

# Circuit Modeling of Non-volatile Memory Devices

M. Sadd, R. Rao, R. Muralidhar  
R. Thoma  
September 16, 2005



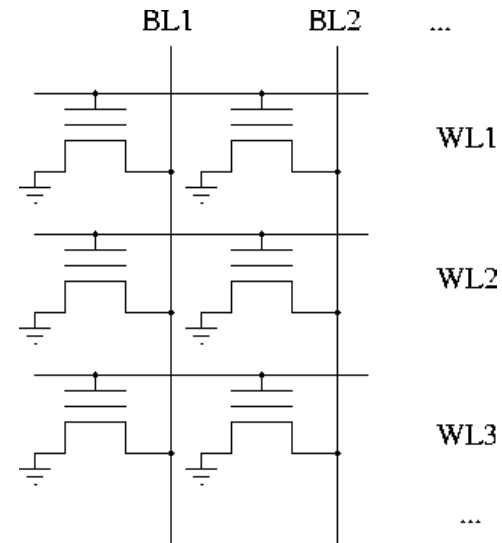
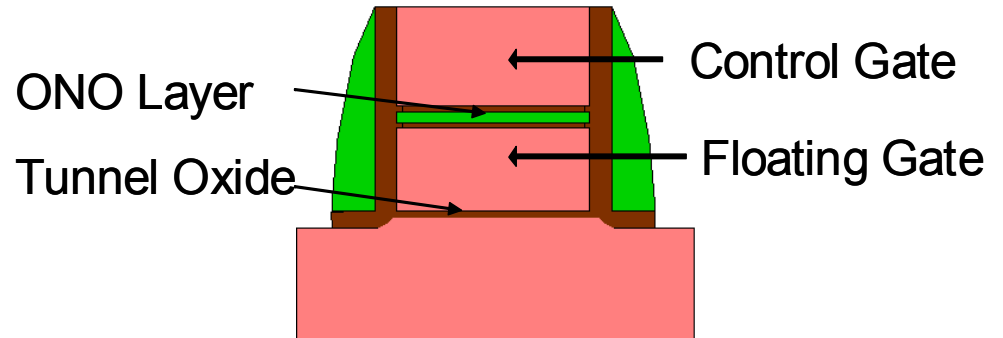
- Introduction to flash
- Capacitor sub-circuit and sense model
- Challenges in Thin film storage model
  - Program/Erase and Endurance
  - 2-bit per cell
  - Reliability

# NVM operates with processes that normally cause failure

Example	NVM Process	Failure Mode
Fe-RAM	Ferro-Electric Hysteresis	$V_t$ instability in High-k dielectrics
SONOS	Charge Trapping in gate stack	Fixed Charge instability
Flash	HCI Programming/ Tunnel Erase	Stress-induced trap creation and charging

- **Need to model effects that are minimized in most other devices!**

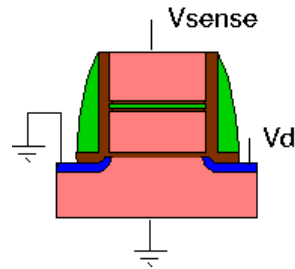
- **Flash Cell** ⇒ most common type of NVM:
- **NOR Array:**
  - The memory becomes “Flash” when organized in an array with “block” erase



