High-Performance SRAM Design

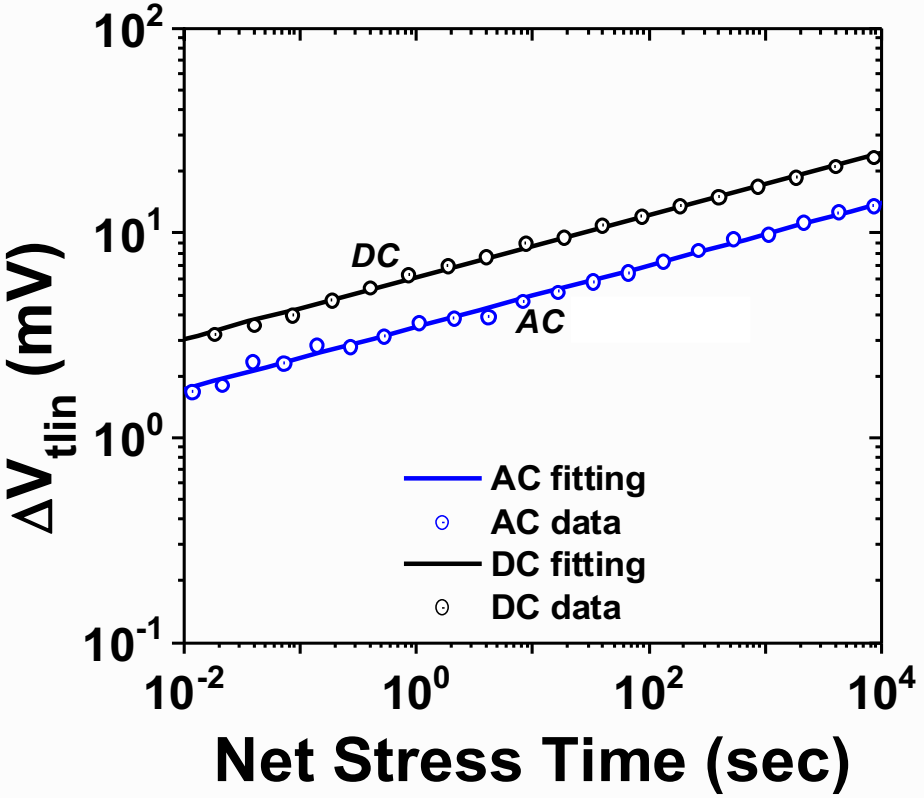
Rahul Rao

IBM Systems and Technology Group

Bias Temperature Instability

FET characteristics and BTI

- Threshold voltage (V_T)
- Current degradation approximately corresponds to V_T degradation
- Known sensitivities
  - PBTI is *more* sens. to Voltage
  - NBTI is *more* sens. to temperature
- Semi-empirical model (for both NBTI and PBTI)

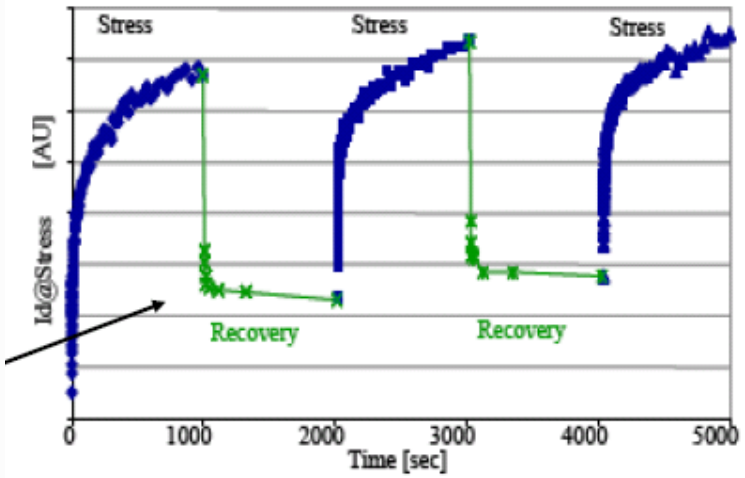


$$\Delta V_T|_{DC\_stress} = A V_{DD}^{a} T^{b} t^{n}$$ *(without including recovery)*

Recovery after stress

Static stress and recovery or "0.001 Hz waveform"

*Ramey et. al., Intel, IRPS 2009



- **Recovery happens when FET is OFF**

$$\Delta V_T\,|_{after\_relax} = FR \mathrm{x} \Delta V_T\,|_{DC\_stress}$$

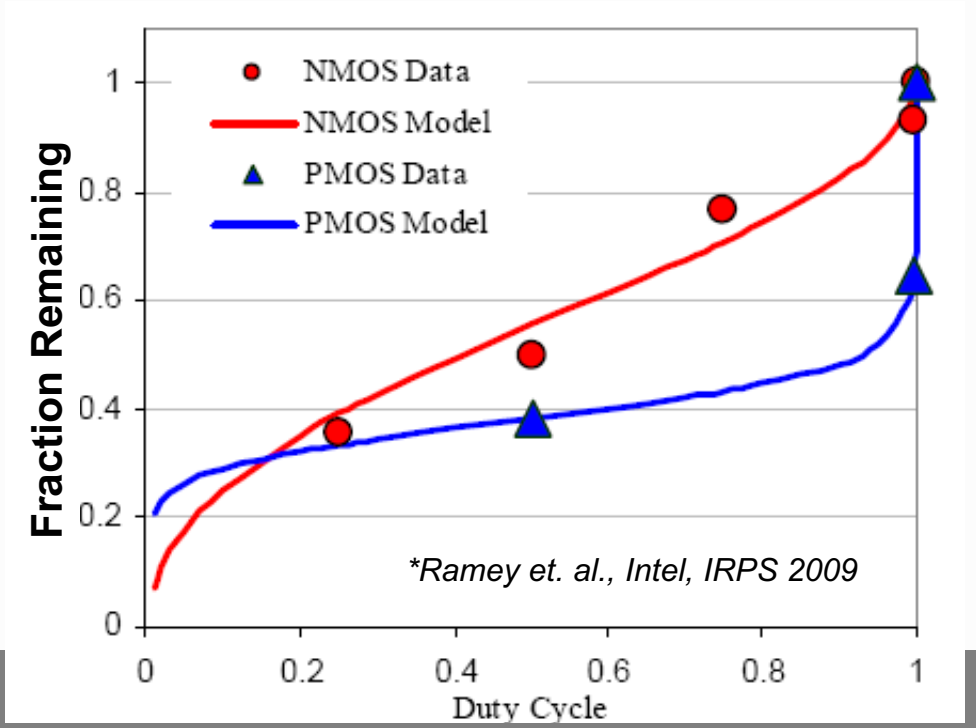
Fraction Remaining

$$FR = \left[1 + \alpha \left(\frac{t_{relax}}{t_{stress}}\right)^n\right]^{-1}$$

