers 8 groups of 3 bits over Row bus due to dual edge clocking (24 bits total)
  - 24 bits in Row Activate command split between device address (6 bits), bank select (4 bits), row select (9 bits), and reserved bits
- There are no chip select lines, internal register holds device address
  - All chips monitor bus - if bus device address matches internal id, then chip is selected.

Row Activate Command



Deep Pipelining => High Latency

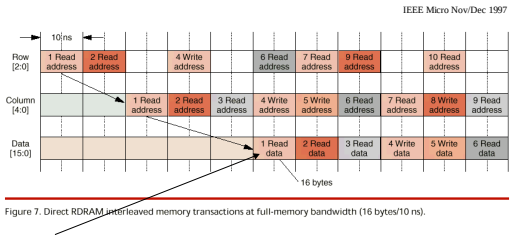


Figure 7. Direct RDRAM interleaved memory transactions at full-memory bandwidth (16 bytes/10 ns).

16 bytes transferred because 4 clocks * 2 edges * 2 bytes/transfer

(external bus is 16 or 18 bits wide). 20 clock latency

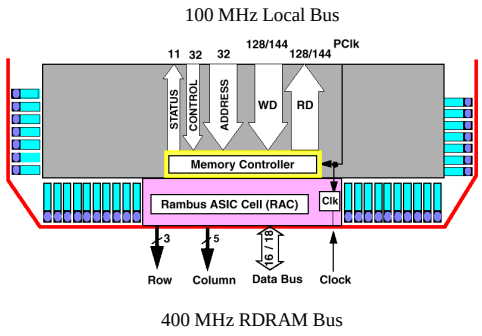
7/10/00 16

Maximum Bandwidth

- Note that maximum bandwidth with one RDRAM controller is 1.6GB/s.
  - Only one RDRAM chip can be active at a time on RDRAM bus.
  - More RDRAM chips increase capacity, not bandwidth.
    - With normal DRAM and SDRAM, can increase bandwidth by just adding more DRAM chips in parallel from same DRAM controller
  - To double the bandwidth, would need two separate RDRAM controllers

7/10/00 17

RDRAM Controller



7/10/00 18

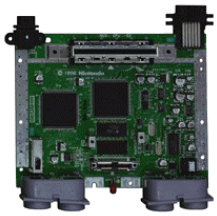
Nintendo 64

4 major chips:

MIPS RS4300i CPU

Reality Engine (Graphics)

Two RDRAMs



Memory bandwidth of 562MB/s, 31 pin interface to Memory controller.

Memory took small amount of board estate, pin count.

Used first generation RDRAMs.

7/10/00 19

The Future of RDRAM

- Intel's chipset (i820) for 800Mhz+ Pentiums currently supports RDRAM.
- At this time (Summer 2000), the promised performance advantage over SDRAM and double-data rate SDRAM (DDR-SDRAM) not been shown yet.
- RDRAM still more expensive than SDRAM, DDR-SDRAM
- Using multiple RDRAM channels, can get extremely high data bandwidths
  - Bandwidth = N * 1.6 GB/s where N is the number of channels

7/10/00 20