Advanced Topics in VLSI

EE6361

Jan 2018

Course Objectives

- Introduce students to some relevant advanced topics of current interest in academia and industry
- Give the students a feel for research topics and what research means
- Make students aware of work happening in India

Current Topics

- Embedded Memory Design
  - Built In Self Test (BIST) – Dr. C. P. Ravikumar, TI
  - SRAMs (Dr. Rahul Rao, IBM India)
  - eDRAMs (Dr. Janakriaman, IITM)

Learning Objectives for BIST

- ❑ Explain how memories are tested
- ❑ Explain how memory testing is different from digital testing
- ❑ Describe various memory faults
- ❑ Propose solutions to various memory faults
- ❑ Explain the need for ECC in memories
- ❑ Elaborate and explain the concept of BIST

Learning Objectives for SRAM

- ❑ Articulate memory hierarchy and the value proposition of SRAMs in the memory chain + utilization in current processors
- ❑ Explain SRAM building blocks and peripheral operations and memory architecture (with physical arrangement)
- ❑ Articulate commonly used SRAM cells (6T vs 8T), their advantages and disadvantages
- ❑ Explain the operation of a non-conventional SRAM cells, and their limitations
- ❑ Explain commonly used assist methods
- ❑ Explain how variations impact memory cells

Learning Objectives for EDRAM

- ❑ Explain the working of a (e)DRAM. What does Embedded mean?
- ❑ Explain the working of a feedback sense amplifier and modify existing designs to improve performance
- ❑ Calculate the voltage levels of operation of various components for an eDRAM
- ❑ Introduce stacked protect devices to reduce voltage stress of the WL driver

Grading

- Assignments – 10%
- Quiz 1– 15%
- Quiz 2 -15%
- Project – 20%
- End Semester – 40%

Course Schedule

- Thursday – 2:00 – 3:00 PM and 5:00 – 6:30 PM
  - Saturday possible for the BIST module?
- ESB 213B

Embedded DRAM

Janakiraman V

Assistant Professor

Electrical Department

IIT Madras

Topics

- Introduction to memory
- DRAM basics and bitcell array
- eDRAM operational details (case study)
- Noise concerns
- Wordline driver (WLDRV) and level translators (LT)
- Challenges in eDRAM
- Understanding Timing diagram – An example
- Gated Feedback Sense Amplifier (case study)
- References

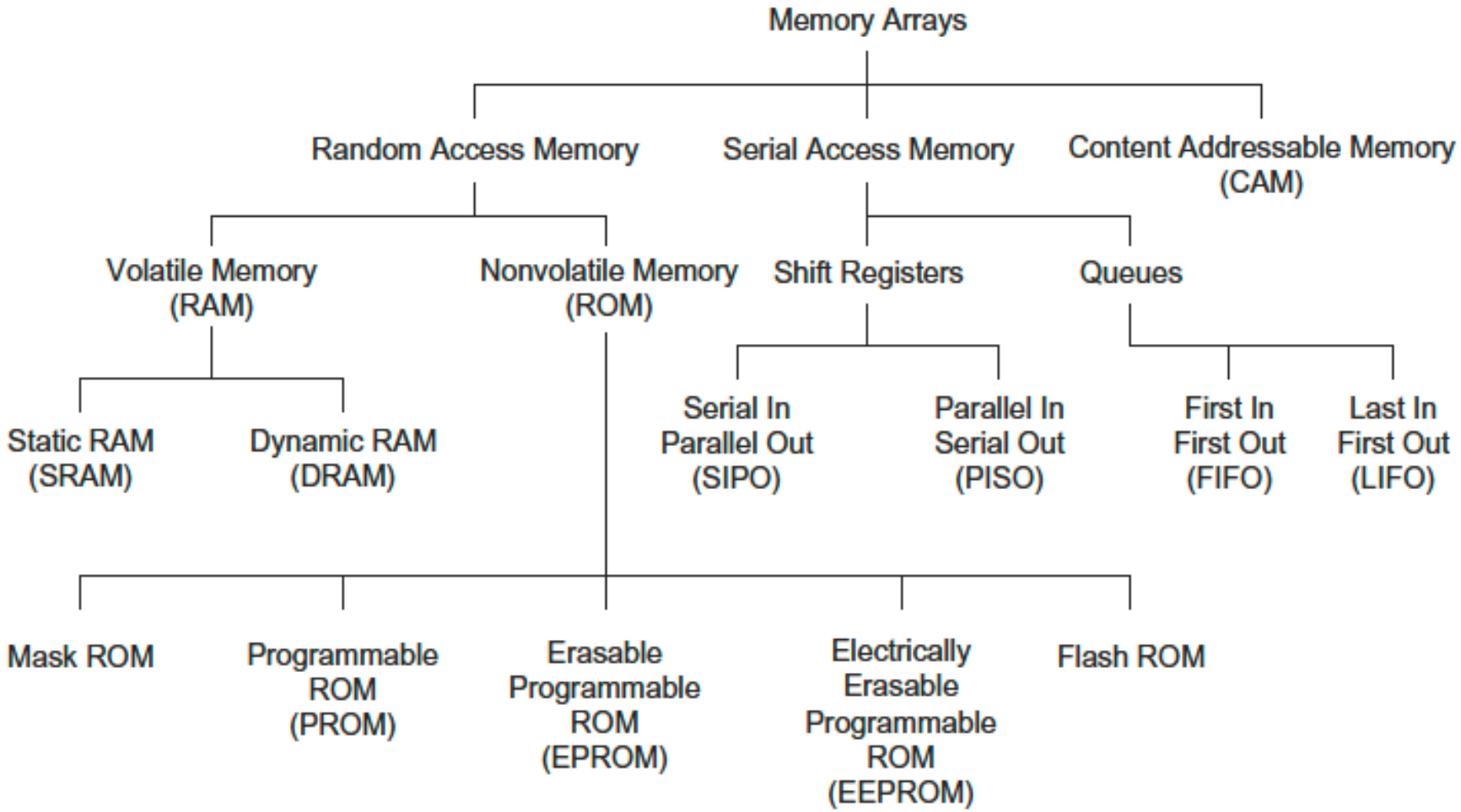
Acknowledgement

- Raviprasad Kuloor (Course slides were prepared by him)
- John Barth, IBM SRDC for most of the slides content
- Madabusi Govindarajan
- Subramanian S. Iyer
- Many Others

Topics

- ❑ Introduction to memory
- ❑ DRAM basics and bitcell array
- ❑ eDRAM operational details (case study)
- ❑ Noise concerns
- ❑ Wordline driver (WLDRV) and level translators (LT)
- ❑ Challenges in eDRAM
- ❑ Understanding Timing diagram – An example

Memory Classification revisited



Motivation for a memory hierarchy – infinite memory

Cycles per Instruction (CPI) = Number of processor clock cycles required per instruction

CPI[∞ cache]

Finite memory speed

CPI = CPI[∞ cache] + FCP

Finite cache penalty

Locality of reference – spatial and temporal

Temporal

If you access something now you'll need it again soon

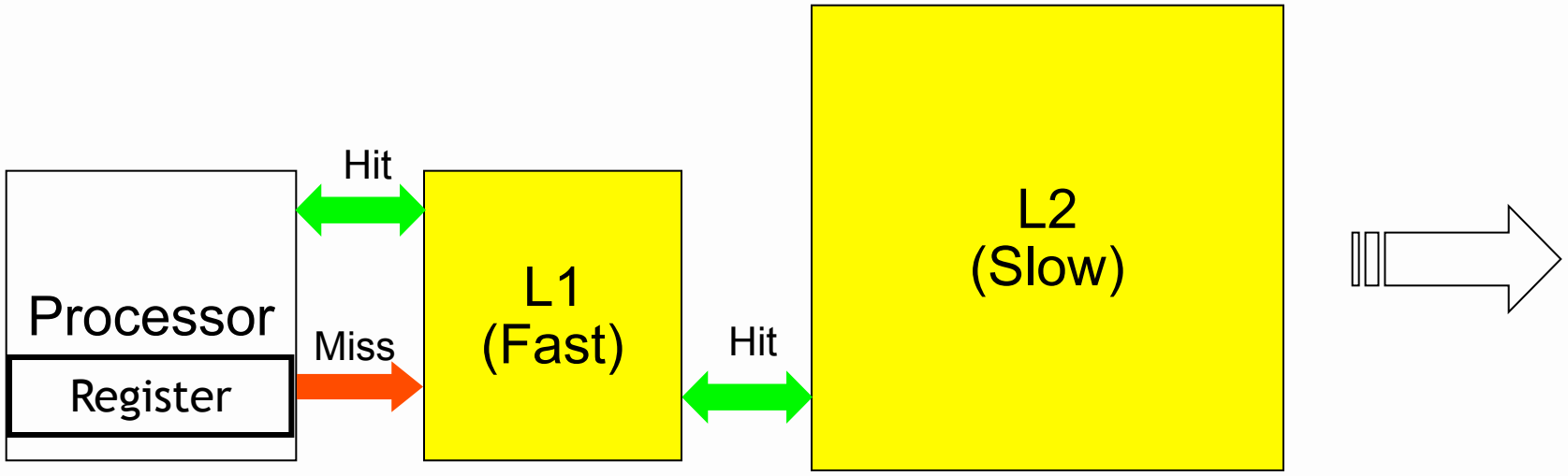
e.g: Loops

Spatial

If you accessed something you'll also need its neighbor

e.g: Arrays

Exploit this to divide memory into hierarchy

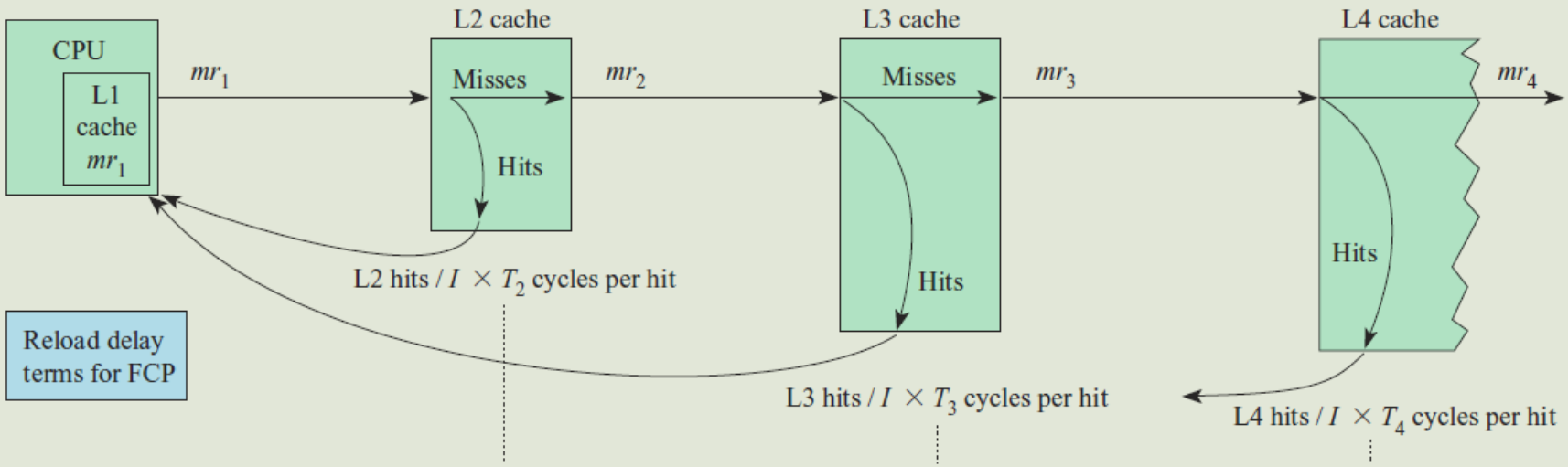


Cache size impacts cycles-per-instruction

Logic-based eDRAM: Origins and rationale for use

R. E. Matick
S. E Schuster

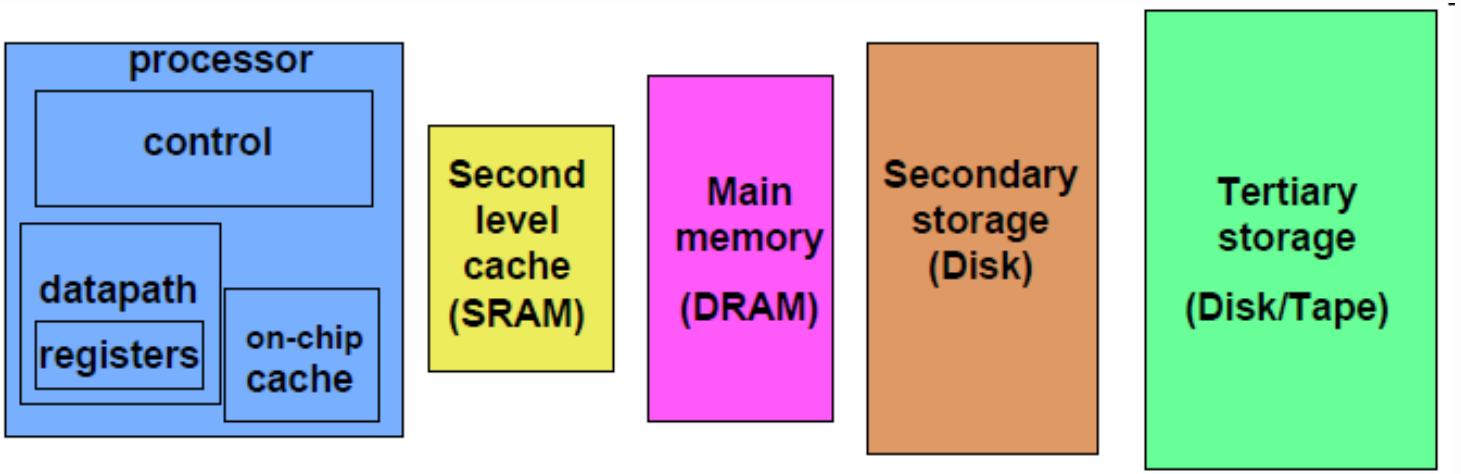
IBM J. RES. & DEV. VOL. 49 NO. 1 JANUARY 2005



Finite cache penalty	=	Delay / I for L2 hits	+	Delay / I for L3 hits	+	Delay / I for L4 hits ...
FCP	=	$(mr_1 - mr_2)\, T_2$	+	$(mr_2 - mr_3)\, T_3$	+	$(mr_3 - mr_4)\, T_4$ + ...
Cycles per instruction	=	Hits per instruction × cycles per hit				

→ Access rate reduces → Slower memory is sufficient

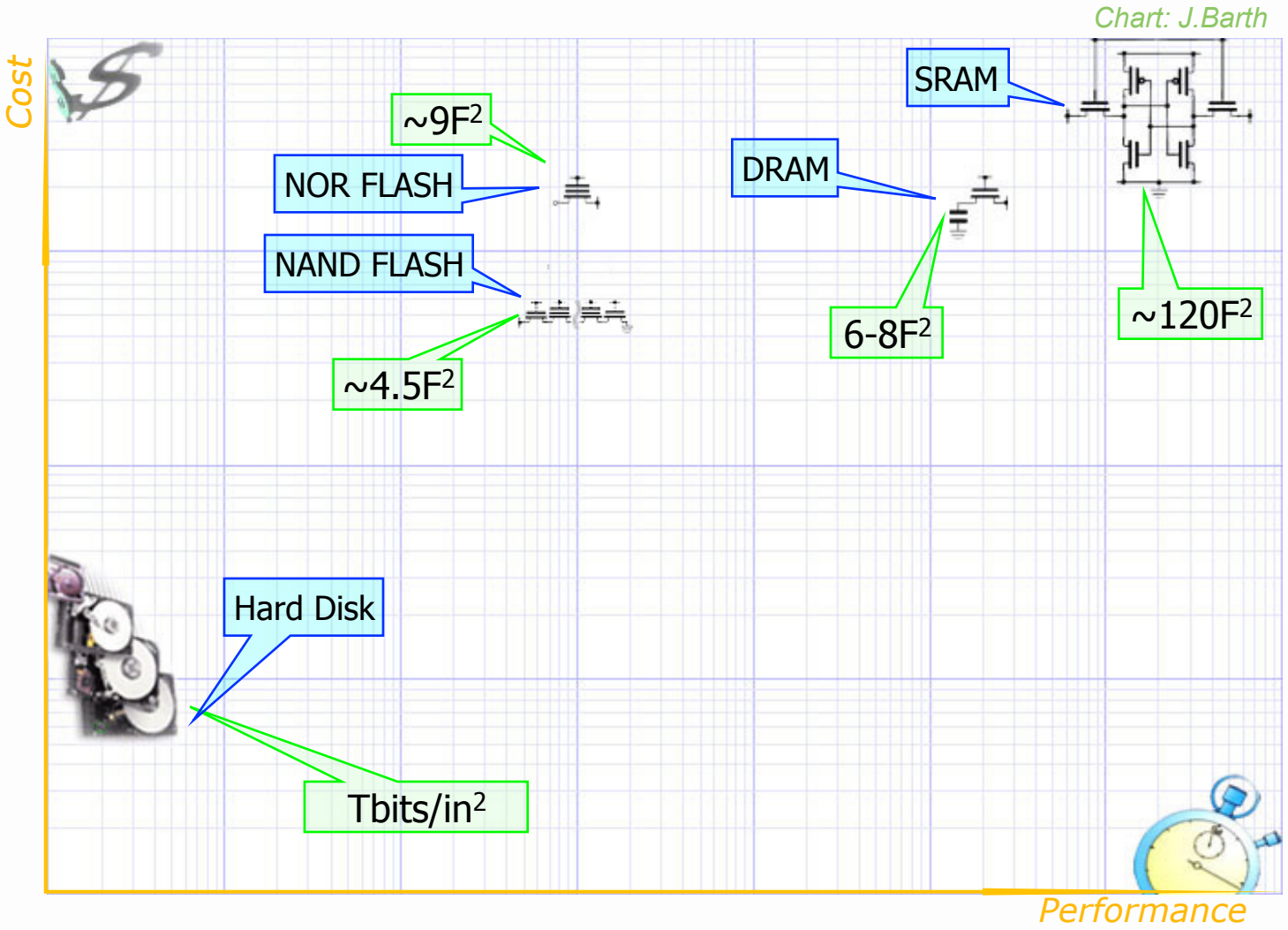
Cache size impacts cycles-per-instruction



Speed	1ns	10ns	100ns	10ms	10sec
Size	B	KB	MB	GB	TB

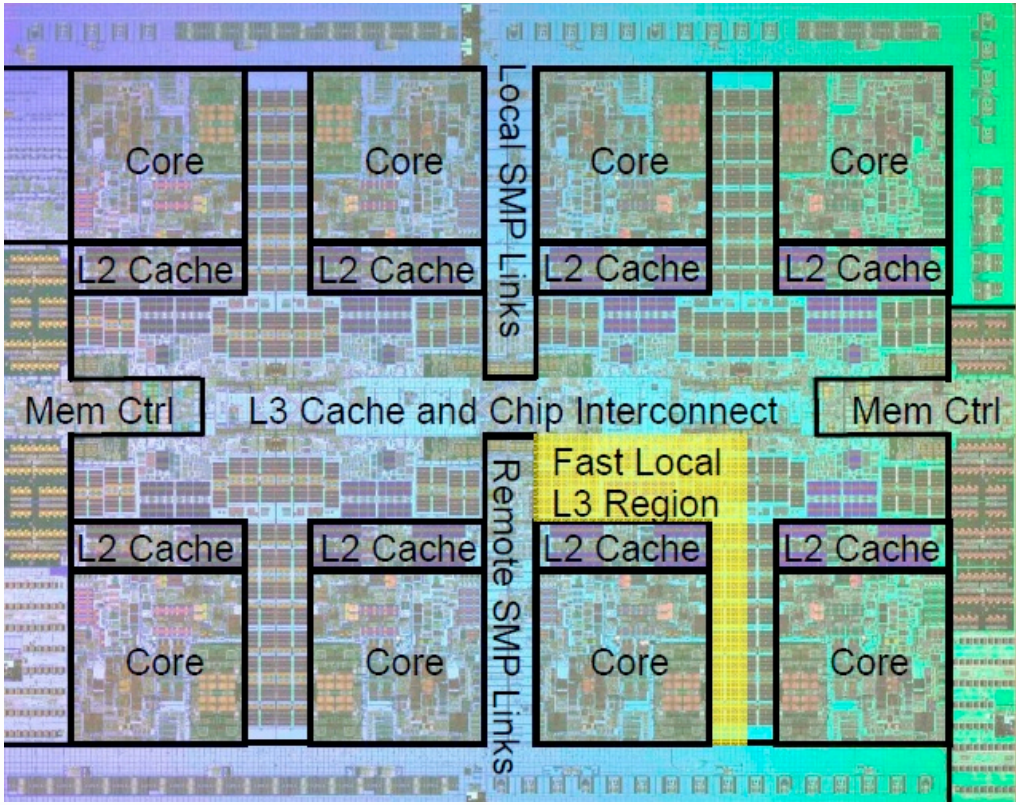
For a 5GHz processor, scale the numbers by 5x

Technology choices for memory hierarchy



eDRAM L3 cache

Power7 processor

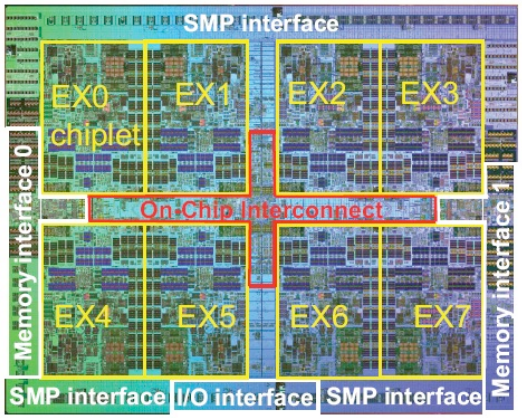


Move L2,L3 Cache inside of the data hungry processor

Higher hit rate → Reduced FCP

Embedded DRAM Advantages

IBM Power7tm
32MB eDRAM L3

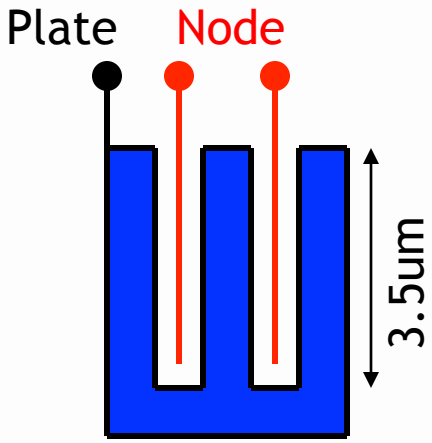


Memory Advantage

- 2x Cache can provide > 10% Performance
- ~3x Density Advantage over eSRAM
- 1/5x Standby Power Compared to SRAM
- Soft Error Rate 1000x lower than SRAM
- Performance ? DRAM can have lower latency !
- IO Power reduction

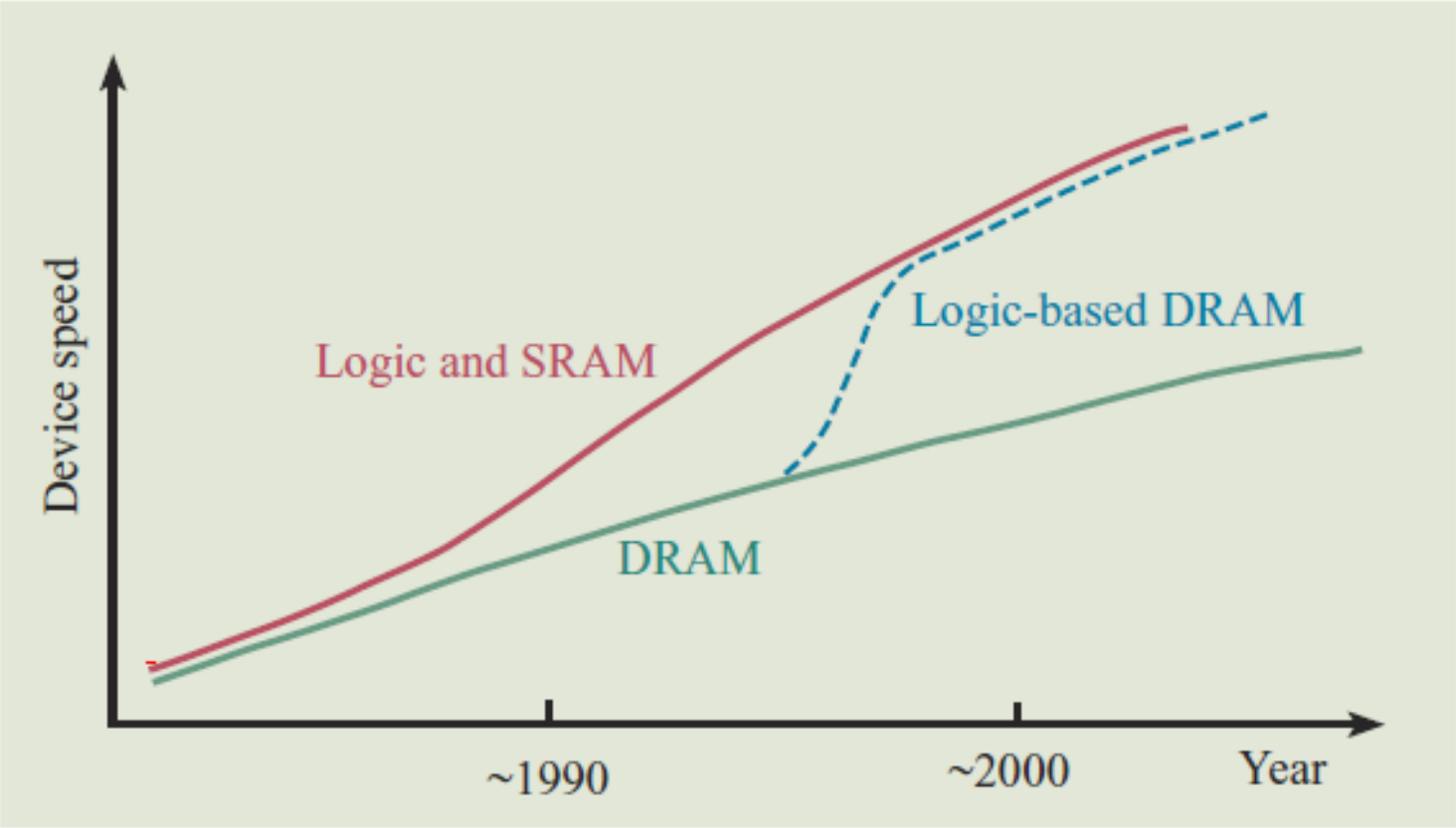
Deep Trench Capacitor

- Low Leakage Decoupling
- 25x more Cap / μm^2 compared to planar
- Noise Reduction = Performance Improvement
- Isolated Plate enables High Density Charge Pump

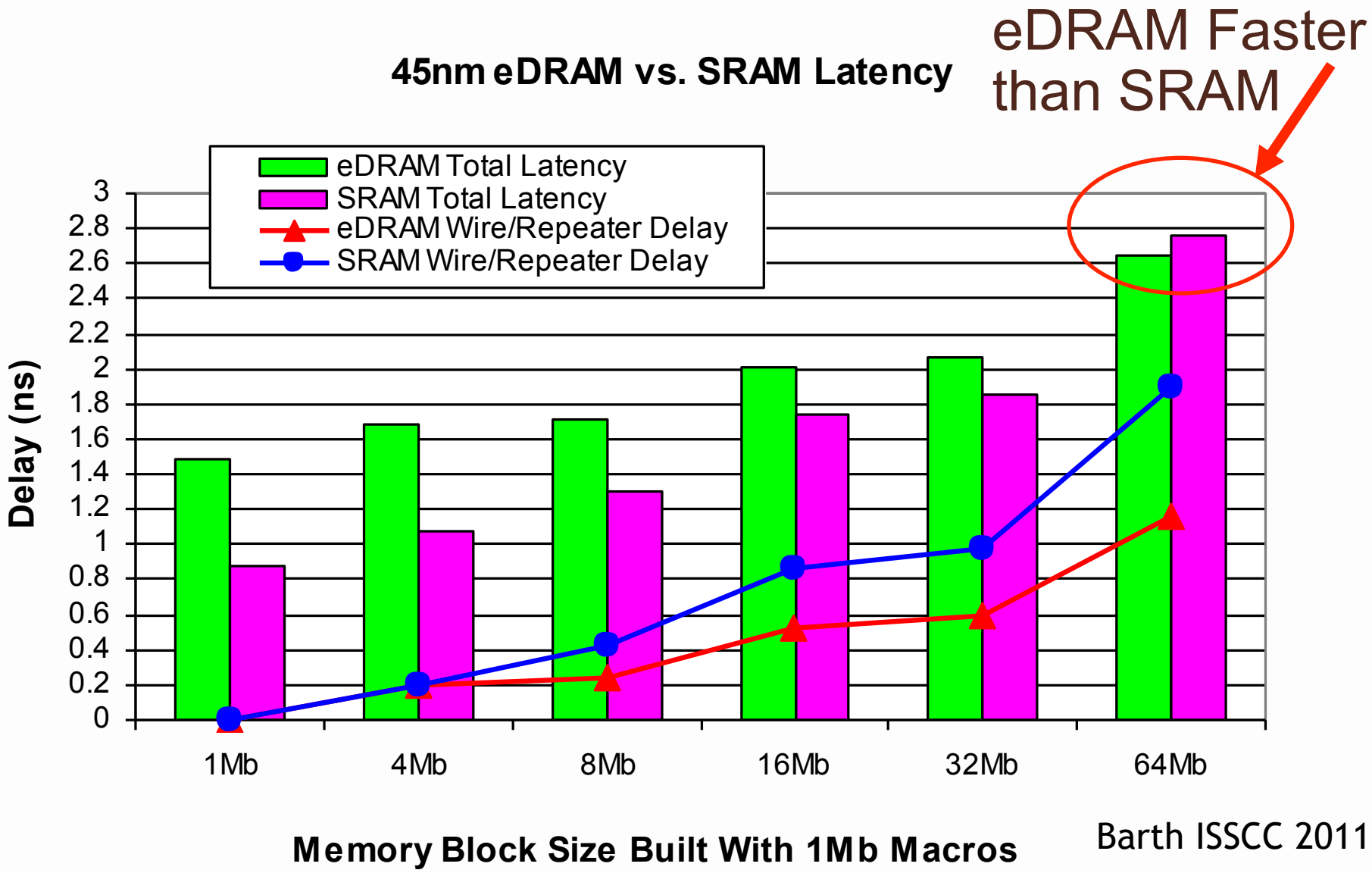


Cache performance – SRAM vs. DRAM

Chart: Matick & Schuster, op. cit.