- In other words, if probability of '1' at the gate of an NFET is say P_1, then,

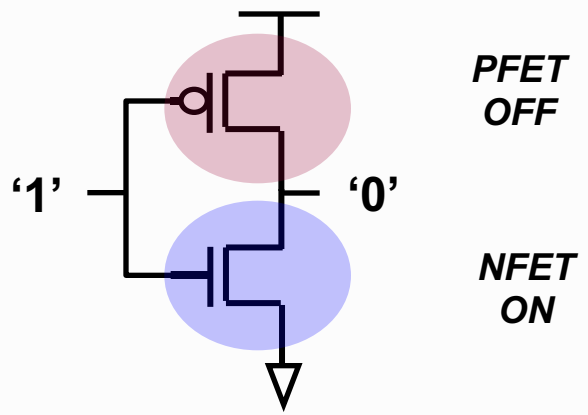
$$\text{duty\_cycle} = \frac{t_{\text{stress}}}{t_{\text{stress}} + t_{\text{relax}}} = P_1$$

- **Biggest benefits of recovery when duty_cycle < 95%**



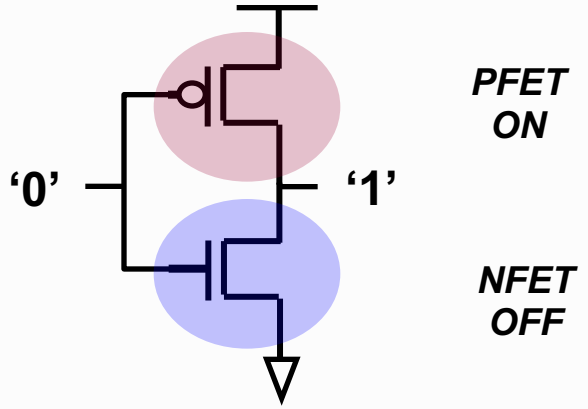
FET stress-recovery

Let’s take an example of simple CMOS inverter



PFET is relaxing: Gate-source HIGH and drain LOW

NFET is stressed: Gate is HIGH and source-drain LOW



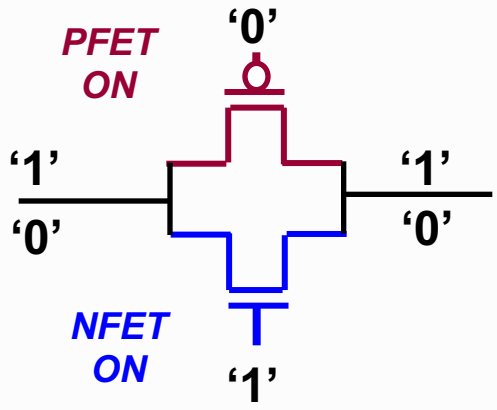
PFET is stressed: Gate is LOW and source-drain HIGH

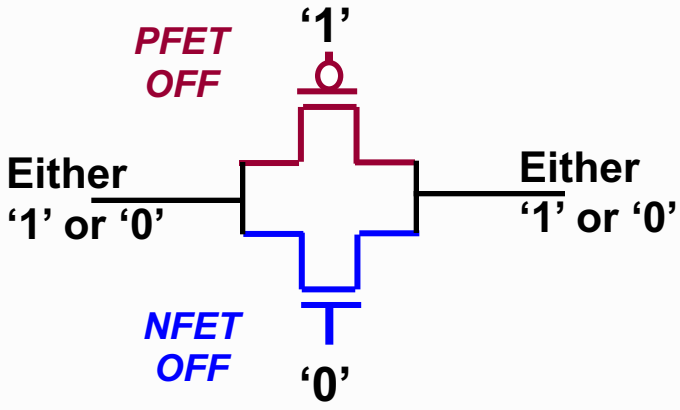
NFET is relaxing: Gate-source LOW and drain HIGH

- If a FET is ON, it’s stressed (for both NFET and PFET)
- If a FET is OFF, it’s relaxed (for both NFET and PFET)

FET stress-recovery

For other circuit types say transmission gate



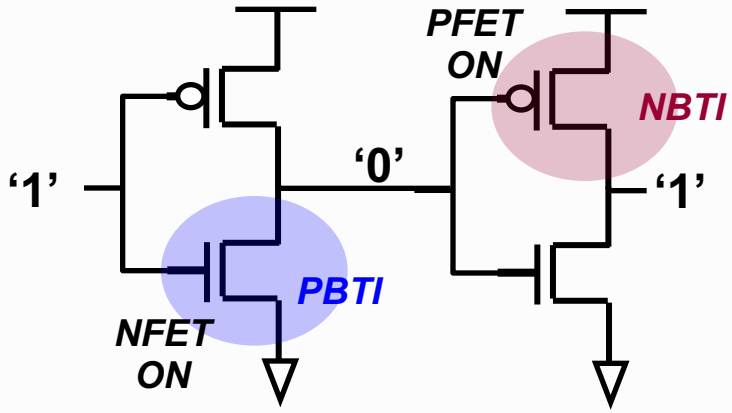


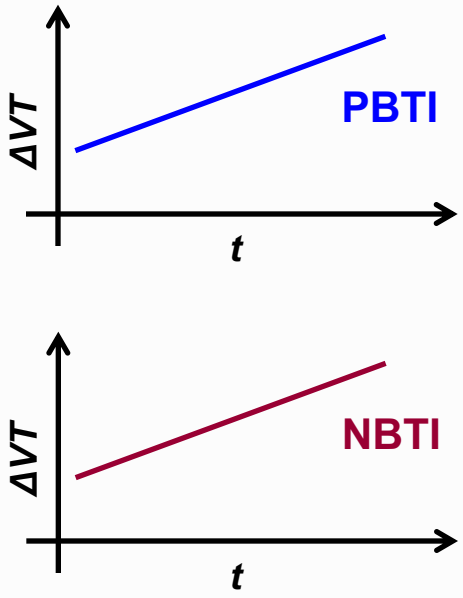
- Both the FETs are either ON or OFF together
- Source-drain voltages during relaxation depends on other circuit blocks on left and right
  - e.g., negative gate-drain or gate-source voltage in NFET may speed up recovery (also true for inverter on previous slide)

Nature of Stresses/recovery

Static or DC stress

– FETs are in the same voltage bias condition during the USAGE





Examples:

–SRAM cells storing the same data for long time

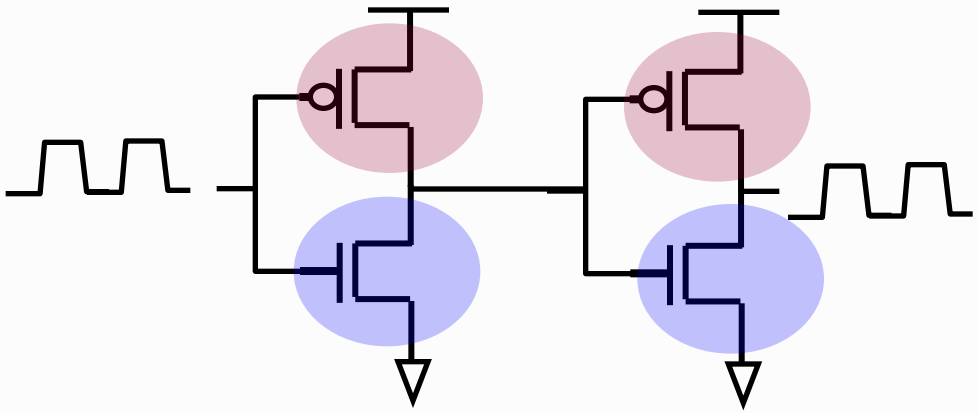
–Circuit paths not used for long time but powered on

