Dynamic Voltage Allocation Based on Mutual Information For NAND Flash Memory

Presenter: Adam Williamson

Additional Contributors: Tsung-Yi Chen, Kasra Vakilinia, Tong Zhang, Richard Wesel

CODESS Workshop, September 19, 2013

Outline

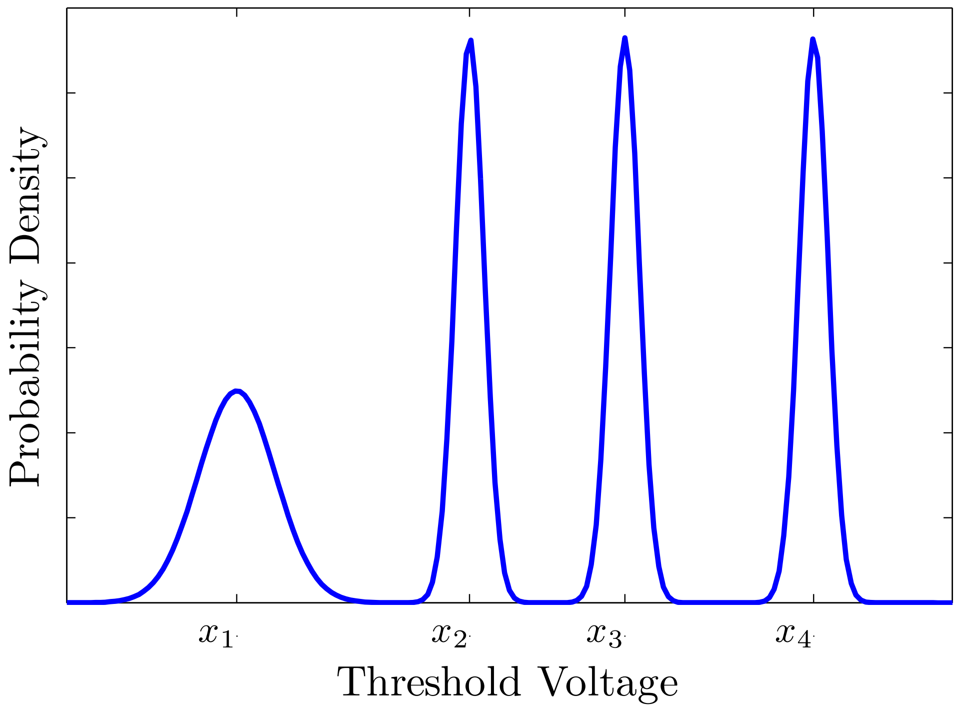
- A three-component noise model
- Learning the channel from enhanced precision
- Dynamic voltage allocation approach

Outline

- A three-component noise model
- Learning the channel from enhanced precision
- Dynamic voltage allocation approach

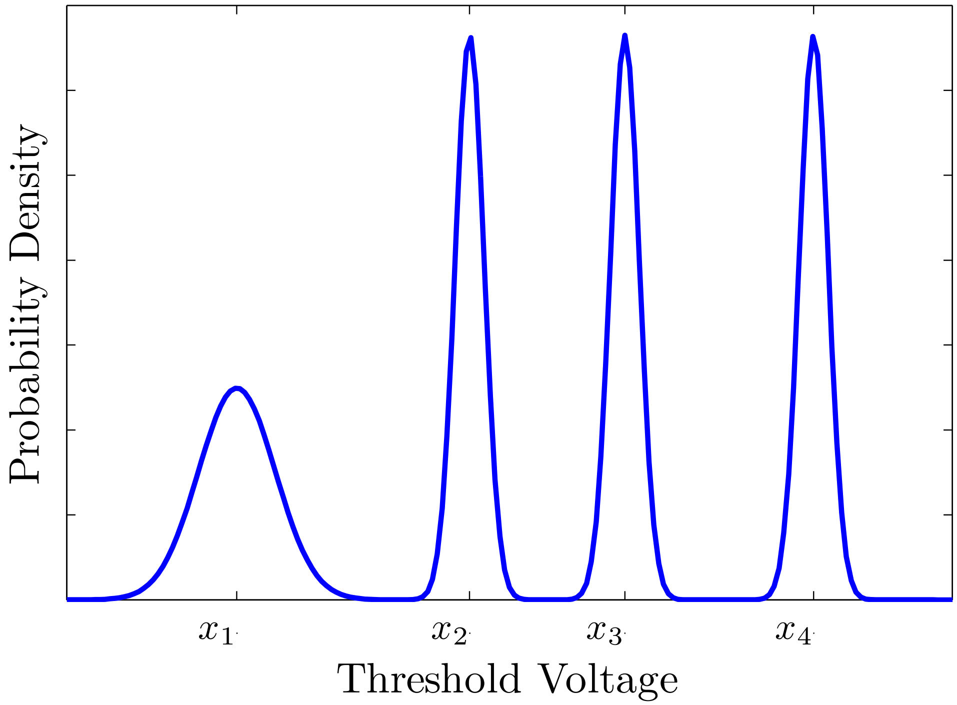
Programming Noise

(a) $f(x+n_p)$



- **Programming noise**: The baseline noise seen when writing to a new, unused cell and read it back immediately.