Embedded DRAM Performance



Barth ISSCC 2011

Topics

- ❑ Introduction to memory
- ❑ DRAM basics and bitcell array
- ❑ eDRAM operational details (case study)
- ❑ Noise concerns
- ❑ Wordline driver (WLDRV) and level translators (LT)
- ❑ Challenges in eDRAM
- ❑ Understanding Timing diagram – An example

Fundamental DRAM Operation

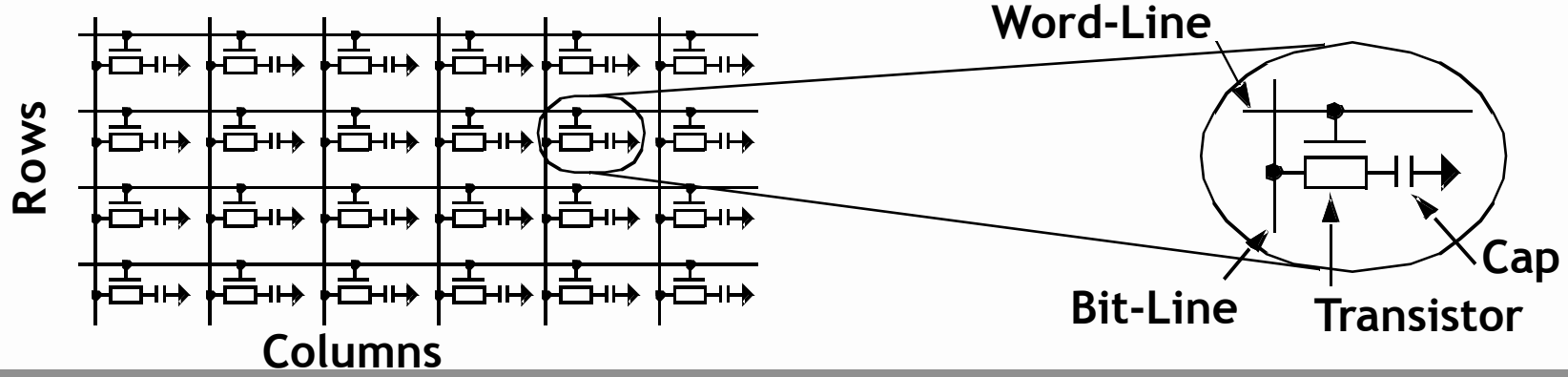
Memory Arrays are composed of Row and Columns

Most DRAMs use 1 Transistor as a switch and
1 Cap as a storage element (Dennard 1967)

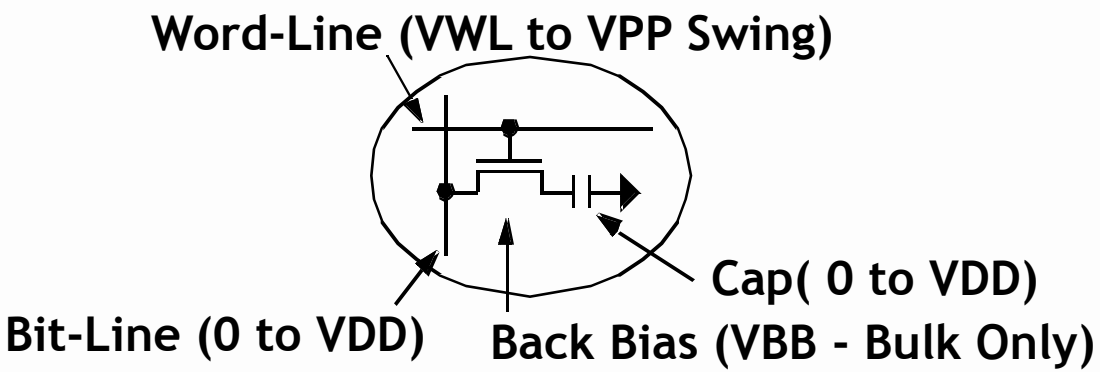
Single Cell Accessed by Decoding One Row / One Column (Matrix)

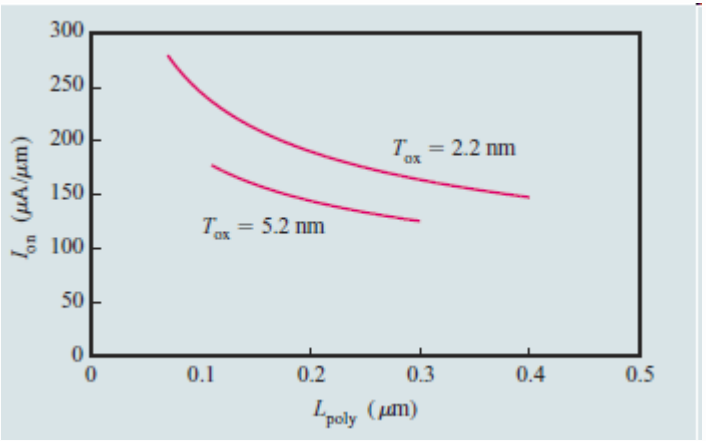
Row (Word-Line) connects storage Caps to Columns (Bit-Line)

Storage Cap Transfers Charge to Bit-Line, Altering Bit-Line Voltage



1T1C DRAM Cell Terminals





VWL: Word-Line Low Supply, GND or Negative for improved leakage

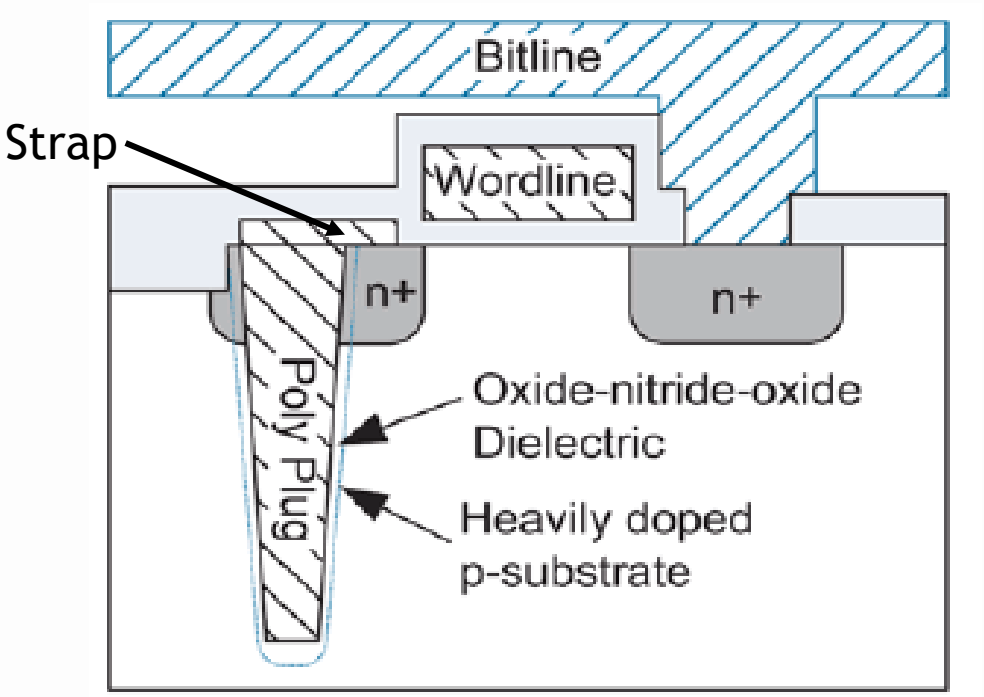
VPP: Word-Line High Supply, 1.8V up to 3.5V depending on Technology
Required to be at least a Vt above VDD to write full VDD

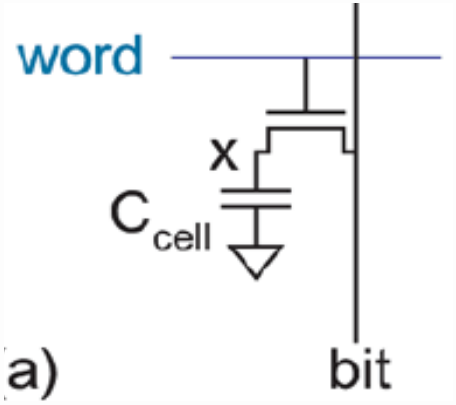
VBB: Back Bias, Typically Negative to improve Leakage
Not practical on SOI

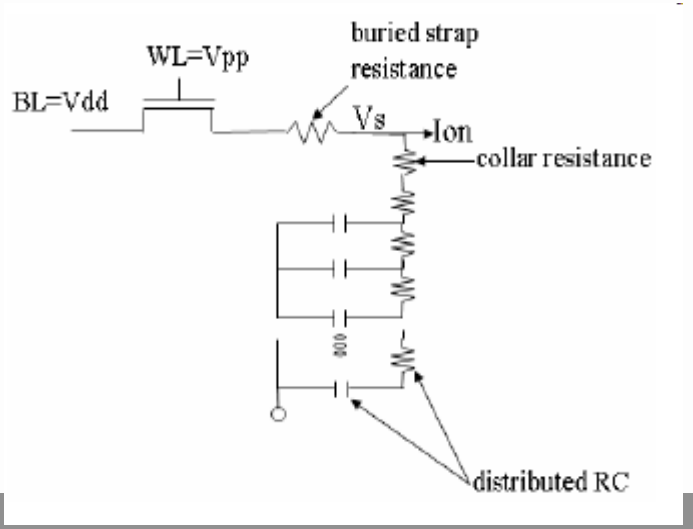
IBM J RES & DEV 2005

DRAM cell Cross section

- Store their contents as charge on a capacitor rather than in a feedback loop.
- 1T dynamic RAM cell has a transistor and a capacitor

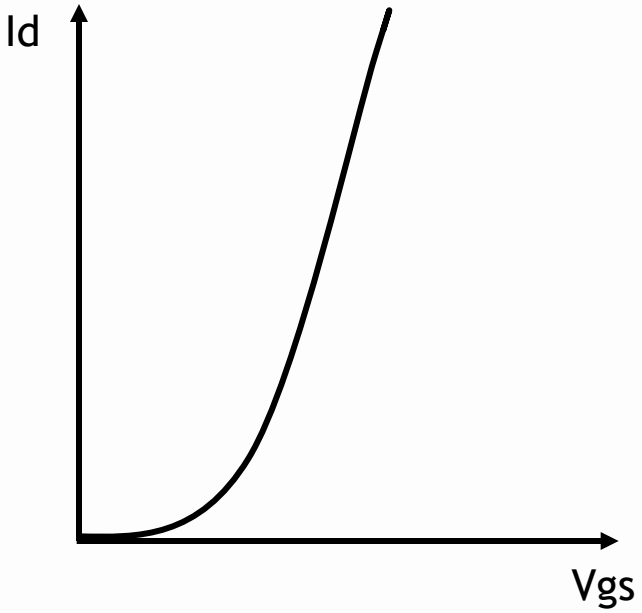


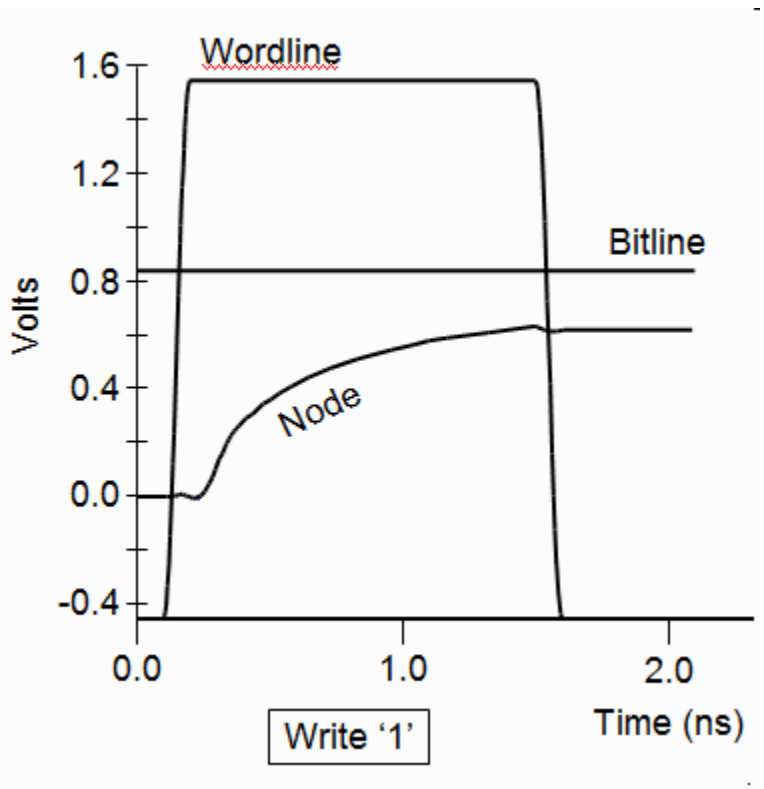




CMOS VLSI design - PEARSON

Storing data ‘1’ in the cell

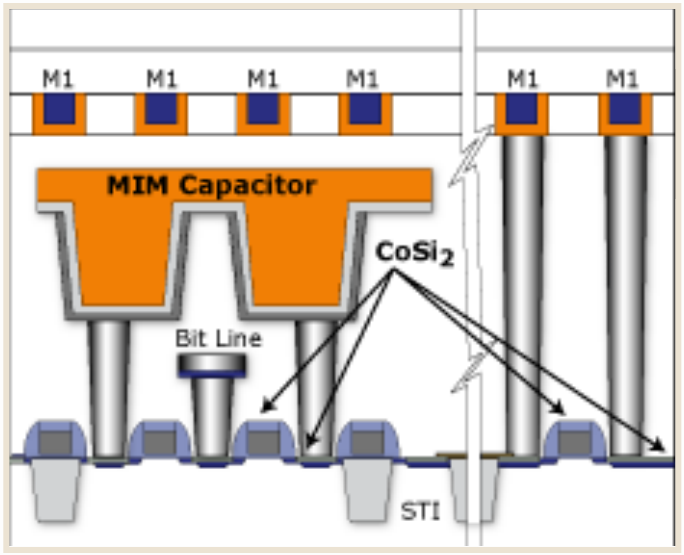




Vgs for pass transistor reduces as bitcell voltage rises, increasing Ron

Why there is a reduction in cell voltage after WL closes? Experiment

MIM Cap v/s Trench



MIM eDRAM Process

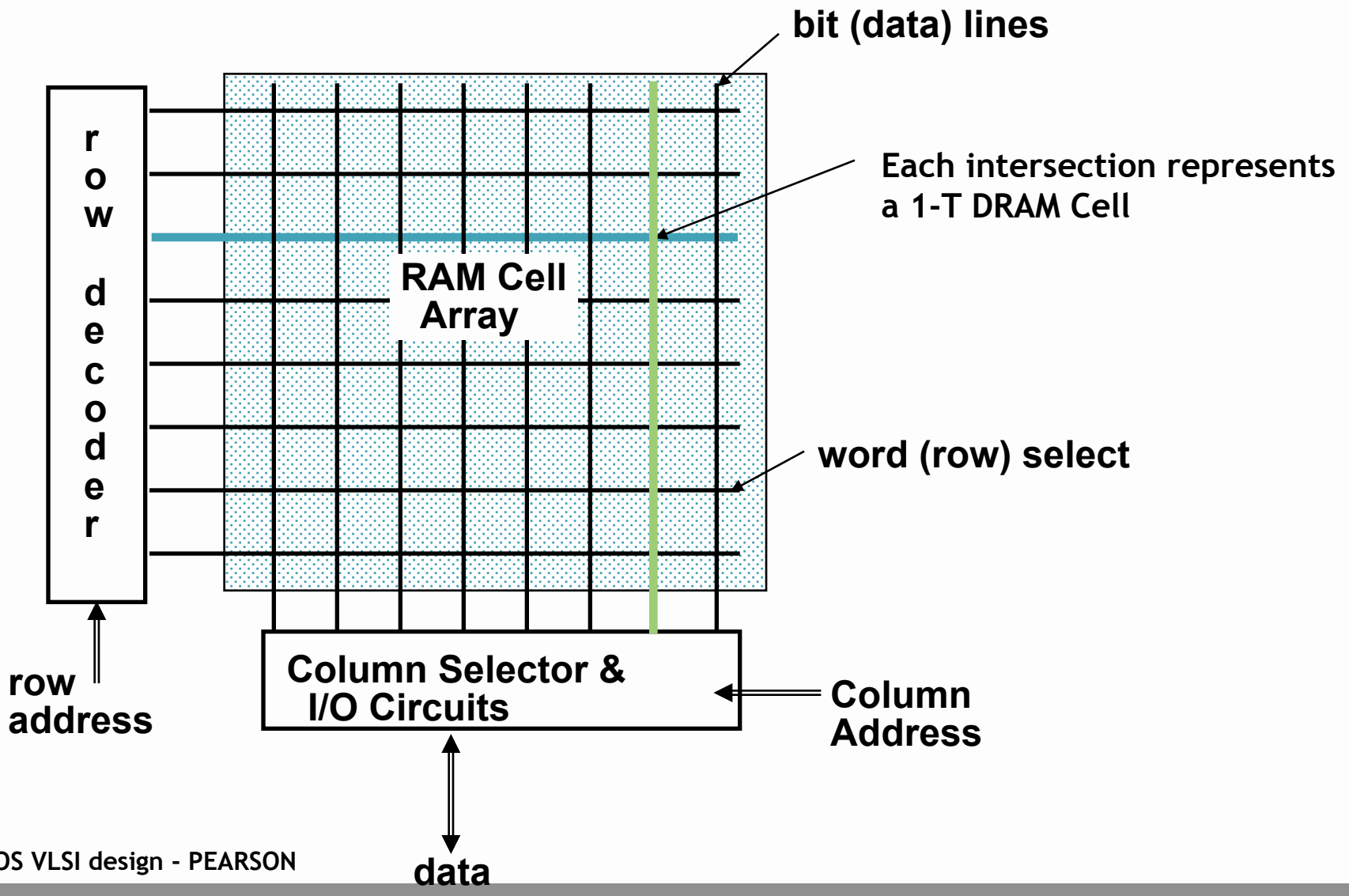
Back End

Front End

Trench eDRAM Process

- Stack capacitor requires more complex process
- M1 height above gate is increased with stacked capacitor
  - M1 parasitics significantly change when wafer is processed w/o eDRAM
  - Drives unique timings for circuit blocks processed w/ and w/o eDRAM
    - Logic Equivalency is compromised – Trench is Better Choice

Classical DRAM Organization



DRAM Subarray

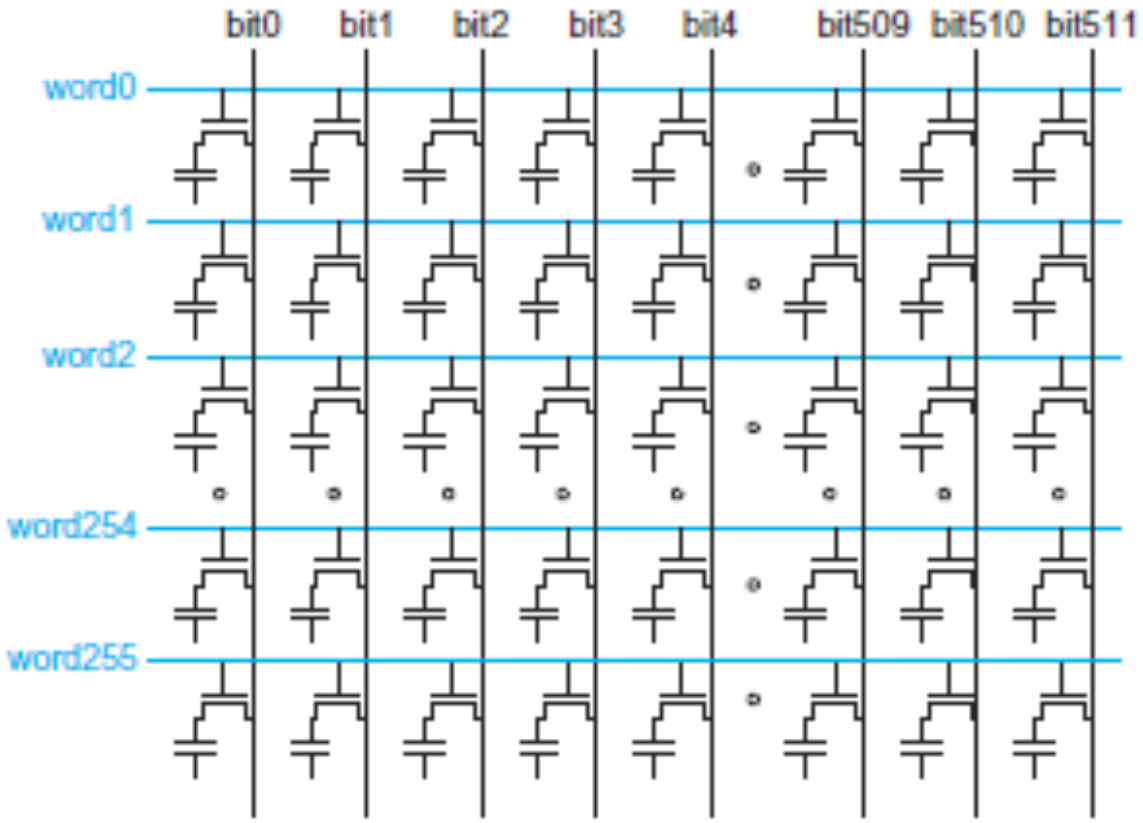
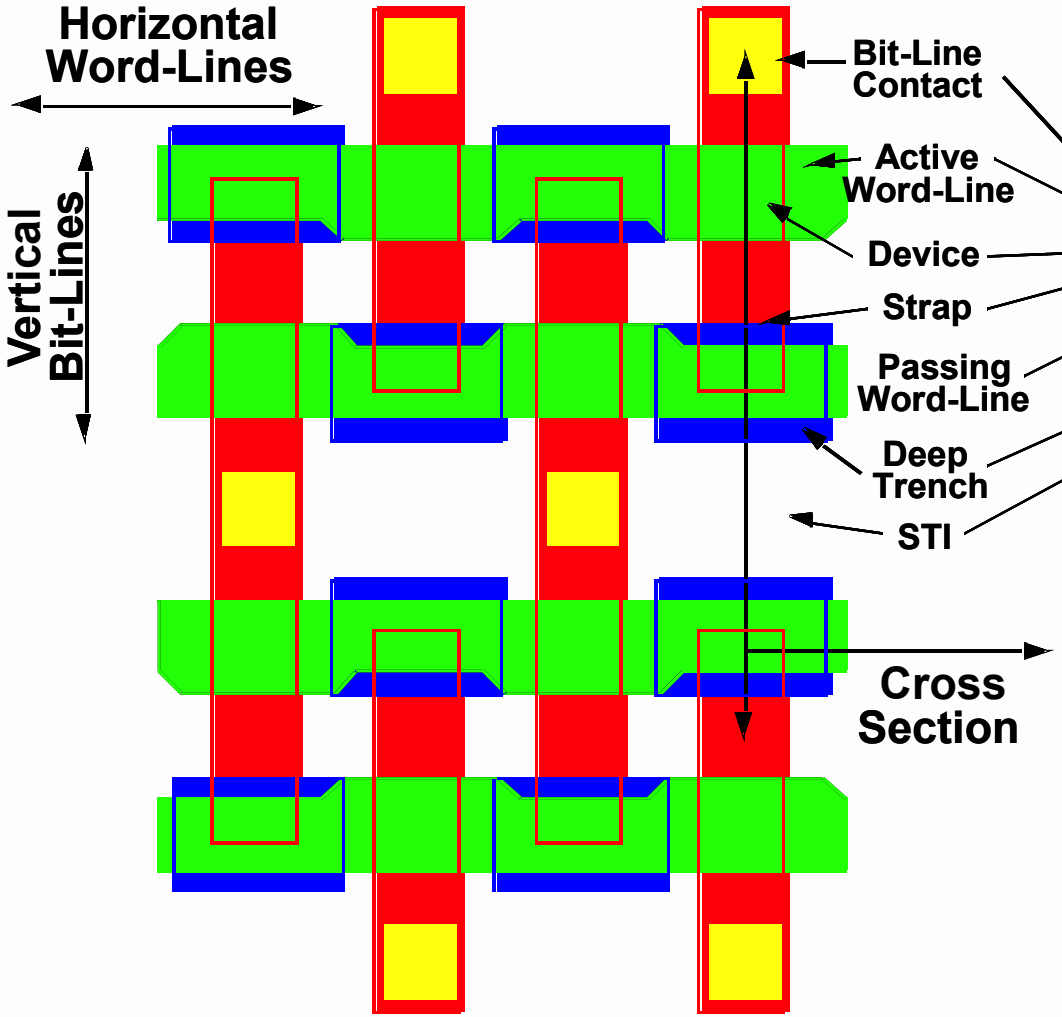
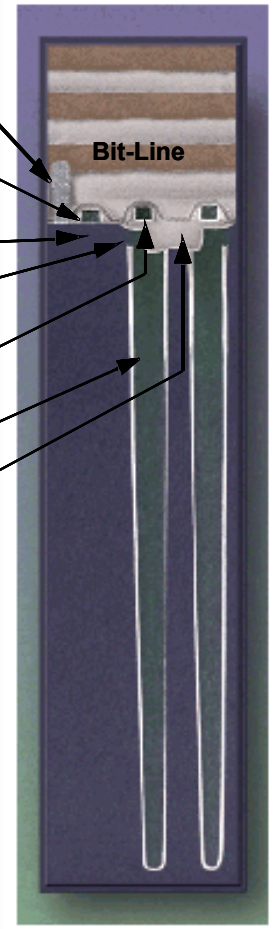


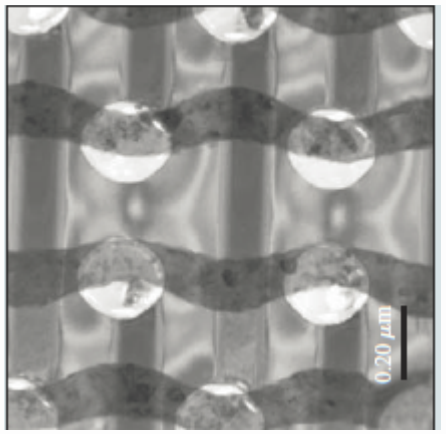
FIGURE 12.43 DRAM subarray

CMOS VLSI design - PEARSON

Trench cell layout and cross-section







Silicon Image

References so far

Barth, J. et al., "A 300MHz Multi-Banked eDRAM Macro Featuring GND Sense, Bit-line Twisting and Direct Reference Cell Write," ISSCC Dig. Tech. Papers, pp. 156-157, Feb. 2002.

Barth, J. et. al., "A 500MHz Multi-Banked Compilable DRAM Macro with Direct Write and Programmable Pipeline," ISSCC Dig. Tech. Papers, pp. 204-205, Feb. 2004.

Barth, J. et al., "A 500MHz Random Cycle 1.5ns-Latency, SOI Embedded DRAM Macro Featuring a 3T Micro Sense Amplifier," ISSCC Dig. Tech. Papers, pp. 486-487, Feb. 2007.

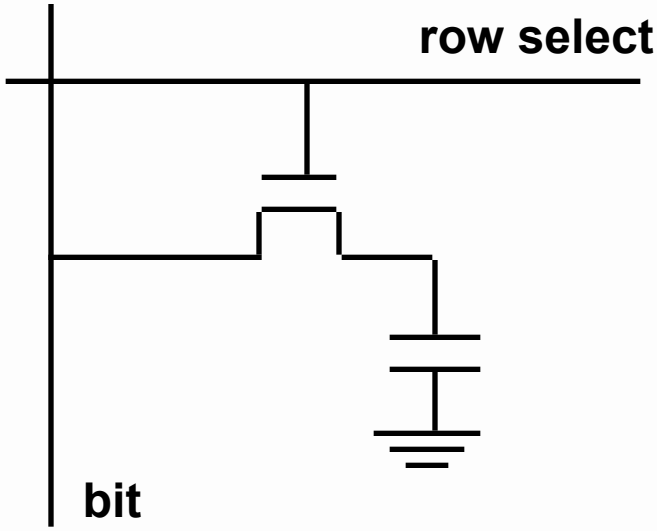
Barth, J. et al., "A 45nm SOI Embedded DRAM Macro for POWER7TM 32MB On-Chip L3 Cache," ISSCC Dig. Tech. Papers, pp. 342-3, Feb. 2010.

Butt,N., et al., "A 0.039um2 High Performance eDRAM Cell based on 32nm High-K/Metal SOI Technology," IEDM pp. 27.5.1-2, Dec 2010.

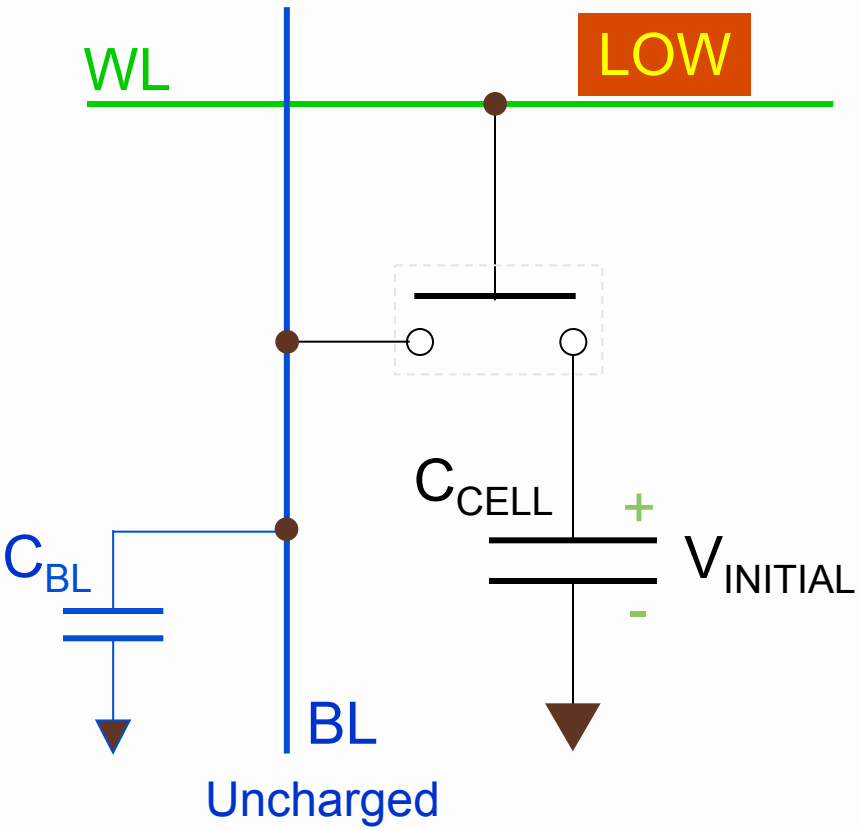
Bright, A. et al., "Creating the BlueGene/L Supercomputer from Low-Power SoC ASICs," ISSCC Dig. Tech. Papers, pp. 188-189, Feb. 2005.