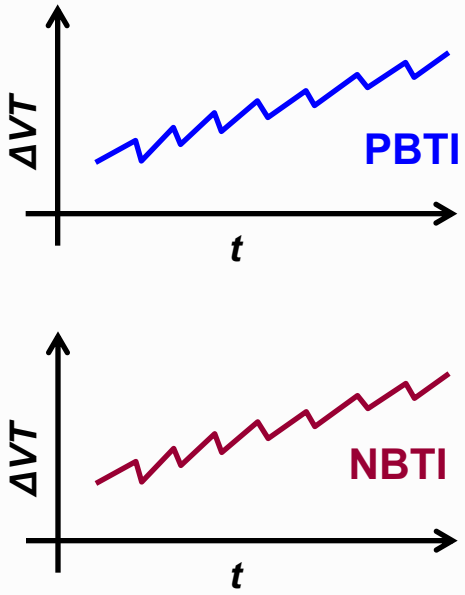
– Word line drivers, local and global eval circuits

Nature of Stresses/recovery

Alternating or AC stress

- FETs turn ON (stress) and OFF (recover) during the USAGE
- It's composed of several DC stress and recovery conditions
- Durations of stress and recovery depend on nature of program running and typically can not be estimated





Examples:

–SRAM cells frequently changing the stored data

–Logic circuit paths doing computations

Question

- The bias temperature instability device degradation in a circuit can be reduced by

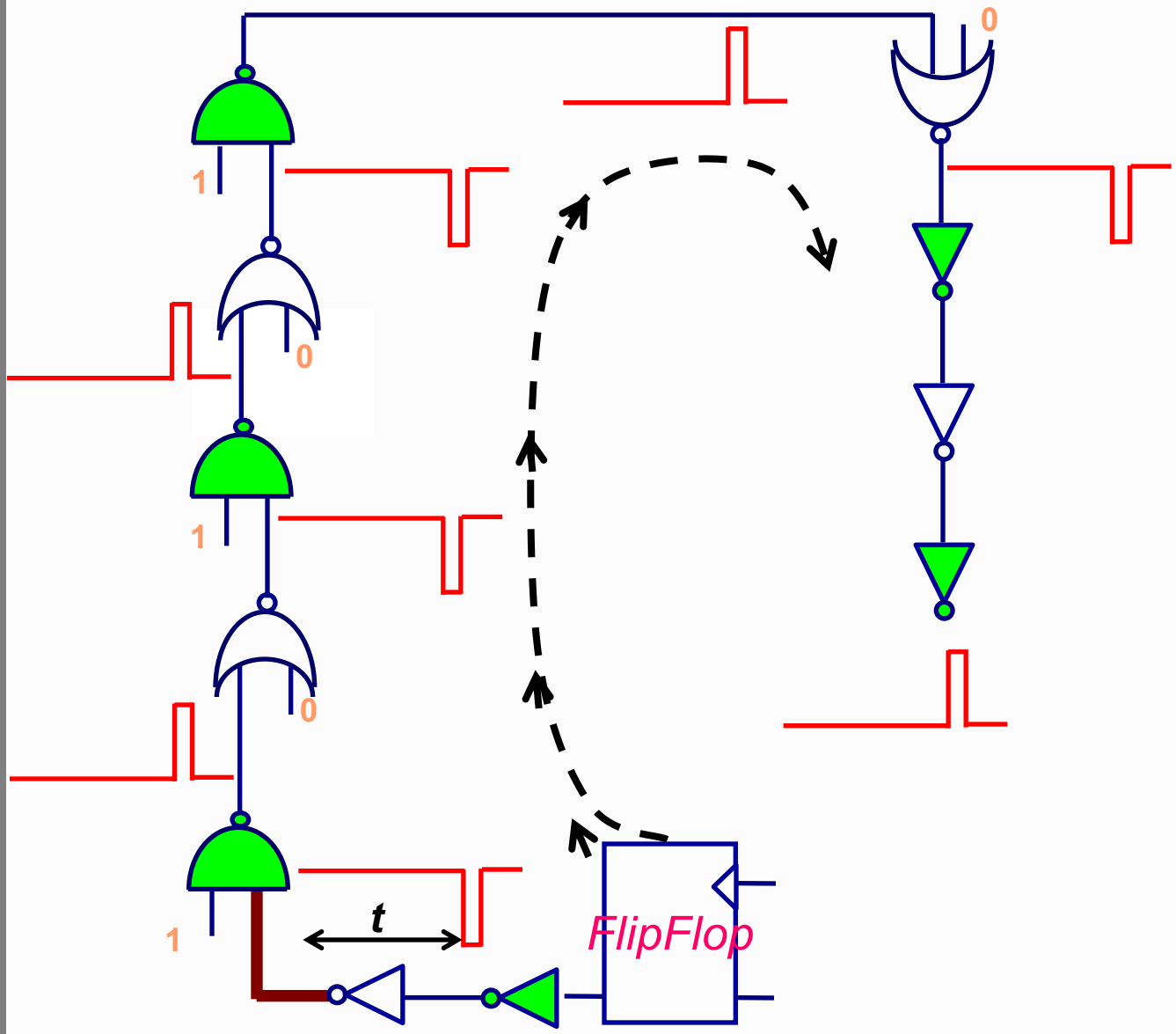
a) Using assymmetric transistors since they will degrade at different rates

