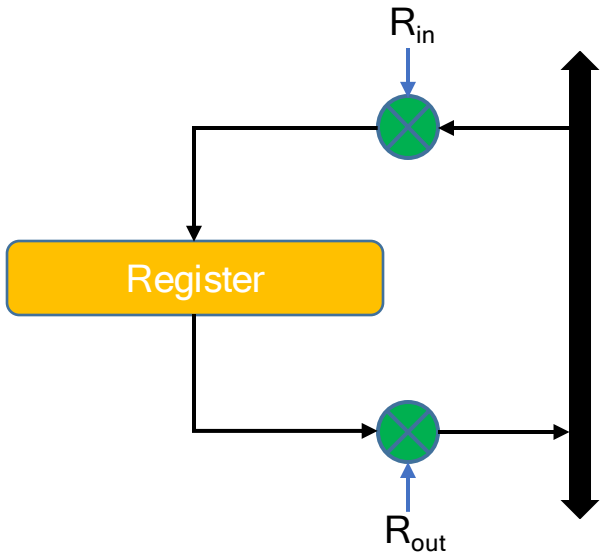
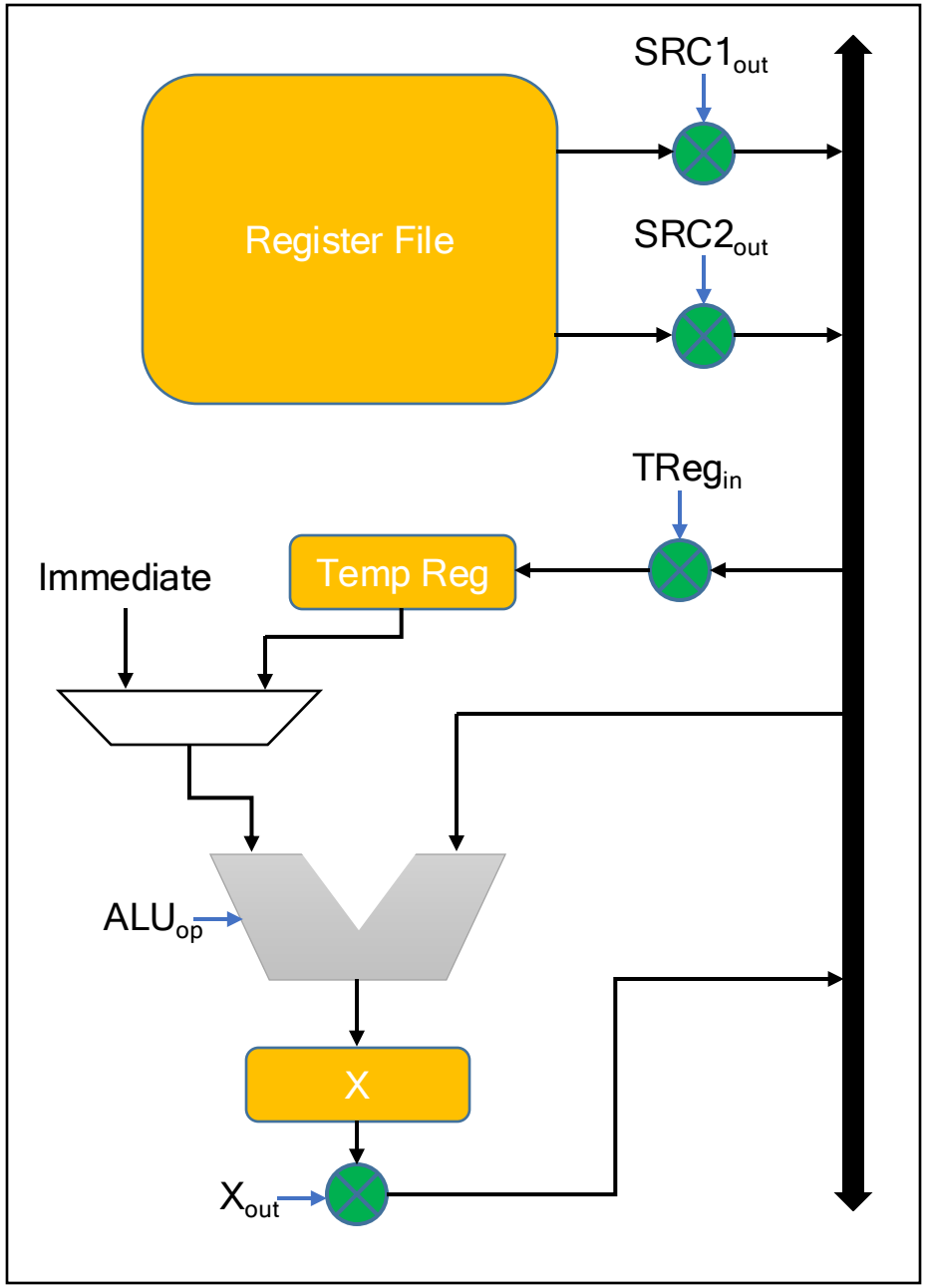
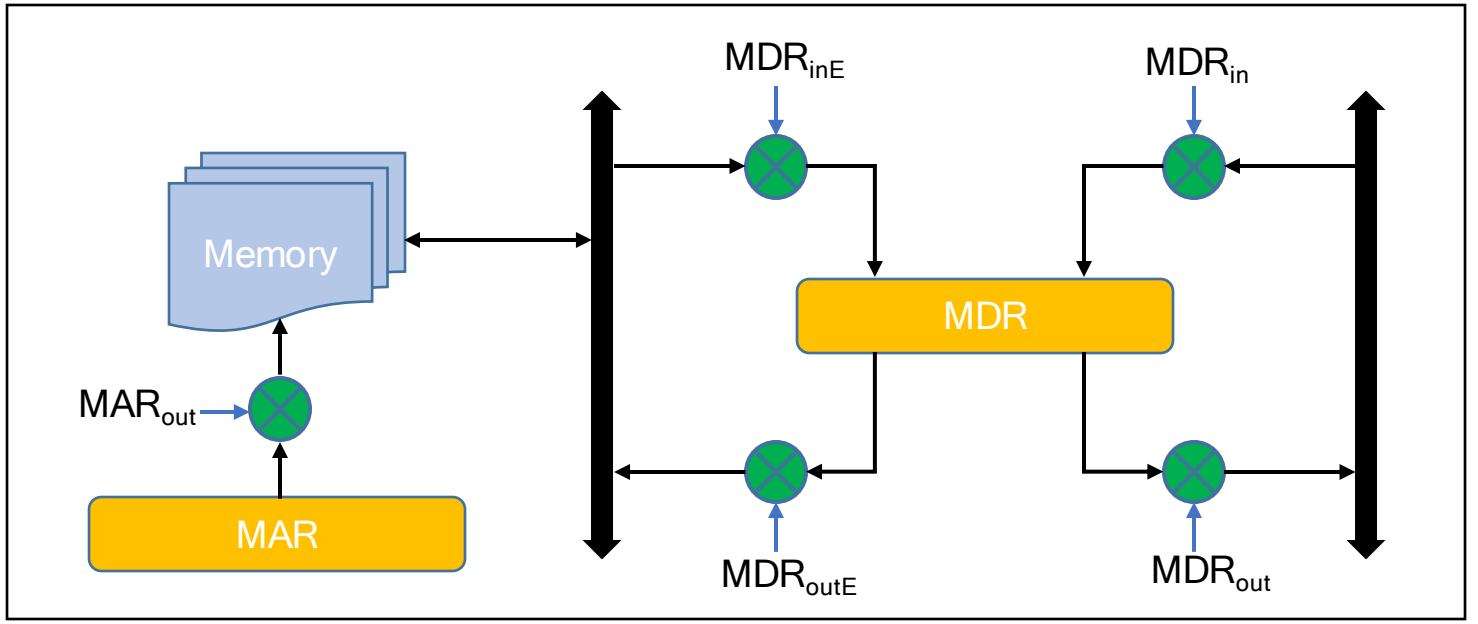
Pipelining