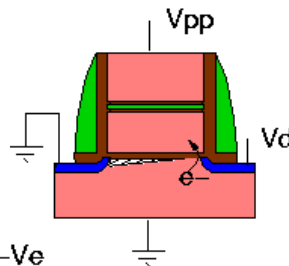
# Flash Cell Operations

- **Operation:**

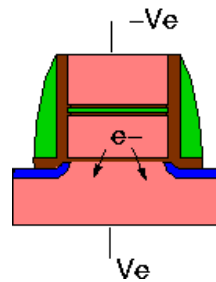
- Sense



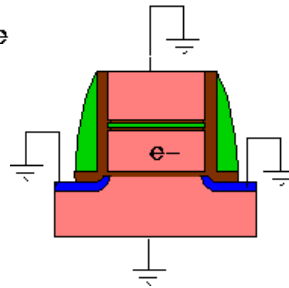
- Program



- Erase



- Retention/Disturb



- **Model Needs to describe:**
  - Variable Threshold voltage

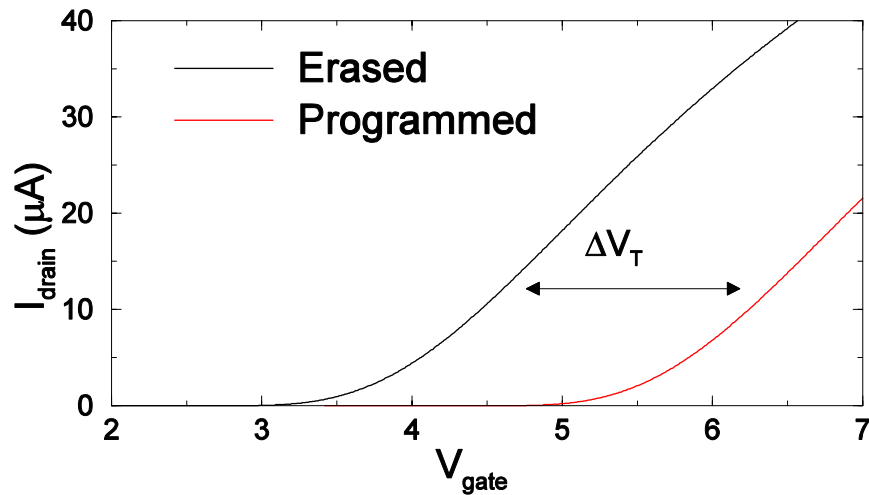
- HCI or Tunneling

- Tunneling

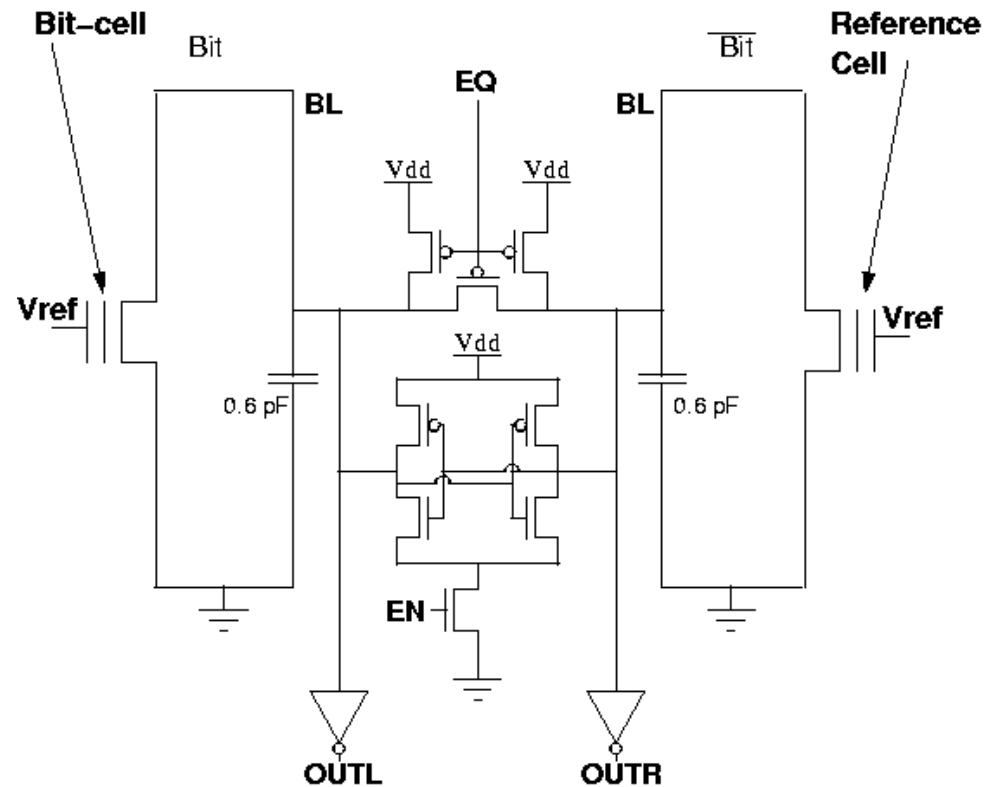
- Charge loss with bias

# Flash Sense Operation

- Memory senses the  $V_t$  shift from stored charge

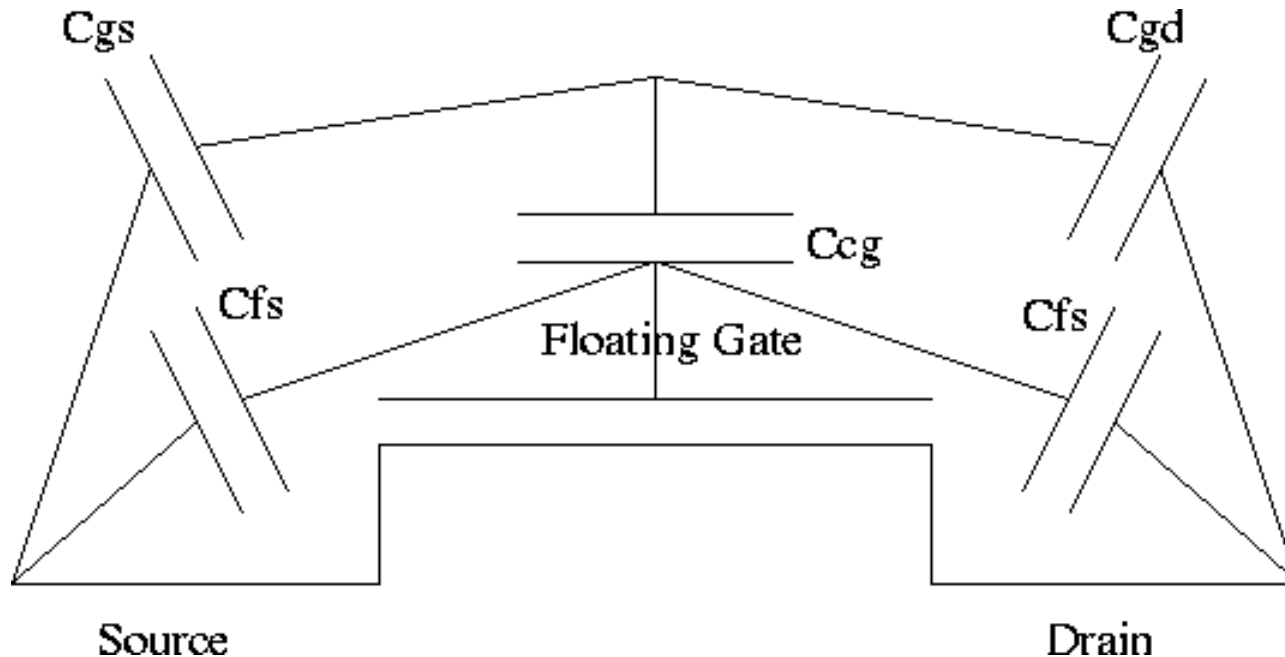


- Basic sense circuit

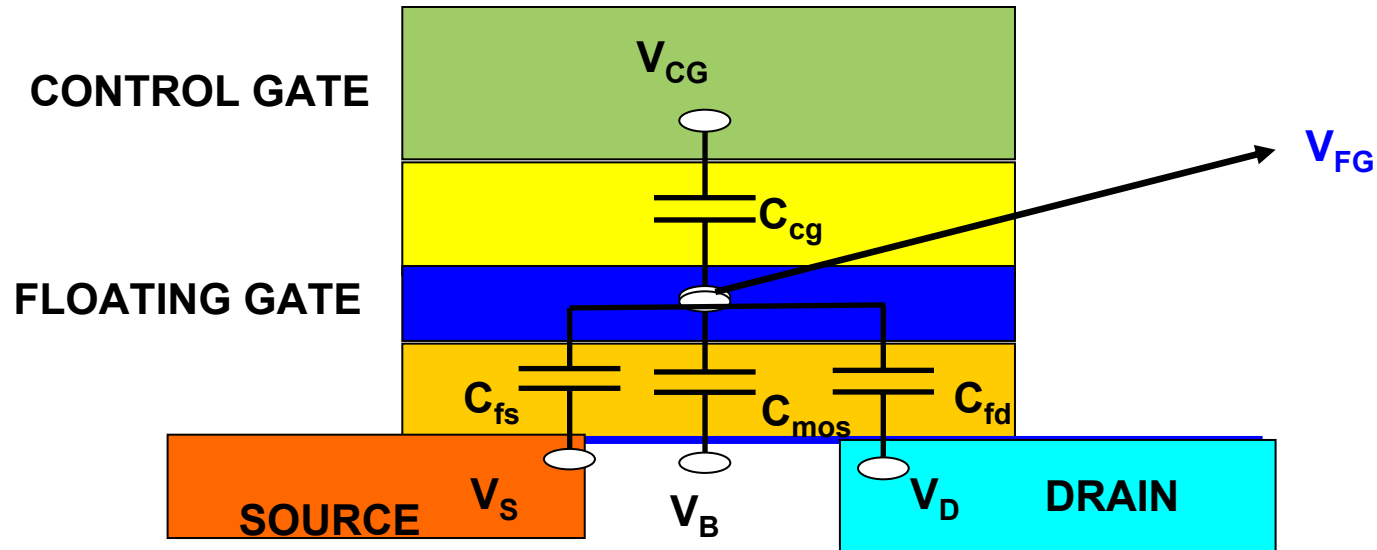


## Simple Approach

- Separate models for program/erase  $V_t$
- More flexible sub-circuit



# Floating Gate Capacitance Model



$$V_{FG} = \alpha_g V_{CG} + \alpha_s V_S + \alpha_d V_D + \alpha_b V_B + Q_{FG} / C_T$$
$$\alpha_g = C_{cg} / C_T ; C_T = C_{cg} + C_{fs} + C_{fd} + C_{mos}$$