b) Avoiding pass gates in the design and always using transmission gates

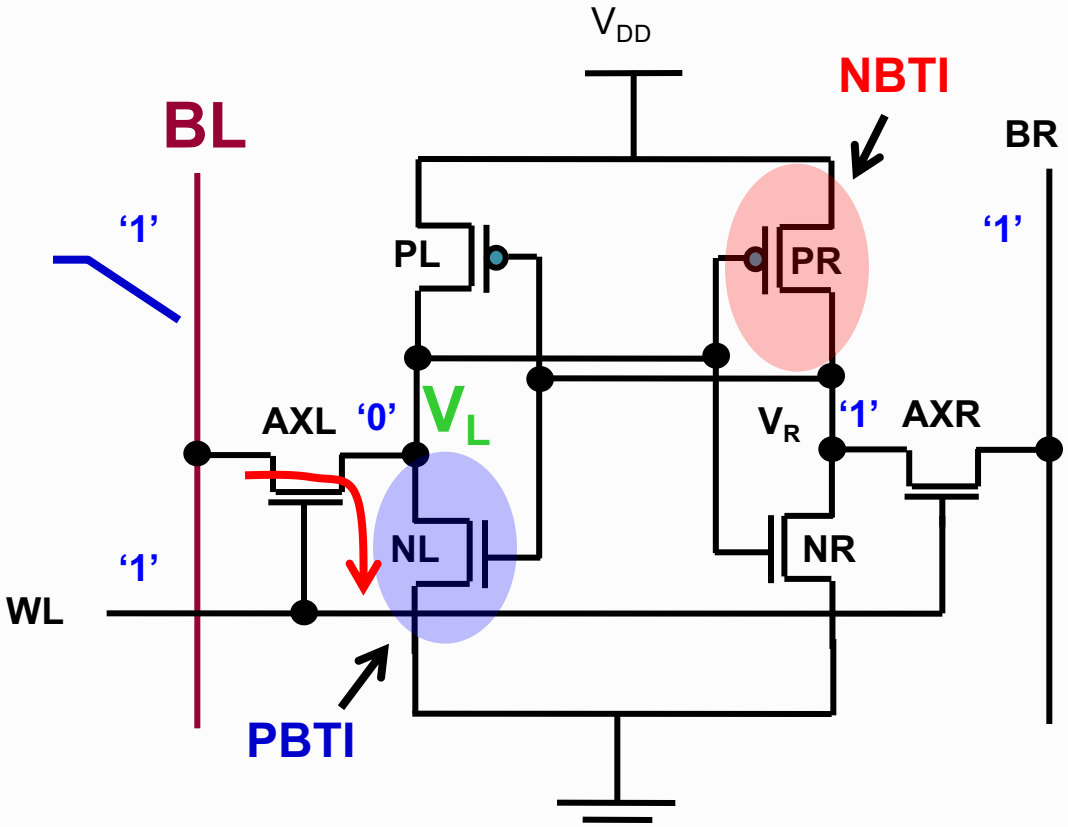
c) Increasing the supply voltage of the circuit

d) Ensuring that all nodes in a circuit switch every N cycles

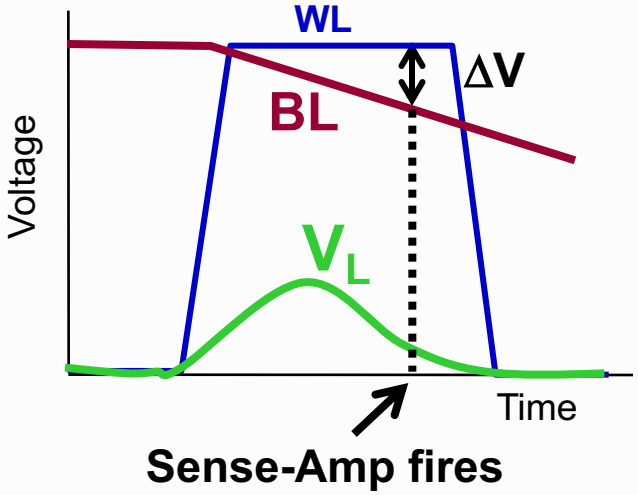
Timing Failure due to BTI



SRAM Operating Mode: READ



BL and BR are pre-charged to VDD and then left hanging

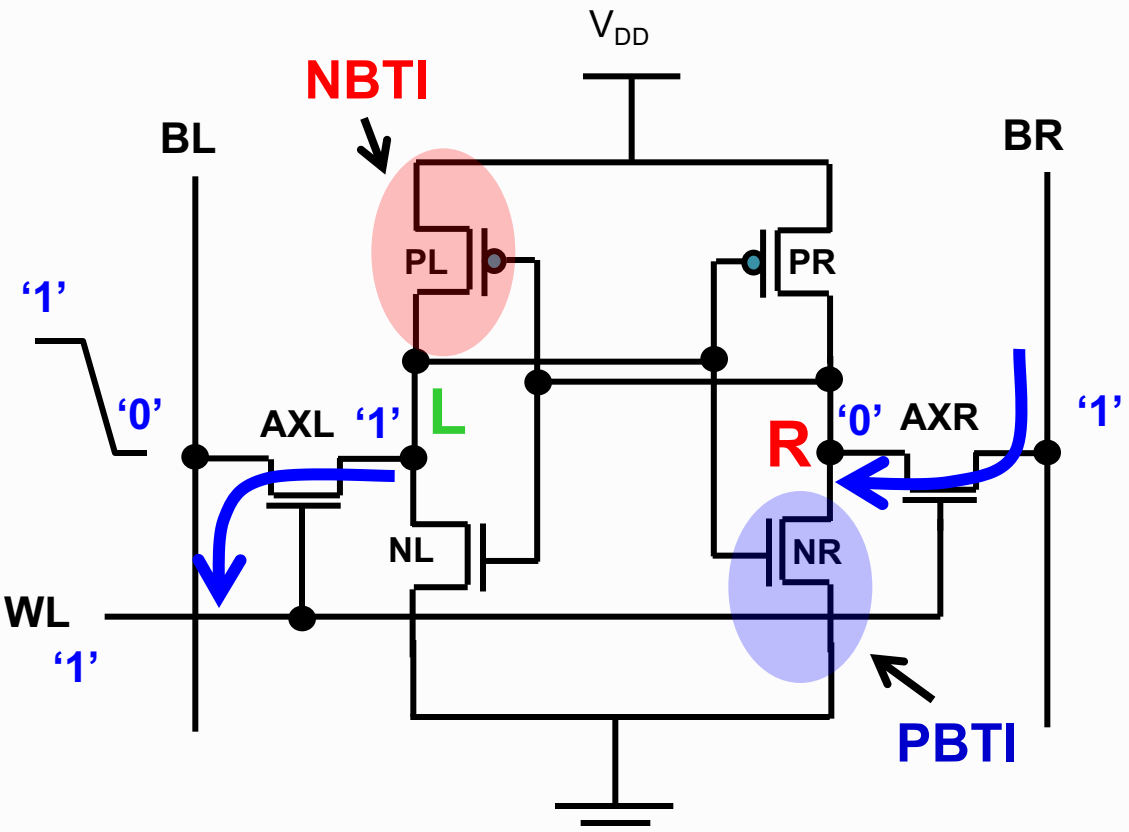


Access FETs (AXL & AXR) are ON for short duration while cell is accessed => assumed negligible degradation

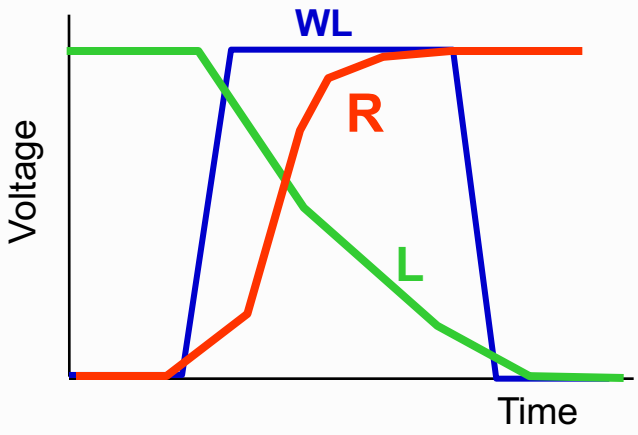
- **The data stored should *not* flip during READ**
  => NL (NR) should be stronger than AXL (AXR): *PBTI can make NL weak (bad!)*
- **Sufficient ΔV to fire SA should be developed while WL = '1'**
  => AXL-NL should fast discharge BL: *Weak NL will slow discharge (bad!)*