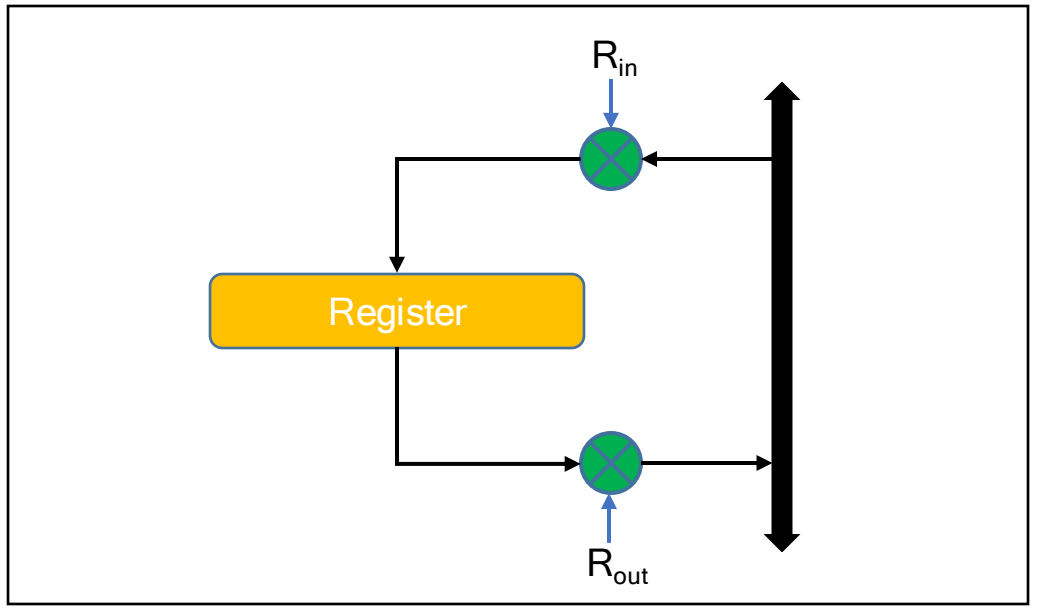
- S

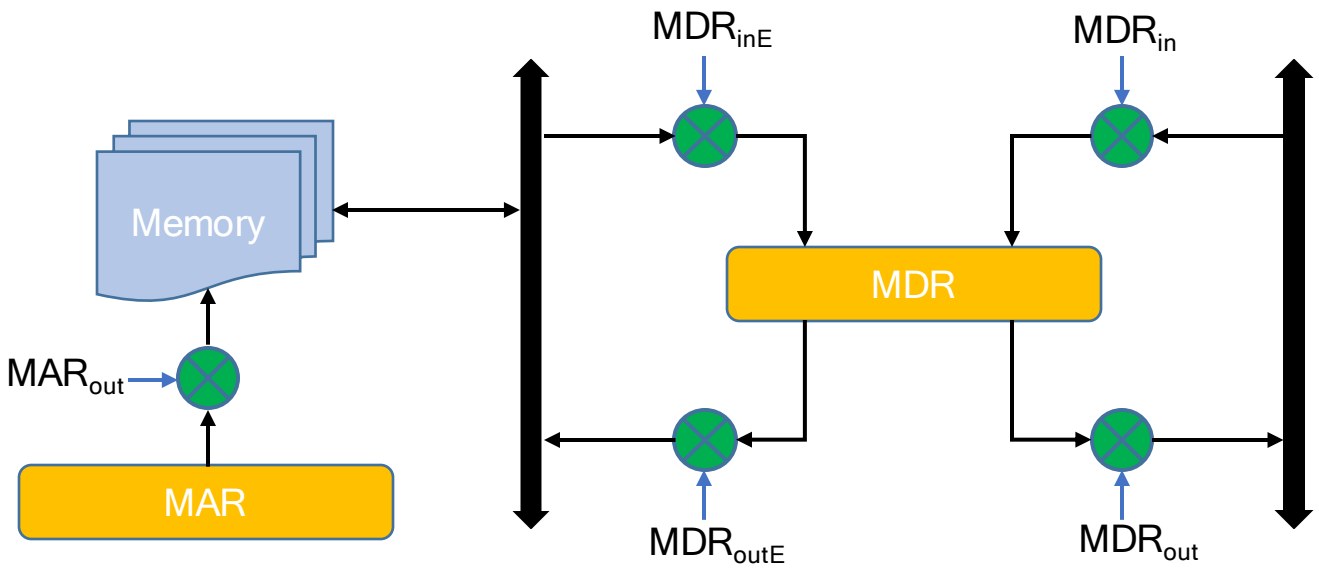

R_{in}
Register
R_{out}

Unpipelined CPU Design







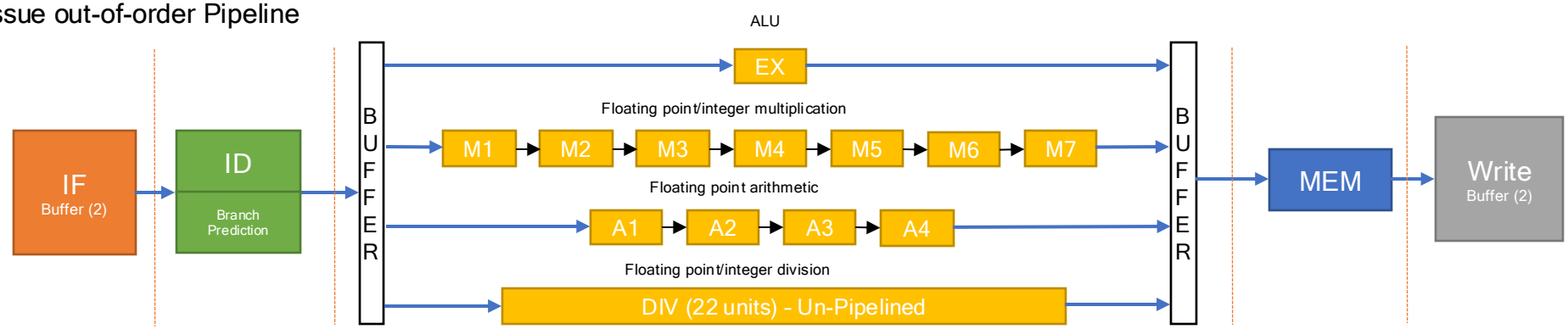

MDR$_{inE}$
MDR$_{in}$
Memory
MDR
MAR$_{out}$
MAR
MDR$_{outE}$
MDR$_{out}$

Pipelining

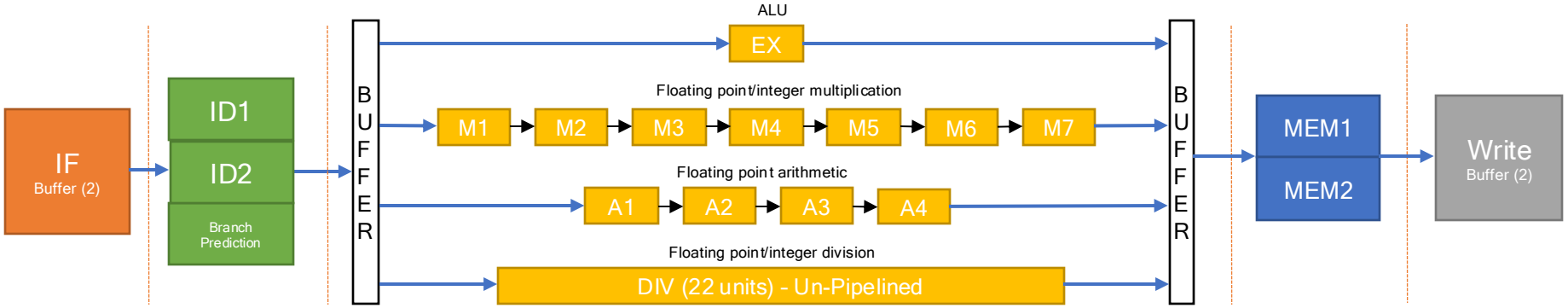
Superscalar processors

In-order Pipeline

Single issue out-of-order Pipeline



Dual issue out-of-order Pipeline



Pipelining Issues

- Not all pipeline stages have uniform latencies
  - Memory access is a critical sub-computation
  - Memory addressing modes should be minimized
  - Fast cache memories should be employed
- Some pipeline stages may not be used
  - Reduce the complexity and diversity of the different instruction types
- Pipeline stall due to dependent instructions
  - Reduce the memory addressing modes because dependency check/detection is difficult
  - Use register addressing mode here, its easy to check the dependancy