Programming Noise Model


(a) $f(x+n_p)$

$$f_{N_p}(n_p) \sim \begin{cases} \mathcal{N}(0,\sigma_p^2) & \text{if } x > 0 \\ \mathcal{N}(0,\sigma_e^2) & \text{if } x = 0 \end{cases} \quad \text{where } \sigma_e^2 > \sigma_p^2$$

Programming Noise Model

- [Compagnoni et al. 2008]

"Ultimate Accuracy for the NAND Flash Program Algorithm due to the Electron Injection Statistics", IEEE Trans. Electron Devices, Oct. 2008.

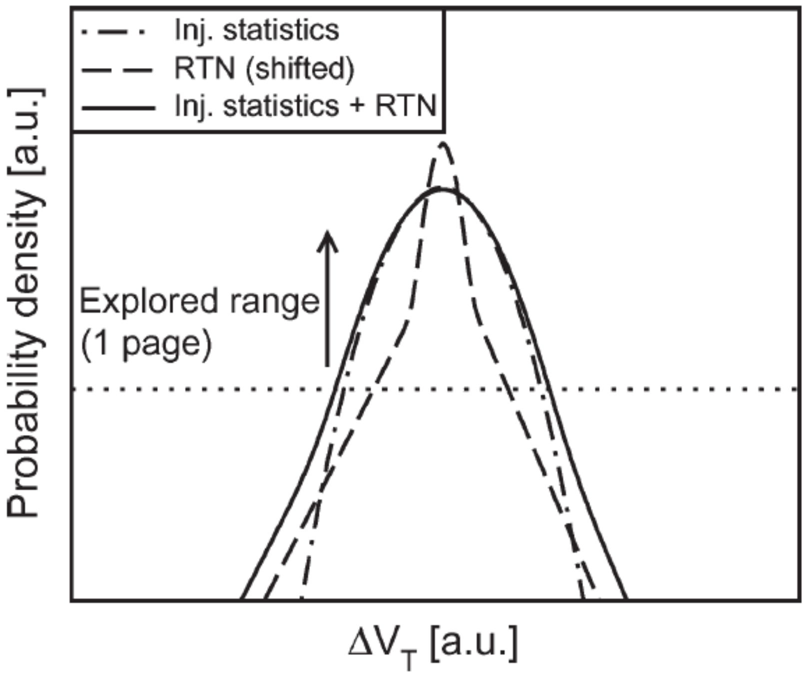
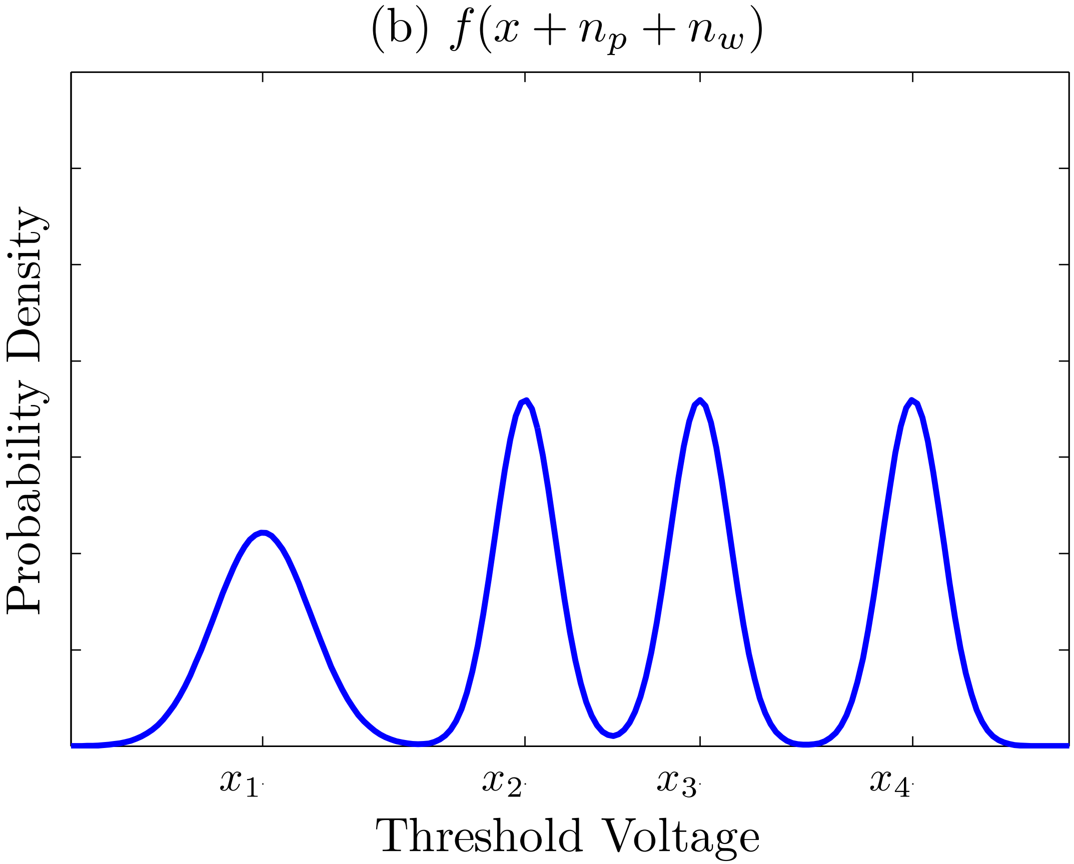