(A. Bansal, MicroReliability 2009)

SRAM Operating Mode: WRITE



BL and BR are FIXED to data



Weak cross-coupled inverters (NL-PL and NR-PR) are good for writing

- **The data stored *must* flip during WRITE**

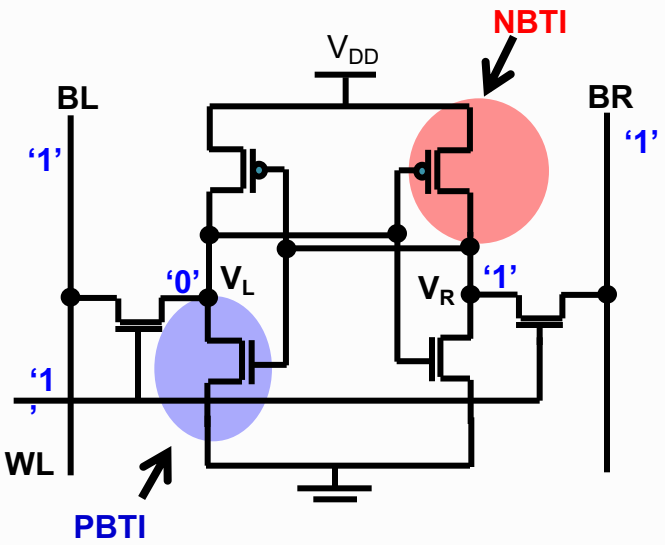
=> AXL (AXR) should be stronger than PL (NR) *NBTI (PBTI) can make PL (NR) weak (good!)*

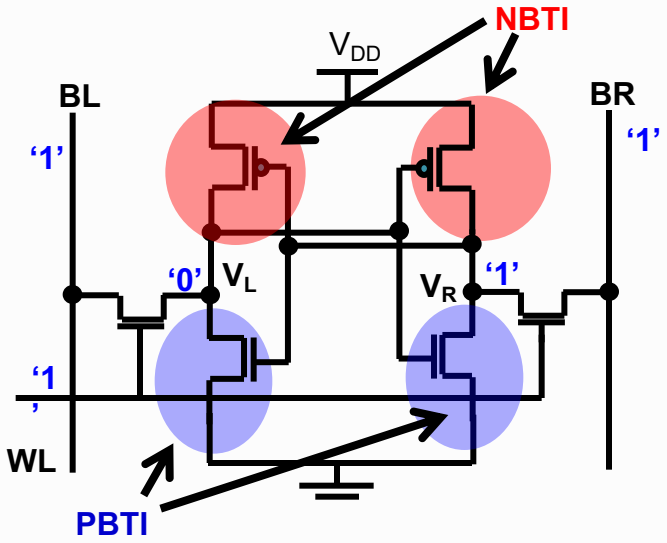
- **Data should flip while WL = '1'**

=> *WL pulse width increased (good)*

(A. Bansal, MicroReliability 2009)

Static and Alternating Stress





Static Stress

- Cell is storing same data for long time => asymmetric
- May be *READ* multiple times but not flipped
- ΔVt for static stress larger than alternating stress (no recovery)
- READ gradually becomes unstable
- Increases READ access time

Alternating Stress

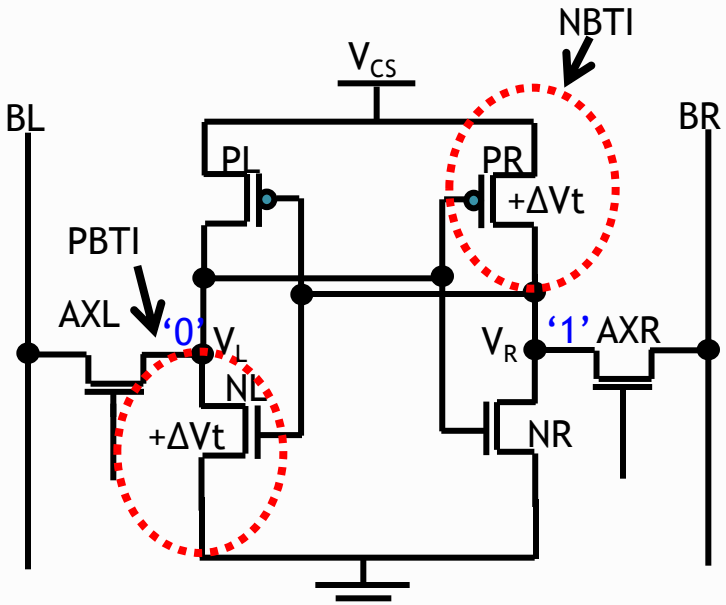
- Cell is regularly flipped => symmetric
- Equal time/relaxation for storing '1' and '0' to maintain symmetry
- ΔVt for same usage is less (low power-on time)
- β-ratio between pull-down and pass-gate FETs varies => PD weakens and READ fail increases
- Increases READ access time

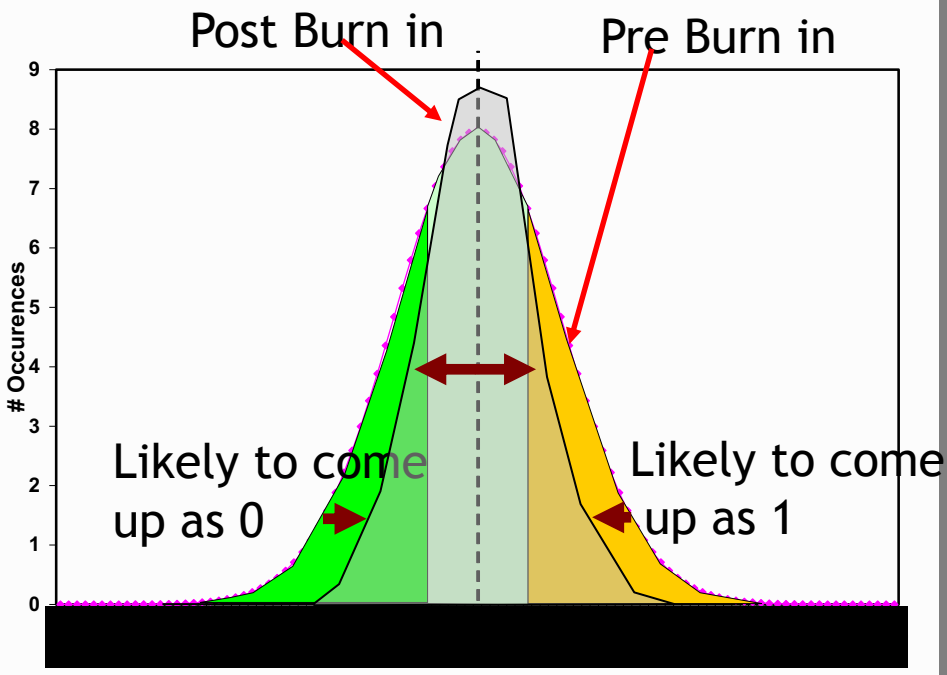
➢ **Typically all cells in between Static to Alternating stress**