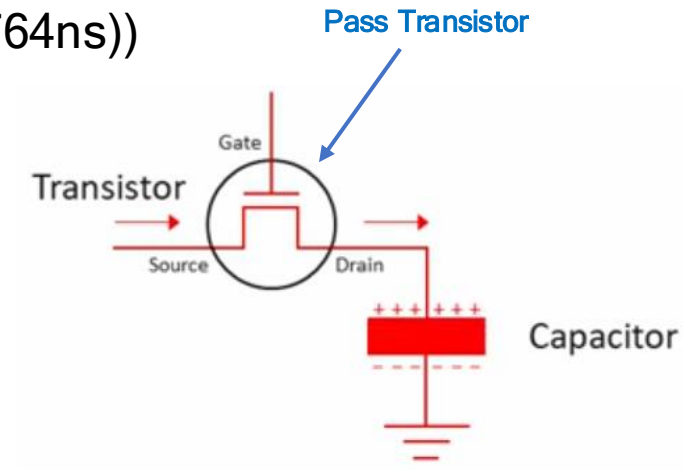
Pipelining Hazards

- Structural hazards
  - Solution 1 – wait
    - Must detect the hazard and have mechanism to stall
  - Solution 2 – Duplicate the HW
    - Multiple such units will help both instructions to progress
- Data hazards
  - RAW
    - Solution - Operand forwarding
  - WAR
    - Issue is due to out-of-order execution
  - WAW
    - Issue is due to out-of-order execution
- Control hazard
  - We will get to know that there is an branch to be taken or not in 4^{th} stage of the pipeline (based on condition)
  - Improve the hardware such that, I should get target address and condition in second stage itself (Branch prediction)
  - How to handle branch hazards?
    - Stall until branch direction is clear
    - Use branch prediction techniques
    - Delayed branch
      - Compilers needs to optimize like some of the instructions needs to be executed either branch is taken or not