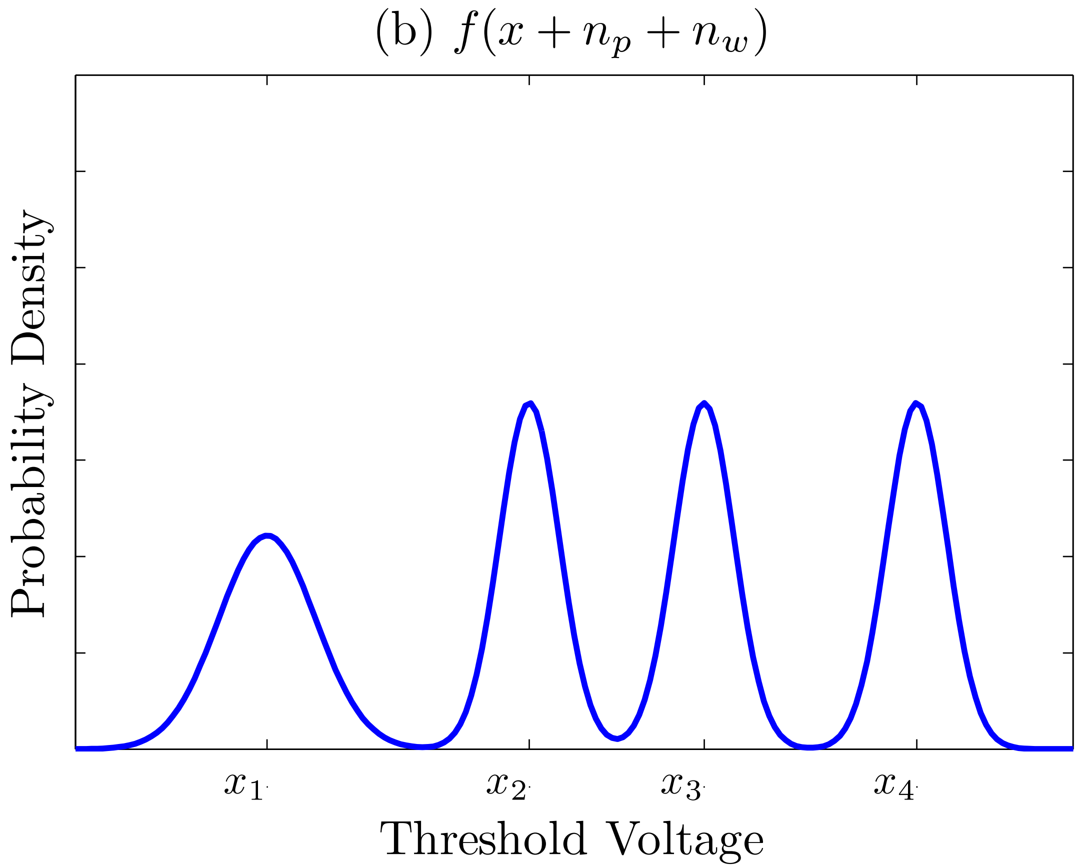
Fig. 5. Probability distribution of $\Delta V_{T,\mathrm{rtn}}$, of ΔV_T in presence of the injection statistics and including both the RTN instability and the injection spread. The $\Delta V_{T,\mathrm{rtn}}$ distribution has been shifted to the same average value of the ΔV_T distribution.

Wear-out Noise



- **Wear-out noise**: Additional noise seen when reading back immediately, which becomes worse as the cell has more charge programmed and erased.