Coupling ratio,  $\alpha_g$  is critical for tunnel program/erase

- As  $\alpha_g$  increases,  $V_{FG}$  emulates more the control gate potential



- **Charge stored on floating node:**

$$Q_{FG} \sim C_{mos} V_{fg} + C_{fs} V_{fg} + C_{fd}(V_{fg} - V_d) + C_{cg}(V_{fg} - V_{cg})$$

- **Define coupling ratios:**

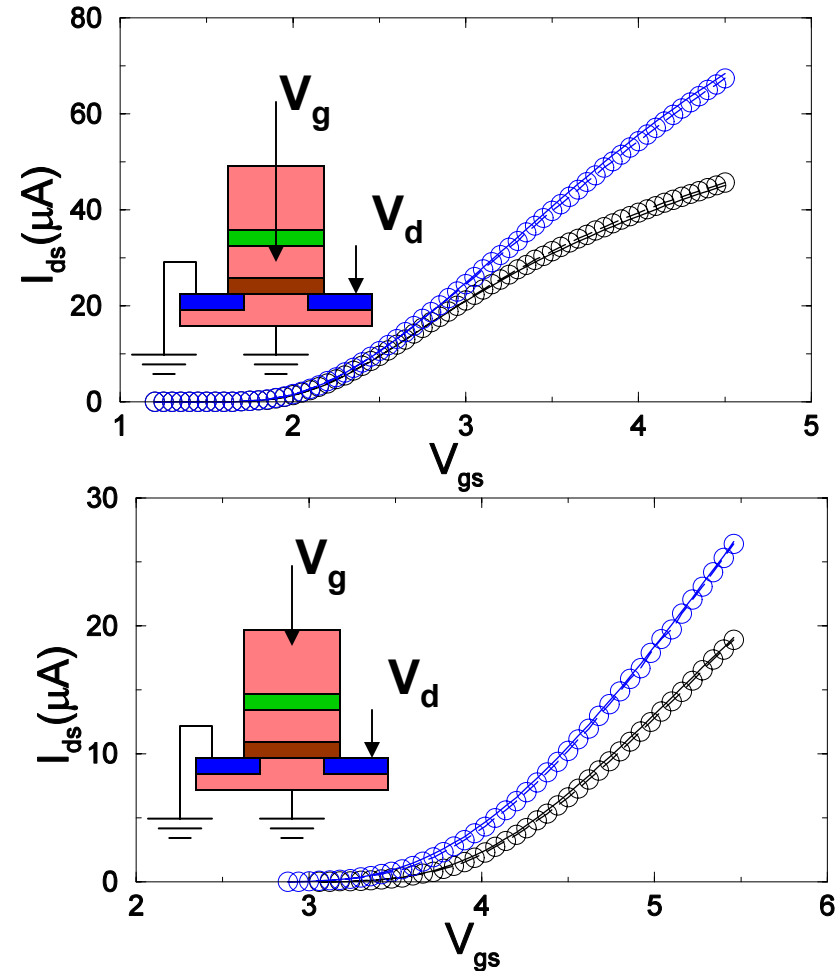
$$\alpha_g = C_{cg} / (C_{cg} + C_{mos} + C_{fd} + C_{fs})$$

$$\alpha_d = C_{fd} / (C_{cg} + C_{mos} + C_{fd} + C_{fs})$$

- $V_T \sim -Q_{FG}/C_{cg} + (1/\alpha_g) V_{T,FG} - (\alpha_d/\alpha_g) V_d$
- **Charge of floating node shifts  $V_t$**
- **Drain coupling to floating gate introduces “DIBL”**
  - Typically  $\alpha_g = 0.5-0.75$  and  $\alpha_d \sim 0.1$

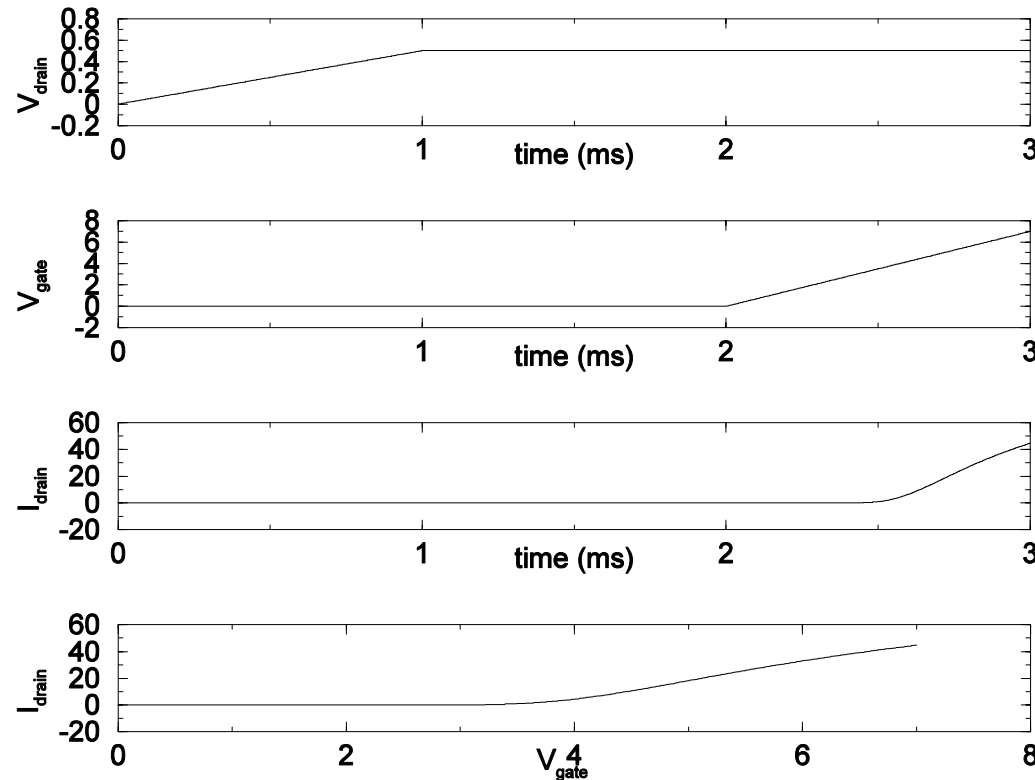
# Sense Model: Extraction

- Extract base MOSFET model by accessing floating gate
- Compare to bit-cell to obtain coupling capacitances
- Requires comparison of two devices  $\Rightarrow$  subject to mis-match errors
- Extraction with bit-cell alone (e.g. ref) = requires erase or program model