(A. Bansal, MicroReliability 2009)

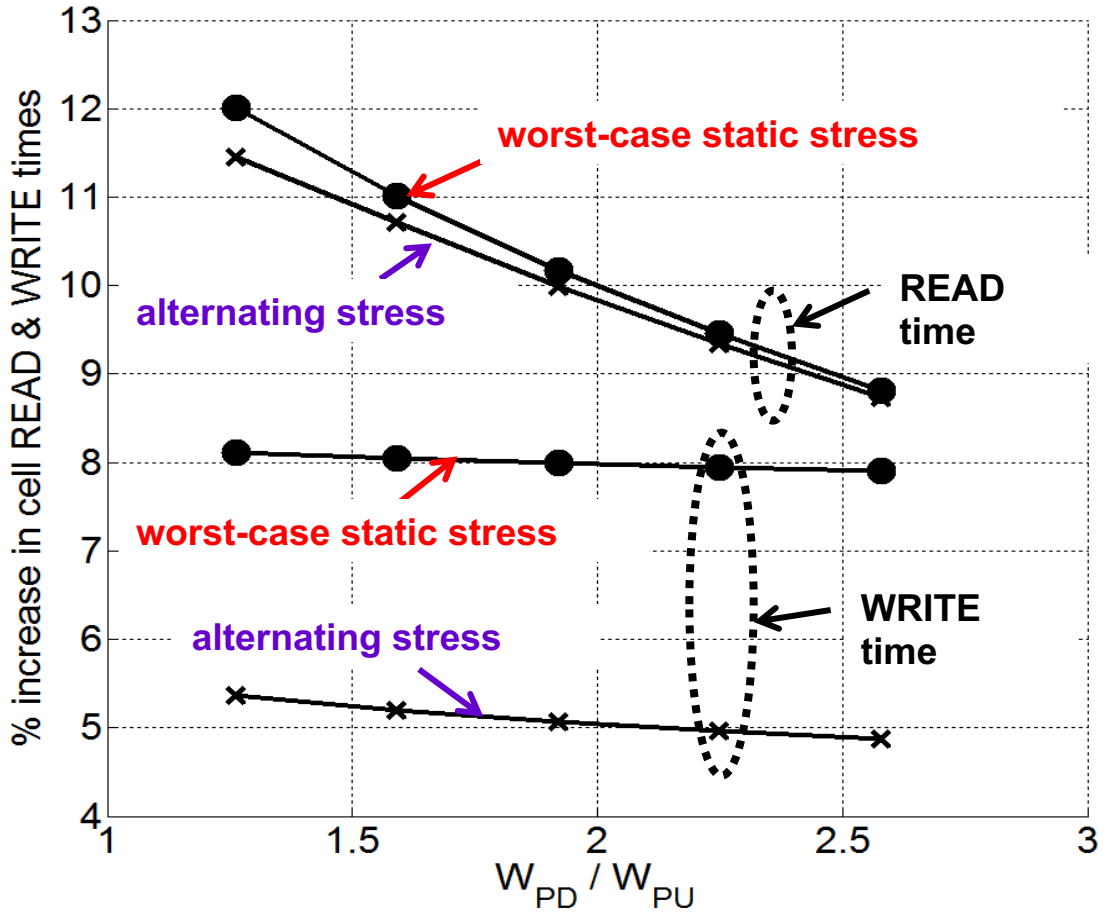
Unbalanced Post-Fabrication SRAM Cells

- Unbalanced cells result in uneven squares in butterfly curve => Reduced noise margin
- Read performance on different read ports become difference (in case of single ended read / 8T cells)
- Bring up state:
  - V_L = 0, NL-PR pair is stronger than PL-NR pair
- Pre-condition sram cell state before burn in (depending on stability or performance need)
  - Burn – in with V_L = 0
- Duration of 'conditional' stress can be based on spread of the initial curve
  - Or by # of 1s / 0s in the initial bring up





Impact on Cell READ and WRITE time

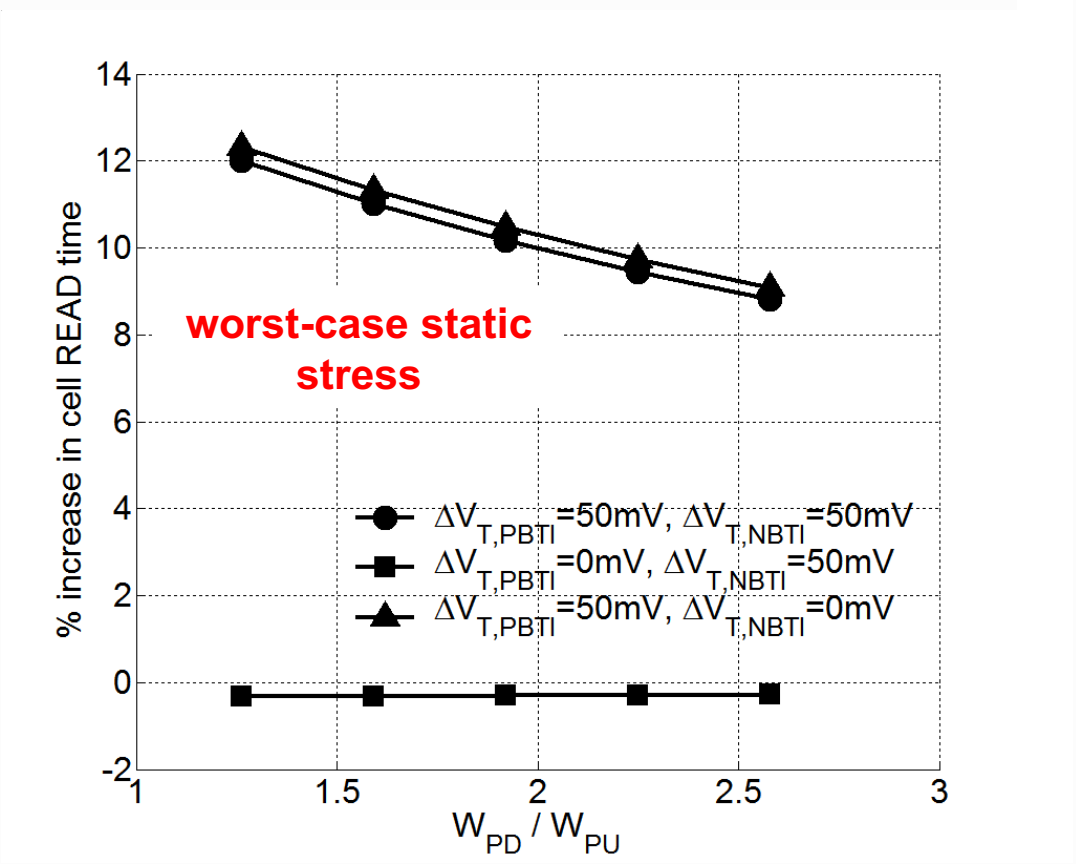


Assuming $\Delta V_T = 50mV$ increase due to NBTI and PBTI at end-of-life

- **READ time reduces as we go from dense to high-performance cell**
- **READ time is *less* dependent on stress condition**