Wear-out Noise Model


(b) $f(x+n_p+n_w)$

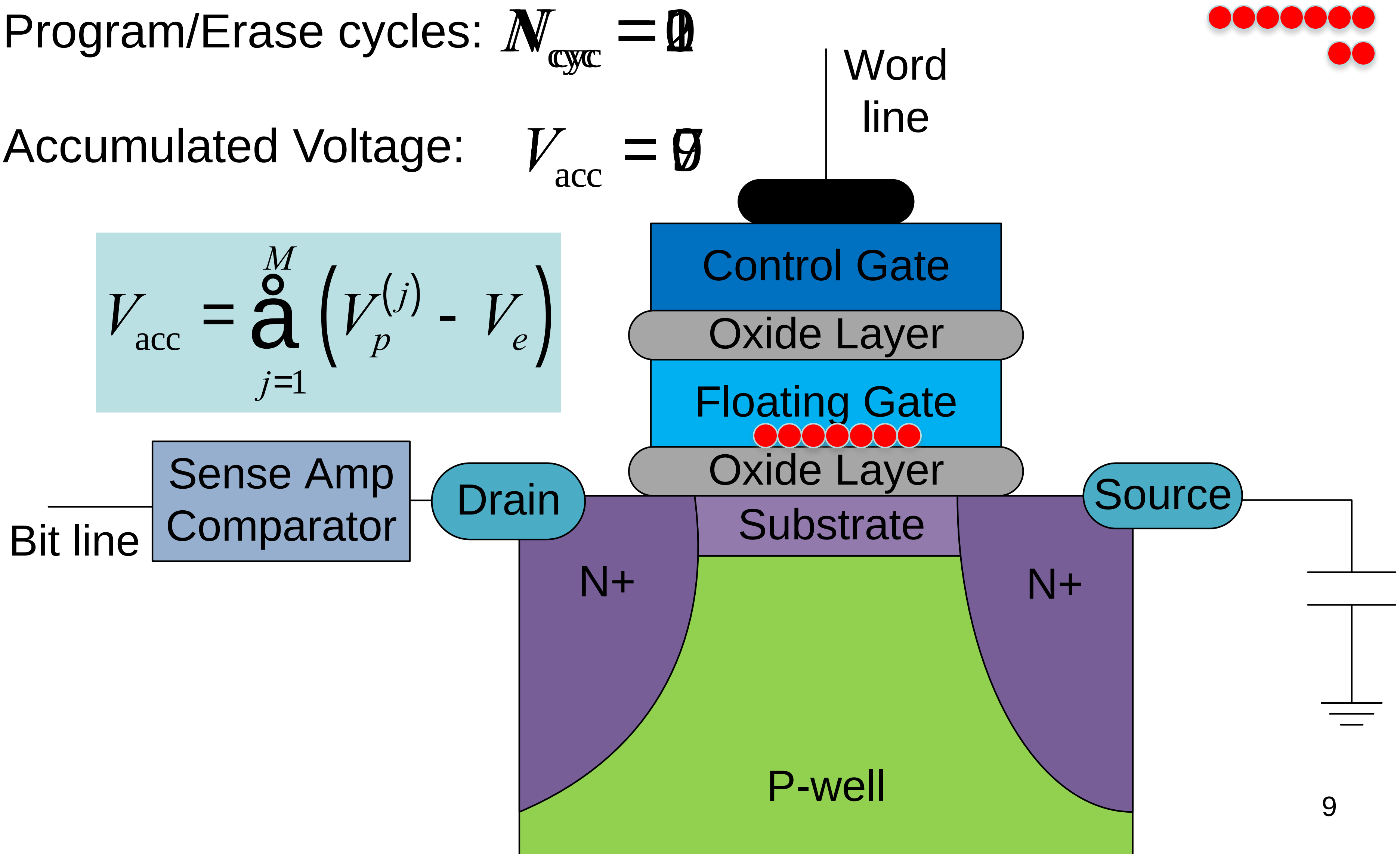
$$f_{N_w}(n_w) = \frac{1}{2\lambda} e^{-\frac{|n_w|}{\lambda}} \qquad \lambda = C_w + A_w \left(\frac{V_{\mathrm{acc}}}{V_{\mathrm{max}}}\right)^{0.62}$$

P/E cycles vs. Accumulated Voltage

Program/Erase cycles: $N_{\text{cyc}} = 0$

Accumulated Voltage: $V_{\text{acc}} = 0$

$$V_{\text{acc}} = \sum_{j=1}^{M} \left(V_p^{(j)} - V_e \right)$$



Wear-out Noise Model

- [Fukuda et al. 2007]

“Random telegraph noise in flash memories – model and technology scaling”, 2007 IEEE Int’l Electron Devices Meeting.

$$f(\Delta Vt_{trap}) = \frac{1}{\sigma}\cdot\exp\left(-\frac{\Delta Vt_{trap}}{\sigma}\right) \qquad (2)$$