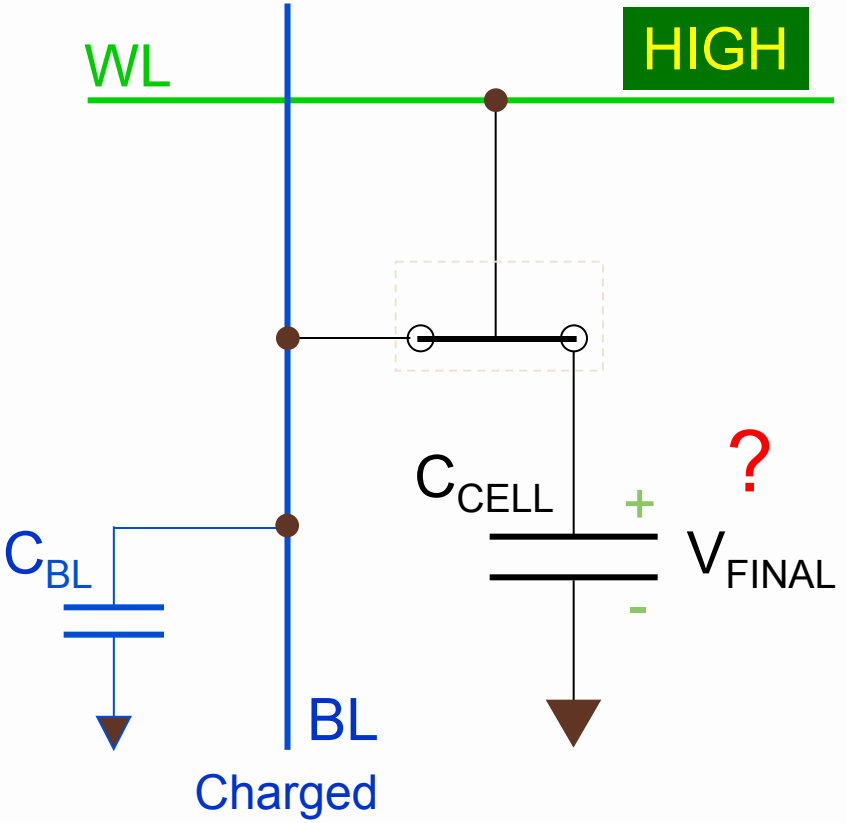
DRAM Read, Write and Refresh

- Write:
  - 1. Drive bit line
  - 2. Select row
- Read:
  - 1. Precharge bit line
  - 2. Select row
  - 3. Cell and bit line share charges
    - Signal developed on bitline
  - 4. Sense the data
  - 5. Write back: restore the value
- Refresh
  - 1. Just do a dummy read to every cell → auto write-back



Cell transfer ratio

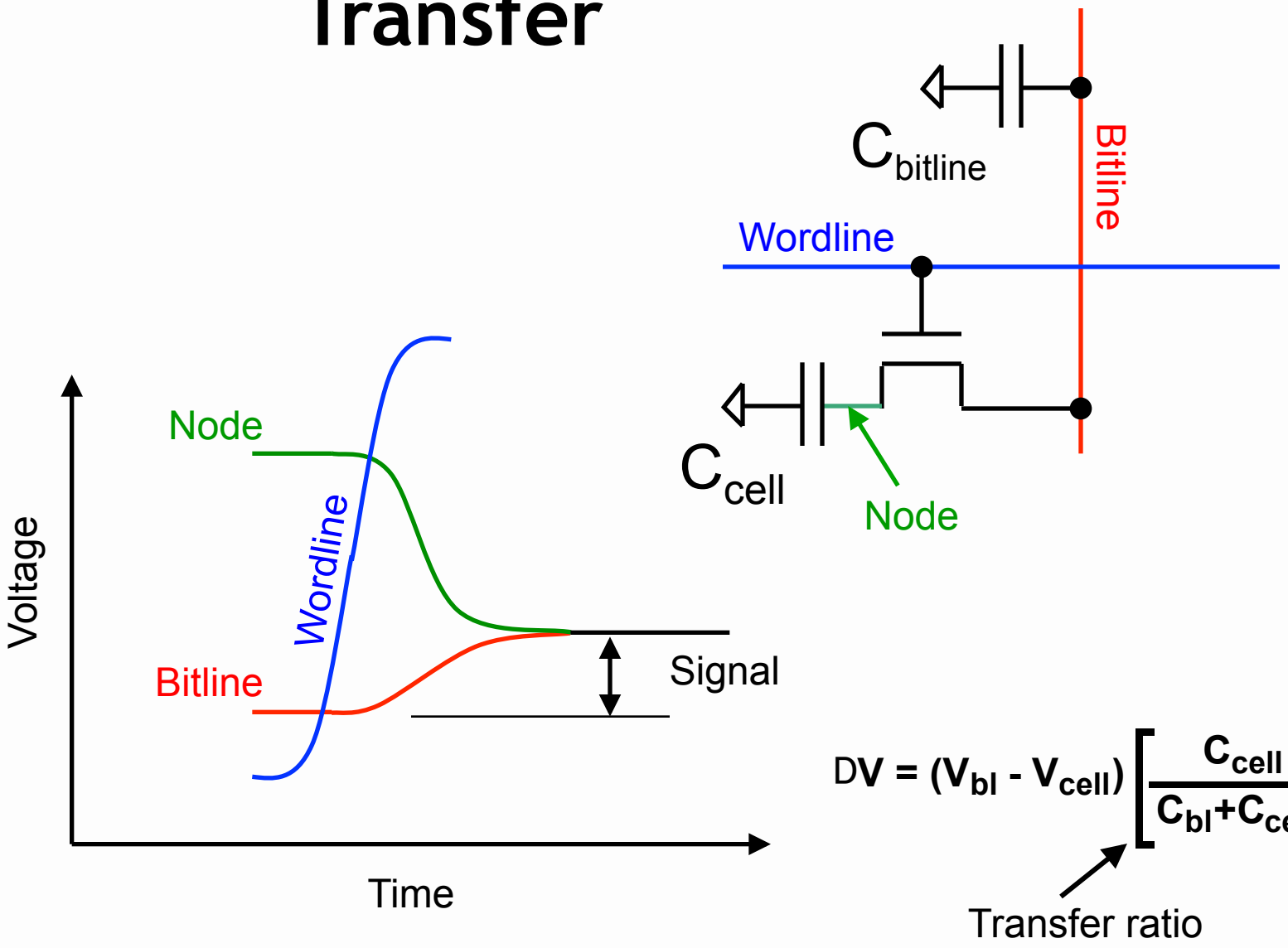




$$C_{CELL} \times V_{INITIAL} = (C_{CELL} + C_{BL}) \times V_{FINAL}$$

$$\text{Transfer ratio} = C_{CELL} / (C_{CELL} + C_{BL})$$

Cell Charge Transfer



$$DV = (V_{bl} - V_{cell}) \left[\frac{C_{cell}}{C_{bl}+C_{cell}} \right]$$

Transfer ratio

Transfer Ratio and Signal

- Bit-Line Voltage Calculated from Initial Conditions and Capacitances:

$$\Delta V = V_{bl} - V_f = V_{bl} - \frac{Q}{C} = V_{bl} - \left[\frac{C_{bl}*V_{bl}+C_{cell}*V_{cell}}{C_{bl}+C_{cell}}\right]$$

$$\Delta V = (V_{bl} - V_{cell})\left[\frac{C_{cell}}{C_{bl}+C_{cell}}\right]$$

Transfer Ratio (typically 0.2)

- Bit-Line Voltage is Amplified with Cross Couple “Sense Amp”

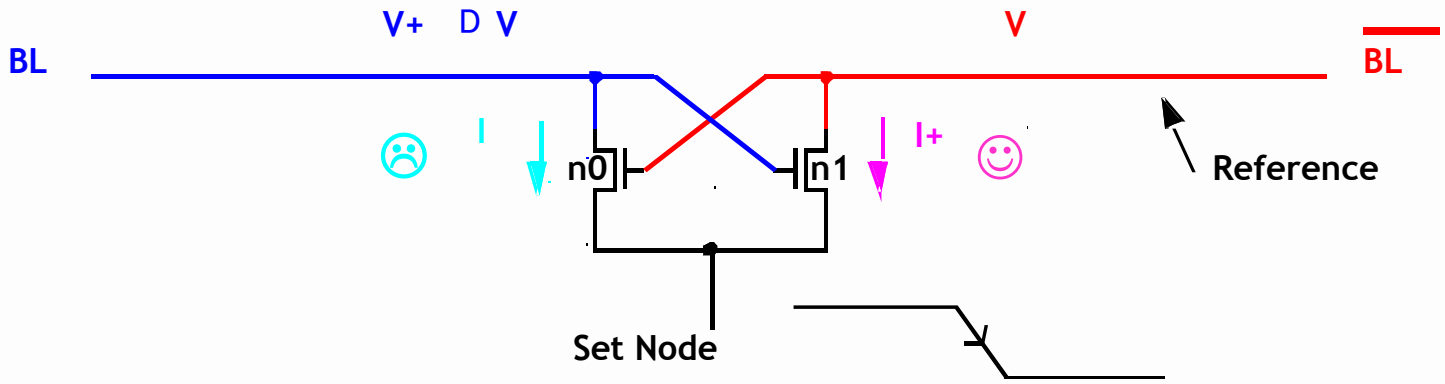
Sense Amp Compares Bit-Line Voltage with a Reference

Bit-Line Voltage - Reference = Signal

Pos Signal Amplifies to Logical ‘1’, Neg Signal Amplifies to Logical ‘0’

Sensing → Signal Amplification

Differential Voltage Amplified by Cross Couple Pair



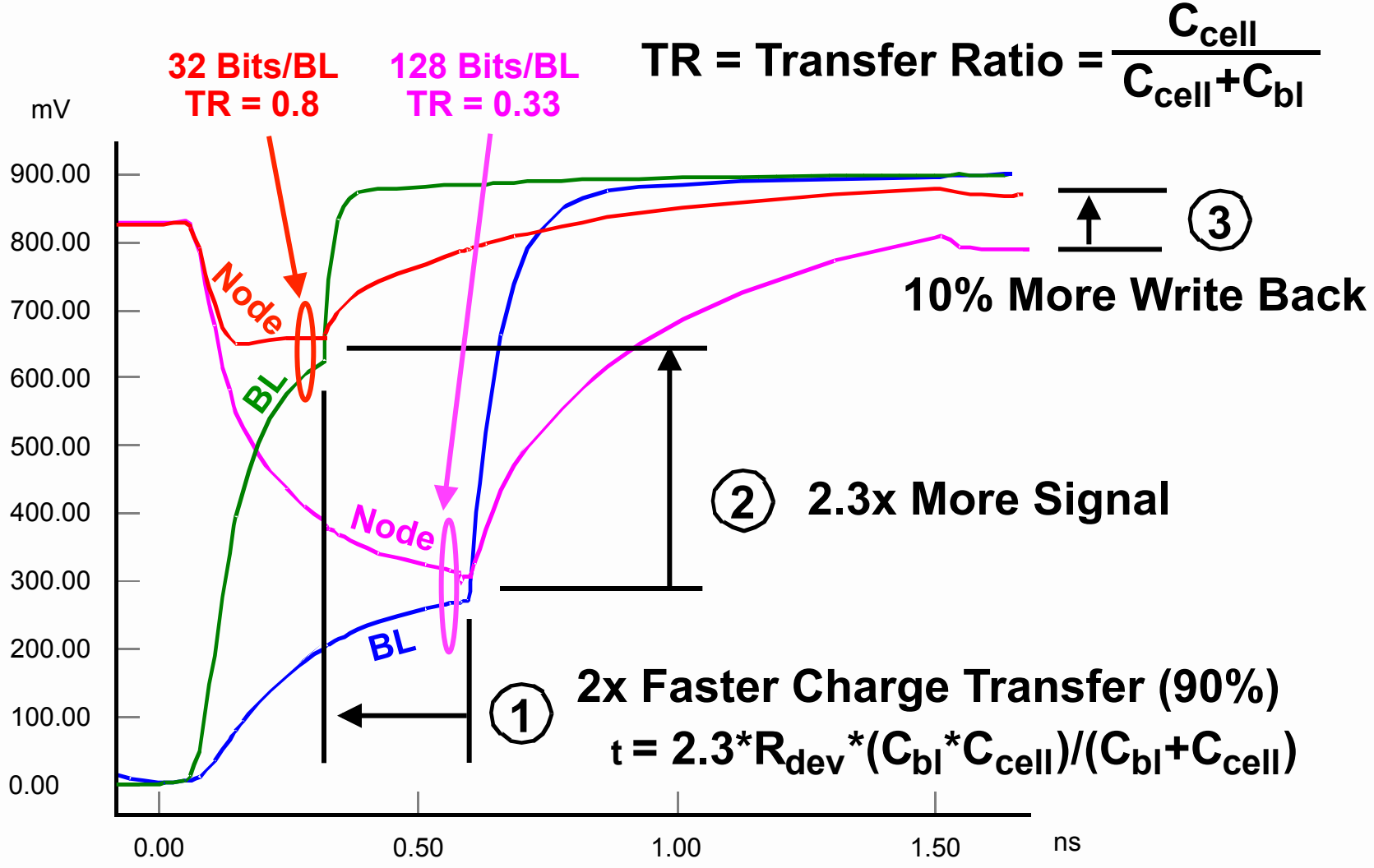
When Set Node < (V+DV) - V_{tn1}, I+ will start to flow (On-Side Conduction)

When Set Node < (V) - V_{tn0}, I will start to flow (Off-Side Conduction)

Off-Side Conduction Modulated by Set Speed and Amount of Signal

Complimentary X-Couple Pairs provide Full CMOS Levels on Bit-Line

Bits per Bit-Line v/s Transfer Ratio



JSSC08

Segmentation

Array Segmentation Refers to WL / BL Count per Sub-Array

Longer Word-Line is Slower but more Area efficient (Less Decode/Drivers)

Longer Bit-Line (more Word-Lines per Bit-Line)

- Less Signal (Higher Bit-Line Capacitance = Lower Transfer Ratio)
- More Power (Bit-Line CV is Significant Component of DRAM Power)
- Slower Performance (Higher Bit-Line Capacitance = Slower Sense Amp)
- More Area Efficient (Fewer Sense Amps)

Number of Word-Lines Activated determines Refresh Interval and Power

- All Cells on Active Word-Line are Refreshed
- All Word-Lines must be Refreshed before Cell Retention Expires
- 64ms Cell Retention / 8K Word Lines = 7.8us between refresh cycles
- Activating 2 Word-Lines at a time = 15.6us, 2x Bit-Line CV Power

Choice of SA

Depending on signal developed SA architecture is chosen

Direct sensing

- Requires large signal development
- An inverter can be used for sensing
- Micro sense amp (uSA) is another option

Differential sense amp

- Can sense low signal developed

This is choice between area, speed/performance

Topics

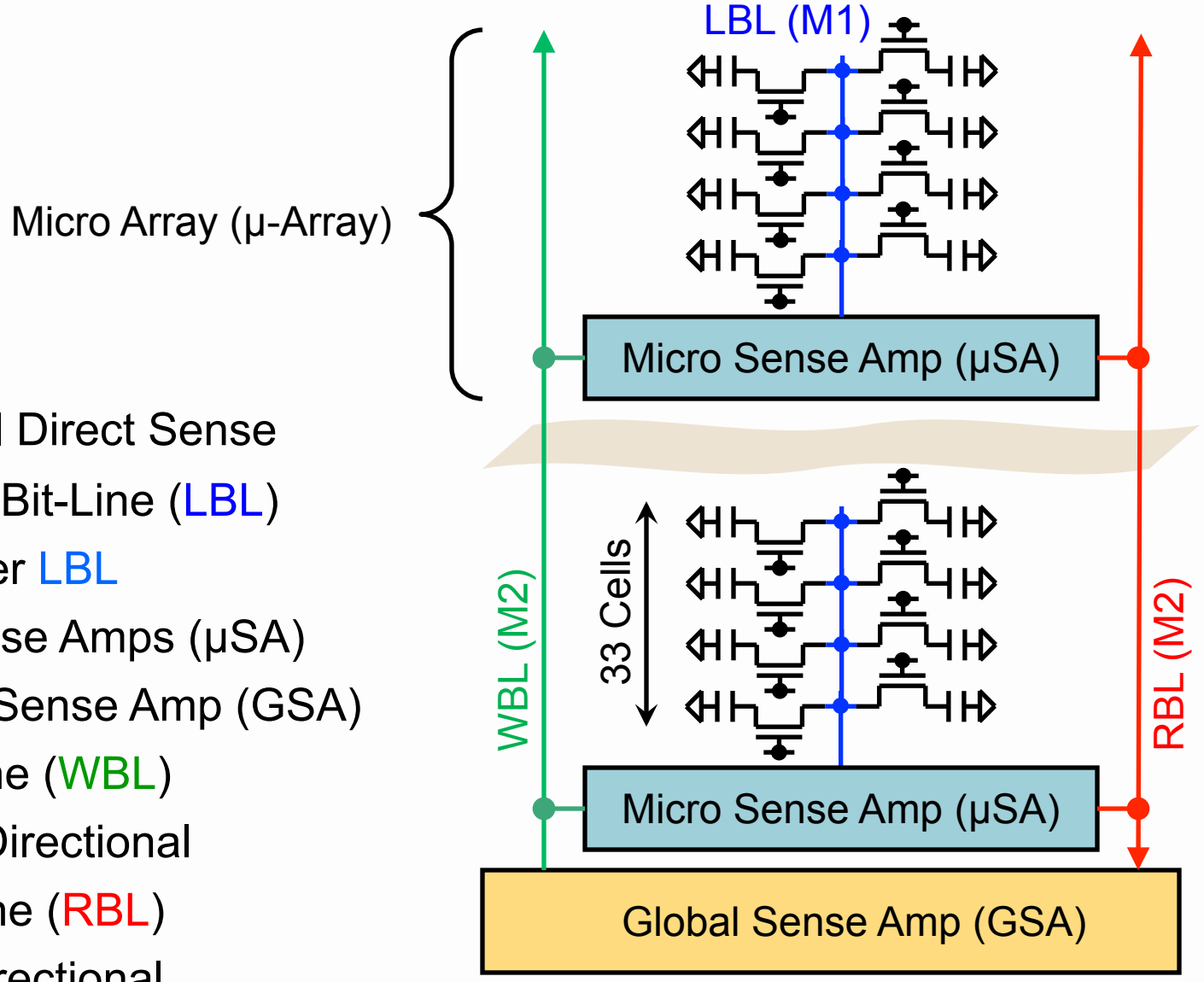
- ❑ Introduction to memory
- ❑ DRAM basics and bitcell array
- ❑ eDRAM operational details (case study)
- ❑ Noise concerns
- ❑ Wordline driver (WLDRV) and level translators (LT)
- ❑ Challenges in eDRAM
- ❑ Understanding Timing diagram – An example

DRAM Operation Details (Case Study)

IEEE JOURNAL OF SOLID-STATE CIRCUITS, VOL. 43, NO. 1, JANUARY 2008

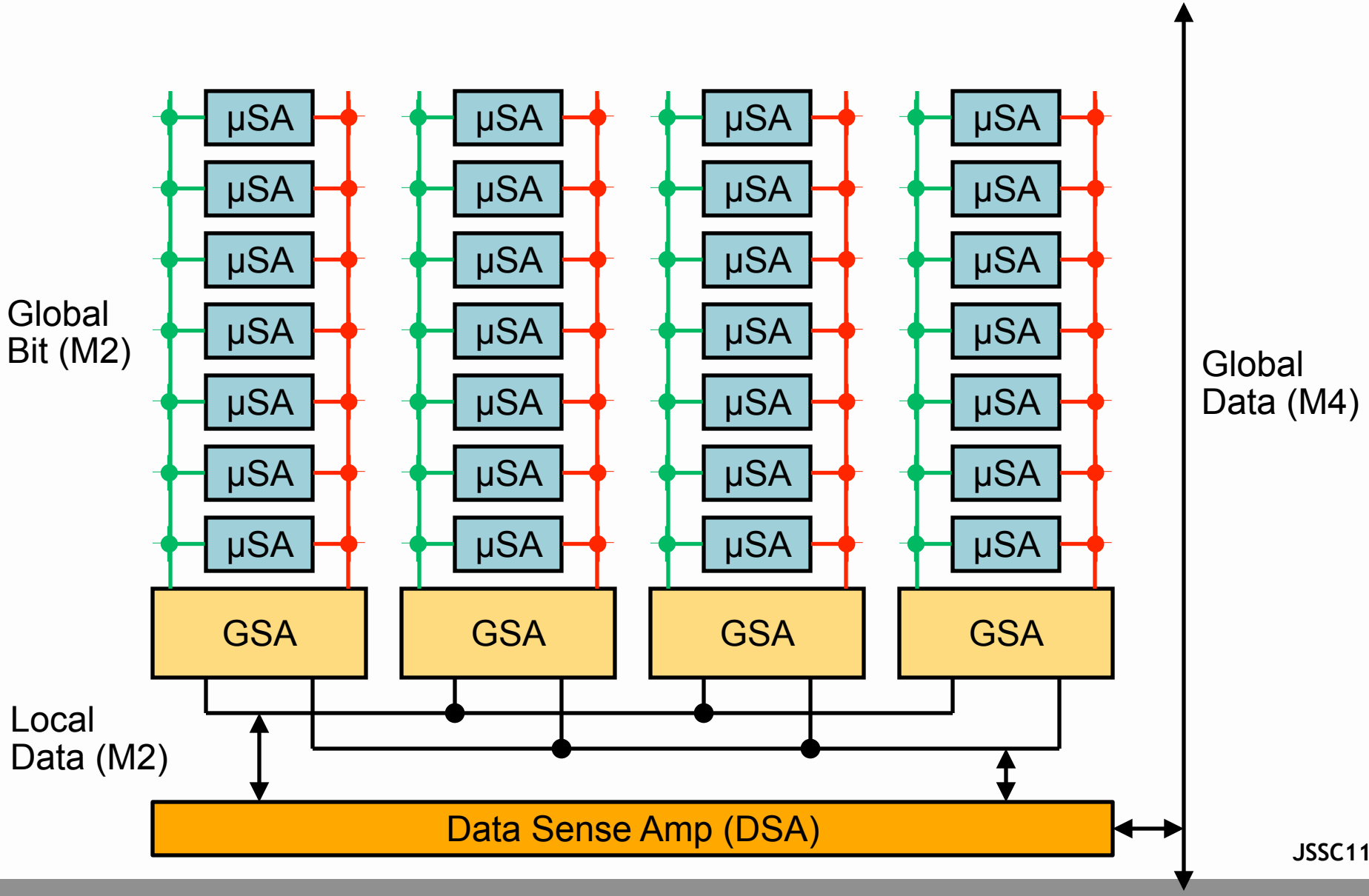
A 500 MHz Random Cycle, 1.5 ns Latency, SOI Embedded DRAM Macro Featuring a Three-Transistor Micro Sense Amplifier (John Barth/IBM)

Micro Sense Architecture



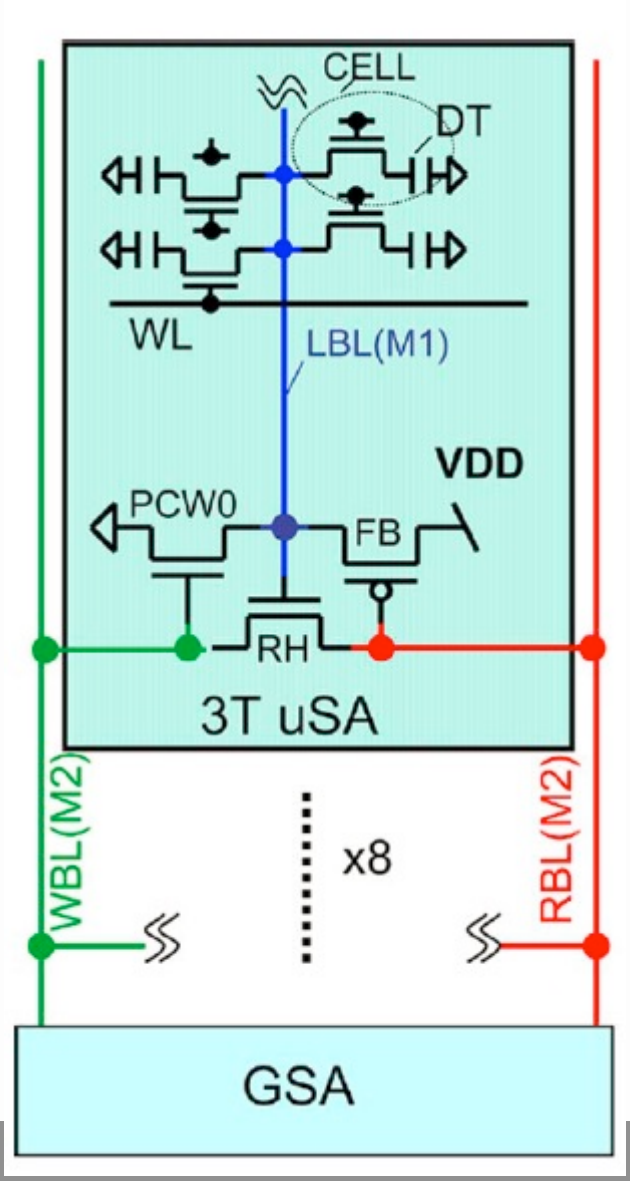
- Hierarchical Direct Sense
- Short Local Bit-Line (LBL)
  - 33 Cells per LBL
- 8 Micro Sense Amps (μSA)
  per Global Sense Amp (GSA)
- Write Bit-Line (WBL)
  Uni-Directional
- Read Bit-Line (RBL)
  Bi-Directional

Micro Sense Hierarchy - Three levels



JSSC11

3T uSA operation



Pre-charge

WL is low. WBL and RBL both pre-charged to HIGH.
Next GSA drives WBL low. **LBL floats to GND level**

Read "0"

LBL remains LOW. RBL is HIGH. Sensed as a "0"

Read "1"

LBL is HIGH. Turns on RH, pulls RBL LOW.
+ feedback as pFET FB turns ON. Sensed as a "1"

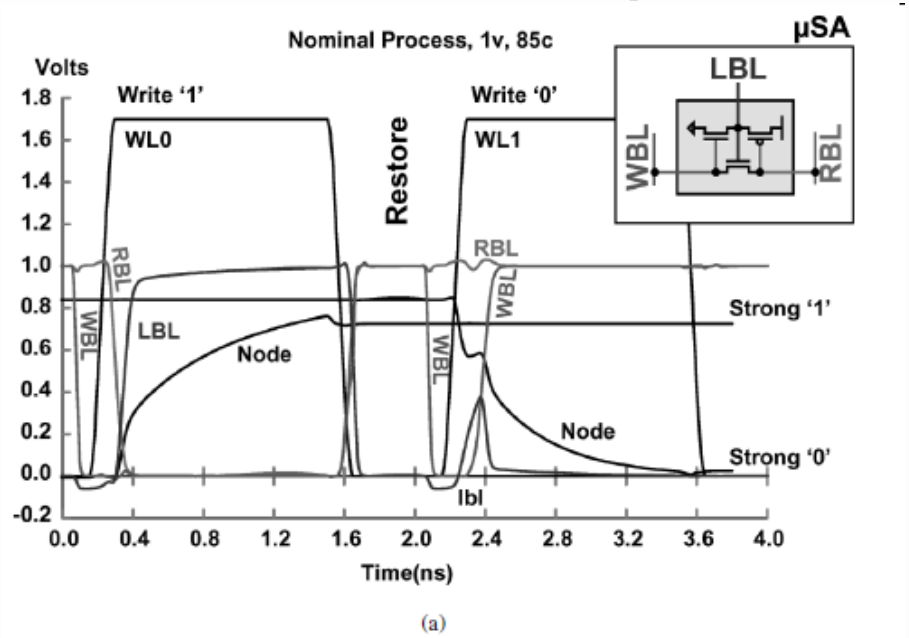
Write "1"

GSA pulls RBL to GND. FB pFET turns ON
Happens while WL rises (direct write)

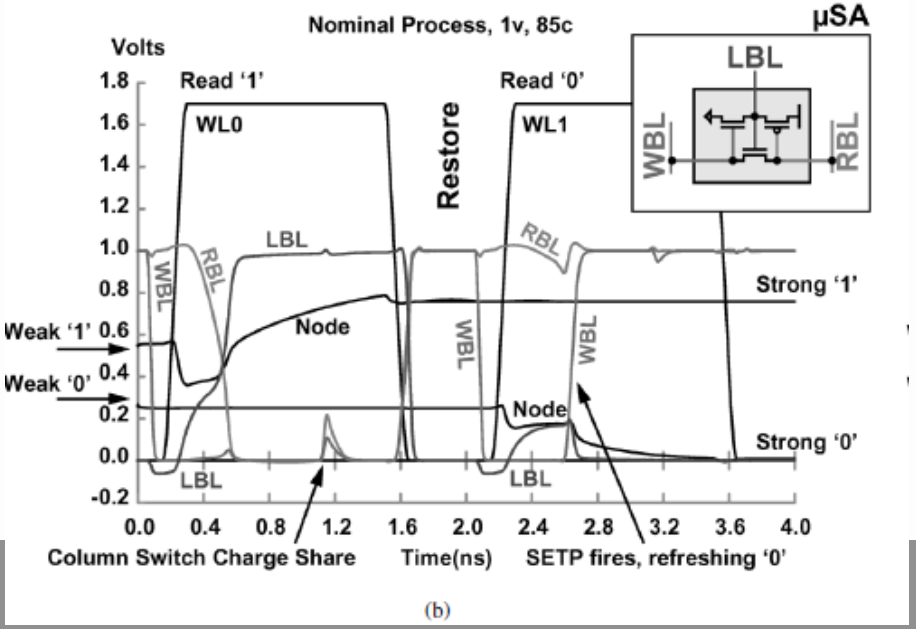
Write "0"

WBL is HIGH, PCW0 ON. Clamps LBL to GND
As WL activates.

Micro Sense Amp Simulations



(a)



(b)

IEEE JOURNAL OF SOLID-STATE CIRCUITS, VOL. 43, NO. 1, JANUARY 2008

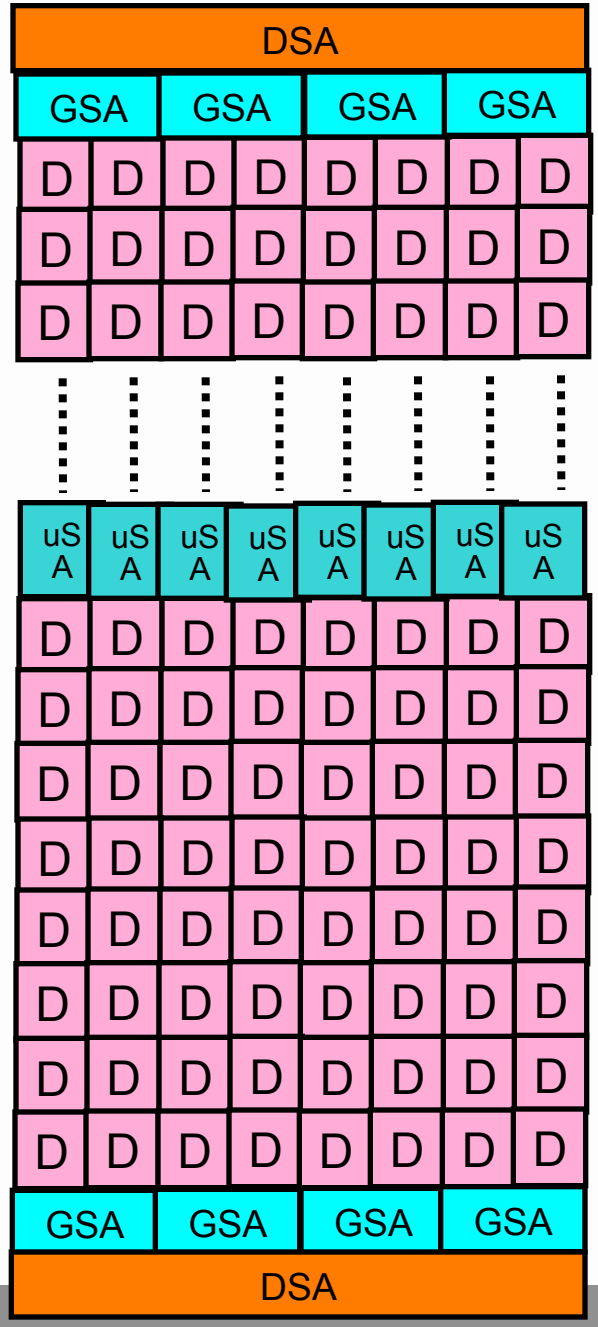
A 500 MHz Random Cycle, 1.5 ns Latency, SOI Embedded DRAM Macro Featuring a Three-Transistor Micro Sense Amplifier