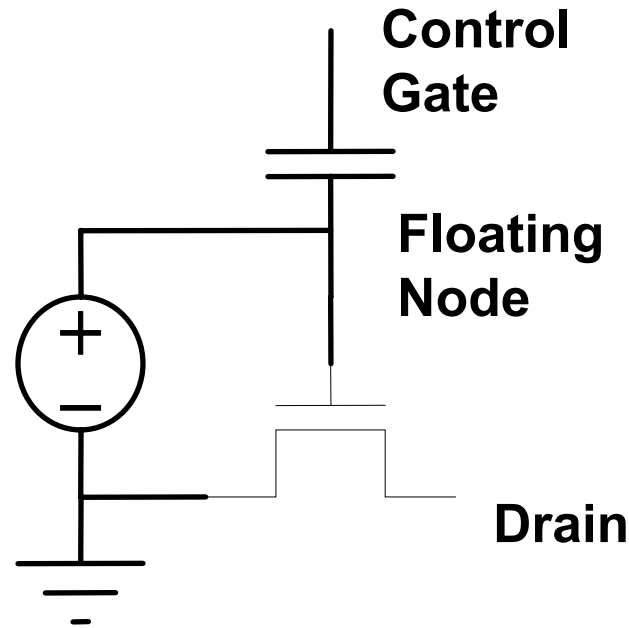
# Flash Sense Model: Use

- Model may only be used for transient simulation
- Example: Generating an Id-Vg curve
  1. Ramp Drain from 0 to Vd
  2. Ramp Gate from 0 to Vg
  3. Compute Idrain
  4. Idrain vs. Vgate
- Ramp slow enough that transient currents ( $C \, dV/dt$ )  $\sim 0$
- Not restrictive: Model used mainly for timing



# Flash Sense Model: DC Model

- May build a DC Flash model:

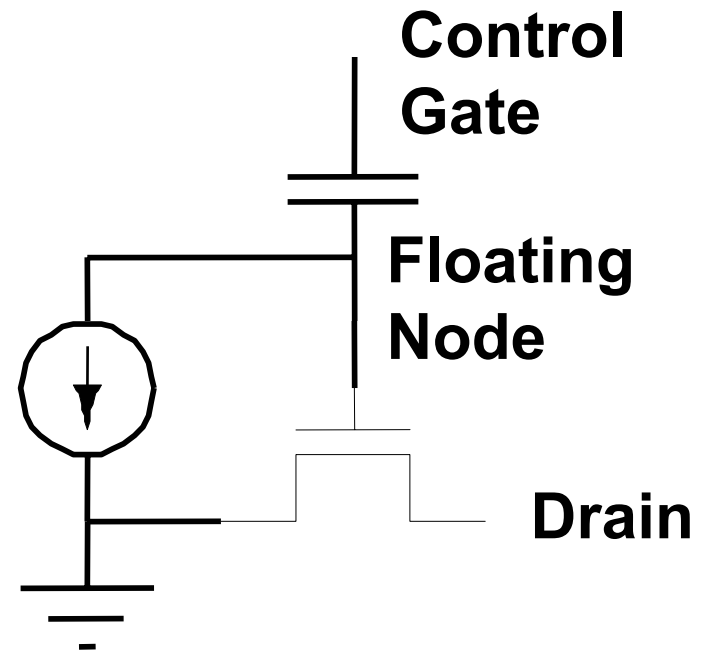


**Solve for  
Floating node  
potential for  
capacitor sub-  
circuit model**

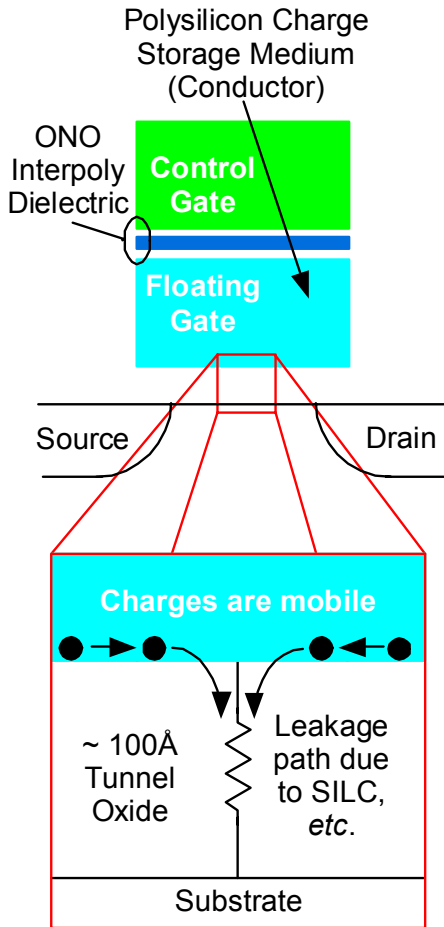
- See:
  - Y. Tat-Kwan, *et. al. IEDM Tech Dig.* p. 157 (1994)
  - L. Larcher, *et. al. IEEE Trans. Elec. Dev.*, 49 p. 301(2002)
- Voltage source sets  $V_{fg}$  such that charge  $Q_{FG}$  is conserved

# Flash Program/Erase Model

- **Multiple Time scales:**
  - Read ~ 10 ns
  - Program ~ 1  $\mu$ s
  - Erase ~ 100 ms
  - Retention/Read Disturb ~ 10 Years
- Read  $\Rightarrow$  tightest timing, so most need for circuit simulation
- Program/Erase  $\Rightarrow$  May need a circuit model (multi-level storage)
- Most models add non-linear resistor or current source