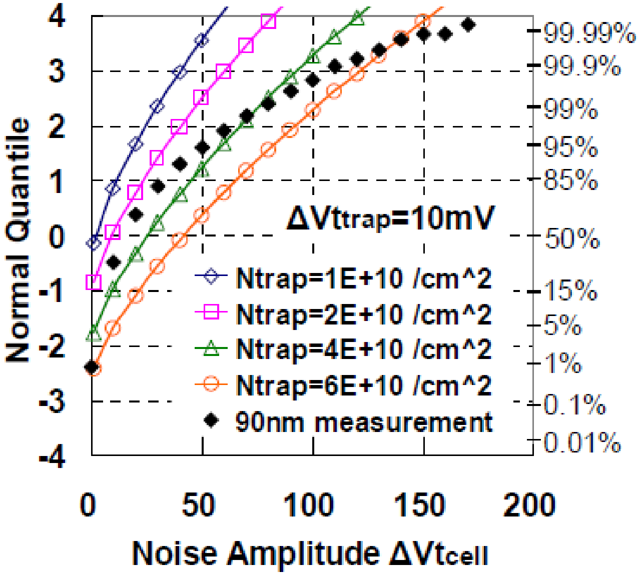
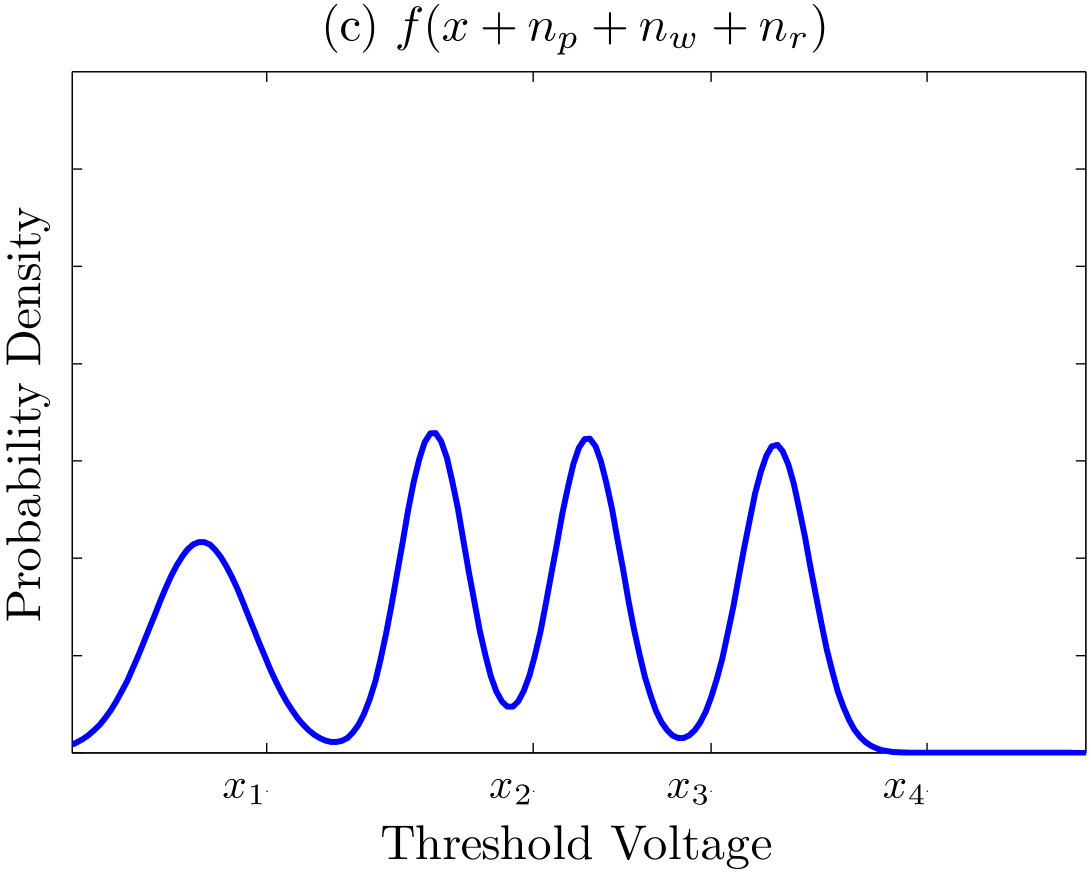


