Server Memory Power Trends

Sam Patel
Principal Member of Technical Staff
Memory Sub-system Architect

AMD

Server Memory Forum Shenzhen 2012
Outline

- Server memory – market demand
- Historic trends for power reduction
- Future trends for power reduction
- Future power optimization focus
- Questions for consideration for future non-charged based RAM
Server Memory Demand is Accelerating...

- Large capacity
- High throughput
- Low latency
- Low Power

- And there is less room for trade-offs...
Power is Cost

- Each bit of data that is transferred to/from memory has a cost
  - xx PJ/bit

- Each bit of data that is stored in memory has a cost
  - xx PJ/bit

- Total power cost can be up to 50 to 60% of total system cost for some type of server systems
Historic Trends to Reduce DRAM Power

Major contributors

- DRAM voltage reduction
- DRAM process technology improvement
DRAM Voltage Reduction

Linear voltage drop

Voltage migration is slowing down
Process Technology (nm)

Speculative Process trends

Process migration is slowing down

Each year new process technology node

DRAM IDDx reduction up to 15% with process technology improvement
Despite non-linear voltage reduction and process technology improvement in future, power trends must continue...
Future Trends to Reduce Memory Power

- **New technology adoption**
  - 3DS technology (elimination of multiple sets of I/Os, DLLs, termination, etc.)
  - Wide I/O bus architecture with 3DS technology

- **Rich feature sets in DRAM, controller and register/buffers**
  - I.e Termination schemes, external voltage supply for charge pumps, page size, etc.
  - Integration of support chip functions; power saving modes and application specific programmability in support chips

- **DRAM density growth**
  - Maintain GB/system growth with fewer DIMMs on a channel (4 → 3 → 2 → 1 and enable non-DIMM application with 3DS/wide I/O bus technology)
Future Power Optimization Focus

• Ability to scale DRAM voltage dynamically
  – Fits nicely with dynamic frequency scaling
  – Similar to P-state transition mechanism in CPUs/memory controllers

• Fast exit from idle & various power management states
  – Longer latency leads to idle power consumption in DRAM, controller and support chips

• Efficient memory bandwidth utilization
  – Hidden housekeeping operation such as refresh, calibration
  – Improvement on bus turn around due to I/O as well as DRAM core timings
  – Improvement on DRAM core timing parameters
Questions to consider

• Process technology migration will eventually lead to limitation in charged based RAM (DRAM)
  – How and when will market exploit other technology in server market in mass volume?
    • Phase Change Memory (PCM RAM)
    • Spin Transfer Torque RAM (STT RAM)
    • Memristor Memory (MRAM)
    • Etc..

• How much power reduction non-charge based RAM offer to servers?

• How does total system memory power trend change when mixture of non-volatile memory and DRAM memory is used simultaneously?
  – Will there be 2x or 5x or 10x or more improvement in power?
Thank You