Memory Technologies - DRAM

DRAM Basics

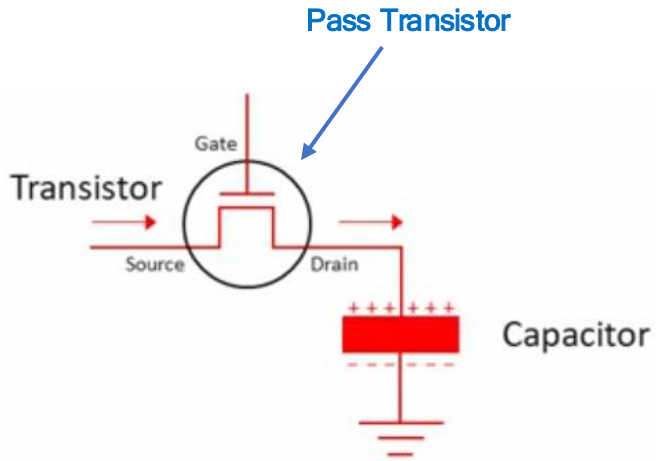
- DRAM cell is composed of a capacitor and a transistor
- Data is stored in capacitor
- Capacitors can loose charge over the time which leads to explicit refreshing (~64ns))
  - Pass transistor is not an perfect switch
  - If dielectric material between the capacitor plate is not perfect then it may leads to allow Some leakage current
  - Higher temperature may accelerate the movement of charge carriers through dielectric
  - Process variability
  - Aging



- Capacitor only takes few Nano seconds to fully charge up
- Act of reading the capacitor is an destructive because charge is capacitor is so small → 30femto (10^{-15})
- The data in the capacitor is always read out and then write back at later time (Precharge)

DRAM Scaling Problems

- DRAM stores charge in a capacitor (Charge based memory)
  - Capacitor must be large enough for reliable sensing
  - Access transistor must be large enough for reliable sensing
- Scaling beyond 17nm is challenging
- Refresh retention time prediction is challenging



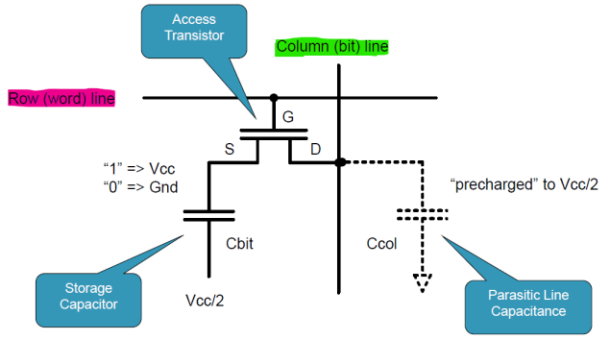
DRAM – Read Write operation

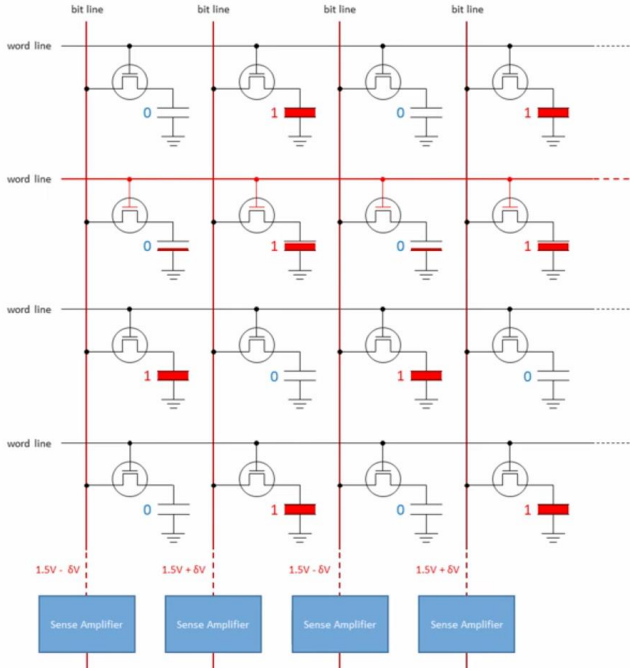
- Bit line is precharged to sense amplifier threshold voltage VDD/2
- Row address strobe signal gets asserted
- If bit line capacitor is charged
  - Bit line voltage gets increased by VDD/2 + β and cell voltage gets reduced by VDD/2 – β
- If bit line capacitor is not charged
  - Bit line voltage gets increased by VDD/2 - β and cell voltage gets reduced by VDD/2 + β
- Sense amplifier senses the voltage difference at the bit line with respect to the

Original voltage and accordingly latched the data in to the row buffer

- Increase or decrease the voltage of the bit line by β according to the row buffer