JSSC08

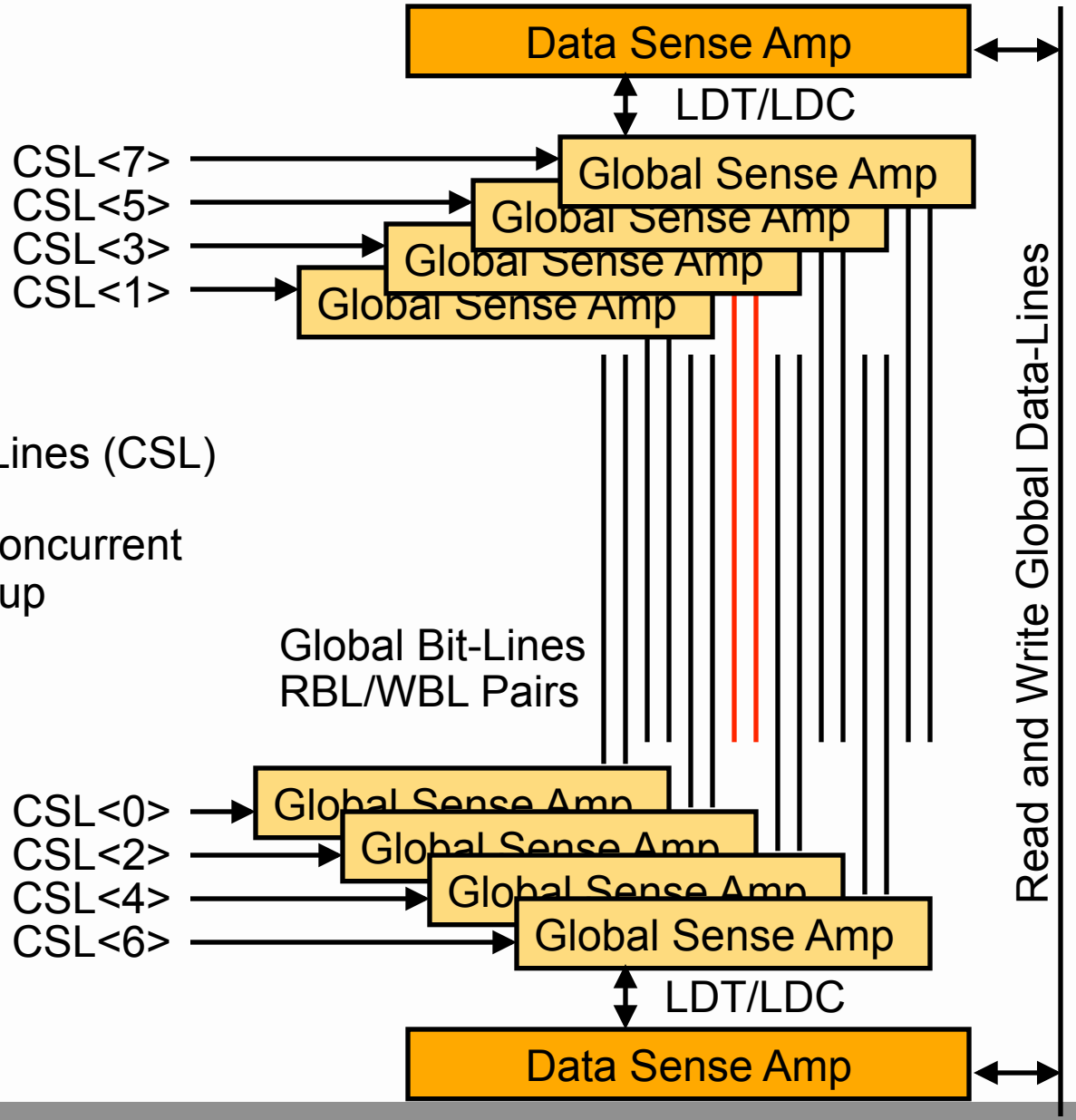
Layout Floor plan of Array+SA

GSA Should fit into the bitcell width or n*bitcell width

Thus, distributed GSA on two sides of bitcell array

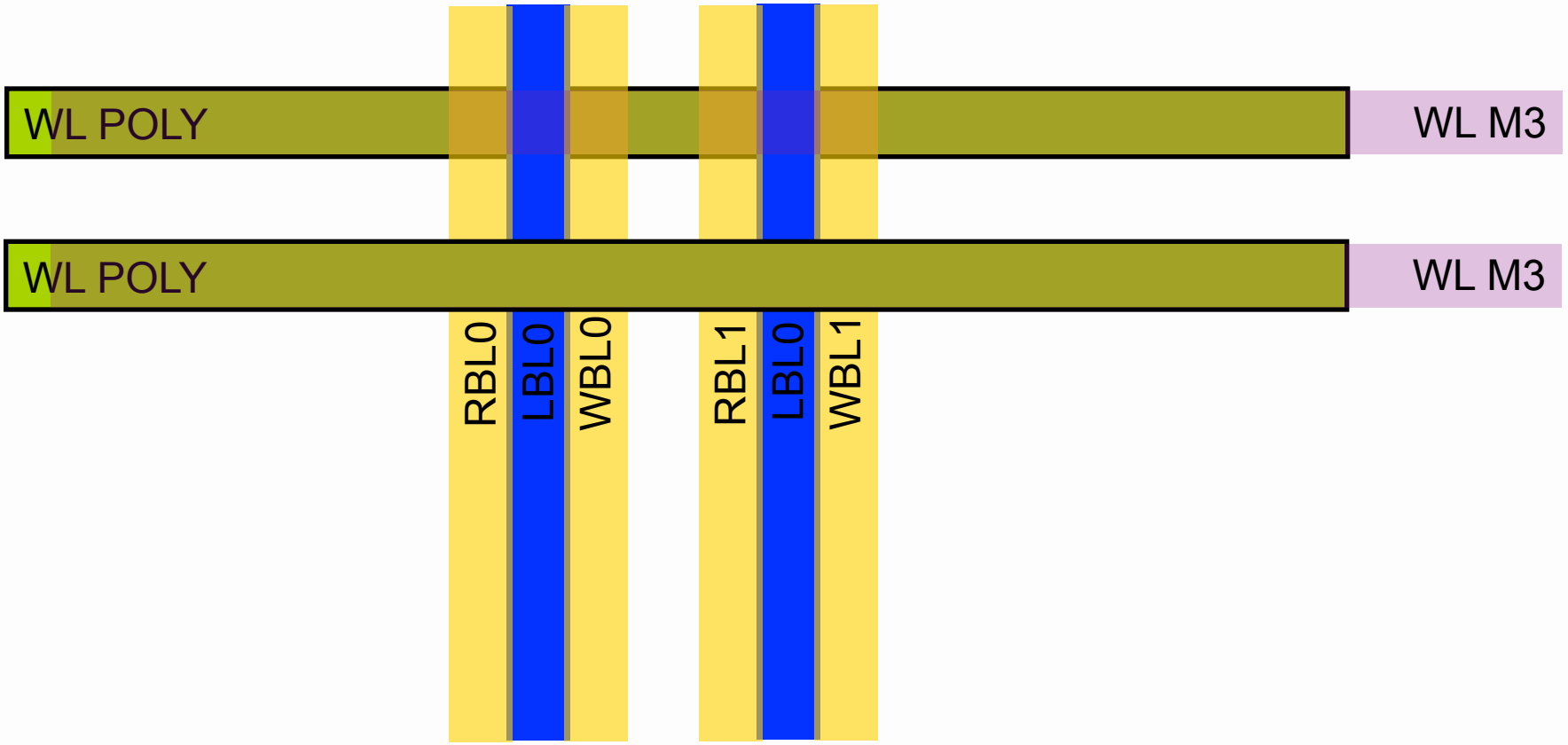


Column Interleave

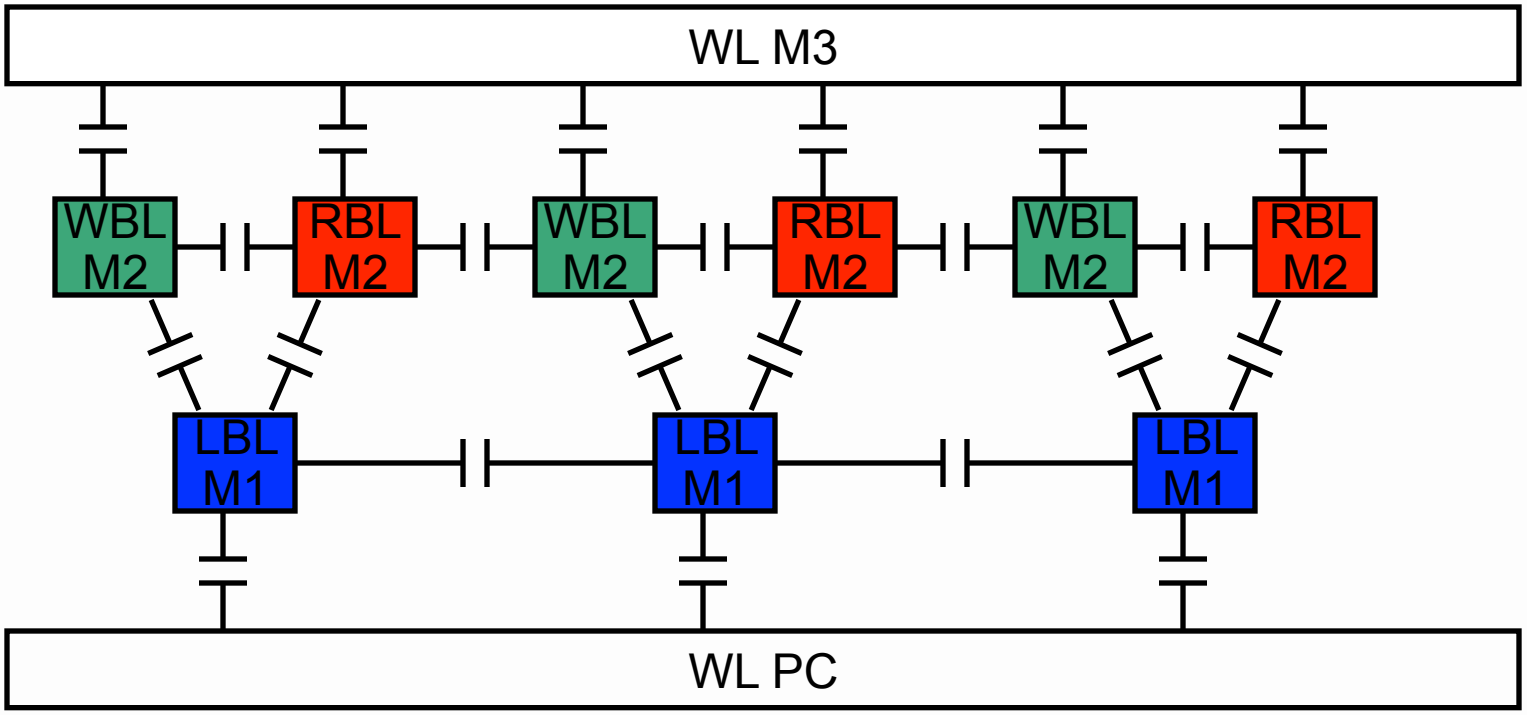


- 1 of 8 Column Select Lines (CSL)
- Fire Early for Write
- Fire Late to Support Concurrent Cache Directory Lookup

LAYOUT of array

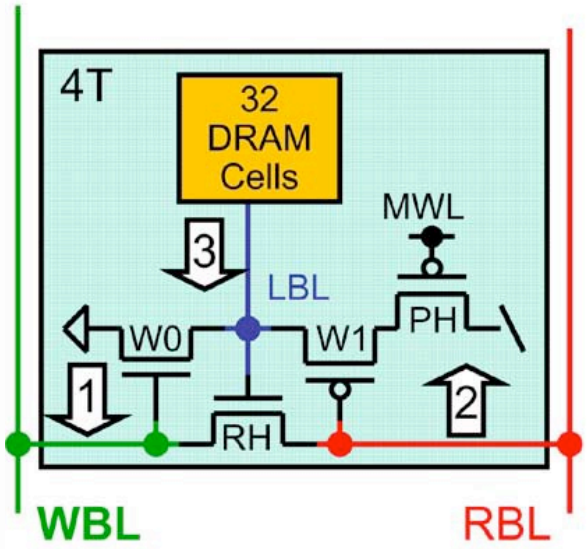


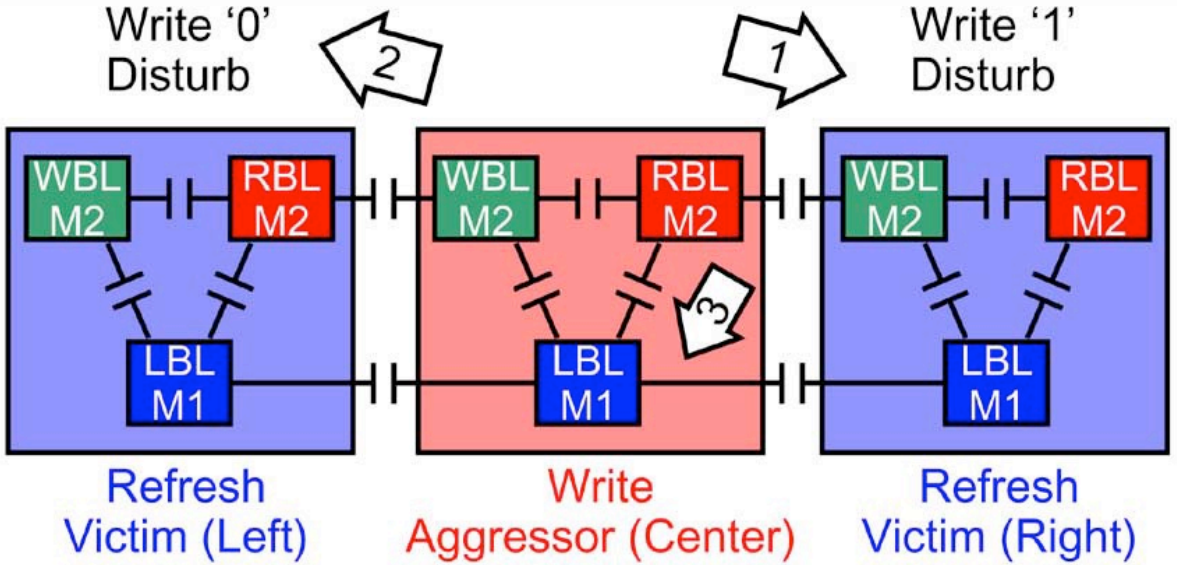
Micro Sense Local Bit-line Cross Section



Single Ended Sense – Twist not effective
Line to Line Coupling must be managed

Micro Sense Coupling Mechanisms

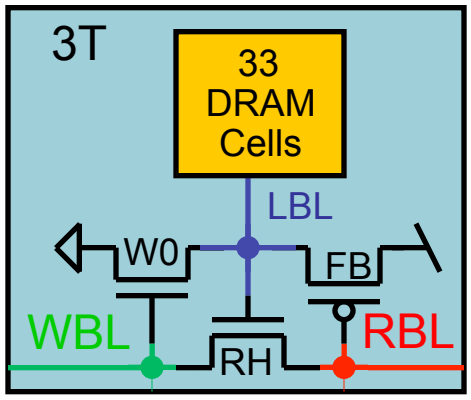




1. Write '1' Couples WBL below Ground
   Increasing RH leakage during Refresh '0'
2. Write '0' Couples RBL above VDD
   Delaying Feedback during Refresh '1'
3. Read '1' Couples Half-Selected LBL Below GND
   Increasing Array Device Sub-VT Leakage

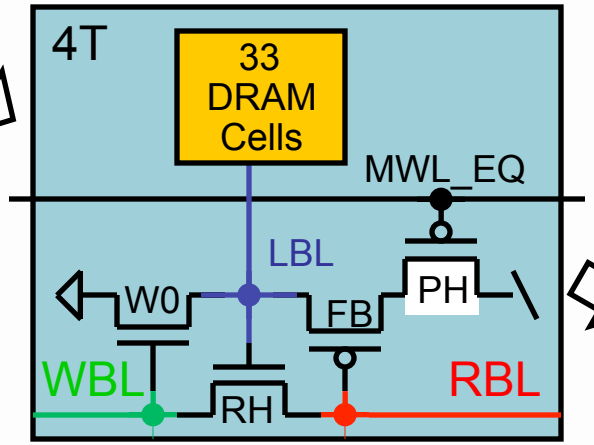
Micro Sense Evolution

1. Write Zero (W0)
2. Read Head (RH)
3. Feed-Back (FB)



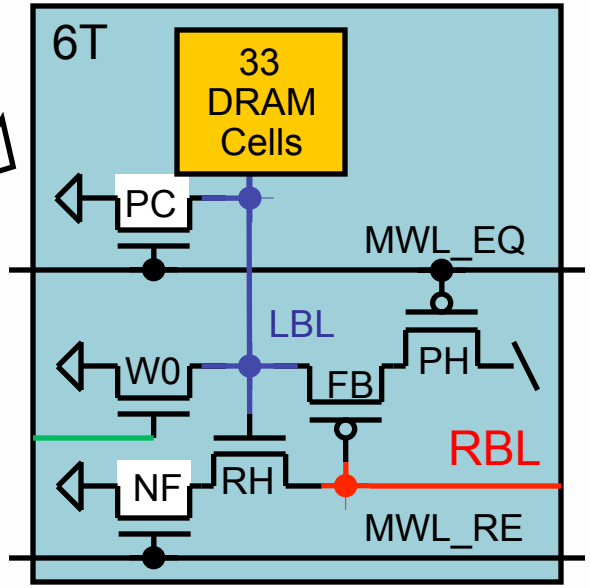
Barth, ISSCC'07

4. PFET Header (PH)
   - LBL Power Gate
   - LBL Leakage



Klim, VLSI'07

5. Pre-Charge (PC)
   - WBL Power (Write '0' Only)
6. NFET Footer (NF)
   - RBL Leakage
   - Decompose Pre-Charge and Read Enable (MWL_RE)



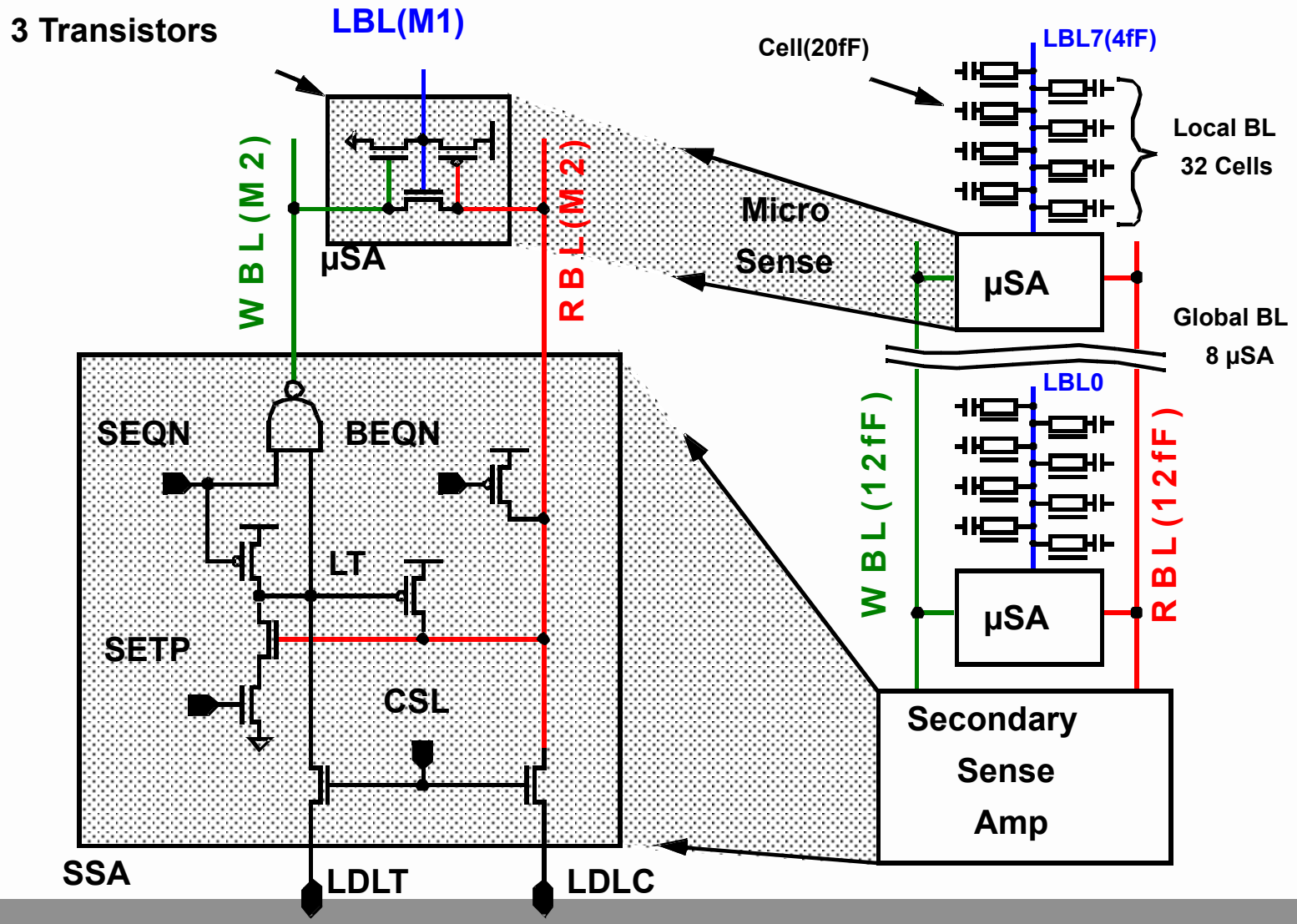
JSSC11

Power Reduction

Power Reduction
Traded for Transistor Count

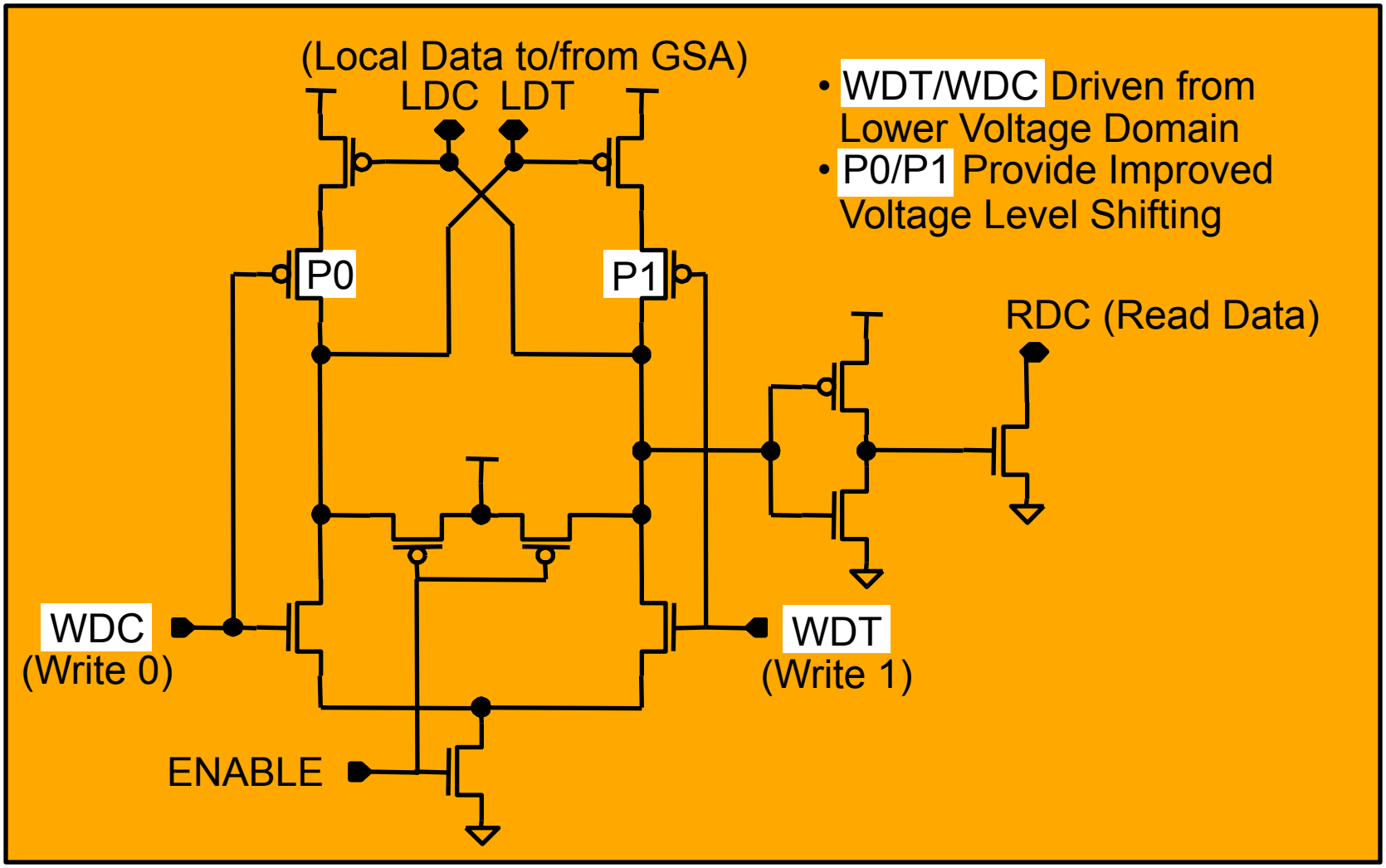
Increased Transistor Count

Micro Sense Architecture (μSA)



JSSC08

Data Sense Amp (DSA)



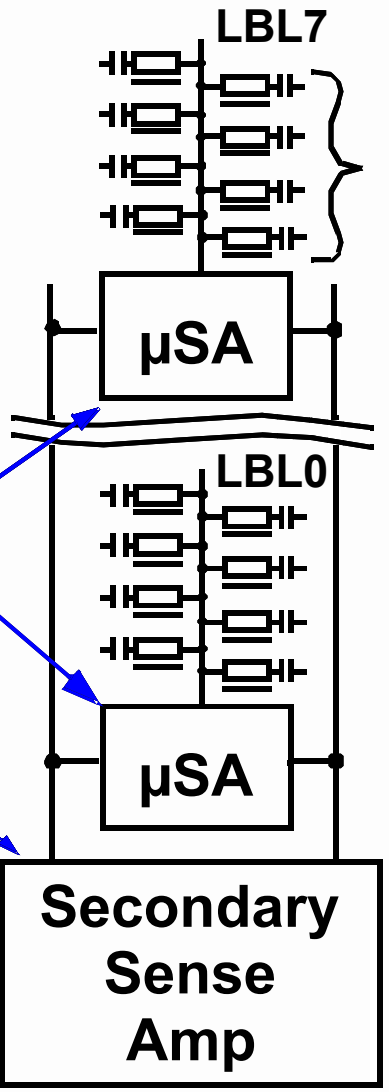
Micro Sense Advantage

Fast Performance of Short Bit-Line

Area Overhead of 4x Longer Bit-Line

Bits/BL	256	128	32
Sense Amp	10%	20%	19%
Reference Cells	2.3%	4%	-
Twist Region	2%	2.6%	-
Second Sense Amp	-	-	8%
Total	14.3%	26.6%	27%

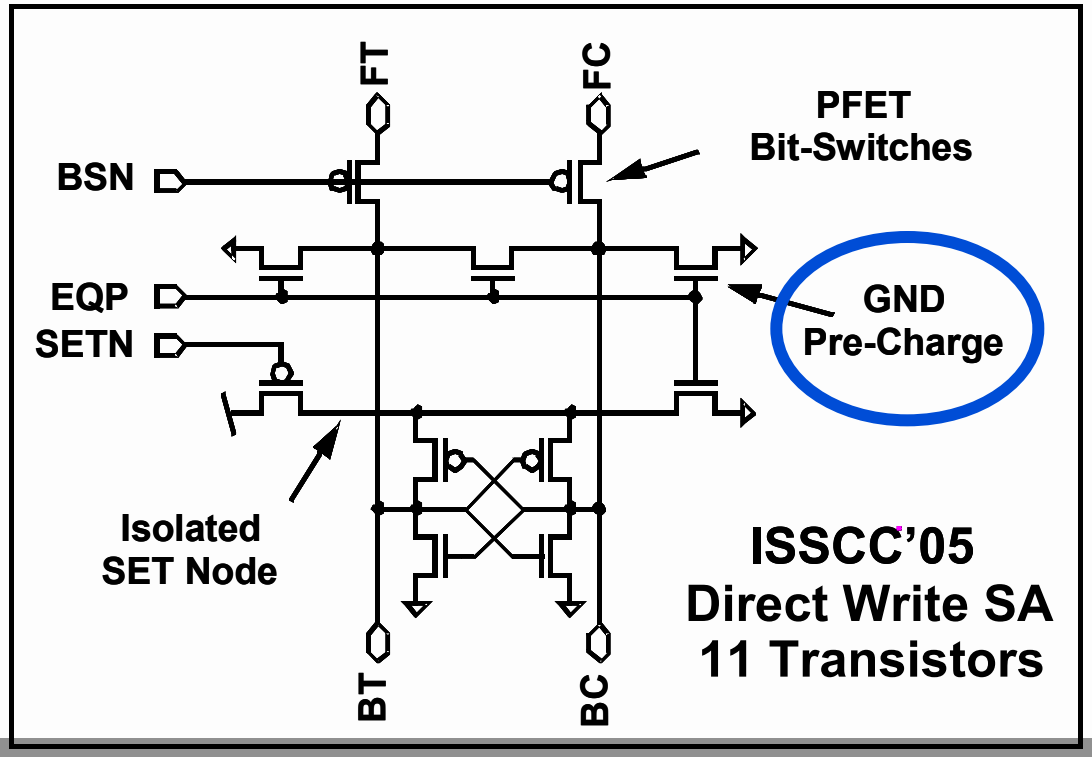
Same Overhead

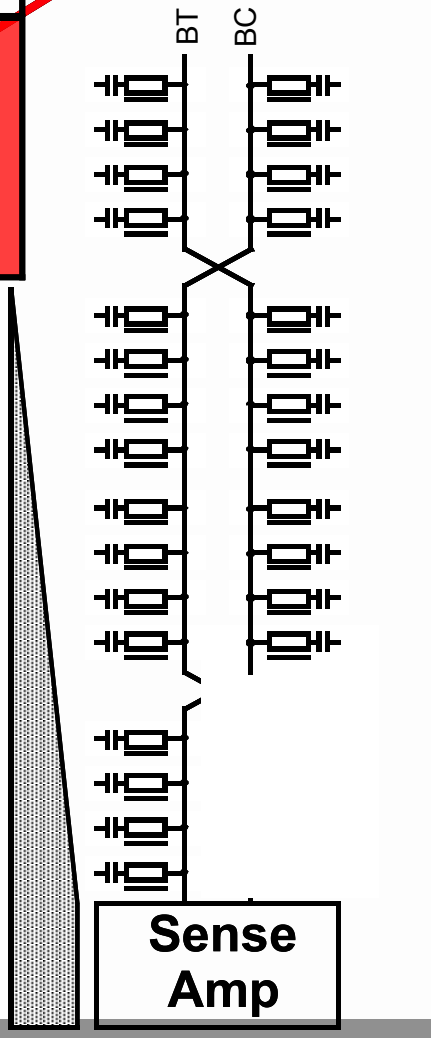


Bit-Line area overhead

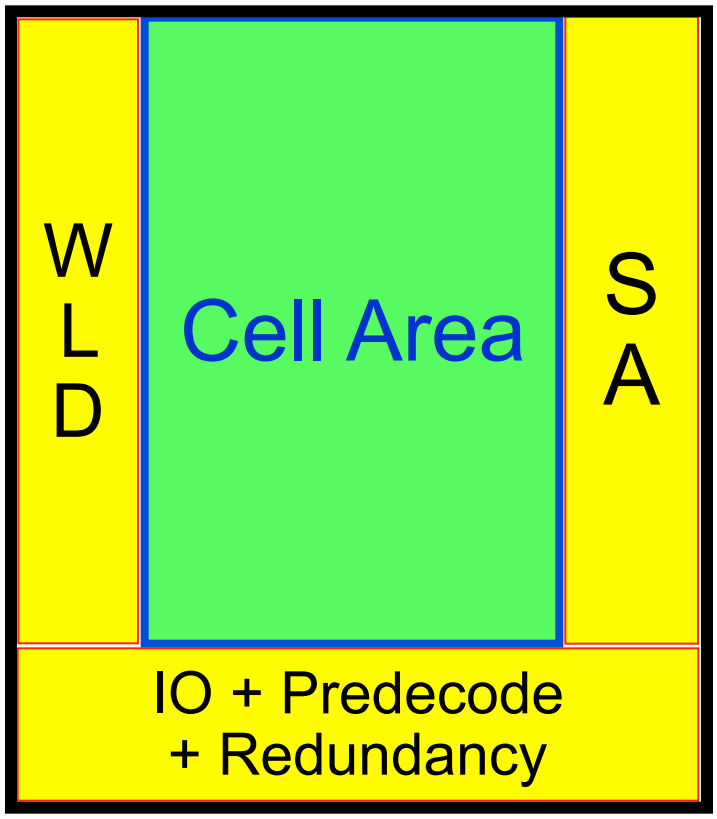
Bits/BL	256	128	32
Sense Amp	10%	20%	> 80%
Reference Cells	2.3%	4%	
Twist Region	2%	2.6%	

Unacceptable





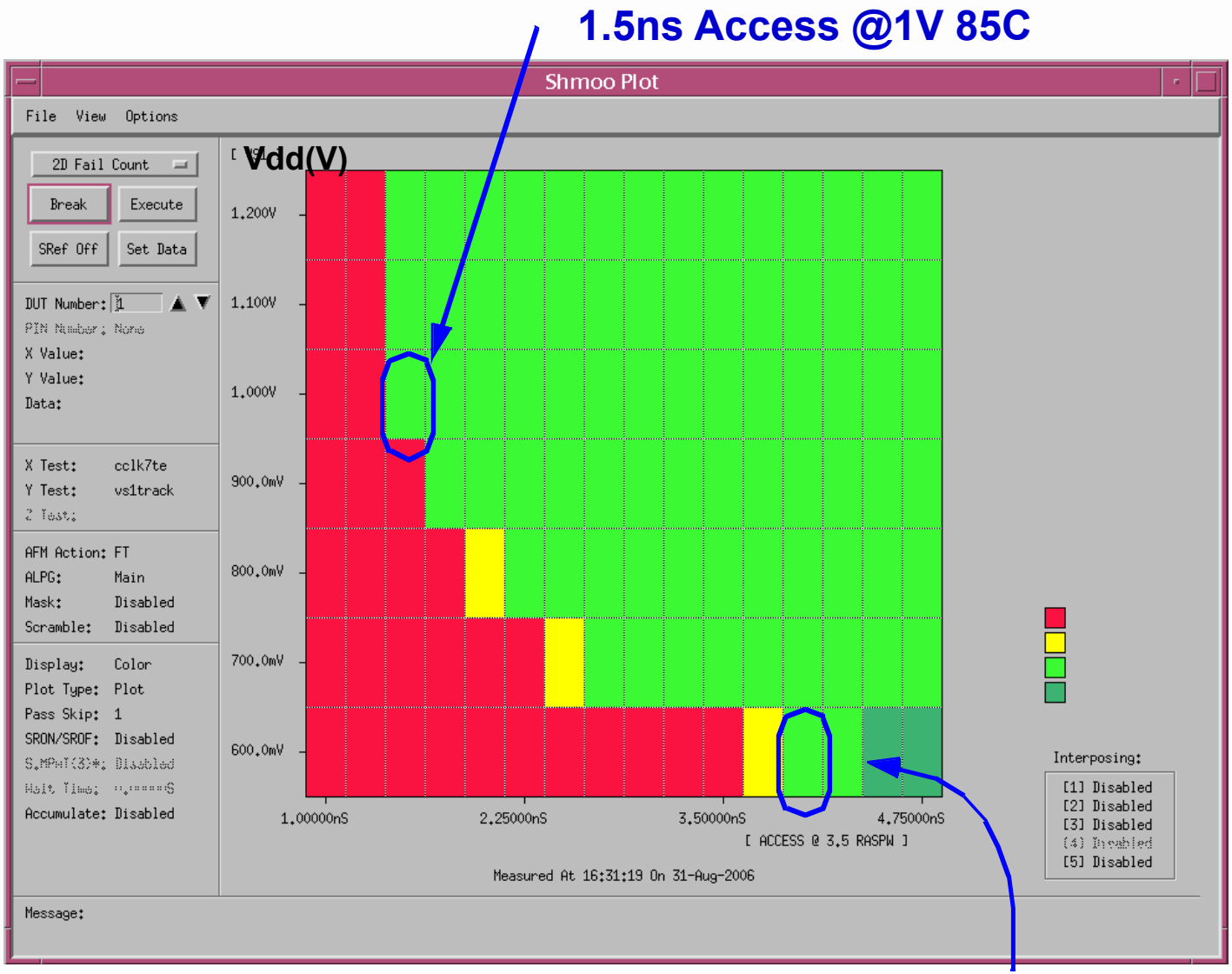
Array utilization



$$\text{Utilization} = \frac{\square}{\square}$$

Mbits/mm^2

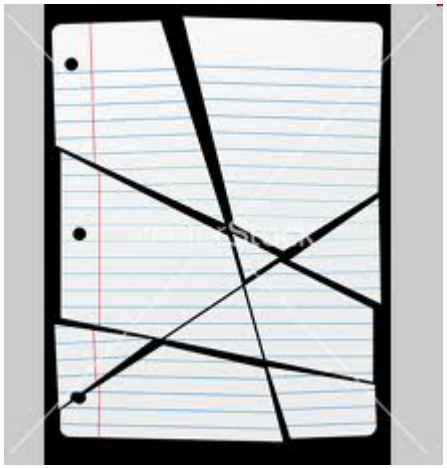
Access Shmoo



Redundancy

Notebook

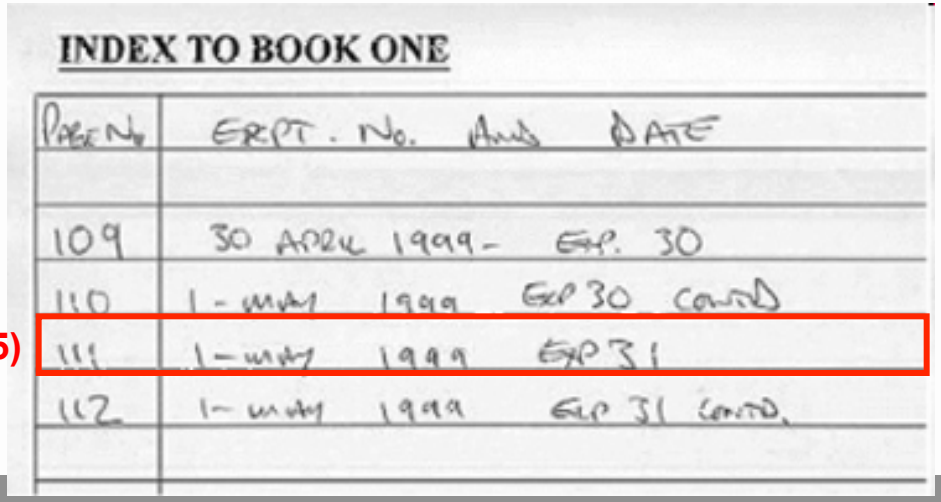
Page 111



Extra Page R05

eFuse based repair table

(see page R05)



INDEX TO BOOK ONE

PAGE No.	EXPT. No. AND	DATE
109	30 APRIL 1999 -	EXP. 30
110	1 - MAY	1999 EXP 30 CONTD
111	1 - MAY	1999 EXP 31
112	1 - MAY	1999 EXP 31 CONTD.

