Some Thoughts on DRAMs

Mark Horowitz

My Background

Start with a disclaimer:

- I am not really a DRAM expect.
    - I am not a DRAM designer, or even worked for a company that makes DRAMs
    - I am not a market analyst -- I do not know how many DRAMs are sold
    - My main research focus is not DRAM related
        - Work in digital IC design, architecture and tools

But:

- I am interested in DRAMs
    - Have an interest in IC fabrication / economics
    - Am involved in Rambus
        - Have worked with companies making DRAMs
- I have an opinion about almost anything

Overview

- DRAM overview
  - Technical Issues
  - Business Issues
- Rambus
  - What it is
  - What was hard
- Thoughts on Merged Processor Memory
  - Need to focus why it might be an advantage
    (just cause it is possible doesn’t mean its a good ideal)

DRAMs

- Dense memory
  - Current generation is 16Mbit
  - Store value on very small capacitors
- New generation every 3-4 years
  - Features shrink 0.7 each generation, 2x in area
  - Density is growing 4x in a generation.
  - Cell efficiency is hard to improve, so chips are getting larger
- Used where cheap memory is needed
  - Not fastest, or easiest to use
  - Comes in very simple (cheap) packages
- Today 50% of all DRAMs go into PCs

Technology

- Storage capacitors must be around 30fF
    - But cells are getting smaller (since cell area is shrinking)
    - Cell capacitance is proportional to L W / t, so it should shrink
        - Cells are getting more 3-D
        - Trenches, or stacked poly structure

    Requires special fabrication technology to make capacitors
- Must maximize the charge you store on the capacitor
    - Eliminate the threshold drop on access device
    - Drive the gate of access transistor ABOVE Vdd
    - Transistors must be able to handle these high voltages
        - Thicker gate oxide than you might expect
- High yield technology
    - Simple as possible
    - Not very many metal layers (4M - 1 metal; 16M is 2)
    - Transistors not very aggressive

Circuits

Non-trivial design

- Small sense signal
  - Ccell/Cbitline

    Partition array to maintain signal
  - Folded bitline for differential sensing

    Twist them for noise reduction
- Destructive readout
  - Sense every bitline
  - No room for sense amps

    Share them

    Keep them very simple
- Dynamic storage
  - No minority carrier injection
  - Substrate bias

    Substrate noise, latch up

Mega-bandwidth Fantasy

- Read 2K (or more) bits at a time
  - If we can use all of them have lots of bandwidth

- But you can’t (unless you change the core)
  - Sense amps are embedded in core
  - Wiring space is pretty limited

- Even if you did change the wiring,
  - Now you have more long wires switching
    Noise issues
    Power

DRAM User

- Wants many things
  - Faster access, more memory bandwidth
- But really only cares about one thing
  - What is the cost per bit
  - Features are generally worth < 10% in main memory business
- Also likes the fact that DRAMs are commodities
  - Do not like single source
  - Want the parts to be off the self (not good at predicting volumes)

    Don’t want to guarantee volumes, but wants access to parts if needed

DRAM Business

- Around 30 Billion dollars this year
  - Greater than 1B units shipped
- A large number of providers (> 10)
  - Largest (Samsung) around 15% of market
  - Very competitive (sell at market price)
- Huge capital investment
  - 2B for a monster fab
  - Not for the meek

    Make (or loose) large amounts of money
- Large throughput, minimize cost per wafer
  - Long latency through the fab, slow response time
  - Want few part types
- A lot of inertia
  - Do not want to screw up the business

DRAM Summary

- Technology is specialize for building capacitors cheaply
  - Both transistors and metals are sacrificed
  - Not best technology for logic

- Circuit optimized for current architecture
  - Getting more bandwidth requires changes to the core
    Add risk and power
  - Noise is a serious concern

- Huge business that changes slowly
  - Technology change is rapid, but architecture of DRAM was set at 4K generation. Changes off the projected course is very slow
  - Both the consumer and supplier are conservative

Rambus

- Started in 1990
- Goal -- Solve the memory bandwidth problem
  - DRAMs were getting larger
  - Systems were getting faster
  - Number of DRAMs/system shrinking
  - Bandwidth of DRAMs growing slowly
- Business model
  - Can't resell DRAMs
  - Can't manufacture DRAMs
  - License the technology

Rambus Idea

- Change the DRAM interface to a high-speed bus
  - 500 Mbit/sec-pin
    Removes the pin bandwidth problem
  - Create a block transfer system
    Get complete cache line from a single DRAM

- Engineer the system level solution

- Place this interface on DRAMs and ASICs

- Result is a 500Mb/sec interface

Reality

- Needed to make modifications to DRAM core
    - Could not supply the data rate that the interface needed
- Power goes up in order to support the modification and the interface
    - Need to worry about still fitting in a plastic package in a fanless box
- DRAM processes are slow
    - Since your section yield must be 100%, must work over all corners
- Things don’t always work on first silicon
    - Often takes a few tries
    - And this was modifying existing cores

But you can make it work, and we did.

Rambus Status

- Widely licensed
  - NEC, Toshiba, Hitachi, Oki, Samsung, LG, etc.

- Available at 500Mb/sec
  - Working on increasing the frequency

- Cost modest
  - But too expensive for main memory at 16M
  - Finding uses in graphics, and small memory systems

Starting to appear in products (yea!)

Merged DRAM Logic

Need to ask one question

- What is the cost advantage?

  Over the best conventional solution

- Cost includes package, test, design
  - Less mm, is not necessarily better
  - Volume can have a large influence on cost

Which is Better?

- Merged DRAM with graphics controller

- Graphics controller with external DRAM?

Bias

I believe that markets decide the fate of products (and ideas)

- Winning ideas are often the least risk way of meeting the constraints
    - Not best method
    - Not best results

- Leads to evolution by capitalism

    -- success gives you money to continue your work

    -- need to have a road map from here to there

- Cost really matters

    -- but cost has many components

    -- volume is a big issue

Processor is Free Myth

What processor are we talking about?

Moderate - and up processor is not free

- Power
  - Processor plus improved DRAM will be hard to cool in plastic
- Design costs
  - Need to modify the DRAM, and create the processor
- Area
  - Need to consider caches, as well as processor
- Testing
  - Need to test processor and memory

Doesn't feel like it will be free to me

What is Free

- Simple low performance processor / micro controller
    - RISCII, MIPSX
    - 3K x 3K features on a side (<1 mm^2)

- Low power

- Low bandwidth requirements

Form a new kind of microcontroller

- But what would you do with it?

Where Merged DRAM Logic is Useful

- Check out the IBM paper at ISSCC

- Use DRAM for SRAM tasks
  - Don’t replace DRAMs (too cost competitive)
  - Use DRAMs instead of more expensive stuff

- Need to use enough DRAM so density difference matters
  - Need to use different fabrication
  - Probably over 50% DRAM

- Build a complete 1MB external cache chip

Summary

- DRAMs are very specialized parts
    - Balance of cost, power technology
    - Enormous business

- Business relies on large volumes
    - Justify capital investment
    - Amortize the design costs

- Less things are free than you might think

- Lower volume parts will be hard to launch