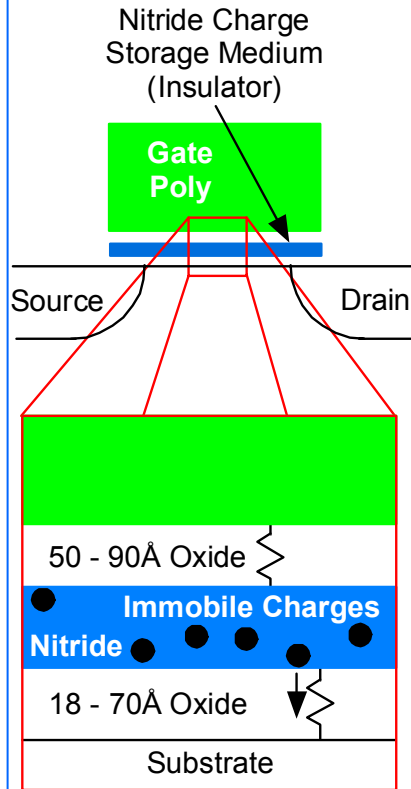


# Floating Gate and Discrete Trapping NVM

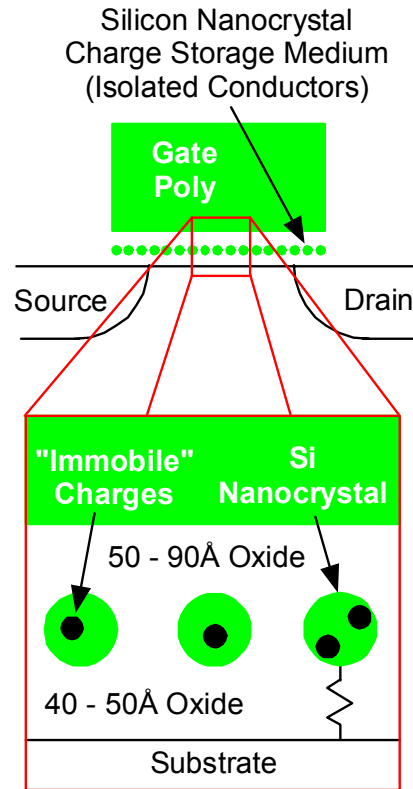


**Floating Gate**

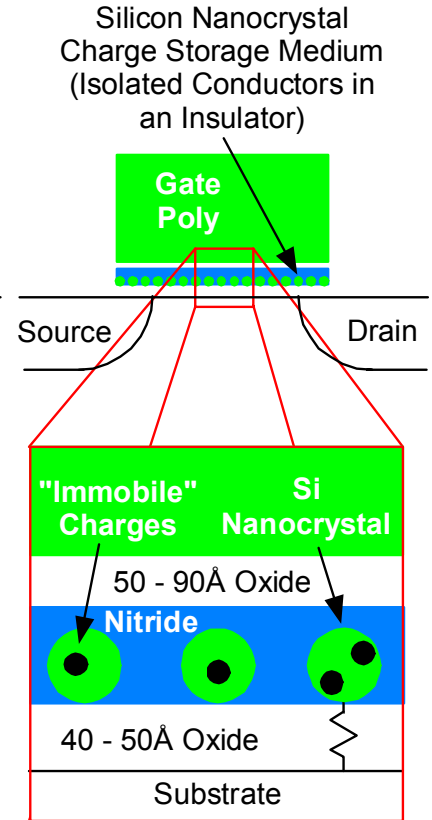
## Discrete Trapping Memories



**SONOS**

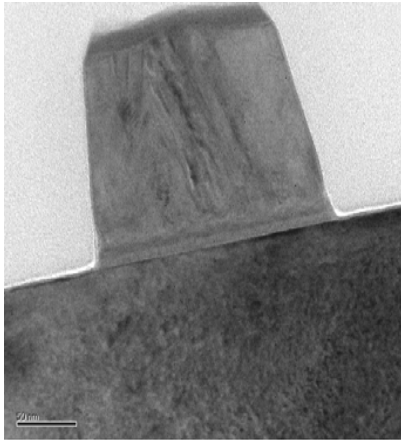


**Nanocrystal NVM**

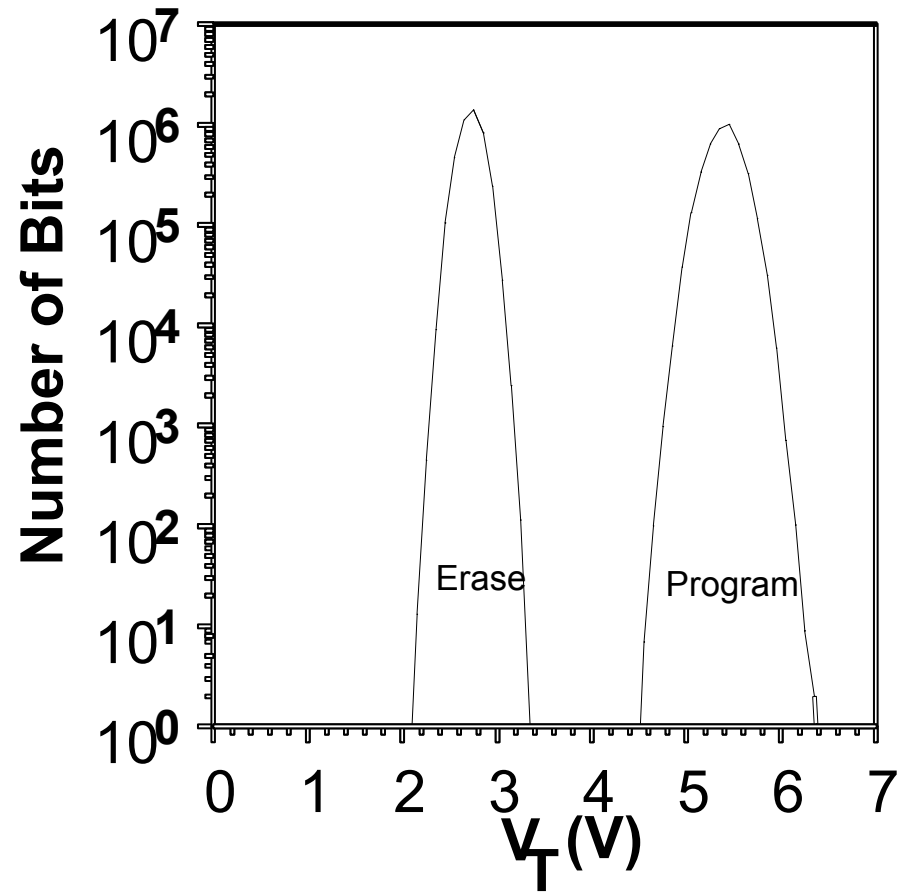
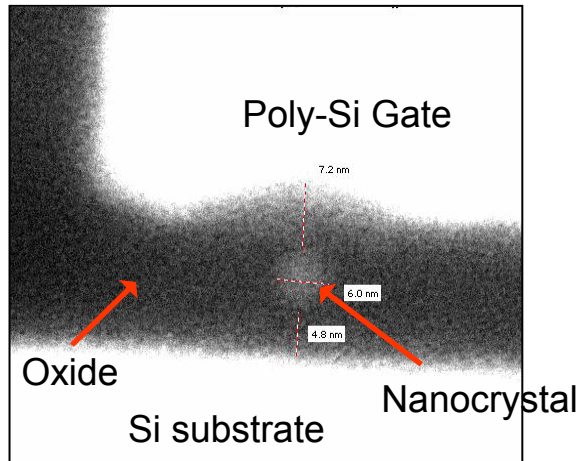