Topics

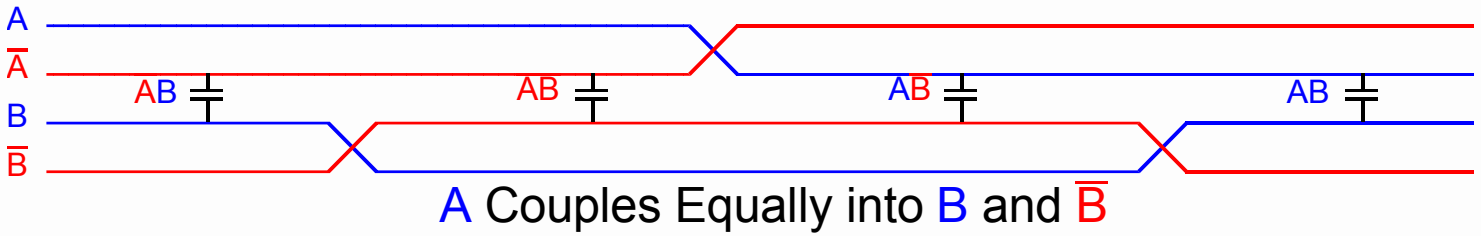
- ❑ Introduction to memory
- ❑ DRAM basics and bitcell array
- ❑ eDRAM operational details (case study)
- ❑ Noise concerns
- ❑ Wordline driver (WLDRV) and level translators (LT)
- ❑ Challenges in eDRAM
- ❑ Understanding Timing diagram – An example

Noise

Coupling and Local Process Variation effectively degrades signal

External Noise (Wire or Sx) Reduced to Common Mode by Folding

Line to Line Coupling Limited by Bit-Line Twisting

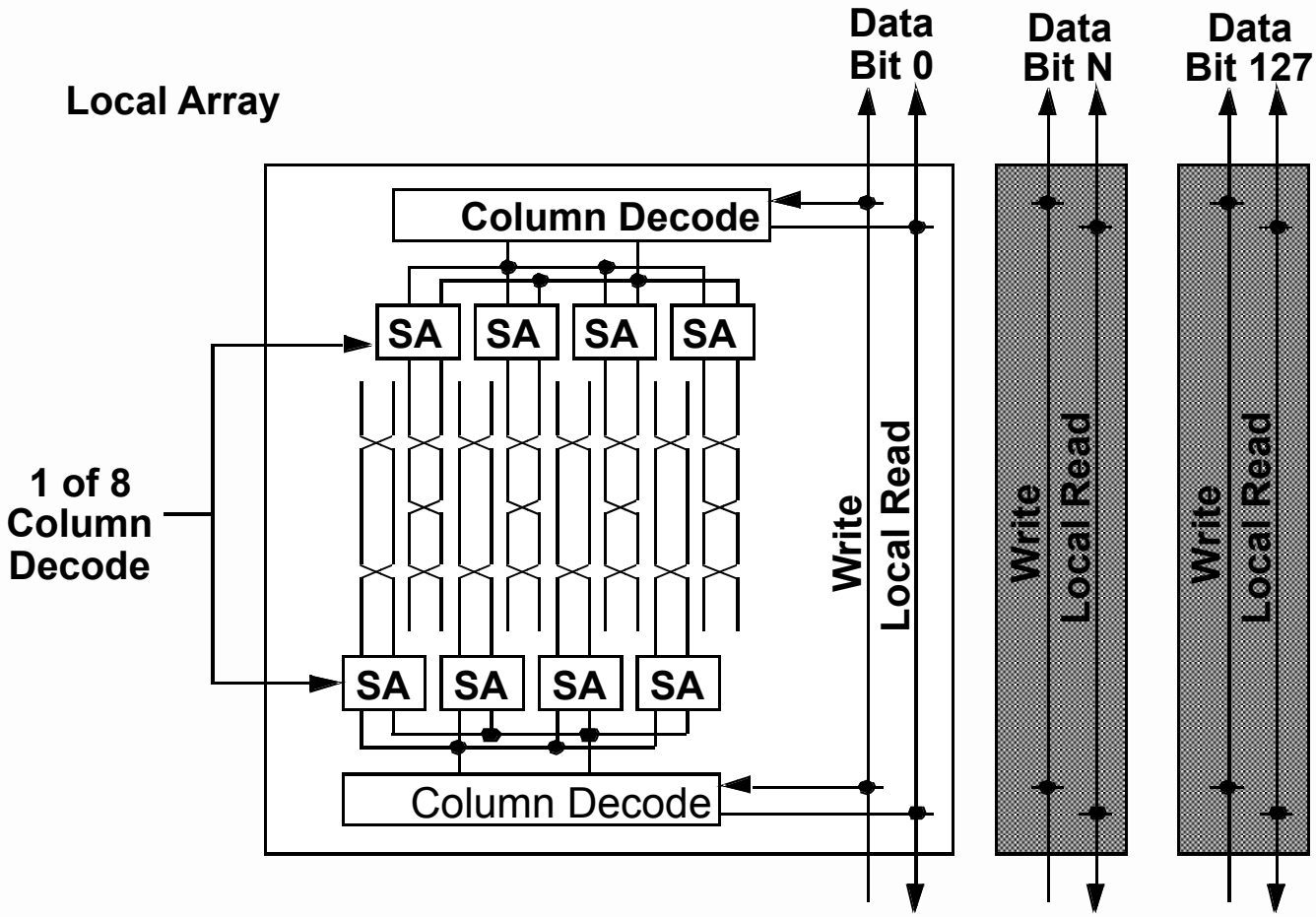


V_t and DL Mis-Match Limited by Longer Channel Length

Overlay Mis-Alignment Limited by Identical Orientation

Capacitive Mis-Match Limited by careful Physical Design (Symmetry)

Interleaved Sense Amp w/ Bit-Line Twist



Open and Folded Bitline Schematic

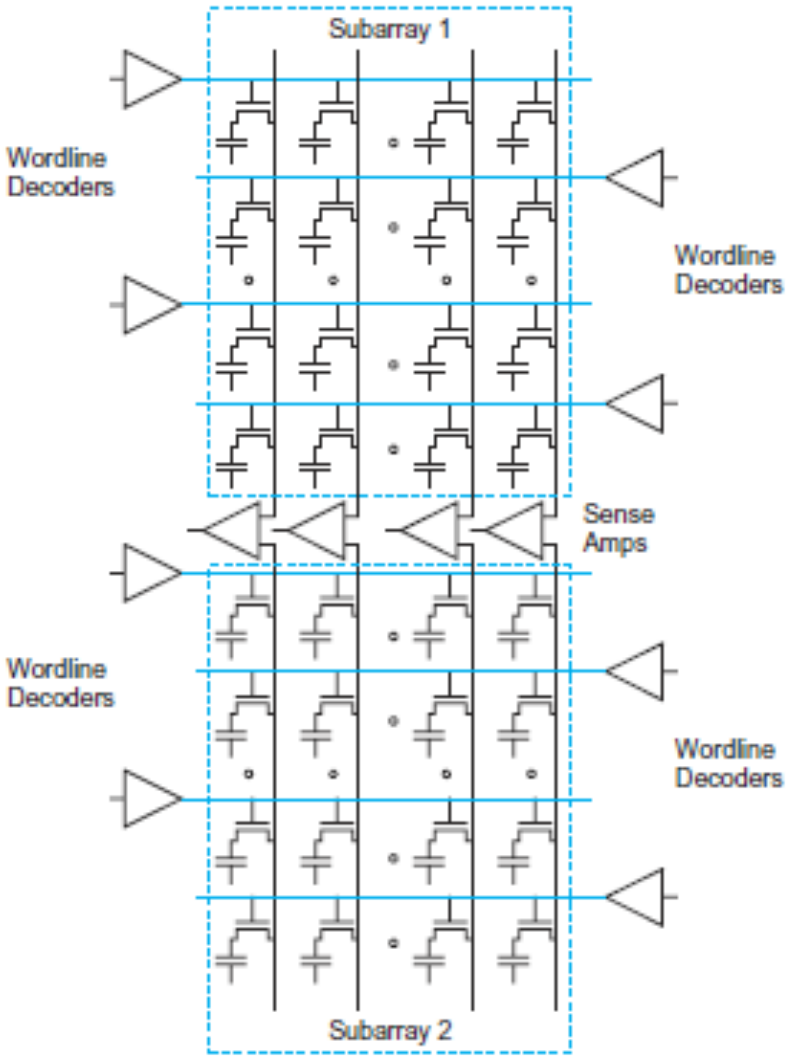


FIGURE 12.44 Open bitlines

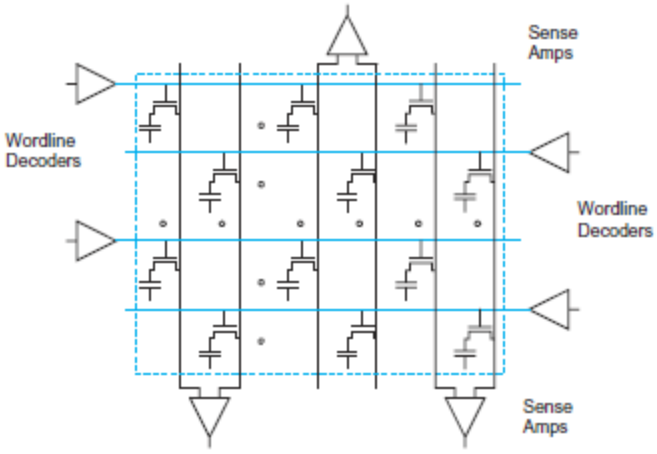


FIGURE 12.45 Folded bitlines

CMOS VLSI design - PEARSON

Folded Bitline Layout

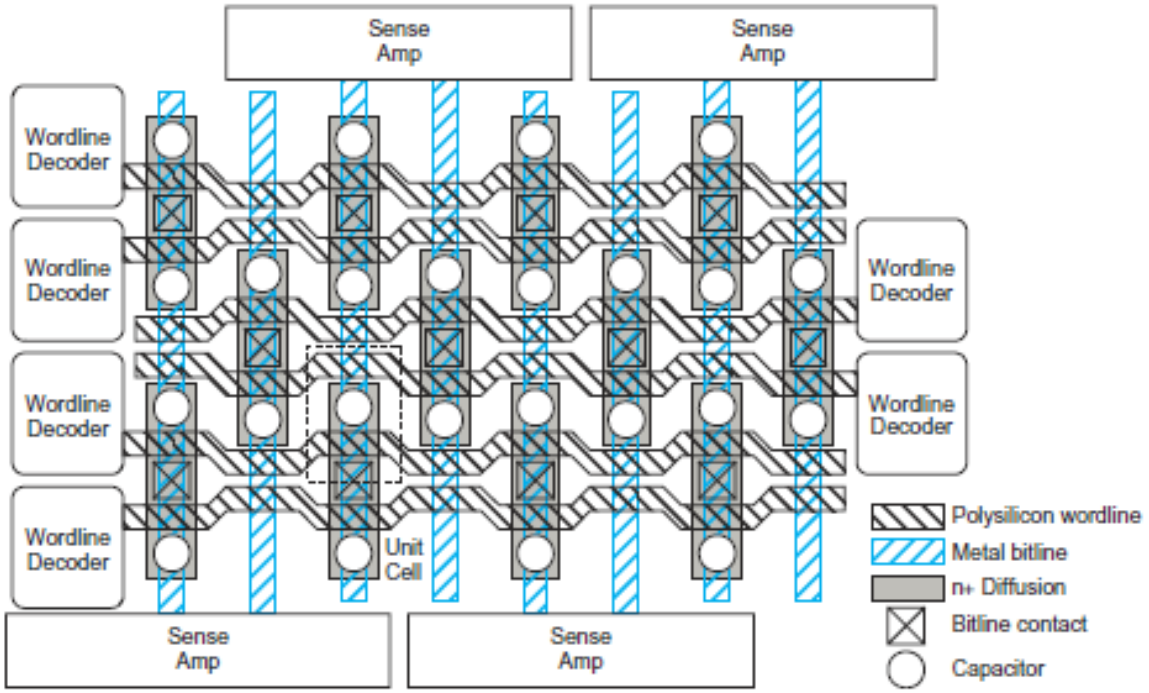


FIGURE 12.46 Layout of folded bitline subarray

Topics

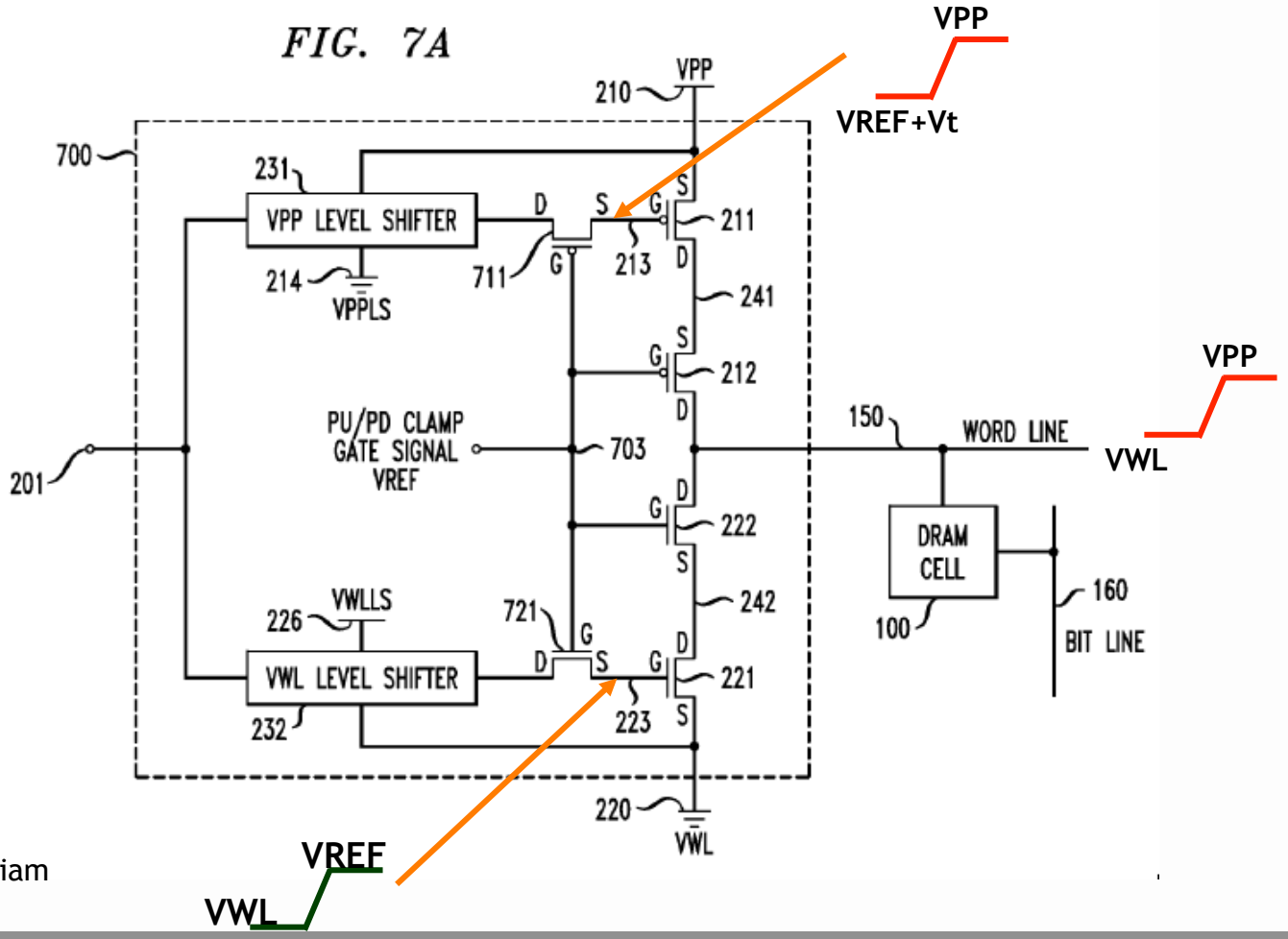
- ❑ Introduction to memory
- ❑ DRAM basics and bitcell array
- ❑ eDRAM operational details (case study)
- ❑ Noise concerns
- ❑ Wordline driver (WLDRV) and level translators (LT)
- ❑ Challenges in eDRAM
- ❑ Understanding Timing diagram – An example

WLDRV

Driver with Low voltage transistors → Logic transistors

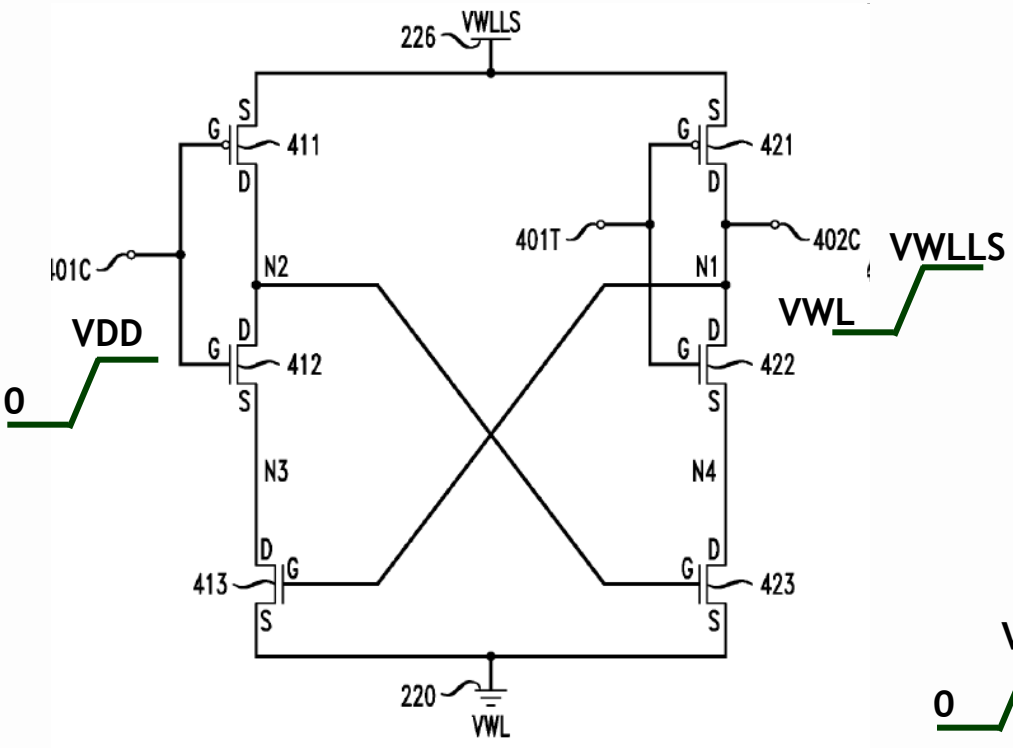
No thick gate oxide transistors required!!

Voltage across any two terminals should not exceed reliability limits

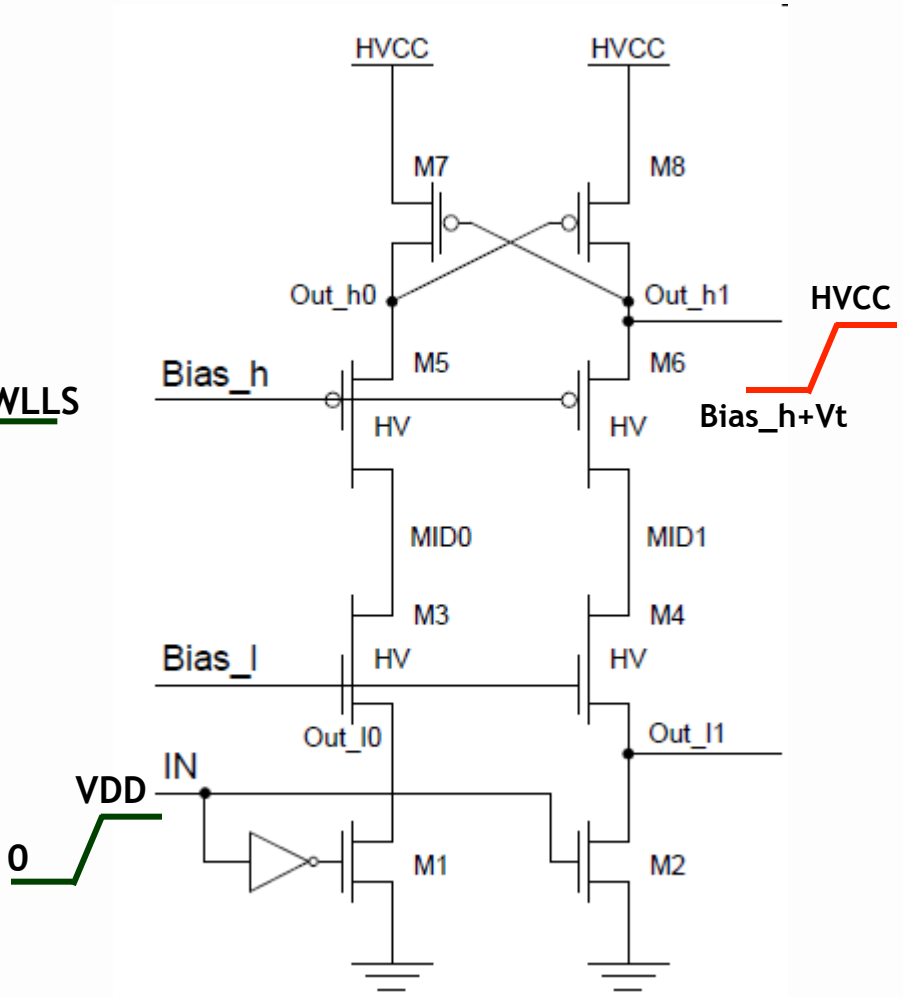


1. US patent No: 8,120,968 → William Robert Reohr, John E Barth

LEVEL Shifter



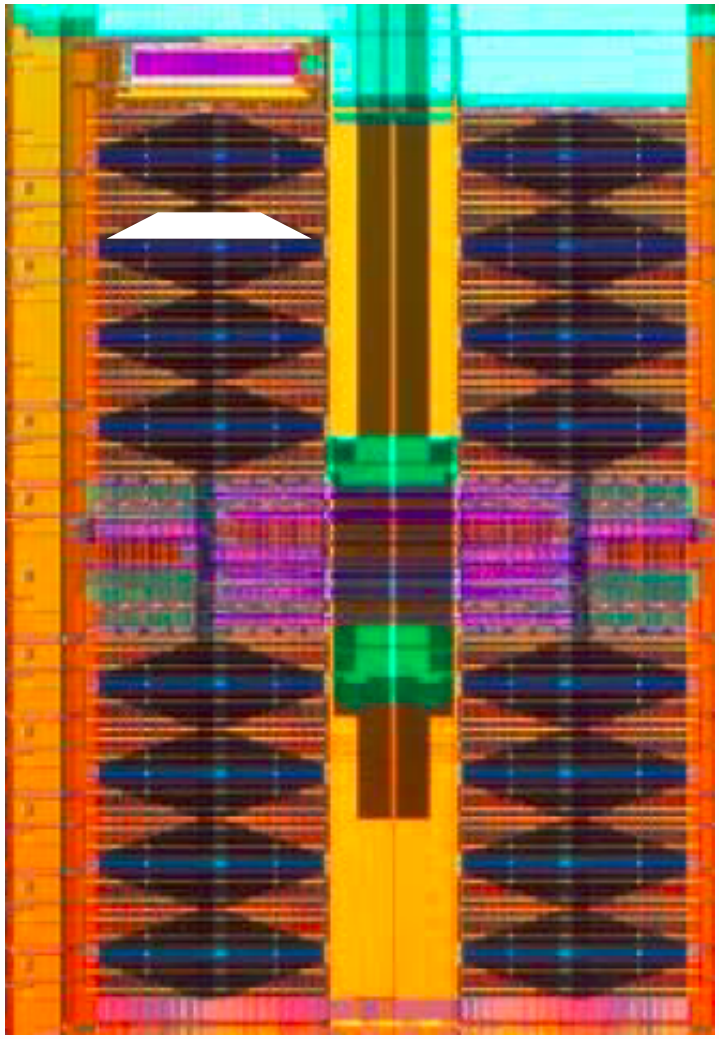
VWL Level shifter

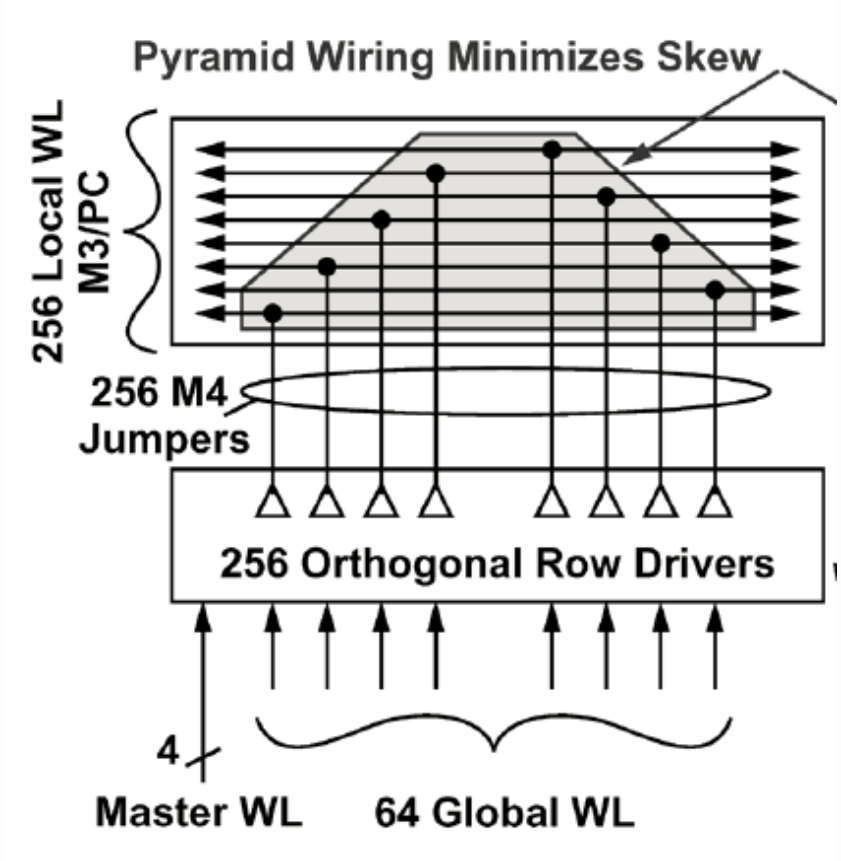


VPP Level shifter

1. US patent No: 8,120,968 → William Robert Reohr, John E Barth
2. A Low Voltage to High Voltage Level Shifter Circuit for MEMS Application → Dong Pan

Orthogonal WLD and pyramid wiring (M3/M4)





JSSC08

Topics

- ❑ Introduction to memory
- ❑ DRAM basics and bitcell array
- ❑ eDRAM operational details (case study)
- ❑ Noise concerns
- ❑ Wordline driver (WLDRV) and level translators (LT)
- ❑ Challenges in eDRAM
- ❑ Understanding Timing diagram – An example

Retention

Transfer Device and Storage Cap are NOT ideal devices: they LEAK
Leakage Mechanisms include: Ioff, Junction Leakage, GIDL,...
Junction Leakage Temperature Dependence = 2x/10C

Cell Charge needs to be replenished (Refreshed), Median Retention Time:

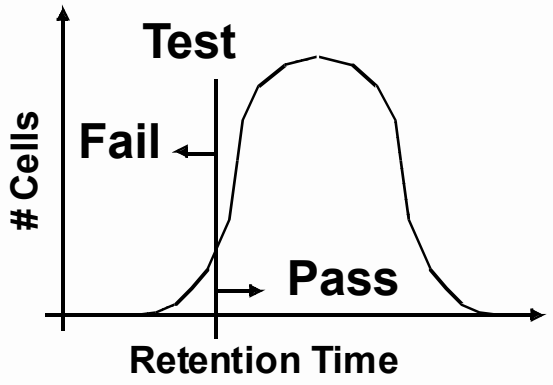
$$T = \frac{CDV}{I_{leak}} = \frac{35fF \times 400mV}{2fA} = 7 \text{ seconds}$$