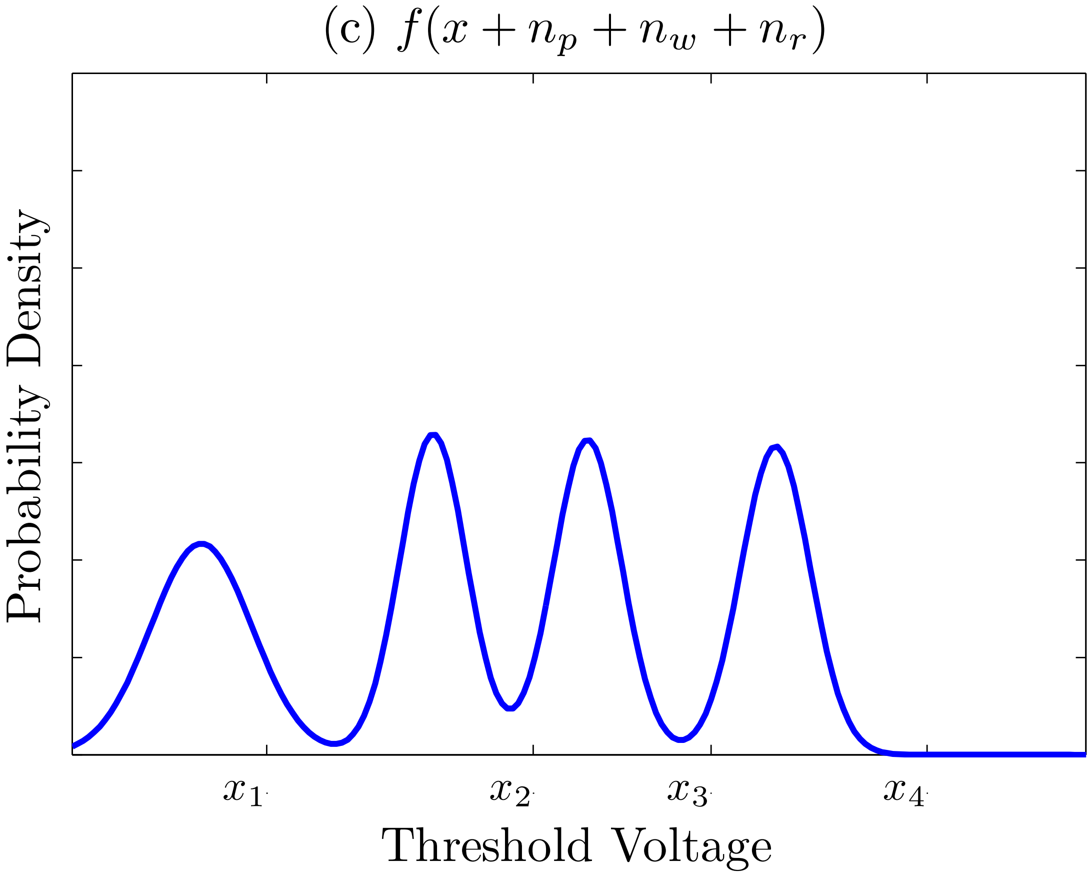
Fig. 2: Measured noise-amplitude distribution of a 90nm NAND flash memory technology. Also shown are noise distributions calculated for various interface trap densities assuming each trap contributes 10mV, Eq.(1).

Retention Noise



- **Retention noise**: the degradation of the threshold voltage integrity due to charge leakage after it is written.

Retention Noise



$$f_{N_r}(n_r) \sim \mathcal{N}(\mu_r, \sigma_r^2)$$

Retention Noise Model

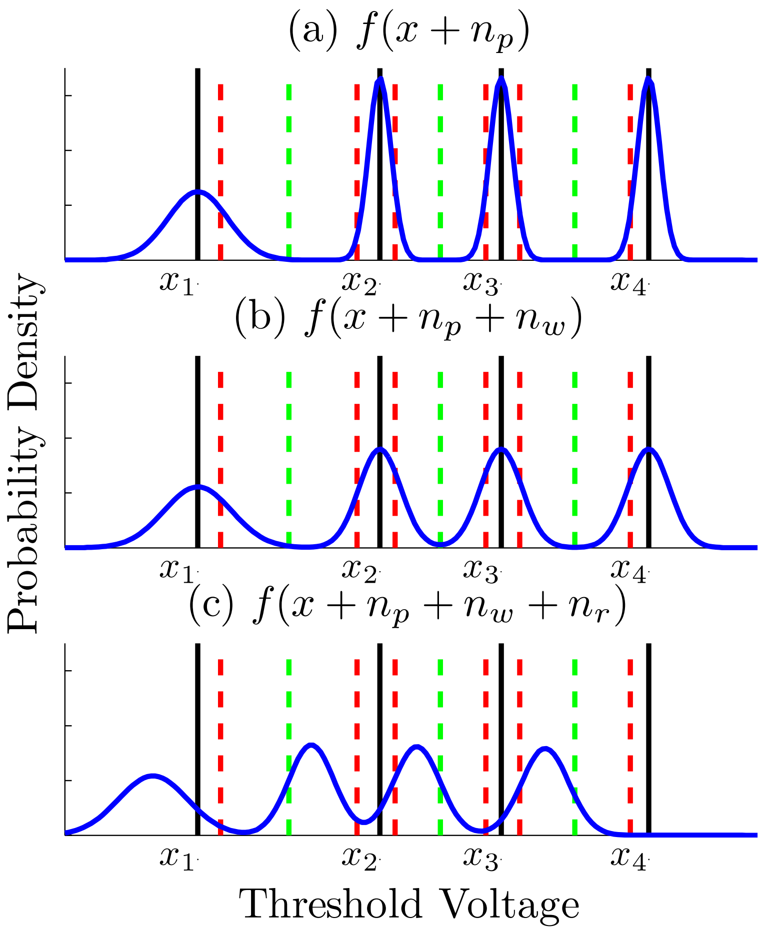
$$m_r\left(x,t,V_{\mathrm{acc}}\right) = -x\ln\left(1+\frac{t}{t_0}\right)\left[A_r\left(\frac{V_{\mathrm{acc}}}{V_{\max}}\right)^{0.62} + B_r\left(\frac{V_{\mathrm{acc}}}{V_{\max}}\right)^{0.3}\right]$$