**Hybrid**

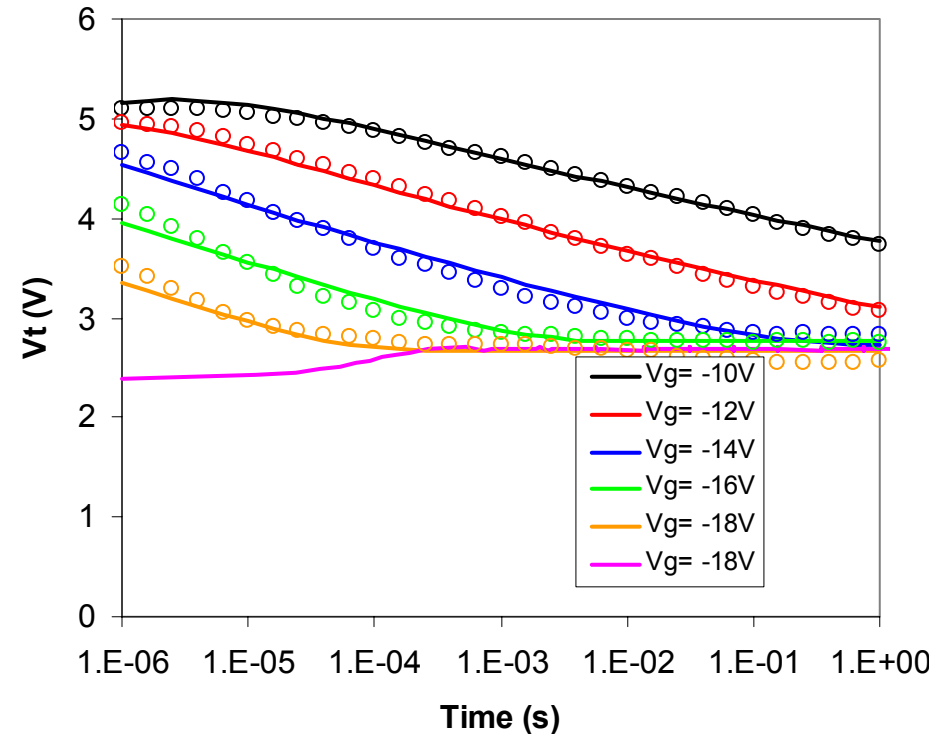
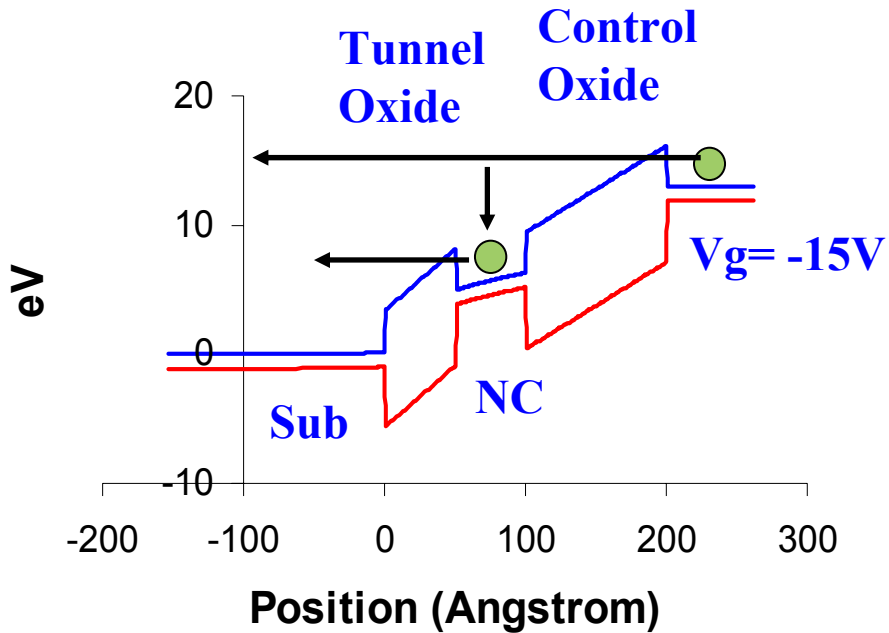
# 4Mb Memory Array



Flash array gate etch process



- **4Mb Nanocrystal Memory arrays fabricated using 90nm CMOS process technology**