(A. Bansal, IRPS 2009)

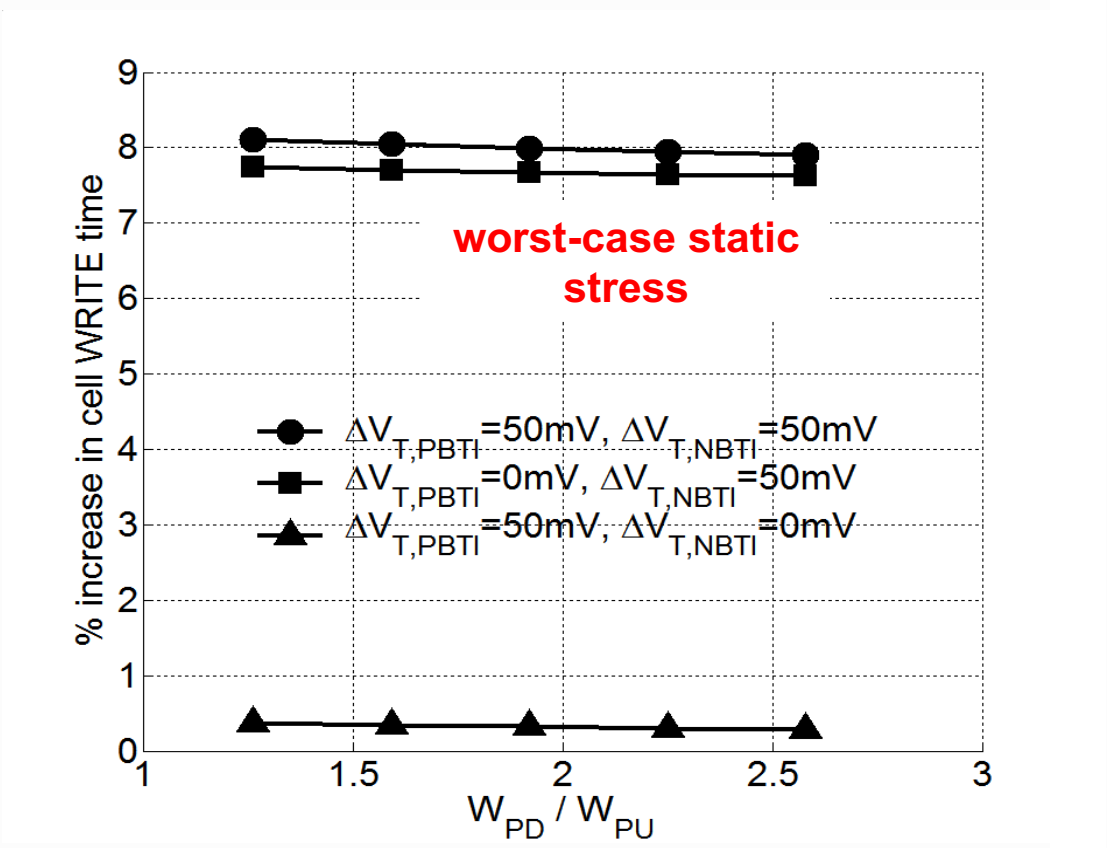
READ access time



- **READ time is *practically immune* to NBTI**
- **READ time degrades ~ 9-12% for 50mV V_T shift due to PBTI**
- ***Stringent PBTI requirements in high-performance cells***

(A. Bansal, IRPS 2009)

WRITE time



- **WRITE time is *practically immune* to PBTI**
- **WRITE time degrades ~ 8% for 50mV V_T shift due to NBTI**
- ***Similar NBTI requirements in high-performance and dense cells***

(A. Bansal, IRPS 2009)

Hot Carrier Injection

- The impact of HCI

a) Is more on access transistors than the pull down and pull up transistors
b) Is more on word-line overdrive (WLOD) assisted cells
c) Is higher on frequently written cells than frequently read cells
d) Improves the read access time of 8T cells

Test

Test

❑ Defect vs Fault

- Defect => A deviation from intended behavior
- Fault => A model for the defect

❑ Structural vs Functional Test

- Structural Test => Testing all nodes of the circuits
- Functional Test => Behaves as desired

❑ Verification vs Test

- Verification => Design is correct
- Test => Hardware is correct

Test

- Stuck At Faults => Cell is stuck at a particular value
- Transition Faults => Cell fails to undergo a particular transition
- Coupling Faults => Cell (v) fails due to Cell (a)
- Pattern Faults => Cell (v) fails due to a multiple set of Cells (a1 – an)

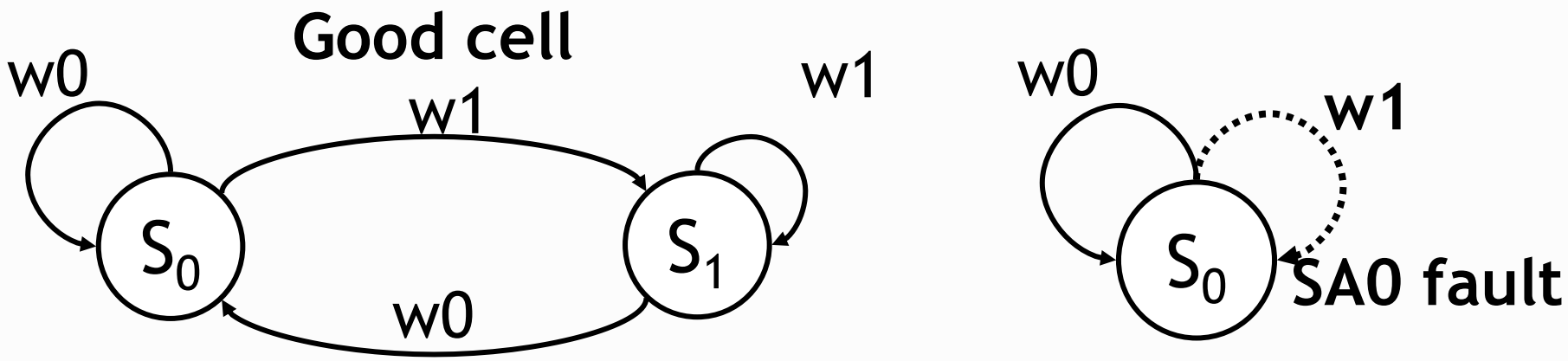
Fault Notation

- <…> describes a fault
- <*S*/*F*> describes a *single-cell fault*
  - *S* describes the state/operation *sensitizing* the fault
    - A fault is sensitized when the fault effect is made present
  - *F* describes the *fault effect* in the *victim cell* (v-cell)
- <*S*;*F*> describes a *two-cell fault* (a Coupling Fault)
  - *S* describes the state/operation of the *aggressor cell* (a-cell) sensitizing the fault
  - *F* describes the fault effect in the v-cell