Where DV is acceptable loss
C is Cell Capacitance
I_{leak} is Total Leakage

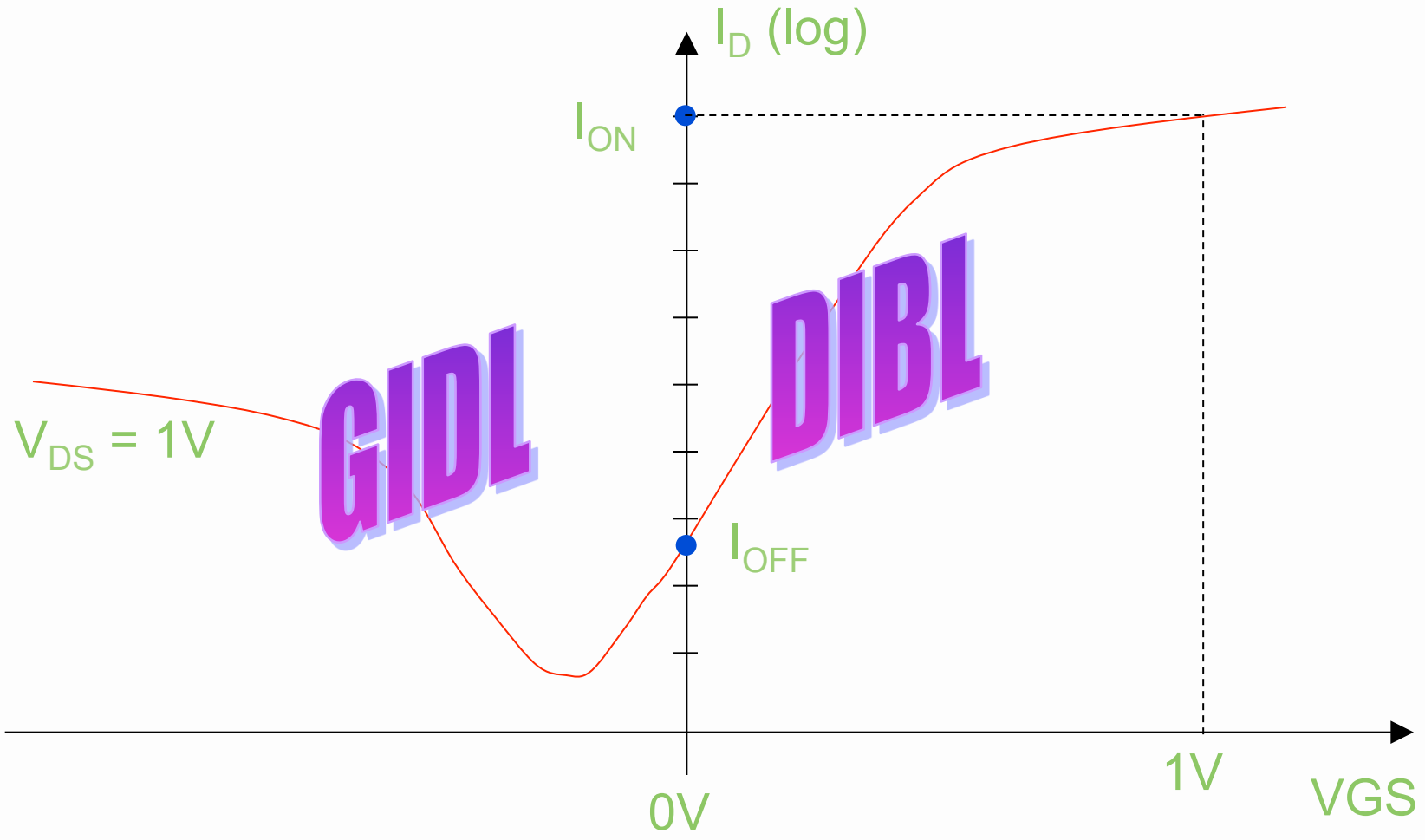
Retention Distribution has Tails
created by Defects and Leaky Cells

Weak Cells Tested out (5x Guardband)
and replaced with Redundancy

Customer issues periodic Refresh Cycle



Pass transistor leakage



Floating Body Effects

Body potential modulated by coupling and leakage

Better source follower vs. bulk during write back (body coupling)

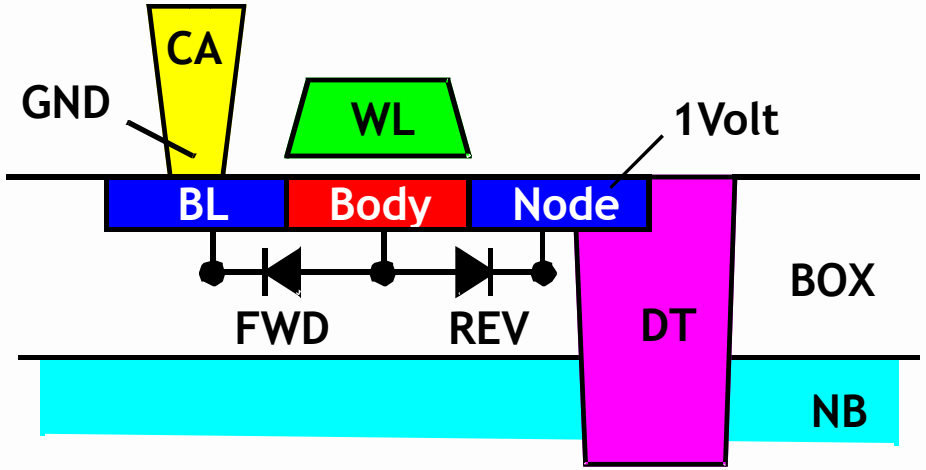
- Þ Improved write ‘1’ cell voltage

Degraded I_{off}/ Retention if body floats high (body leakage)

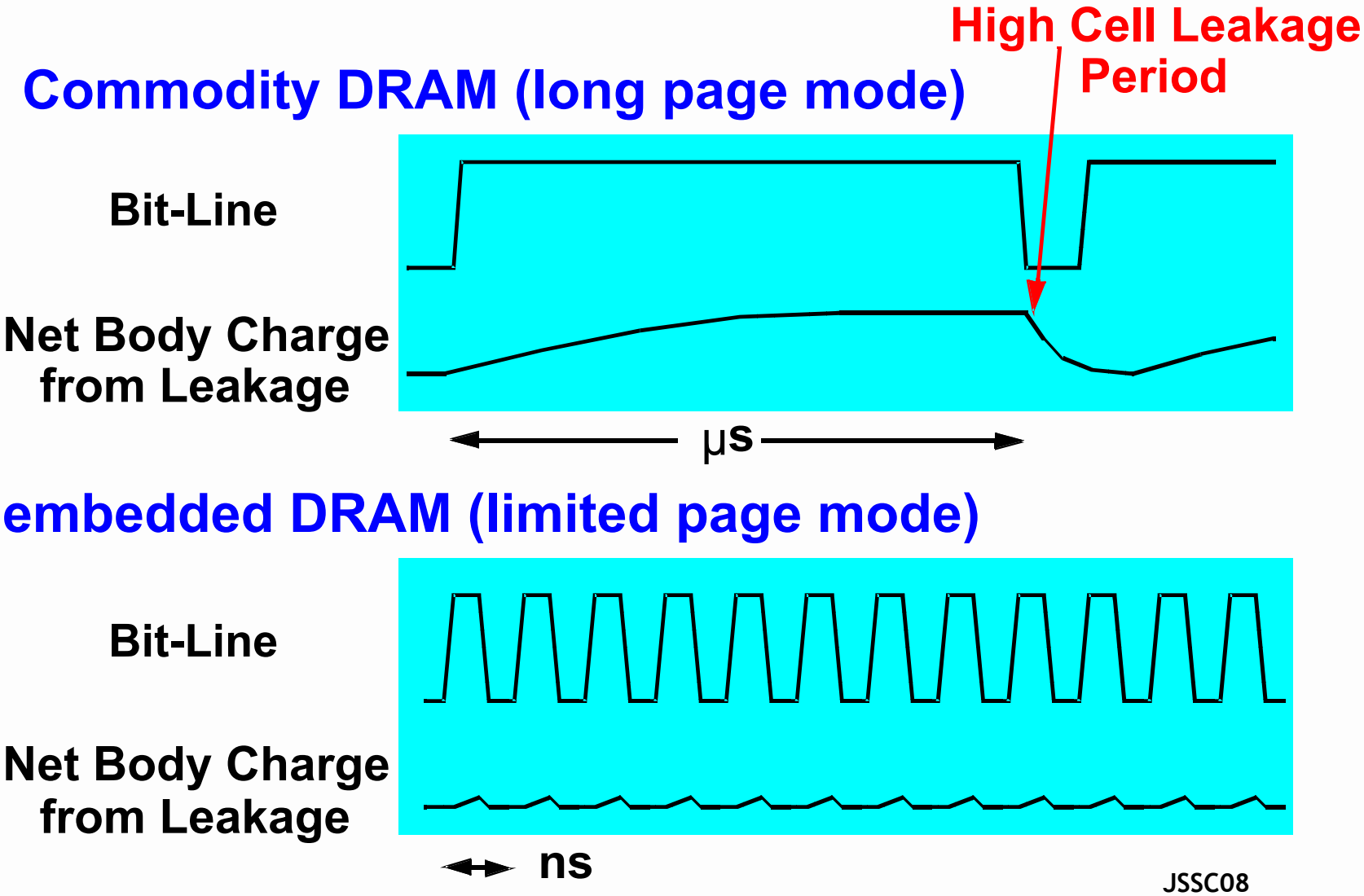
- Þ GND pre-charge keeps body low
- Þ Eliminate long periods with BL high (limit page mode)

$ILeak_{FWD} > ILeak_{REV}$

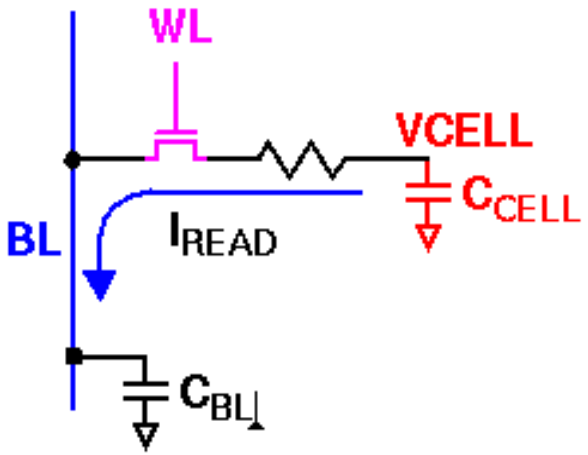
When BL = GND
Body Þ GND



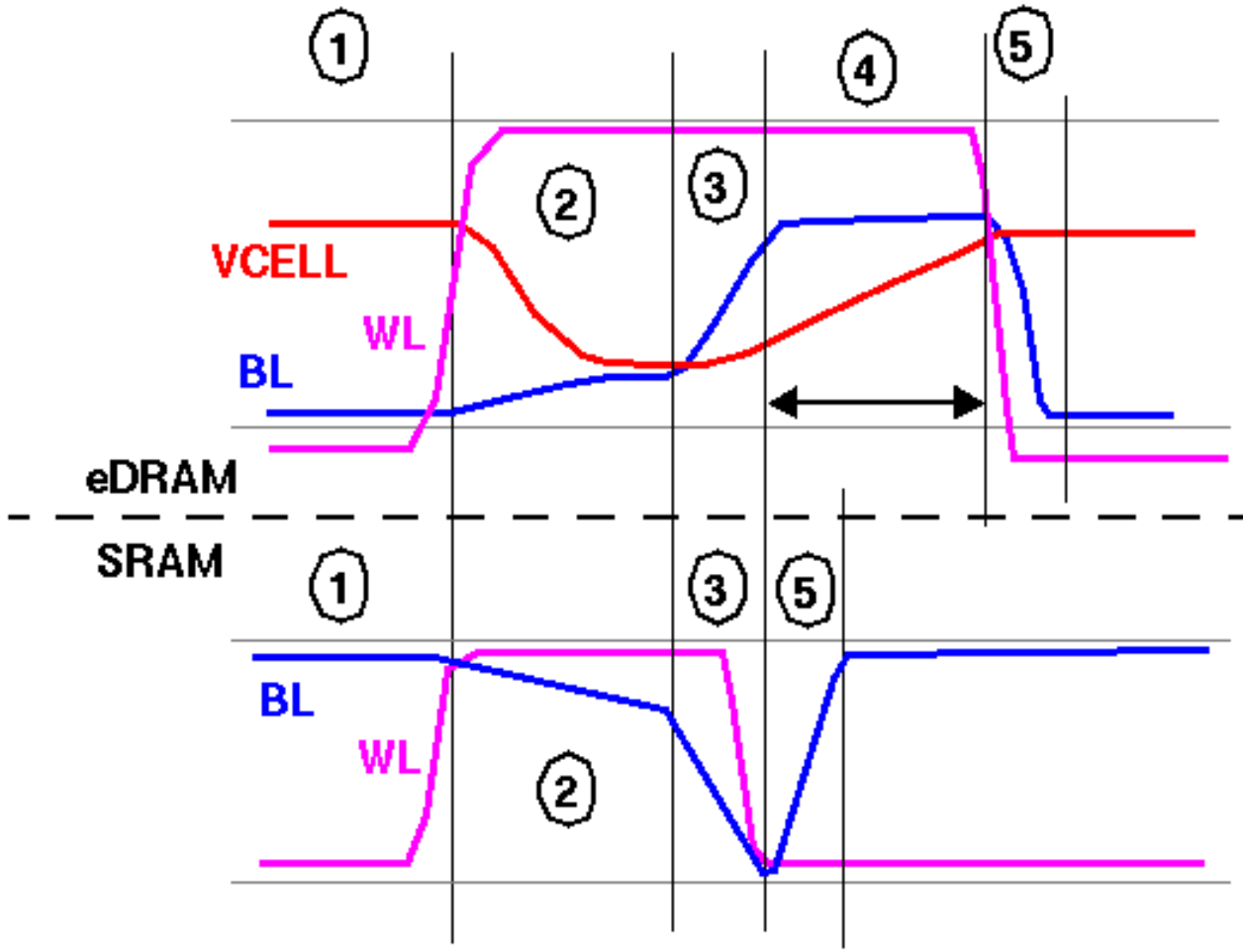
Array Body Charging



eDRAM vs. SRAM Cycle-Time Comparison



1. WL Activation
2. Charge Transfer to Bit-Line
   (I_{READ} Similar to SRAM)
3. Amplification
4. Write-Back
5. Precharge

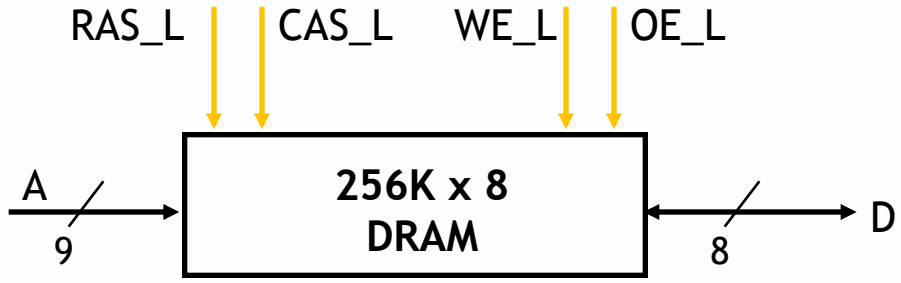


NET: SRAM Random Cycle will continue to lead!

Topics

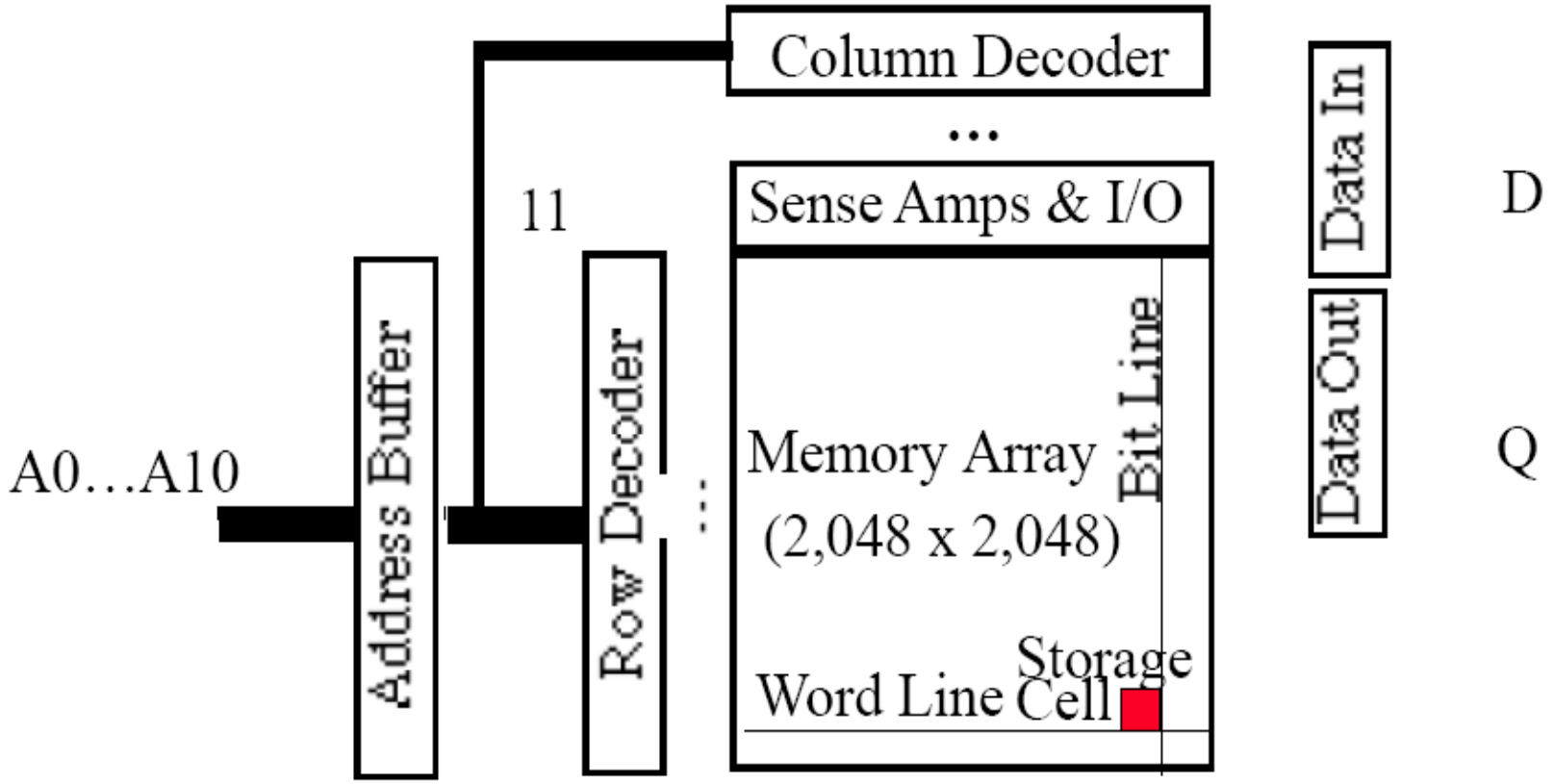
- ❑ Introduction to memory
- ❑ DRAM basics and bitcell array
- ❑ eDRAM operational details (case study)
- ❑ Noise concerns
- ❑ Wordline driver (WLDRV) and level translators (LT)
- ❑ Challenges in eDRAM
- ❑ Understanding Timing diagram – An example

Logic Diagram of a Typical DRAM



- Control Signals (RAS_L, CAS_L, WE_L, OE_L) are all active low
- Din and Dout are combined (D):
  - WE_L is asserted (Low), OE_L is disasserted (High)
    - D serves as the data input pin
  - WE_L is disasserted (High), OE_L is asserted (Low)
    - D is the data output pin
- Row and column addresses share the same pins (A)
  - RAS_L goes low: Pins A are latched in as row address
  - CAS_L goes low: Pins A are latched in as column address
  - RAS/CAS edge-sensitive

DRAM logical organization (4 Mbit)

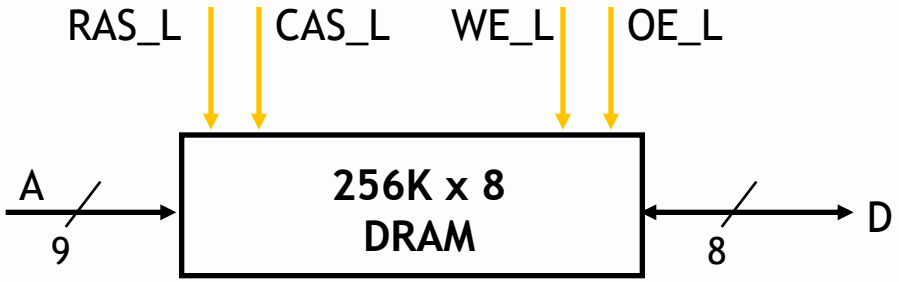


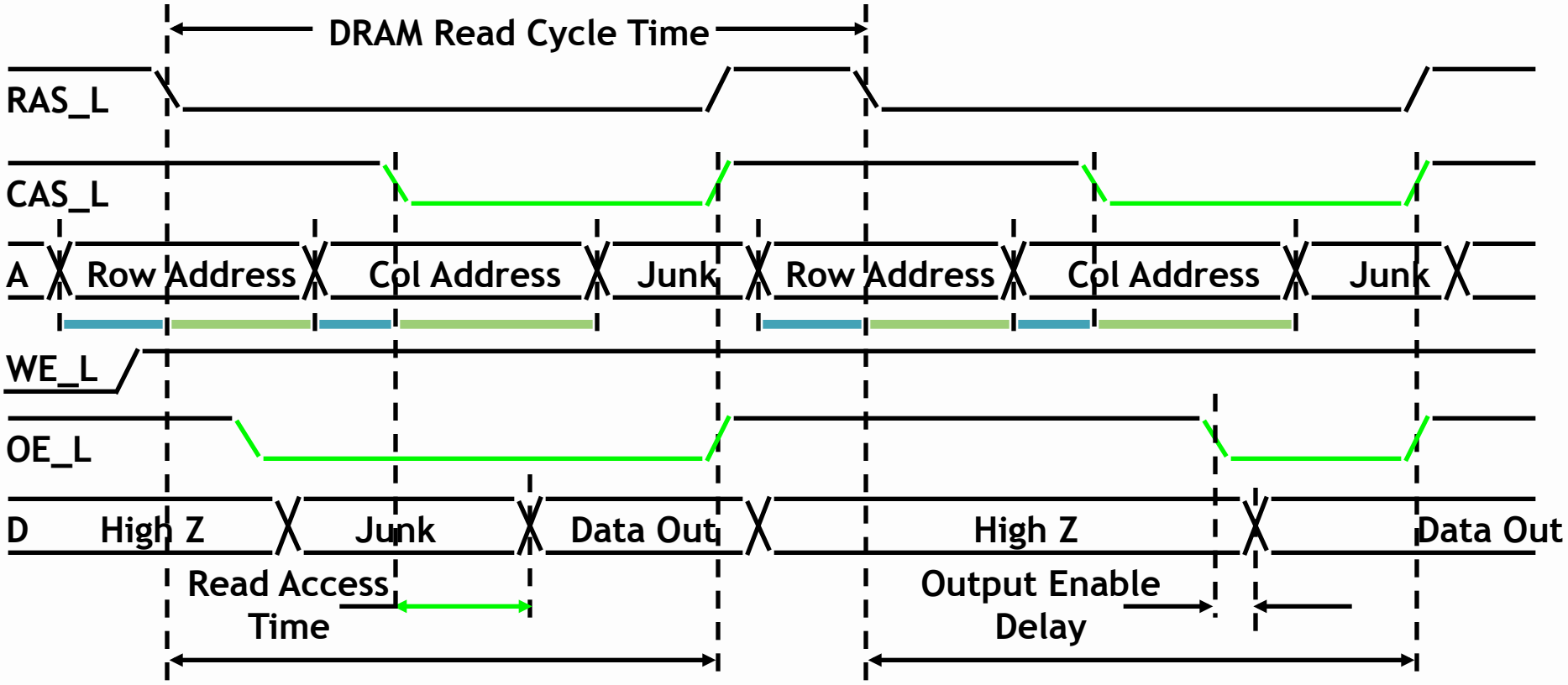
- **Square root of bits per RAS/CAS**

Din Dout can be clubbed together with a BiDi buffer

DRAM Read Timing

- Every DRAM access begins at:
  - The assertion of the RAS_L
  - 2 ways to read:
    early or late v. CAS



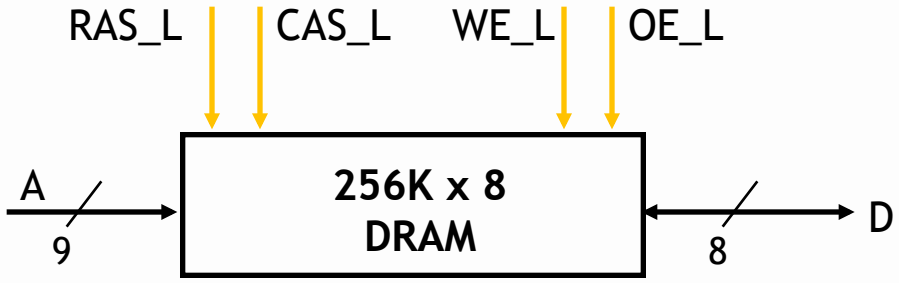


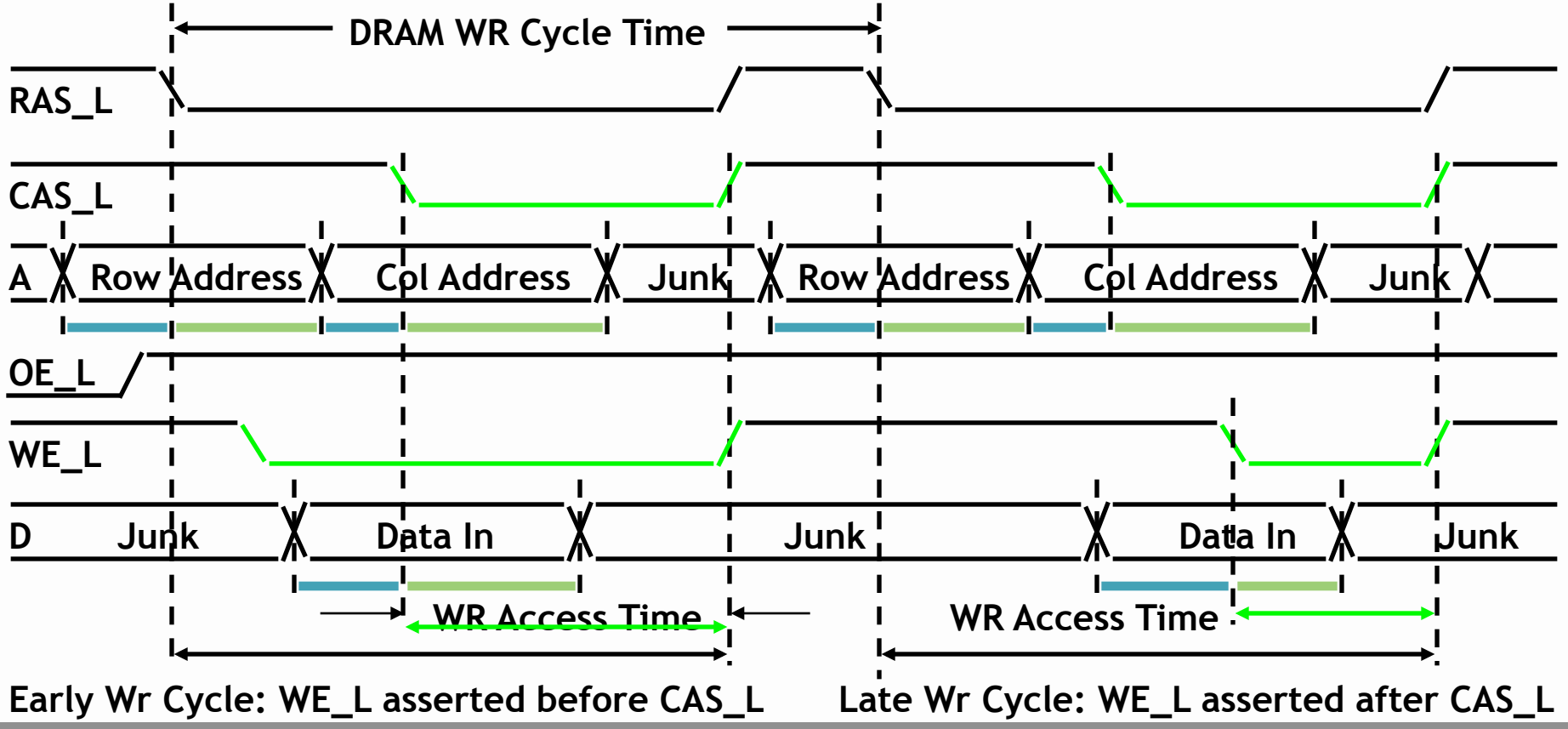
Early Read Cycle: OE_L asserted before CAS_L

Late Read Cycle: OE_L asserted after CAS_L

DRAM Write Timing

- Every DRAM access begins at:
  - The assertion of the RAS_L
  - 2 ways to write:
    early or late v. CAS





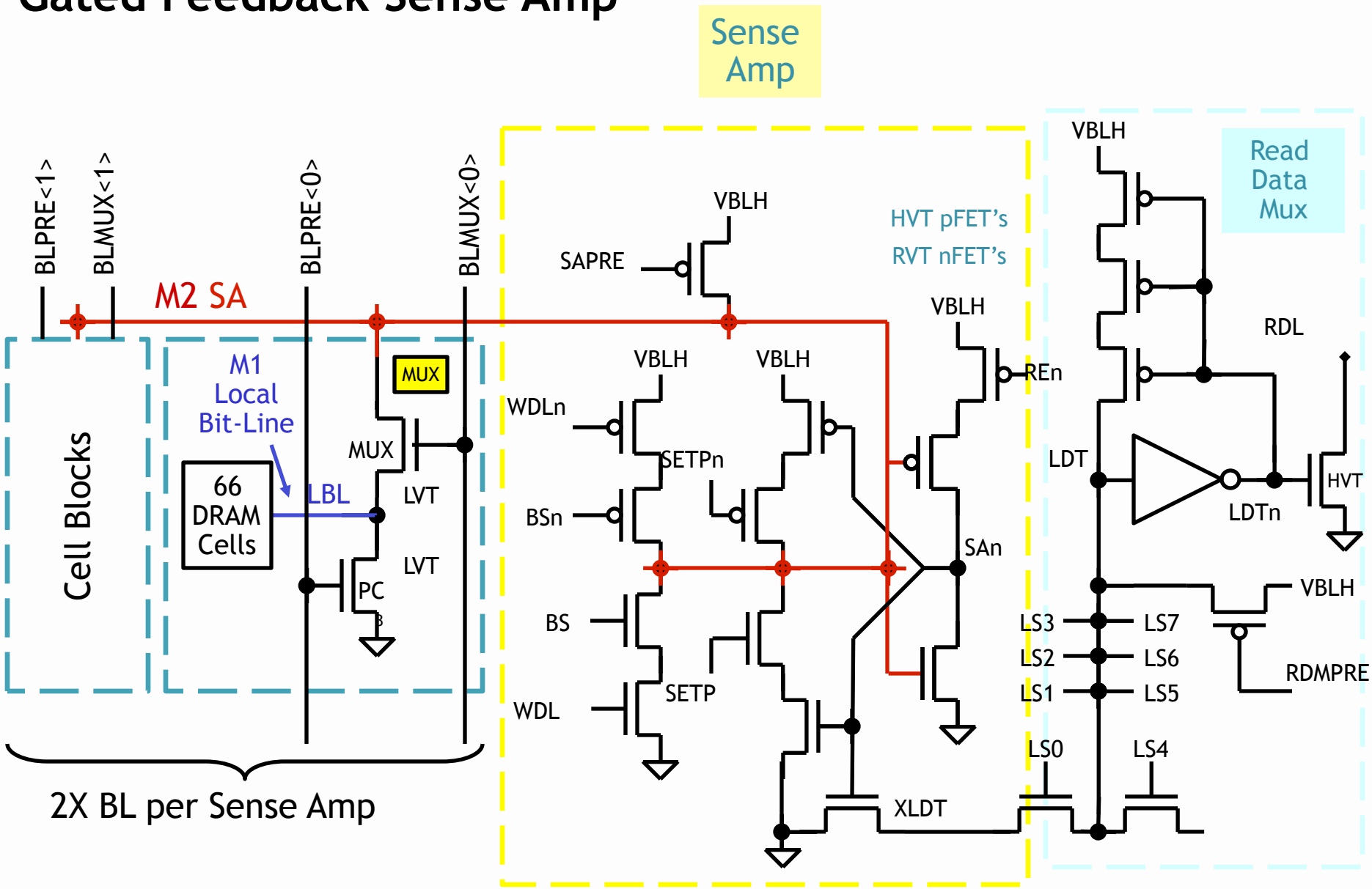
A Fast Sense-Amp (Case Study)