Value which will increase or decrease the voltage of the cell capacitor

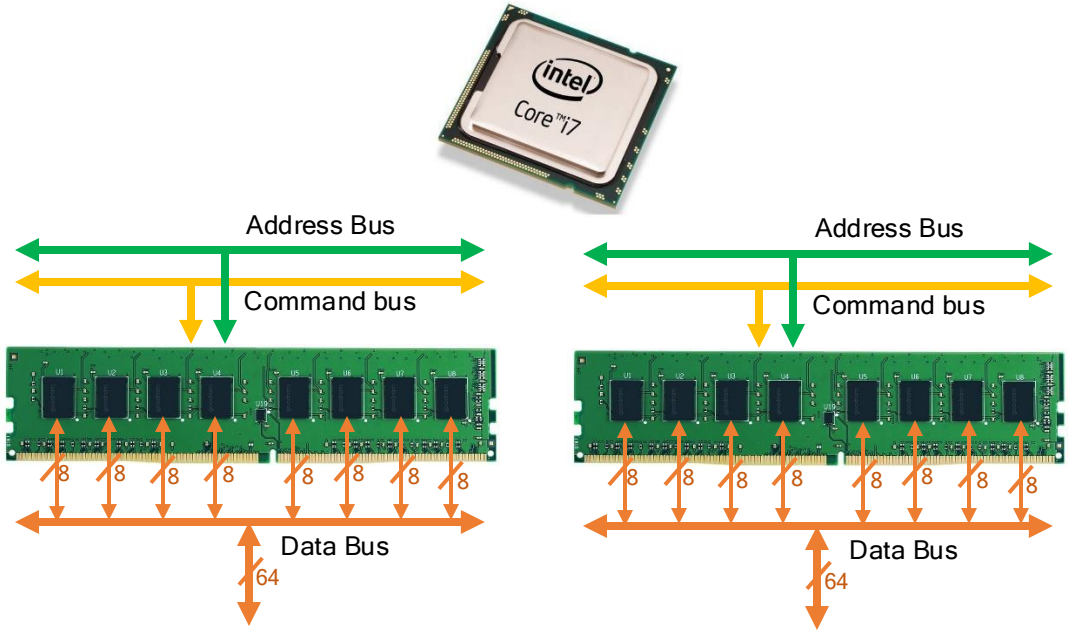


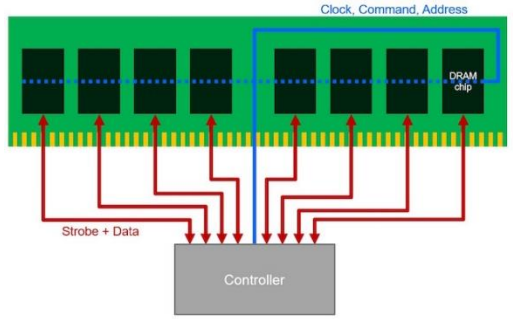


DRAM Basics – How to increase the speed of the chip

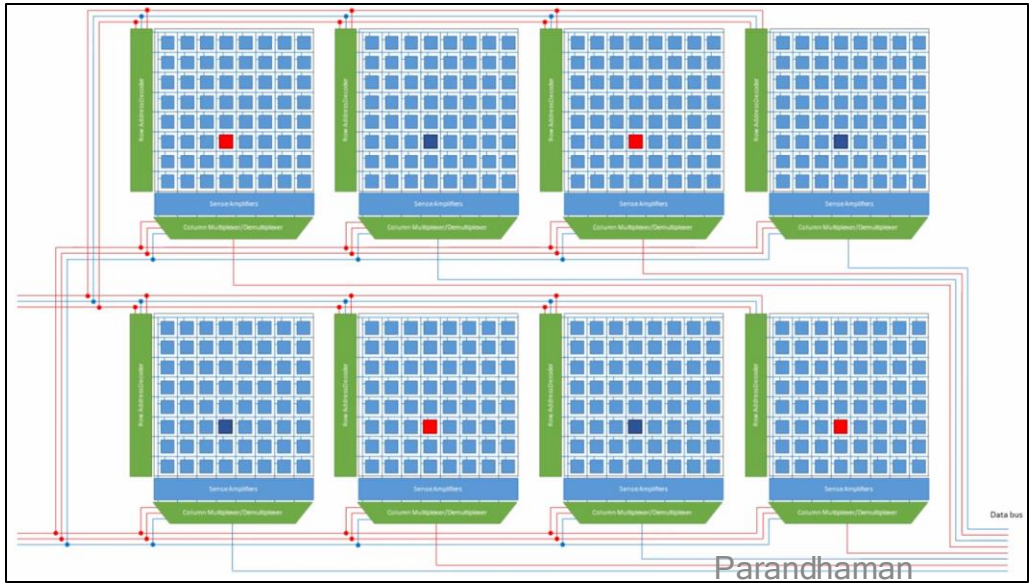
- Split the single big array if memory cells into multiple banks
- This allows more banks to active at the same time
- This can effectively hide the activate/precharge and other timing related to single bank implementations
- Refresh timing is not affected since all the banks are refreshed at the same time

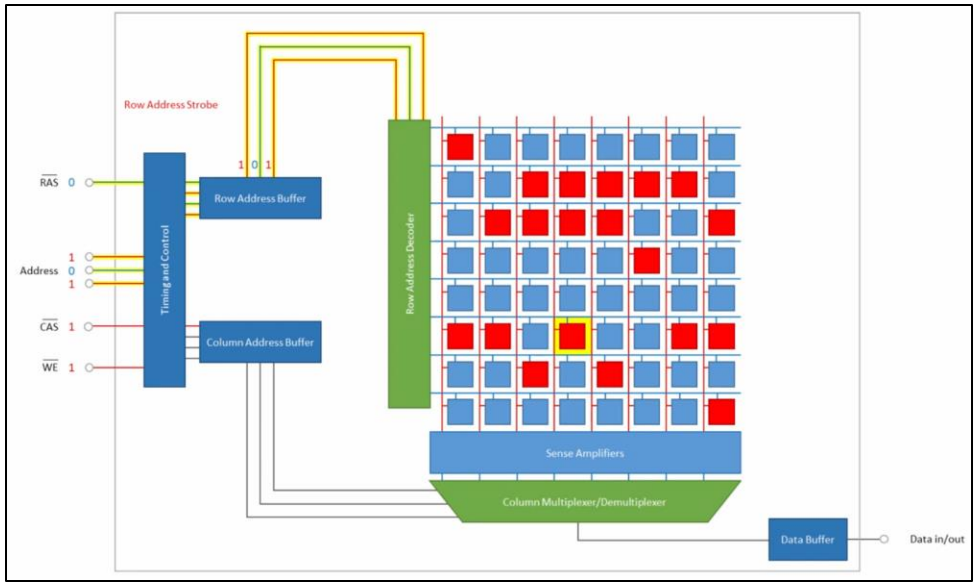
DRAM Organization





Front side of the DIMM is called Rank0
Back side of the DIMM is called Rank1





DDR Memory Comparisons

	SDRAM	DDR1	DDR2	DDR3
Performance	63-133MT/s	200-400MT/s	400-800MT/s	800-1600MT/s
V_{DDQ}	3.3V	2.5V	1.8V	1.5V
V_{TT}	NA	½ V_{DDQ}	½ V_{DDQ}	½ V_{DDQ}
Organization	x4, x8, x16	x4, x8, x16	x4, x8, x16	x4, x8, x16
Density	16Mb-512Mb	64Mb-2Gb	256Mb-4Gb	512Mb-8Gb
Number of banks	4	4	4 (256Mb-512Mb) 8 (1Gb-4Gb)	8
Prefetch	1	2	4	8
Burst length	1,2,4,8,page	2, 4, 8	4, 8	8 (chop 4)
Clock	Single ended	Differential	Differential	Differential
Strobes	NA	Single ended	SE or Differential	Differential