- Examples
  - <∀/0>: a SA0 fault
  - <∀/1>: a SA1 fault
  - <↑/0>: an ↑ TF
  - <↓/1>: a ↓ TF
  - <↑;0>: a CF
  - <↑;1>: a CF
  - <↓;0>: a CF
  - <↓;1>: a CF

Stuck Faults

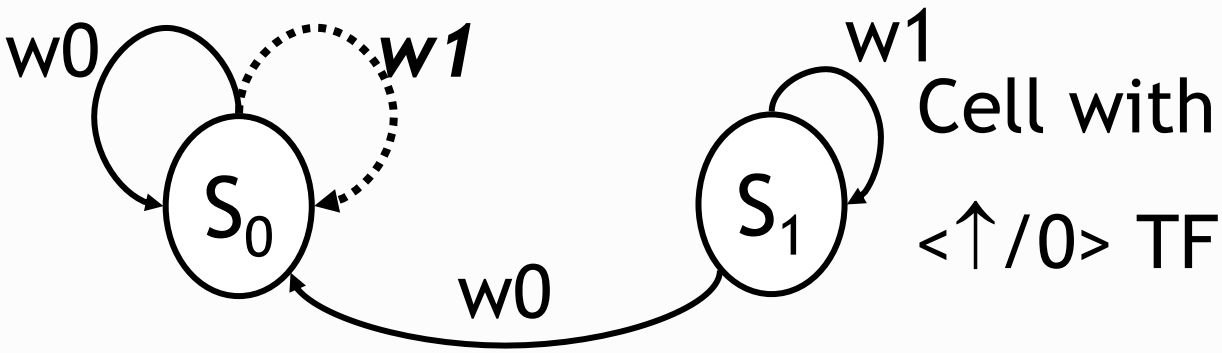
- Stuck At Faults => Cell is stuck at a particular value



- Stuck at Open Fault
  - e.g Word line is broken => When read is performed both BL and BLB remain high

Transition Fault

❑ Transition Faults => Cell fails to undergo a particular transition

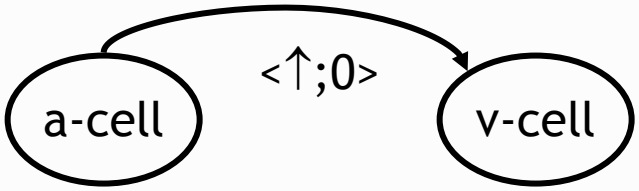


Test: All cells should have a ↑ and ↓ transition and read

❑ Retention Faults => Cell loses value after a while

Coupling Faults

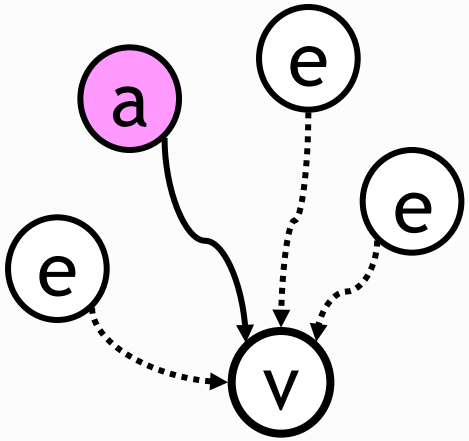
- Coupling Faults => State (or Transition) of Cell A causes a failure in Cell V



- Coupling Transition Faults: <↑;0>, <↑;1>, <↓;0> and <↓;1>
- Coupling state faults: <1;0>, <1;1>, <0;0> and <0;1>
- Coupling inversion fault: <↑;↕> and <↓;↕>

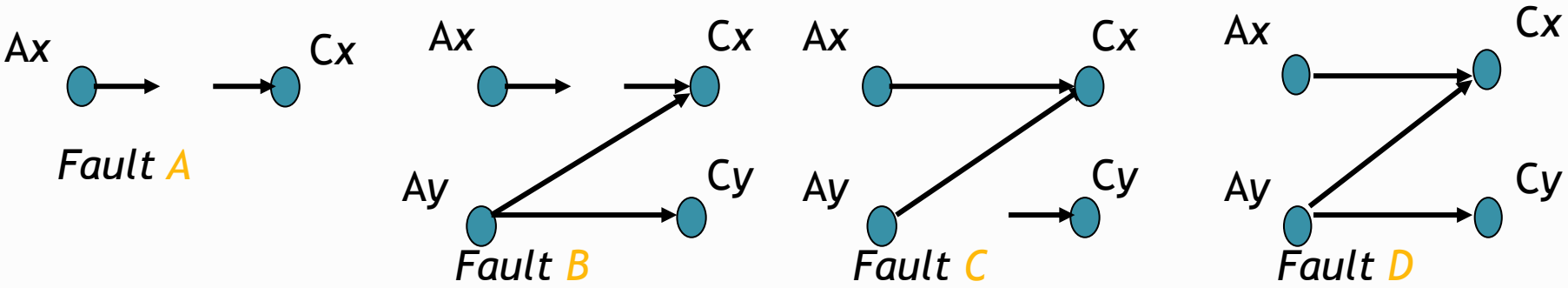
Pattern Sensitive Faults

- The state (transition) or K cells causes a failure in an ith cell
  - NPSF => The k cells are adjacent



Address Faults

- Address (A1) cause no cell to be accessed
- Address (A1) accesses multiple cells
- Cell (C1) is accessed with multiple addresses
- Cell (C1) is accessed by its own and another address



March Tests

- A sequence of operations applied in a particular order aimed at targeting the various fault models
- March C

{⇕(w0);⇑(r0,w1);⇑(r1,w0);⇕(r0);⇓(r0,w1);⇓(r1,w0);⇕(r0)}

- Intermediate r0 shown to be redundant => March C-

{⇕(w0);⇑(r0,w1);⇑(r1,w0);⇓(r0,w1);⇓(r1,w0);⇕(r0)}

Power Computation

- Apply sequence of patterns aimed at separating power components
  - Initialize array with equi probable 0s and 1s (make A0 = all 0s)
  - Hold operation (P_H) => Leakage + Clock power
  - Write all 0s to Address A0 (P_{w0}) => Power of a write where no cells is actually written => Power of address decode + write drivers (peripherals)
  - Read from Address A0 (P_{r0}) => Should read all 0s, i.e all bit-lines will discharge (in a single ended read scenario)
  - Write 111 to Address A0 (P_{w1}) => Power of a write operation where N cells are actually written
  - Read from address A0 (P_{r1}) => Power of a read operation where no bit lines will discharge (in a single ended read scenario)

- Power of a write operation where N of M bits actually flip = $P_H + P_{w0} + (P_{w1} - P_{w0})*N/M$