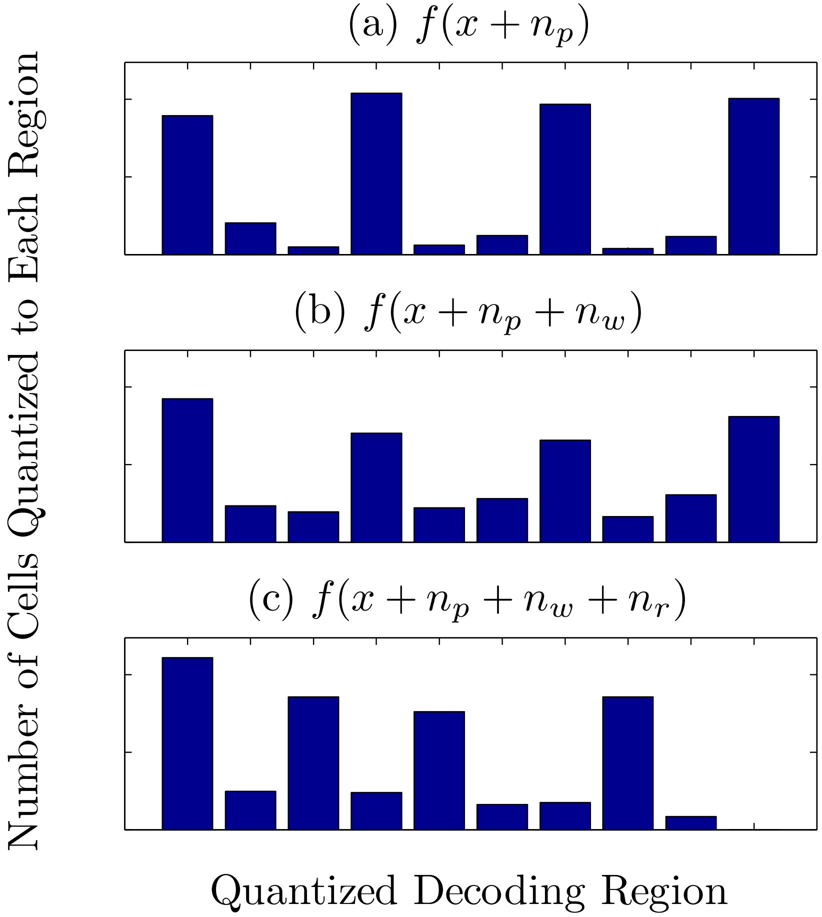
$$s_r^2\left(x,t,V_{\mathrm{acc}}\right) = (0.1)x\ln\left(1+\frac{t}{t_0}\right)\left[A_r\left(\frac{V_{\mathrm{acc}}}{V_{\max}}\right)^{0.62} + B_r\left(\frac{V_{\mathrm{acc}}}{V_{\max}}\right)^{0.3}\right]^2$$

Outline

- A three-component noise model
- Learning the channel from enhanced precision
- Dynamic voltage allocation approach