Reading data sheet

- 32Meg x 4 x 4 banks
  - Density x Data width x banks

DDR Memory Signals

	DDR1	LPDDR1	GDDR1	DDR2	LPDDR2	GDDR2	DDR3	LPDDR3	GDDR3	DDR4	LPDDR4	GDDR4	DDR5	LPDDR5	GDDR5	DDR6	LPDDR6	GDDR6
CK, CK# or CK_t, CK_n																		
CKE																		
CS																		
RAS#, CAS#, WE#																		
DM																		
BA																		
Address																		
DQ																		
DQS																		
VDDQ, VSSQ																		
VDD, VSS																		
VDDL, VSDL																		
VREF																		
TPD																		
TQ																		
ODT																		

LPDDR

LPDDR Acronyms

- tRP → Row Precharge Time
- tWR → Write Recovery Time
- tRCD → Row to Column Delay
- tMRD → Mode Register Delay
- tRFC → Refresh cycle time
- tCK → Clock cycle time
- tREFI → Refresh Interval

1	VDD and VDDQ Ramp: CKE must be held high
2	Apply stable clocks
3	Wait at least 200 µs with NOP or DESELECT on command bus
4	PRECHARGE ALL
5	Assert NOP or DESELCT for t_{RP} time
6	Issue two AUTOREFRESH commands each followed by NOP or DESELECT commands for t_{RFC} time
7	Configure Mode Register
8	Assert NOP or DESELECT for t_{MRD} time
9	Configure Extended Mode Register
10	Assert NOP or DESELECT for t_{MRD} time
11	LPDDR SDRAM is ready for any valid command

LPDDR Features

- Double data rate architecture – Two transfers per clock cycle
- Differential clock inputs
- 4 internal banks for concurrent operation
- Commands are entered each positive edge of the CK
- Bidirectional DQS signal used to capture the data at receiver end
  - Data and data mask are sampled at both the edges of the DQS
  - DQS is edge aligned with data for READ and center aligned with data for WRITE
  - Input data is registered at both the edges of the DQS
  - Output data is referenced to both edges of DQS, as well as to both edges of CK

LPDDR Basics

- LPDDR architecture is an 2n prefetch architecture
- LPDDR is designed to transfer 2 words per clock cycle (Double Data Rate)
- Single read/write data consists of,
  - 1 clock cycle to move data from core
  - ½ clock cycle to move data from I/O pin
- During READ operation DQS is asserted by the DRAM and during write operation DQS signal is asserted by the Memory Controller
- At crossing points of CK and CK#, when CK is going high and CK# is going low is termed as positive edge of the CK
  - Commands (Address and control signals) are registered during each positive edge of the CK
  - All the DRAM internal clocks are derived from CK
- Auto Precharge can be initiated at the end READ/WRITE transactions
- The pipelined multibank architecture of the LPDDR allows the operations to be executed in parallel
  - Provides high effective bandwidth by hiding row precharge and activation times
- Packages
  - Single die – 1CS, 1CKE
  - Dual-die – 2CS, 2CKE
- LPDDR comes with x16, x32 variants only

LPDDR signals

- DM is applicable only for write transactions with x16 and x32 variants
  - X16 – LDM and UDM
  - X32 – DM0, DM1, DM2 and DM3
- DQS is used to capture the data from DQ lines
  - X16 – LDQS and UDQS
  - X32 – DQS0, DQS1, DQS2 and DQS3
- VDD – Core power
- VDDQ – I/O power
- TQ – Asynchronous LV-CMOS output
  - High – when device temperature equals or exceeds 85’C
  - Low - when device temperature less than 85’C

LPDDR Burst transactions

- Read and Write accesses to the LPDDR is burst oriented
  - Accesses start at a selected location and will continue for the programmed number of locations in programmed sequence
- Programmable Read/Write burst lengths of 2, 4, 8 or 16

LPDDR Low power features

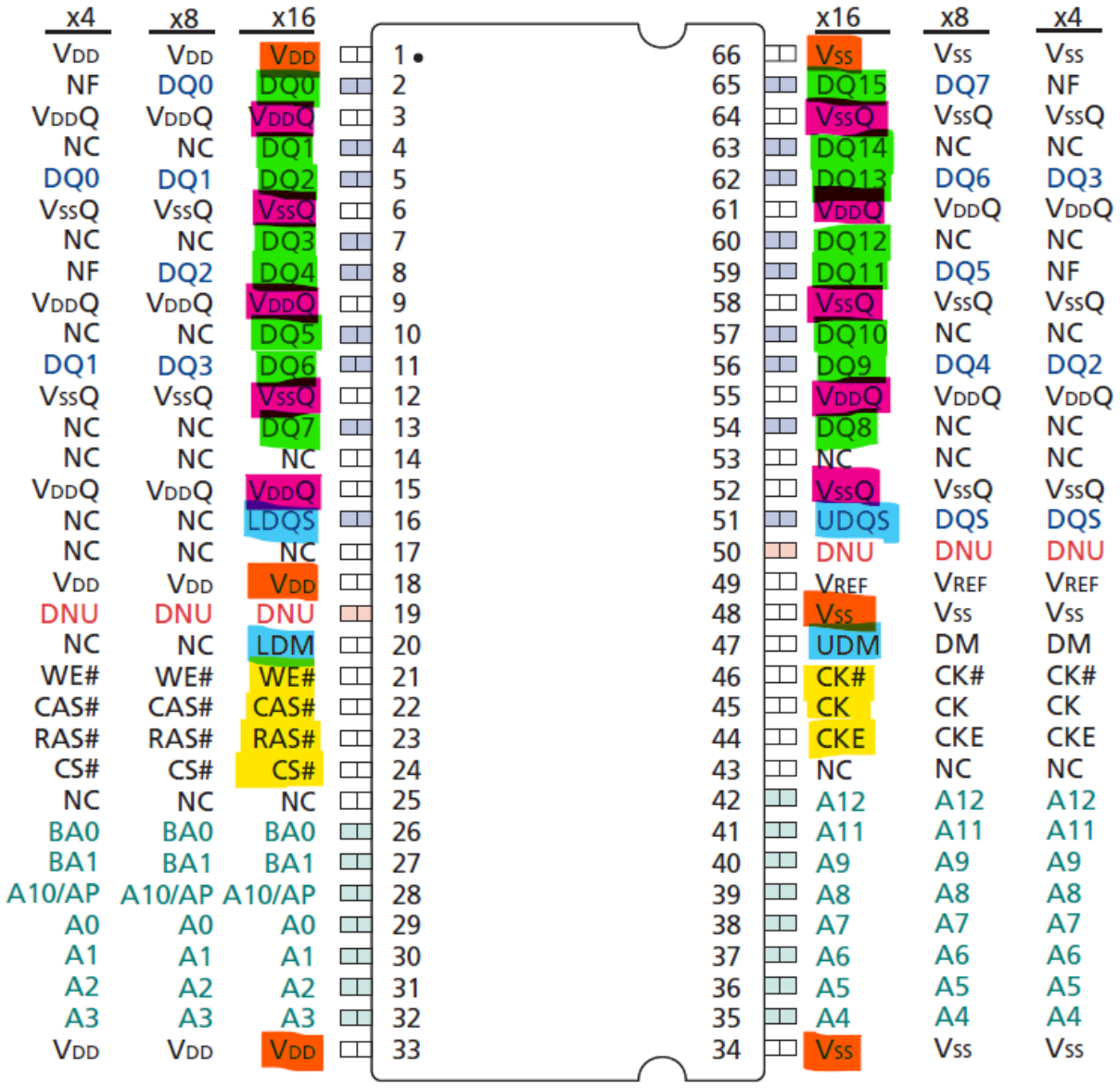
- Auto refresh mode is provided, along with power saving power down mode
- Self refresh mode may have Temperature Compensated Self Refresh (TCSR) and Partial Array Self Refresh (PASR) options, which allows users to achieve additional power saving
  - TCSR and PASR options can be programmed via extended mode register
- All inputs are LV-CMOS compatible
- VDD and VDDQ supply of 1.8V
- Power down
  - Looks like same of self refresh entry
  - 2 possible power down
    - All banks are idle - Means no row is open → Precharge power down
    - Banks are active - Means row is open → Active power down
- Deep power down
  - Deep power down is entered with burst terminate command except that CKE is registered low
  - All the data in DRAM is lost
  - ToDo - Understand entry and exit sequence
- Clock stop
  - Stopping clock during idle periods are effective way to reduce the power consumption
  - ToDo - Understand entry and exit sequence