G. Fredeman et al., “A 14 nm 1.1 Mb Embedded DRAM Macro With 1 ns Access,” in IEEE Journal of Solid-State Circuits, vol. 51, no. 1, pp. 230-239, Jan. 2016. doi: 10.1109/JSSC.2015.2456873

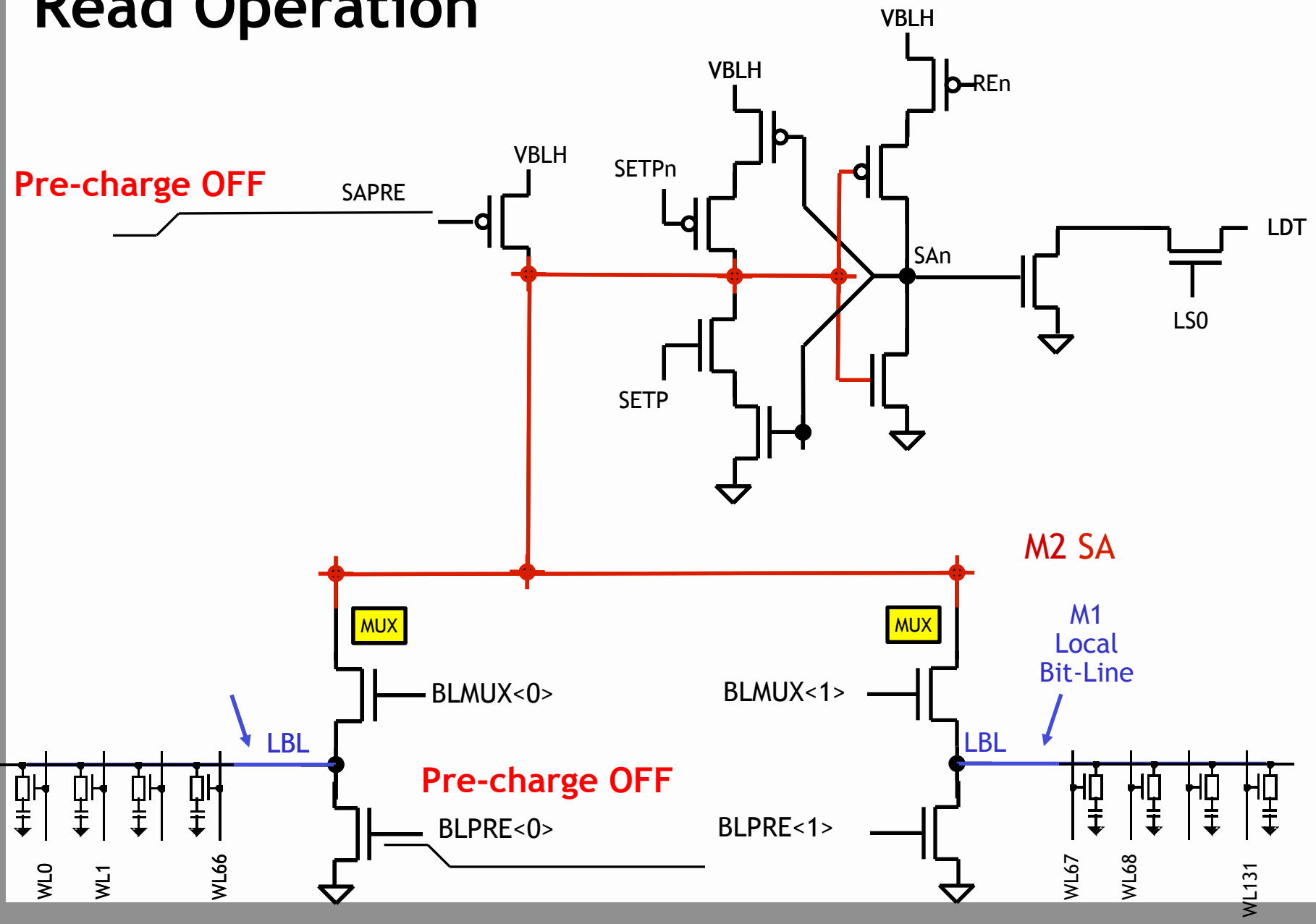
Problems with Micro Sense Amp

- By default the Sense Amp reads a 0
- Access transistor has to pull the LBL HIGH to read 1
  - Asymptotic charge up to High since Vgs keeps reducing
  - Very slow by nature
  - Need to minimize the WLs per BL(33) for performance reasons
- Cannot pre-charge LBL to High
  - Floating Body Effect affects retention
- NMOS (Access Device) is very fast when pulling down to zero
  - Can we make a Sense Amp that reads a one by default?
  - This will allow more WLs per BL

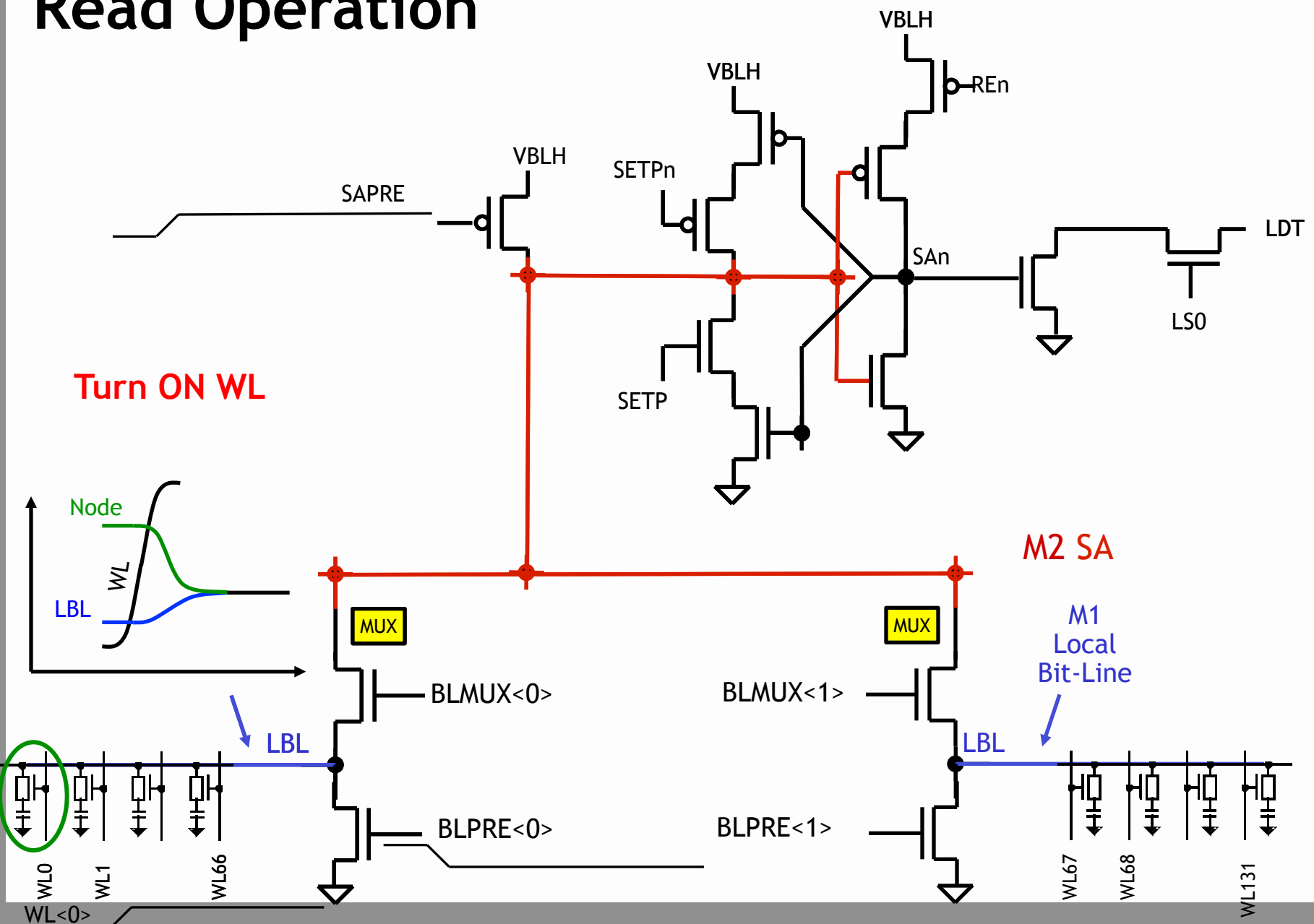
Gated Feedback Sense Amp



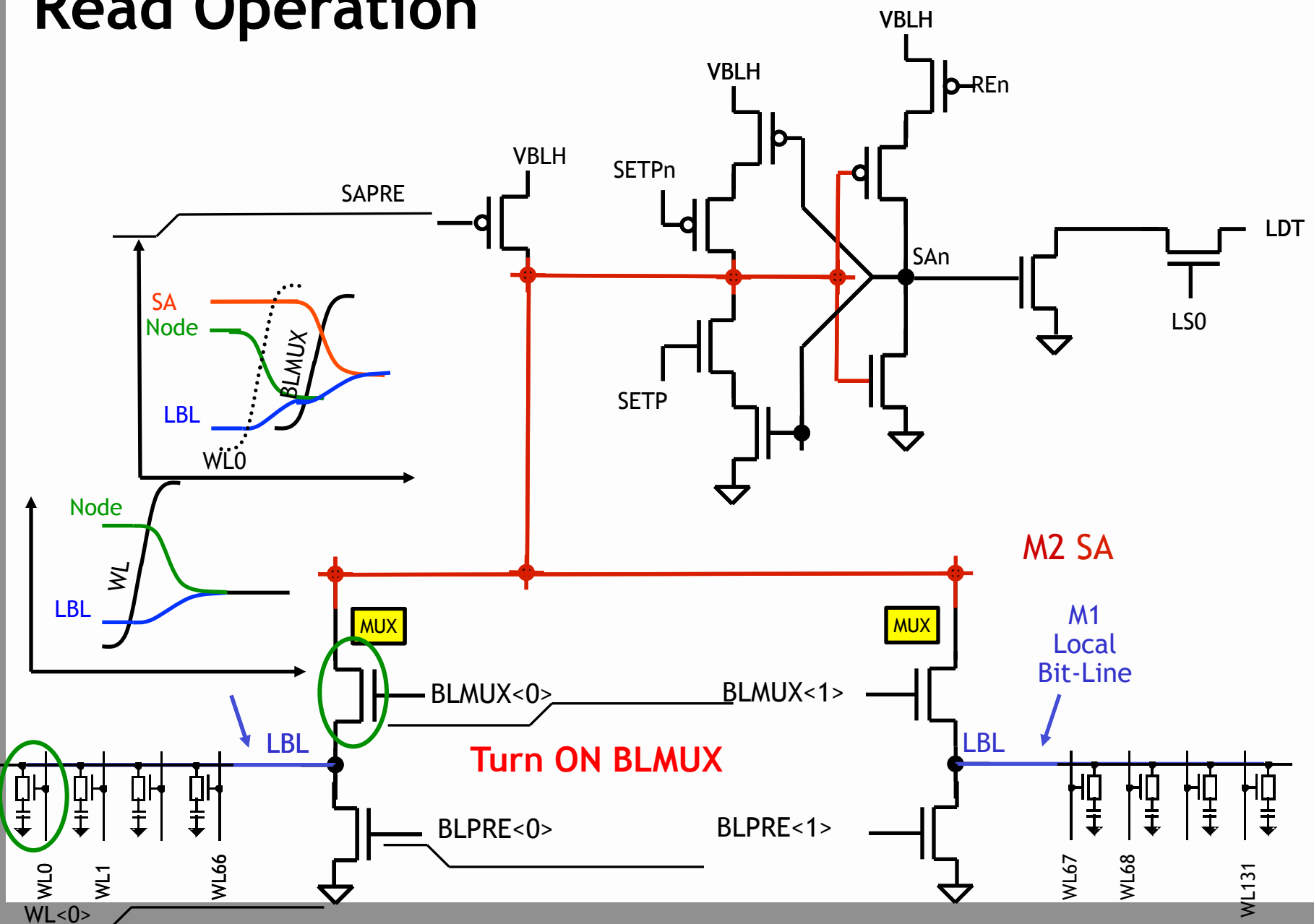
Read Operation



Read Operation

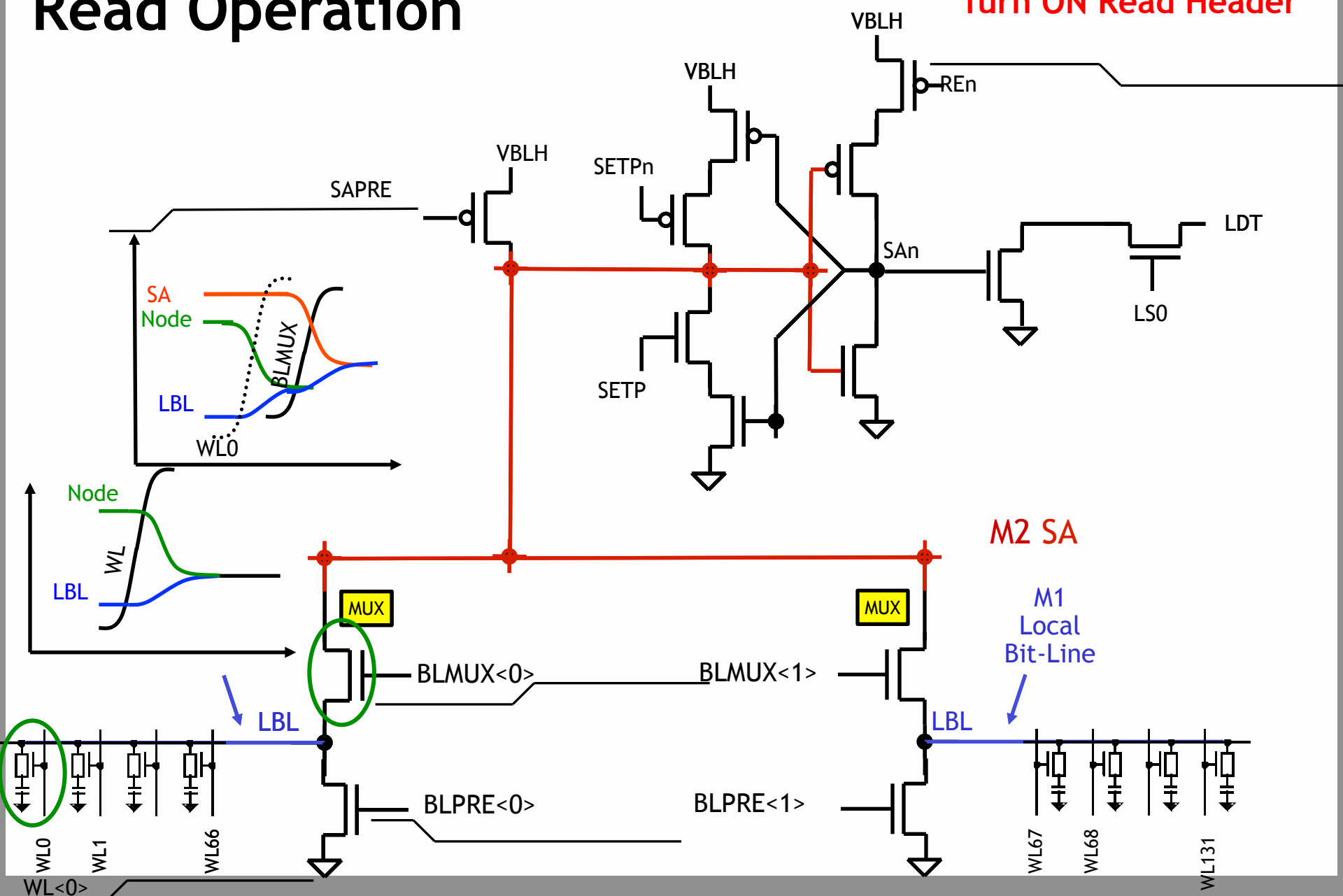


Read Operation

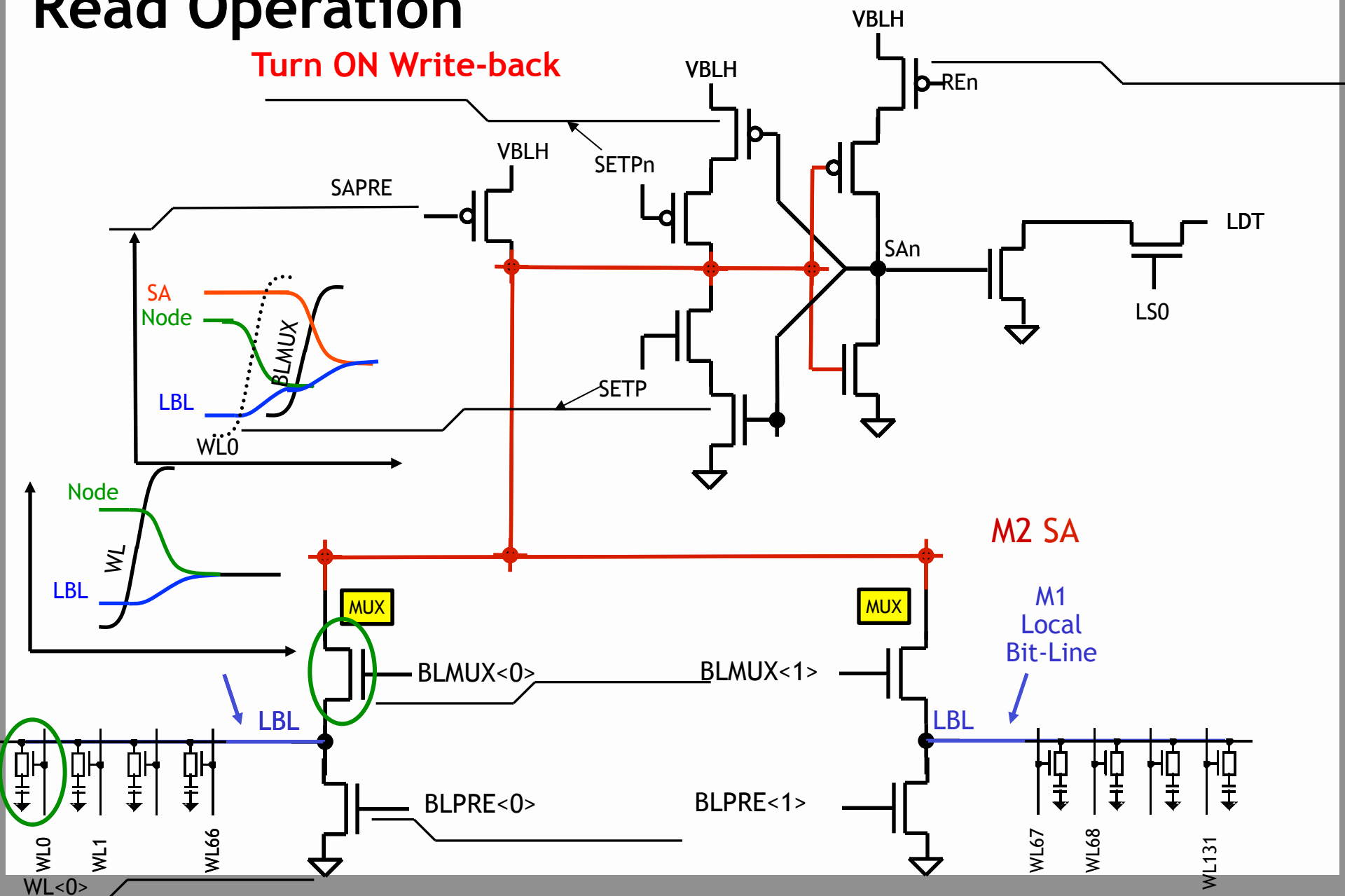


Read Operation

Turn ON Read Header

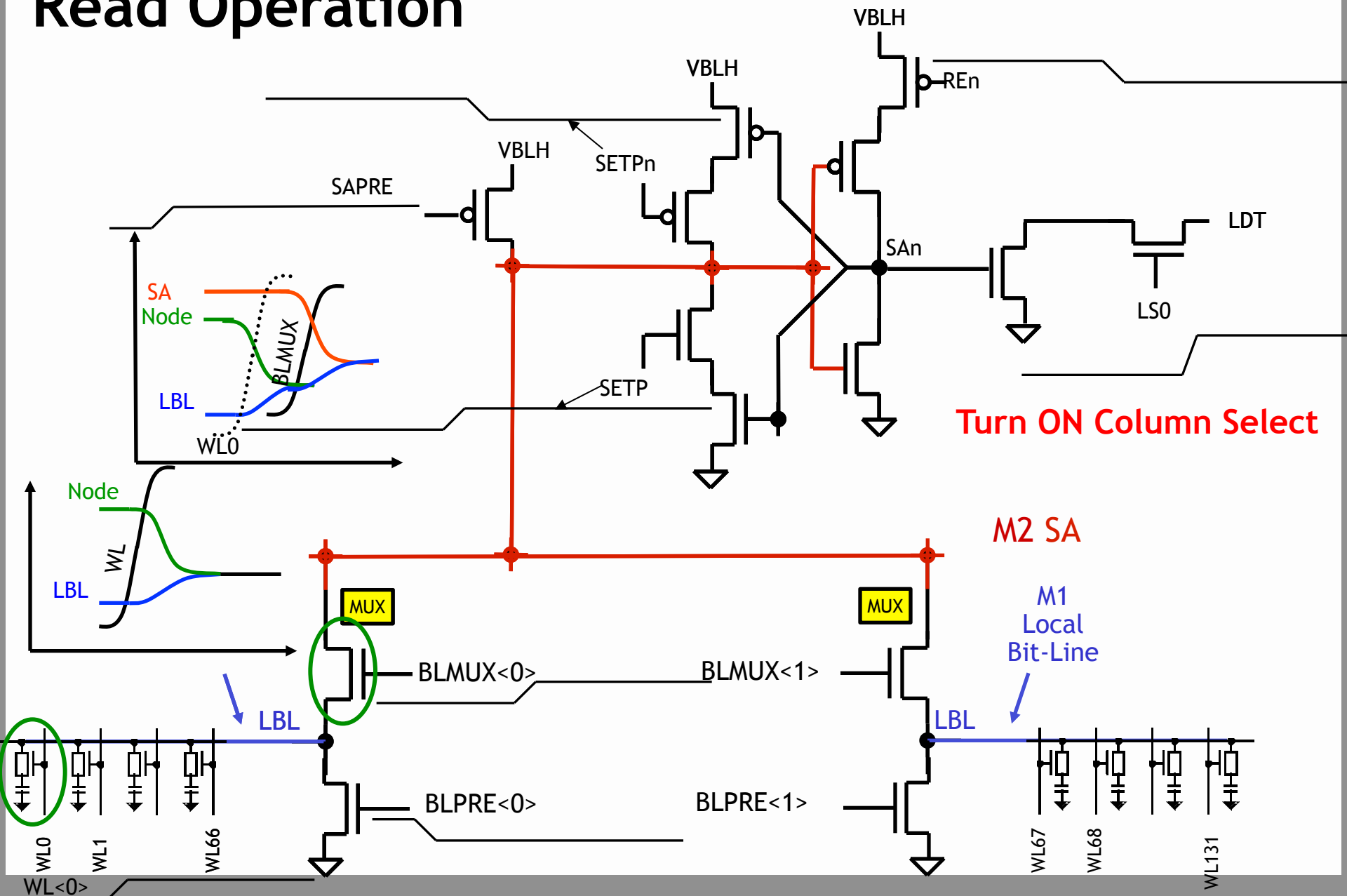


Read Operation

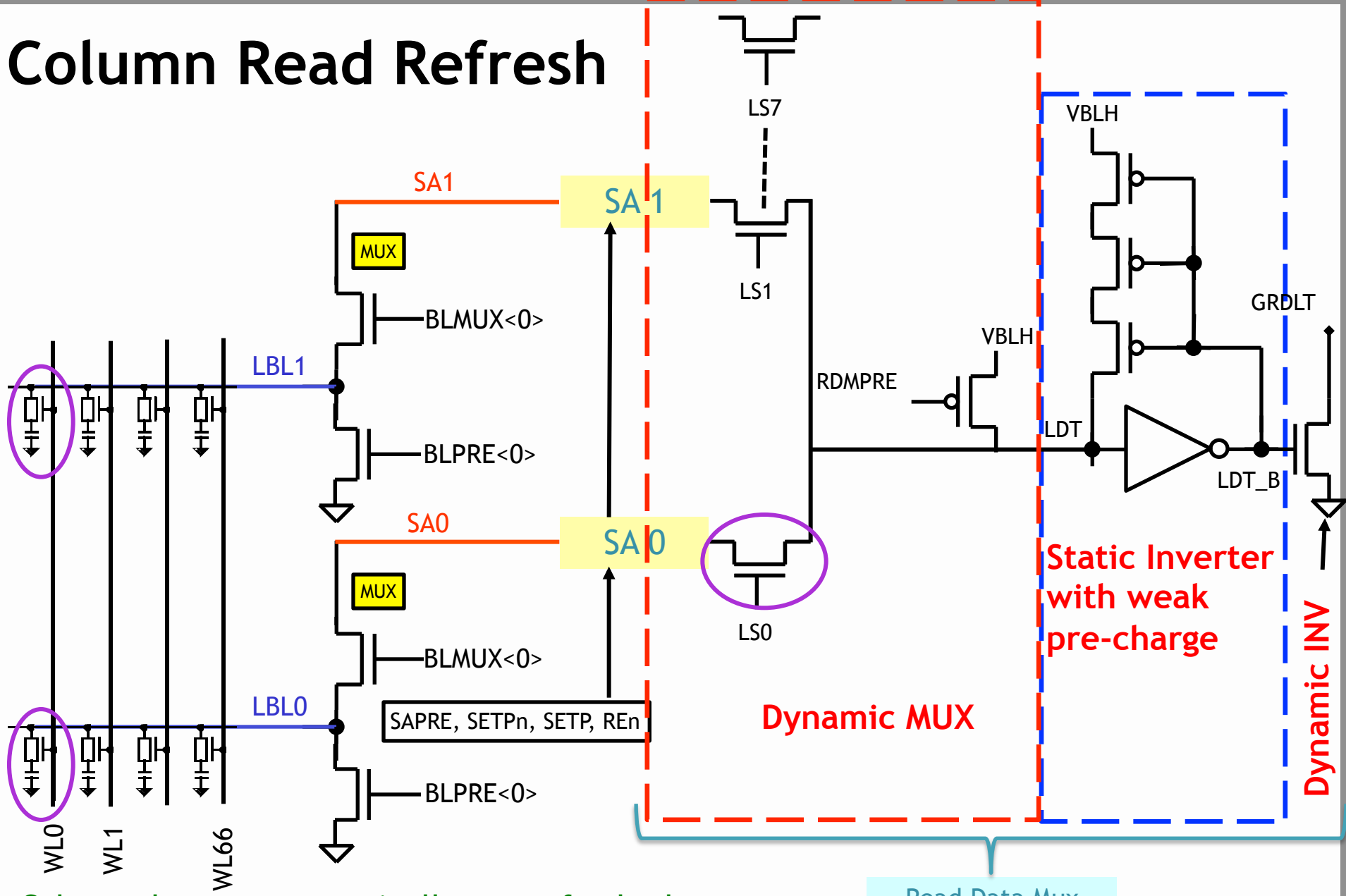


Animation

Read Operation

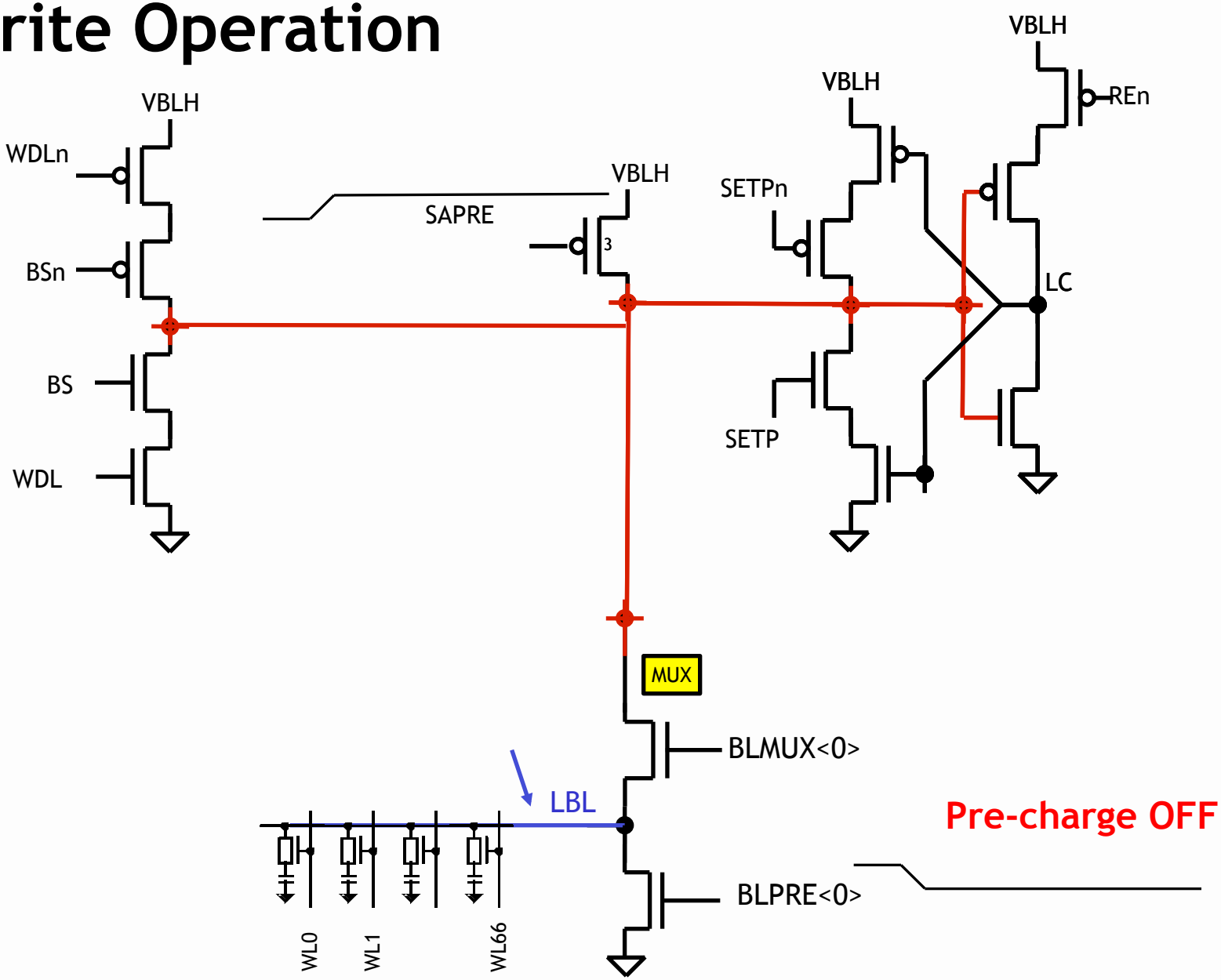


Column Read Refresh



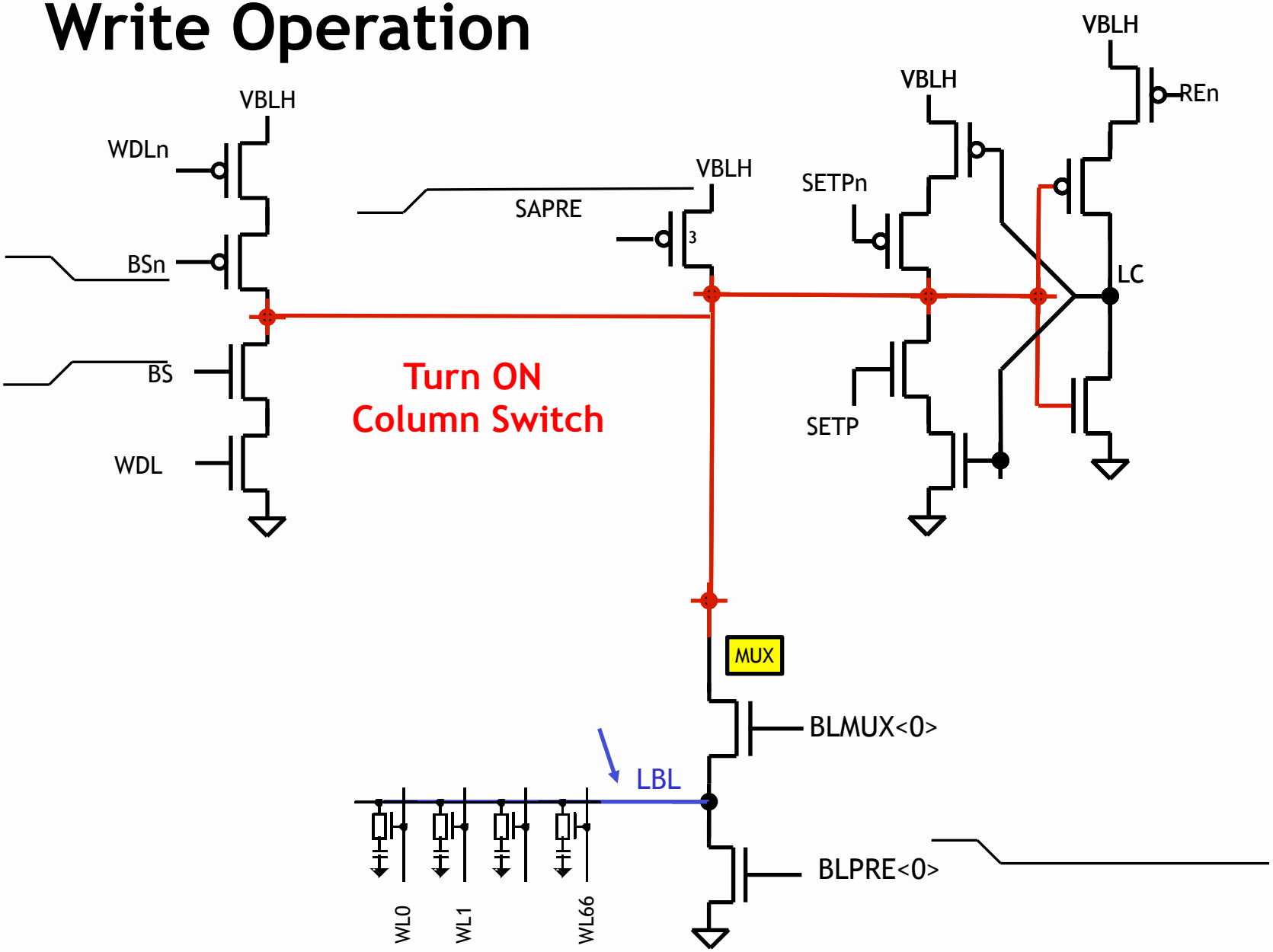
Other columns automatically get refreshed