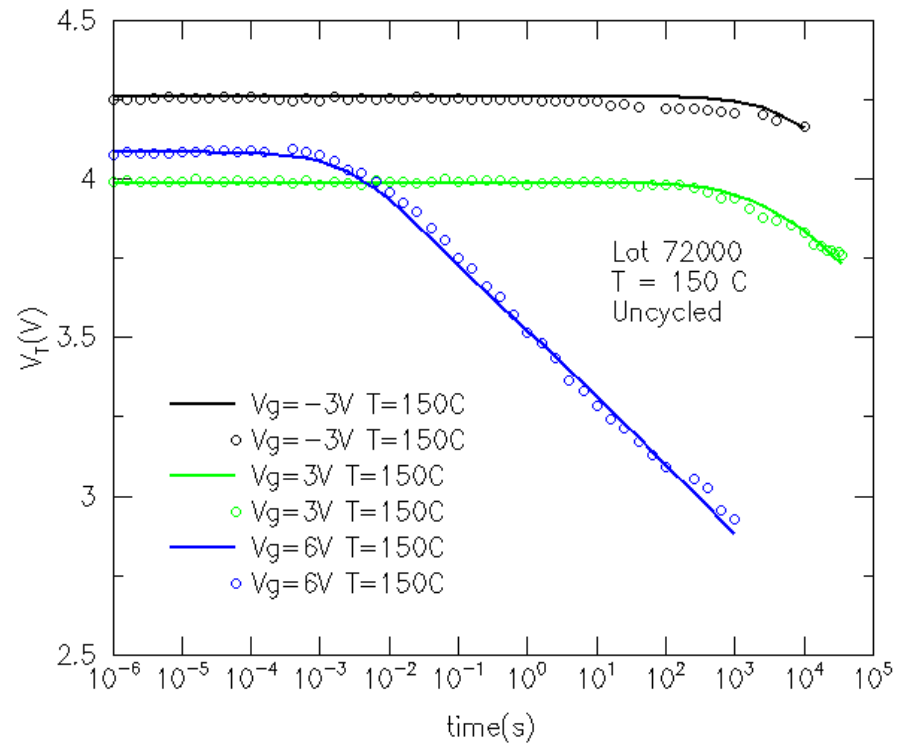
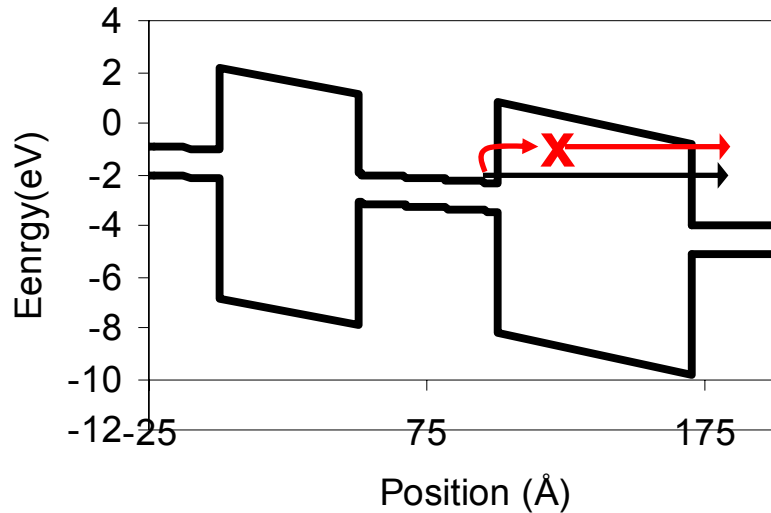


- **WKB Tunneling current from nanocrystal, field dependent gate injection current and Coulomb blockade**
- **Model matches experimental results**



# Nanocrystal Memory: Read disturb of Program State

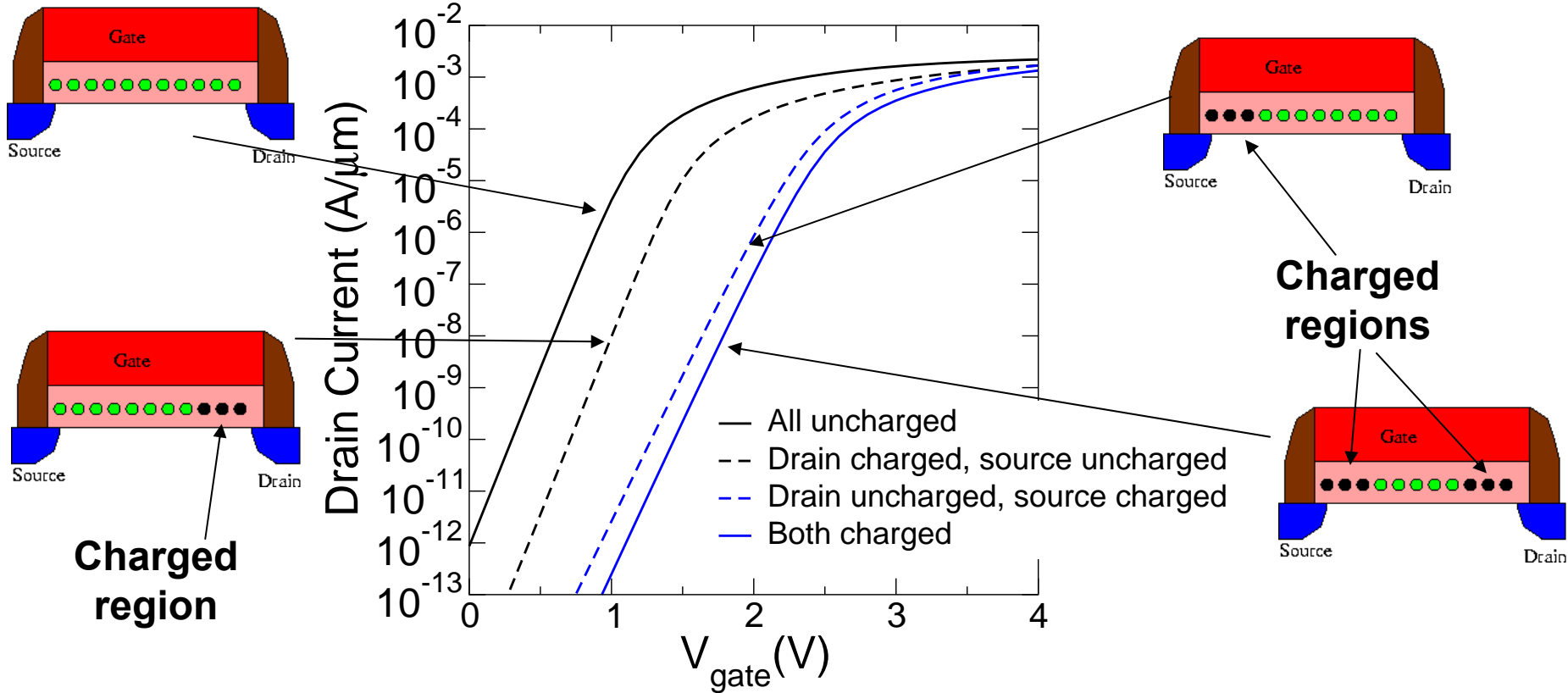
## Additive temperature activated component:



- $\uparrow$ Temp  $\uparrow$  Accessible Defects
- Possible extrinsic effects
- Model matches experimental results

# Non-uniform Charge Storage

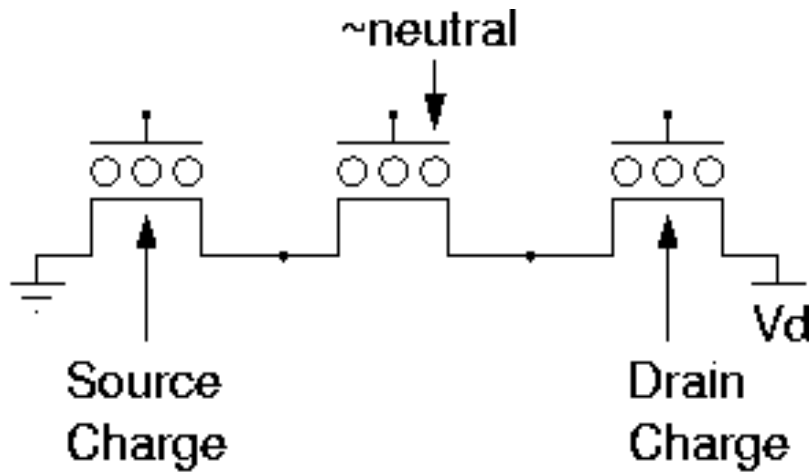
Hot-carrier injection stores charge locally near diffusions:



$V_T$  a function of charge location  $\Rightarrow$  basis of 2-bit/cell

# Charge-Trapping NVM: 2 Bit Storage

- A simple circuit model:



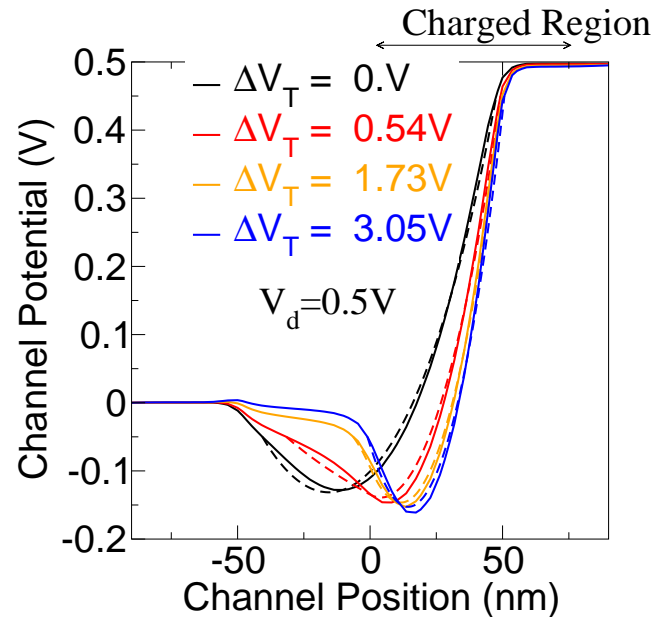
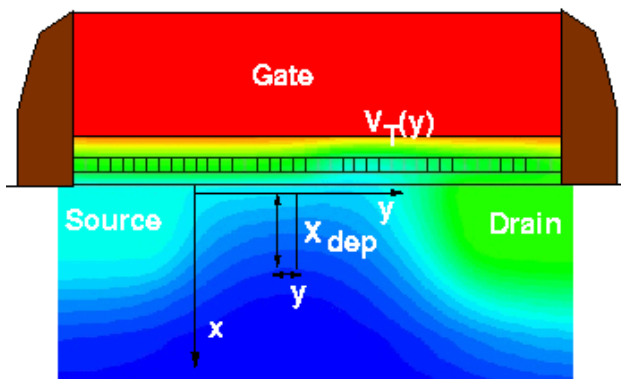
Forward $V_t$	Reverse $V_t$	State
High	High	11
High	Low	10
Low	High	01
Low	Low	00

# Quasi-2D Approximation for Sub-threshold

Approximate 2-D Poisson equation including  $\Delta V_t(y) \propto \sigma(y)$ :

$$\ell^2 \frac{\partial^2 \psi_s(y)}{\partial^2 y} + V_{gs} - V_{t0} - \Delta V_t(y) + 2\phi_f - \psi_s(y) = 0$$

$V_G$  when minimum at  $2\phi_F \Rightarrow V_T$



Solid  $\Rightarrow$  Numerical device simulation

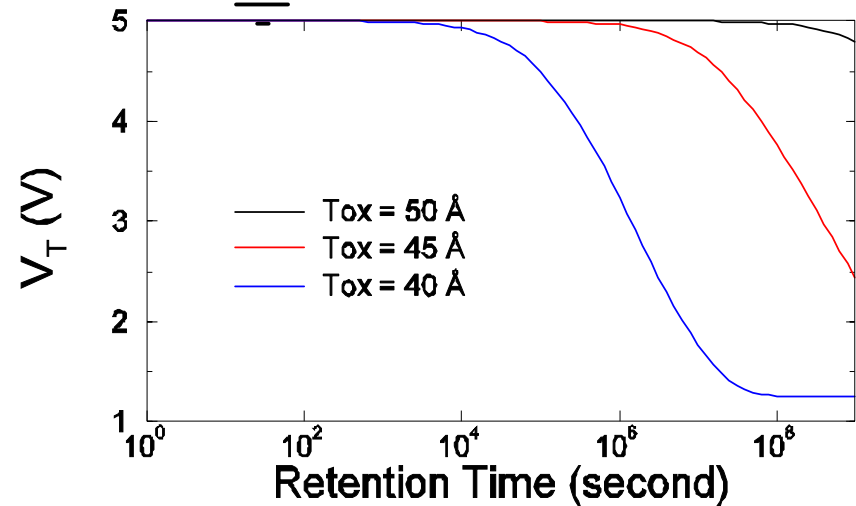
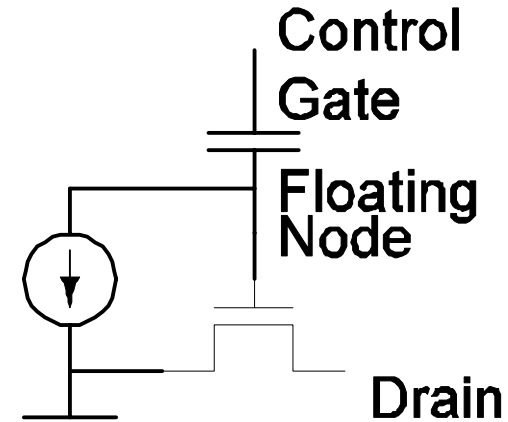
Dashed  $\Rightarrow$  Quasi-2D model

# Reliability Model: For Retention and Read Disturb

- **Non-linear current source**  $\Rightarrow$  model charge loss:
- Integrate in  $\log(t)$
- $dQ/d(\log(t)) = t \, dQ/dt = t \, I_{\text{tunnel}}(V)$

$\Rightarrow$  May calculate long-time loss:

- **Physics of charge loss (tunneling) is lumped into the non-linear current source**



- **Capacitor sub-circuit** ⇒ **foundation for flash model**
- **Appropriate for timing simulation**
- **May be augmented to model:**
  - **Program and erase**
  - **Vt drift due to P-E cycling**
  - **Reliability (charge loss or gain)**
  - **Device asymmetry (2-bit storage)**

**Michael Sadd: M.Sadd@freescale.com**

**Rajesh Rao: Rajesh.Rao@freescale.com**

**R. Muralidhar: RA4479@freescale.com**