LPDDR Refresh requirements

- SDRAM must require to refresh every row at each 64ms
- Accomplished by explicit AUTO REFRESH command or internally timed event in SELF REFRESH mode
- Self refresh entry
  - Register SELF REFRESH command (CKE brought LOW)
  - User may halt the external clock after the one clock of SELF REFRESH command is triggered (CKE brought LOW)
  - The minimum time the device must be in Self refresh mode is defined as tRFC (Refresh Cycle Time)
- Self refresh exit
  - CK must be stable before CKE brought high
  - Wait until tXSR (Self refresh exit time) to allow the completion of any internal refresh in progress

DDR

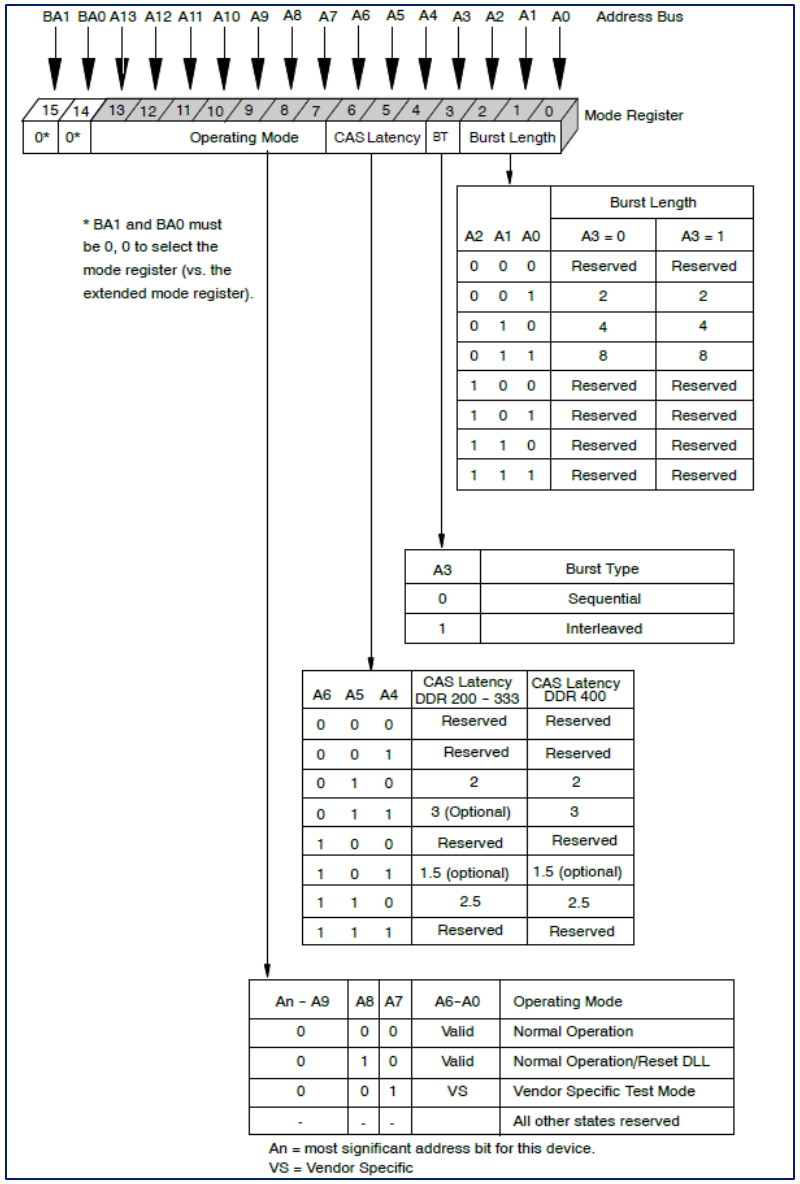
DDR Pin Diagram



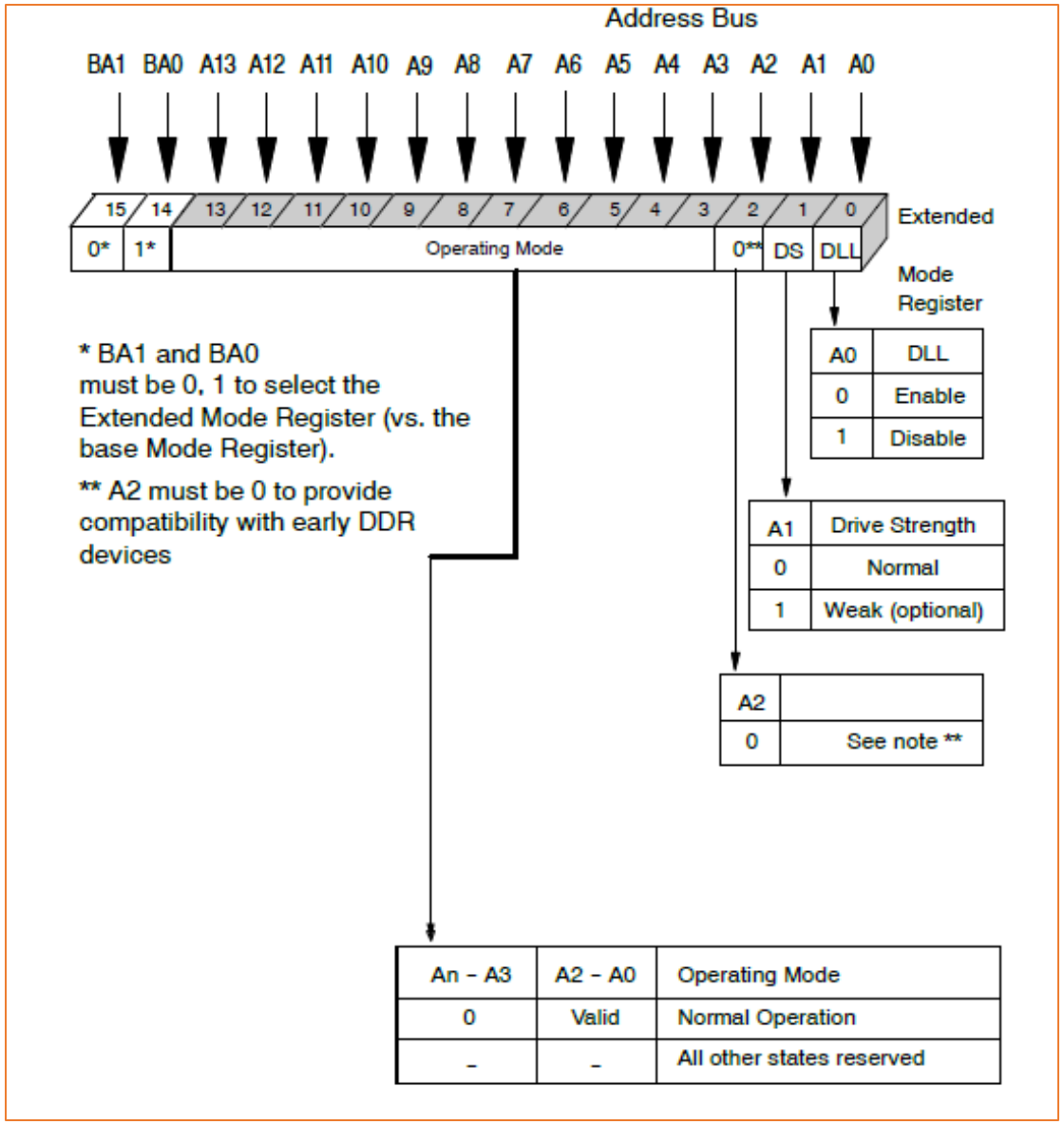
DDR Truth Table

Command	CS#	RAS#	CAS#	WE#	Address	Comments
DESELECT (NOP)						
No Operation (NOP)						
ACTIVE					Bank/Row	
READ					Bank/Column	
WRITE					Bank/Column	
BURST TERMINATE						
PRECHARGE					Code	
SELF REFRESH						
MODE REGISTER SET					Opcode	

DDR Mode and Extended Register



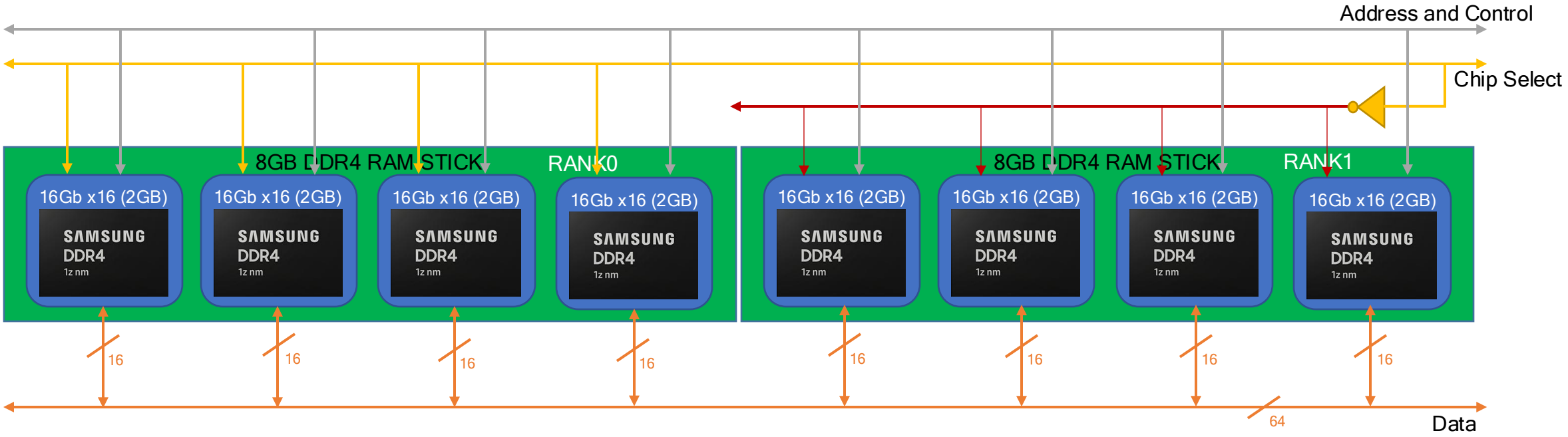
Mode Register



Extended Mode Register

DDR2

16GB DDR with 2 Rank configuration



16GB DDR with 1 Rank configuration