Write Operation



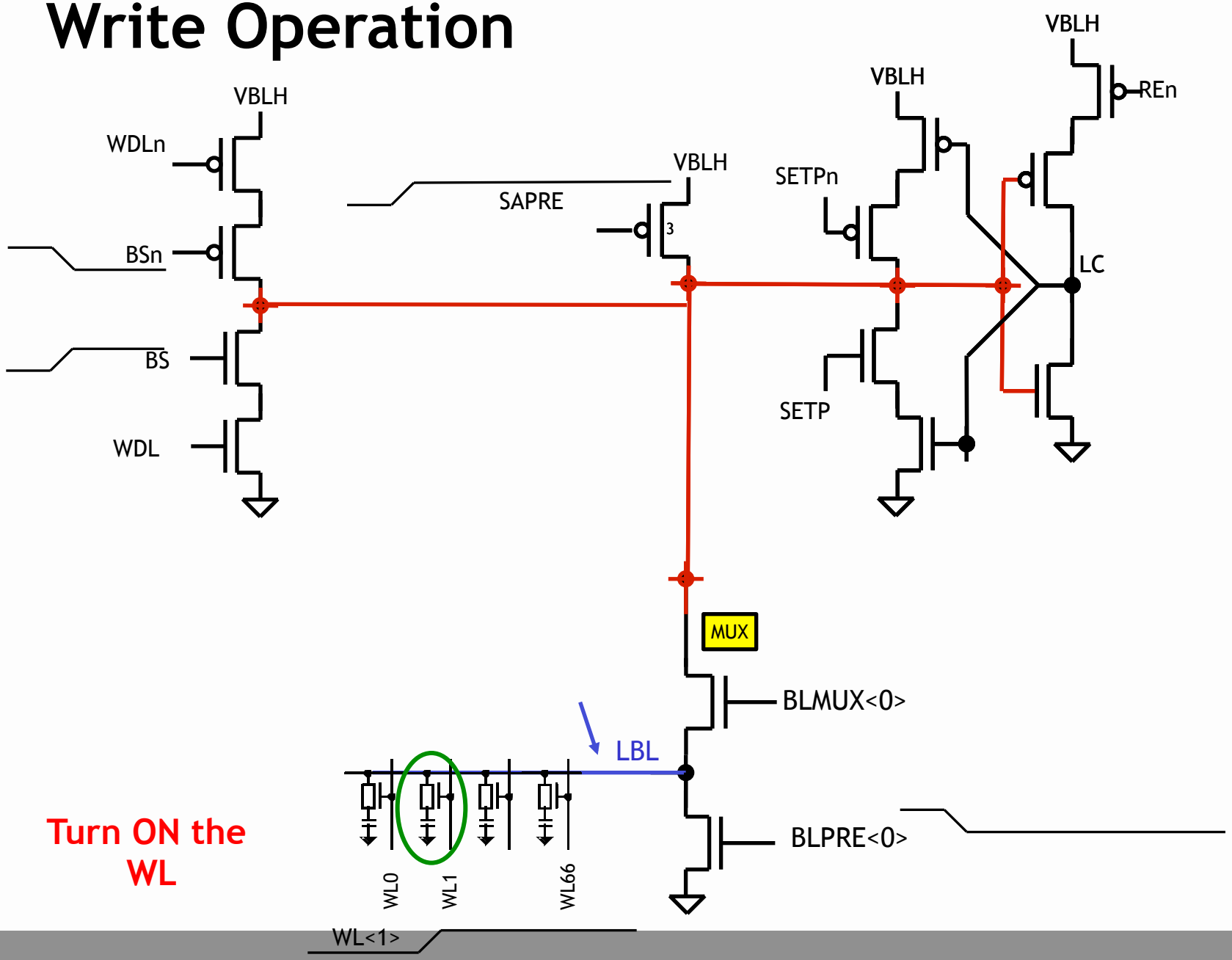
Animation

Write Operation

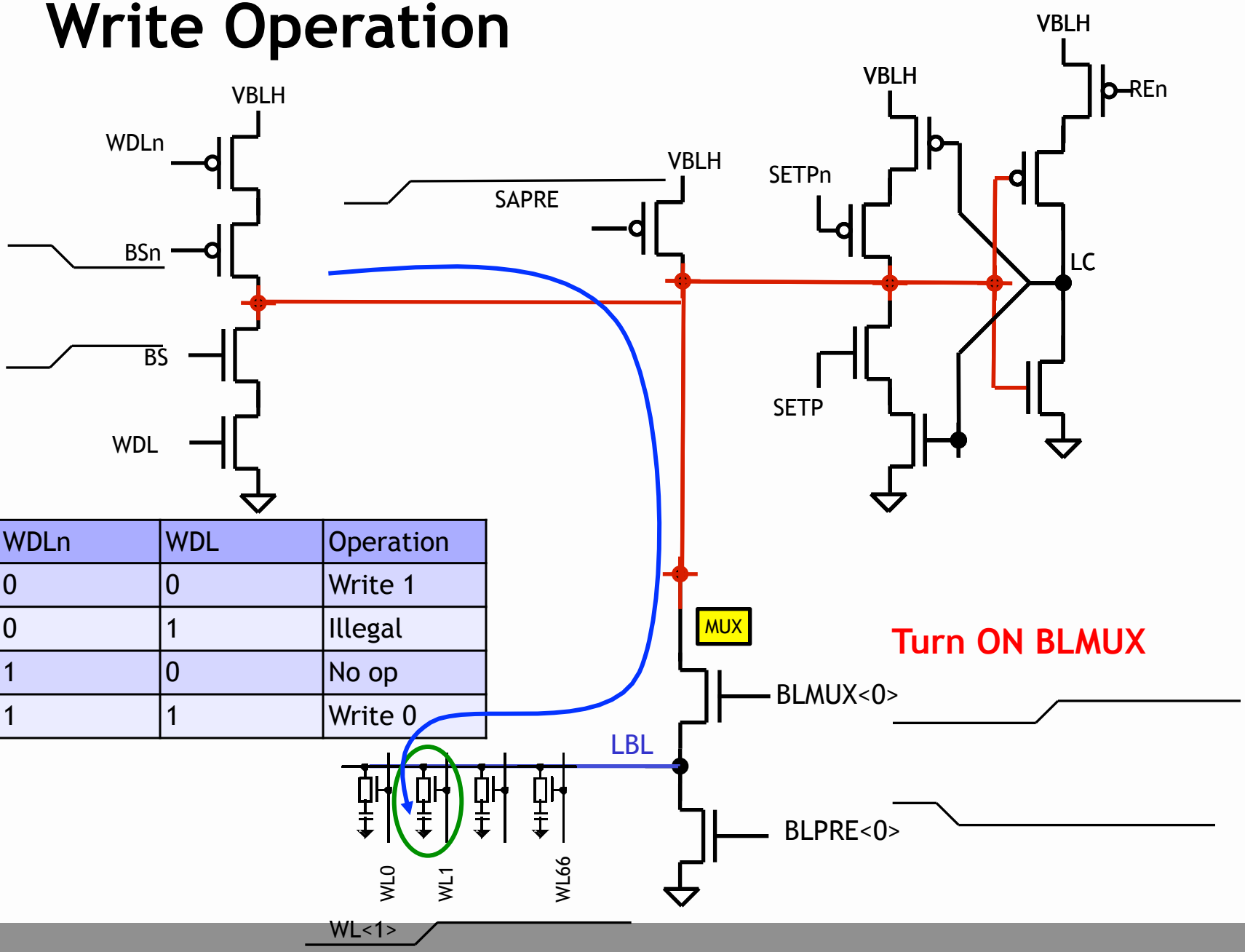


Animation

Write Operation



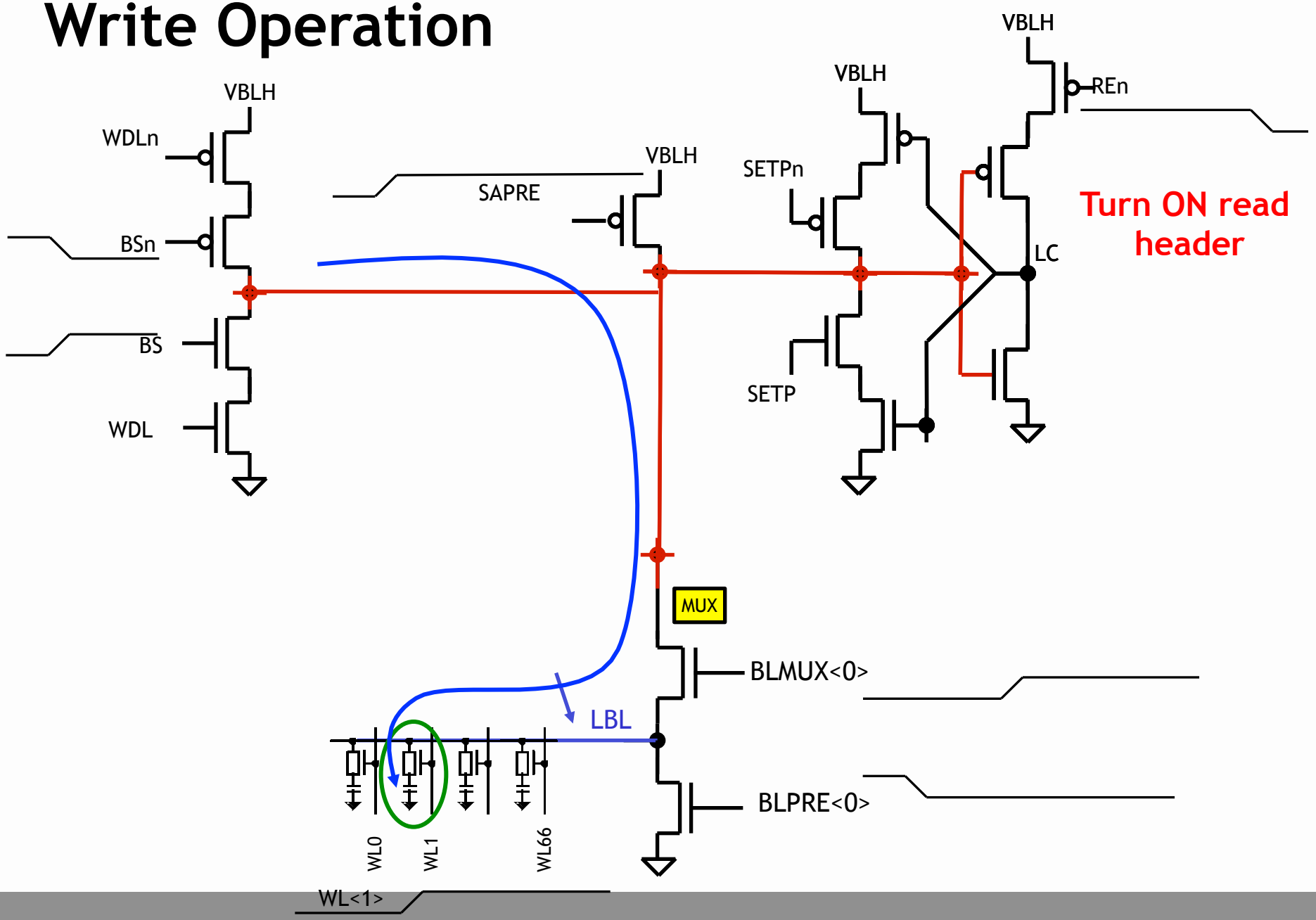
Write Operation



WDLn	WDL	Operation
0	0	Write 1
0	1	Illegal
1	0	No op
1	1	Write 0

Animation

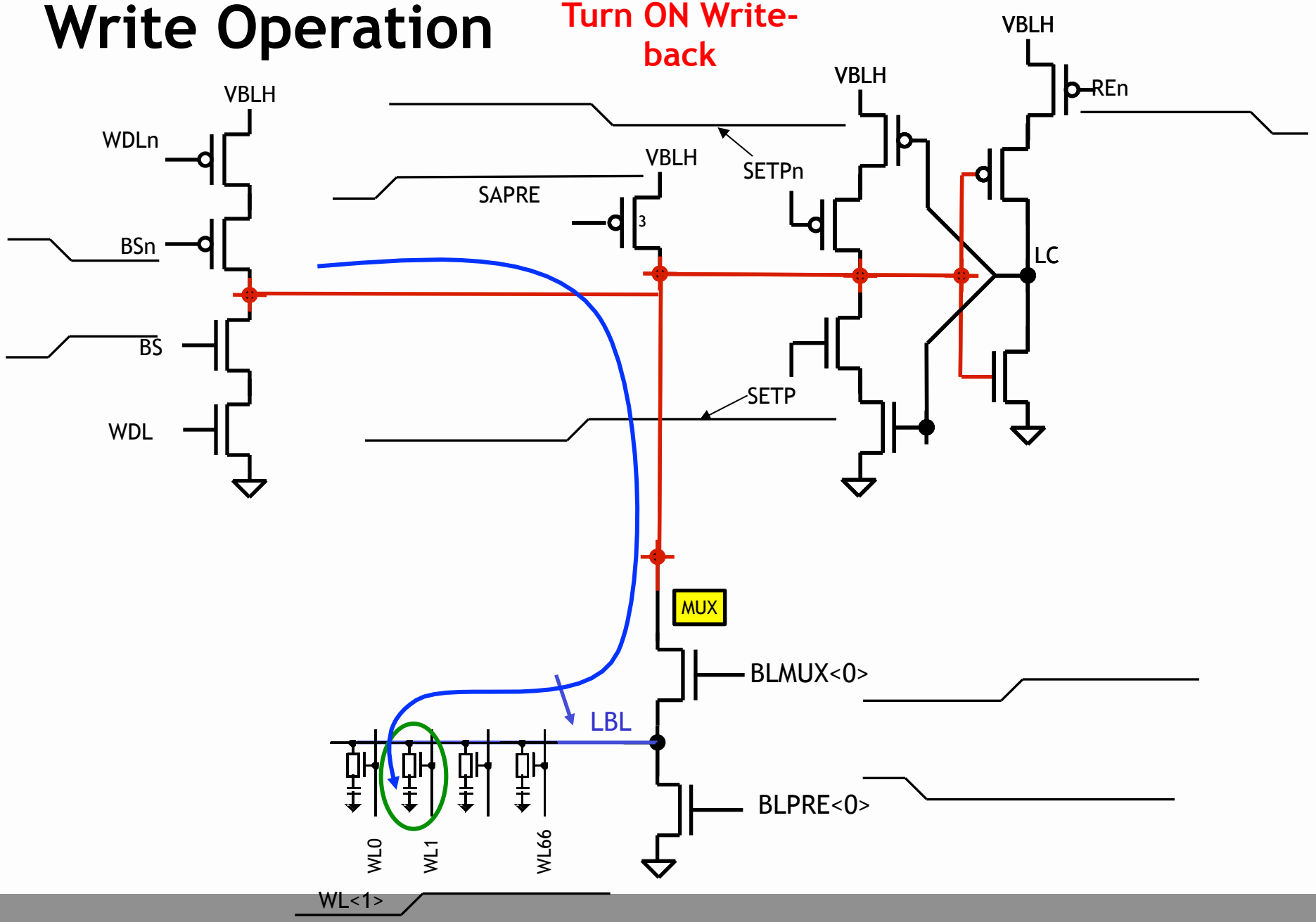
Write Operation



Animation

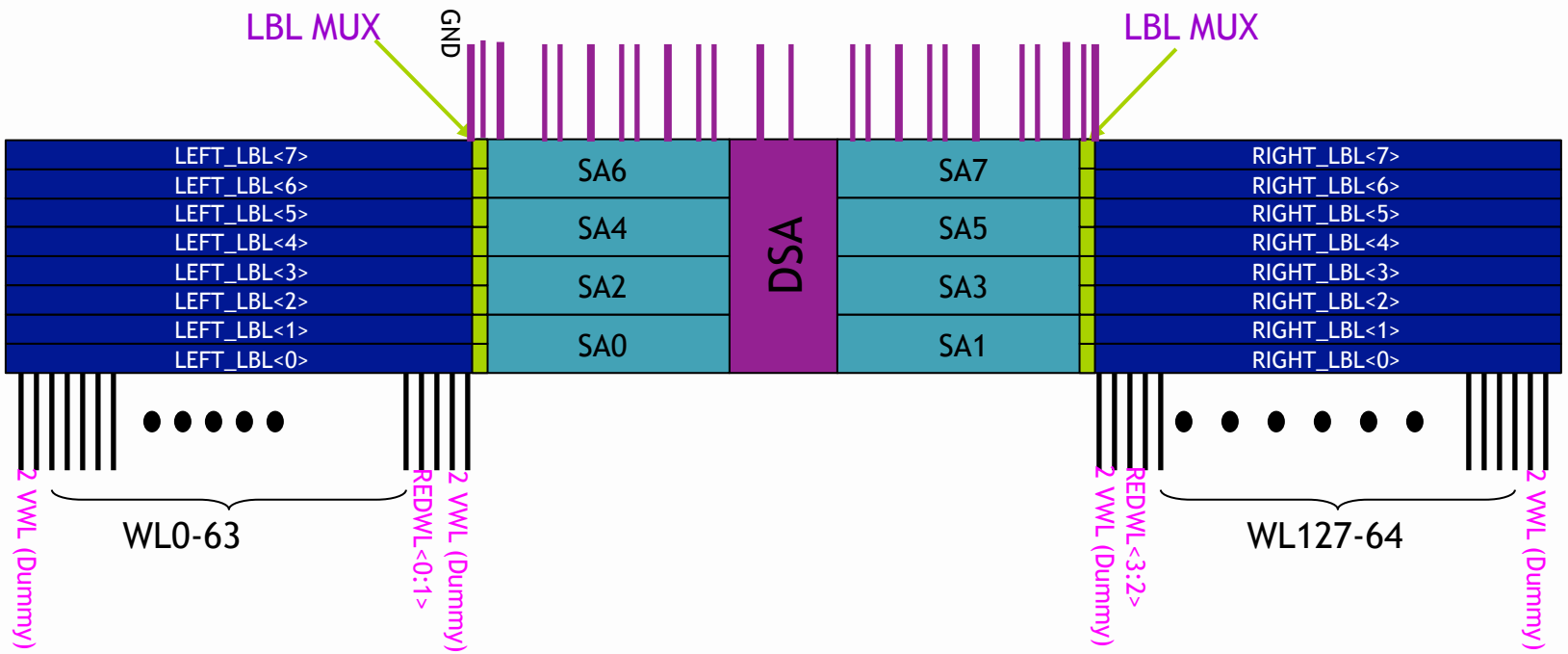
Write Operation

Turn ON Write-back



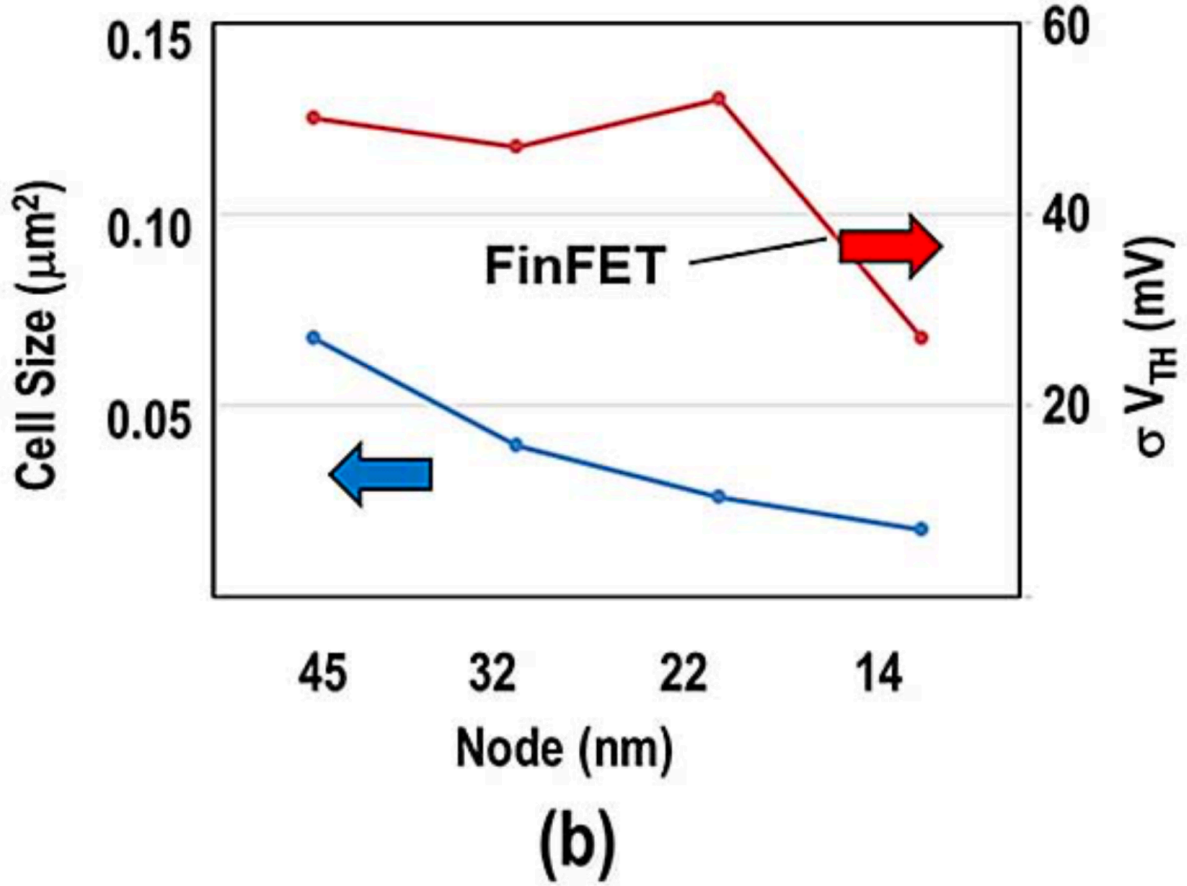
Animation

One Data Line Organization



- Single bit can be read out/ written into by selecting one of 128 rows and one of 8 columns
- The components are sized and arranged to make the layout nice and rectangular
- Repeat this structure as many as there are Data-lines

14nm FinFET Advantage

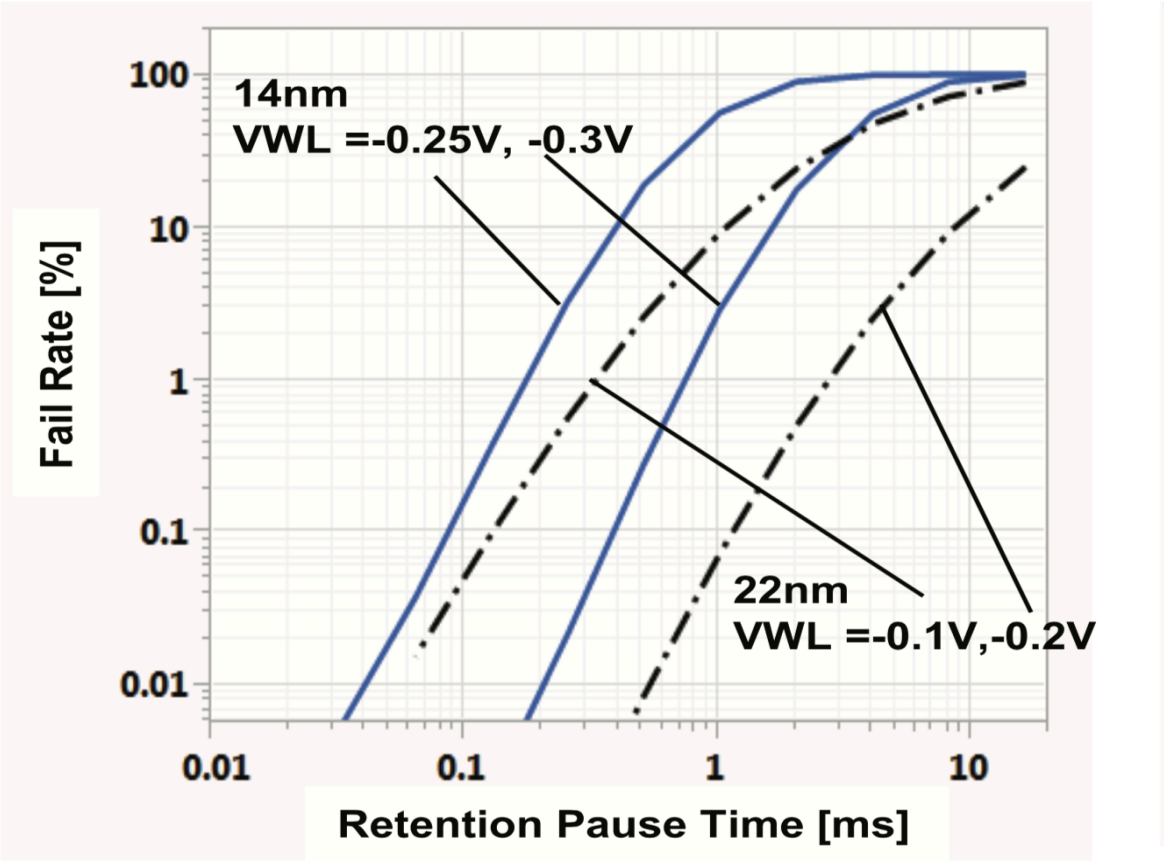


14nm Access Device is 2.5X stronger than the 22nm planar device due to

- 50% more effective width
- 42% shorter channel length
- Lower target Vth

Lower VT variation due to undoped channel

Lower Vth Variation Effect on Retention



- Write a 1 into all the cells
- Read the cells after a pause time
- Ideally (with no local variations) there should be an step jump in the #fail
  - With variations, steeper the slope lesser the variations

Conclusion

- Pulling more DRAM cache (L2,L3) inside the processor improves overall performance
- eDRAM design using logic process is a challenge
- Case study is done, covering many of the eDRAM design aspects
- Sense amp has to read a 1 by default to provide performance improvement
  - Achieved in the Gated Feedback Sense Amp

References

Matick, R. et al., “Logic-based eDRAM: Origins and Rationale for Use,” IBM J. Research Dev., vol. 49, no. 1, pp. 145-165, Jan. 2005.

Barth, J. et al., “A 500MHz Random Cycle 1.5ns-Latency, SOI Embedded DRAM Macro Featuring a 3T Micro Sense Amplifier,” ISSCC Dig. Tech. Papers, pp. 486-487, Feb. 2007.

Barth, J. et al., “A 500 MHz Random Cycle, 1.5 ns Latency, SOI Embedded DRAM Macro Featuring a Three-Transistor Micro Sense Amplifier,” IEEE JOURNAL OF SOLID-STATE CIRCUITS, VOL. 43, NO. 1, JANUARY 2008.

Barth, J. et al., “A 45nm SOI Embedded DRAM Macro for POWER7TM 32MB On-Chip L3 Cache,” ISSCC Dig. Tech. Papers, pp. 342-3, Feb. 2010.

Barth, J. et al., “A 45 nm SOI Embedded DRAM Macro for the POWER™ Processor 32 MByte On-Chip L3 Cache,” IEEE JOURNAL OF SOLID-STATE CIRCUITS, VOL. 46, NO. 1, JANUARY 2011.

S. Iyer et al., “Embedded DRAM: Technology Platform for BlueGene/L Chip,” IBM J. Res. & Dev., Vol. 49, No. 2/3, MARCH/MAY 2005, pp.333-50.

Barth, J. et al., “A 300MHz Multi-Banked eDRAM Macro Featuring GND Sense, Bit-line Twisting and Direct Reference Cell Write,” ISSCC Dig. Tech. Papers, pp. 156-157, Feb. 2002.

Barth, J. et. al., “A 500-MHz Multi-Banked Compilable DRAM Macro With Direct Write and Programmable Pipelining,” IEEE JOURNAL OF SOLID-STATE CIRCUITS, VOL. 40, NO. 1, JANUARY 2005.

Butt,N., et al., “A 0.039um2 High Performance eDRAM Cell based on 32nm High-K/Metal SOI Technology,” IEDM pp. 27.5.1-2, Dec 2010.

Bright, A. et al., “Creating the BlueGene/L Supercomputer from Low-Power SoC ASICs,” ISSCC Dig. Tech. Papers, pp. 188-189, Feb. 2005.

Blagojevic, M. et al., “SOI Capacitor-Less 1-Transistor DRAM Sensing Scheme with Automatic Reference Generation,” Symposium on VLSI Circuits Dig. Tech. Papers, pp. 182-183, Jun. 2004.

References

Karp, J. et al., “A 4096-bit Dynamic MOS RAM” ISSCC Dig. Tech. Papers, pp. 10-11, Feb. 1972.

Kirihata, T. et al., “An 800-MHz Embedded DRAM with a Concurrent Refresh Mode,” IEEE Journal of Solid State Circuits, pp. 1377-1387, Vol. 40, Jun. 2003.

Luk, W. et al., “2T1D Memory Cell with Voltage Gain,” Symposium on VLSI Circuits Dig. Tech. Papers, pp. 184-187, Jun. 2004.

Luk, W. et al., “A 3-Transistor DRAM Cell with Gated Diode for Enhanced Speed and Retention Time,” Symposium on VLSI Circuits Dig. Tech. Papers, pp. 228-229, Jun. 2006.

NEC eDRAM Cell Structure (MIM Capacitor): http://www.necel.com/process/en/edramstructure.html

Ohsawa, T. et al., “Memory Design using One-Transistor Gain Cell on SOI,” ISSCC Dig. Tech. Papers, pp. 152-153, Feb. 2002.

Pilo, H. et al., “A 5.6ns Random Cycle 144Mb DRAM with 1.4Gb/s/pin and DDR3-SRAM Interface,” ISSCC Dig. Tech. Papers, pp. 308-309, Feb. 2003.

Taito, Y. et al., “A High Density Memory for SoC with a 143MHz SRAM Interface Using Sense-Synchronized-Read/Write,” ISSCC Dig. Tech. Papers, pp. 306-307, Feb. 2003.

Wang, G. et al., A 0.127 mm2 High Performance 65nm SOI Based embedded DRAM for on-Processor Applications,” International Electron Devices Meeting, Dec. 2006.

G. Fredeman et al., “A 14 nm 1.1 Mb Embedded DRAM Macro With 1 ns Access,” in IEEE Journal of Solid-State Circuits, vol. 51, no. 1, pp. 230-239, Jan. 2016. doi: 10.1109/JSSC.2015.2456873