(Limited) Enhanced Precision Reveals Channel



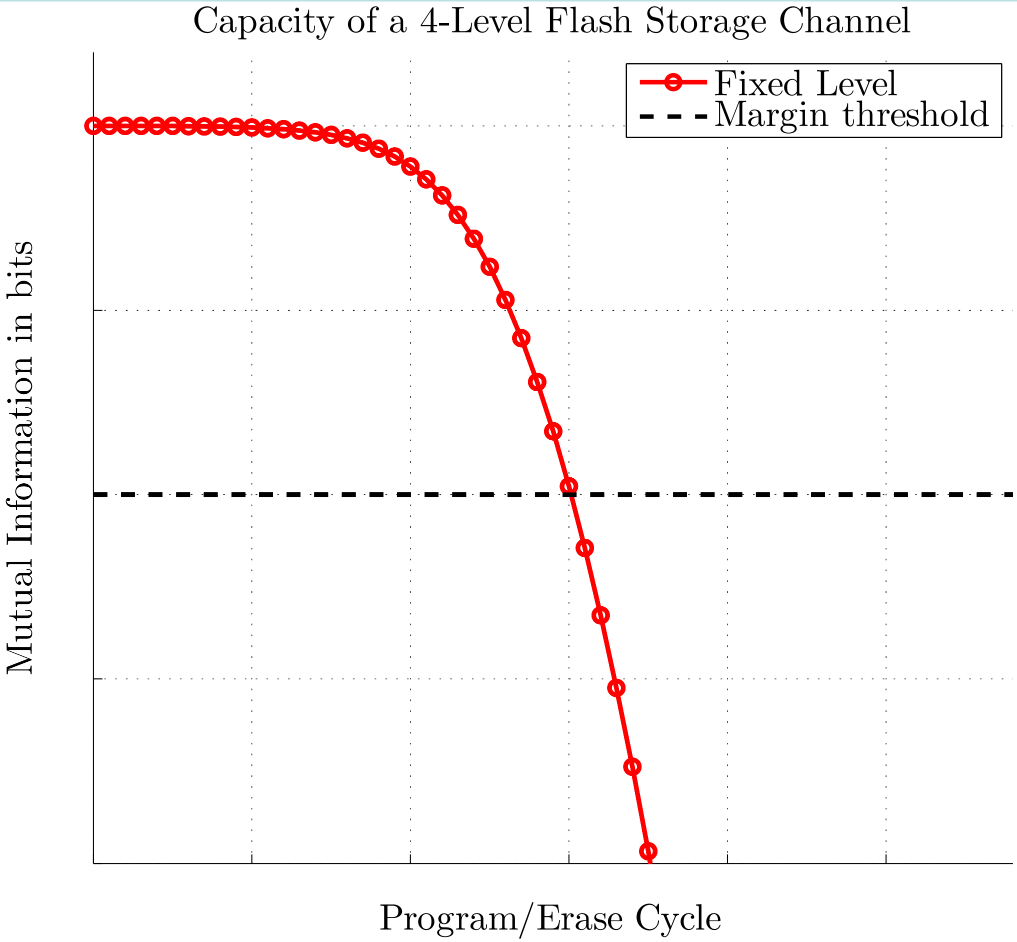


Outline

- A three-component noise model
- Learning the channel from enhanced precision
- Dynamic voltage allocation approach

Computation of Capacity

$$f_{Y|X}(y|x) = \frac{e^{\sigma^2/2\lambda^2}}{2\lambda}\left[\exp\left(\frac{y-\mu}{\lambda}\right)Q\left(\frac{y-\mu}{\sigma}+\frac{\sigma}{\lambda}\right)+\exp\left(\frac{-y+\mu}{\lambda}\right)Q\left(\frac{-y+\mu}{\sigma}+\frac{\sigma}{\lambda}\right)\right]$$

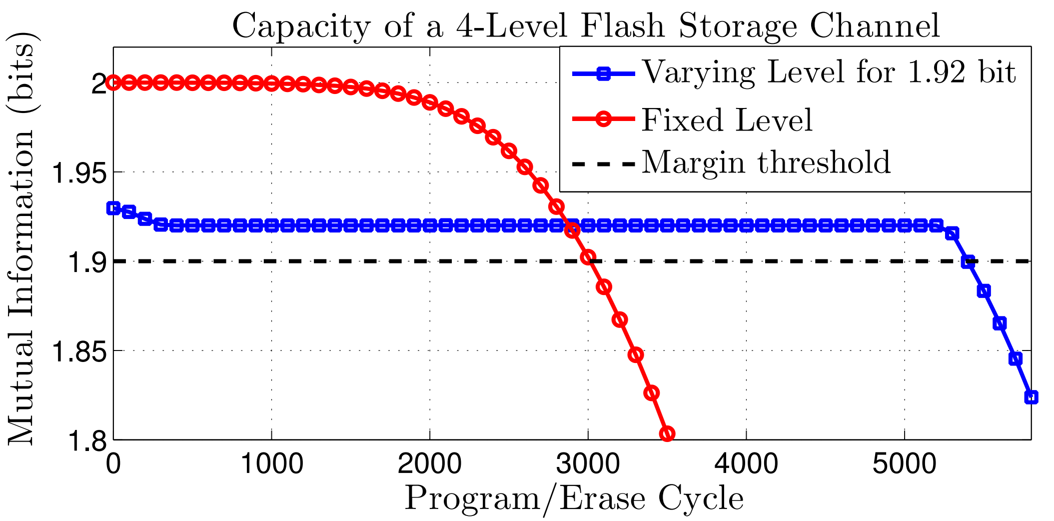


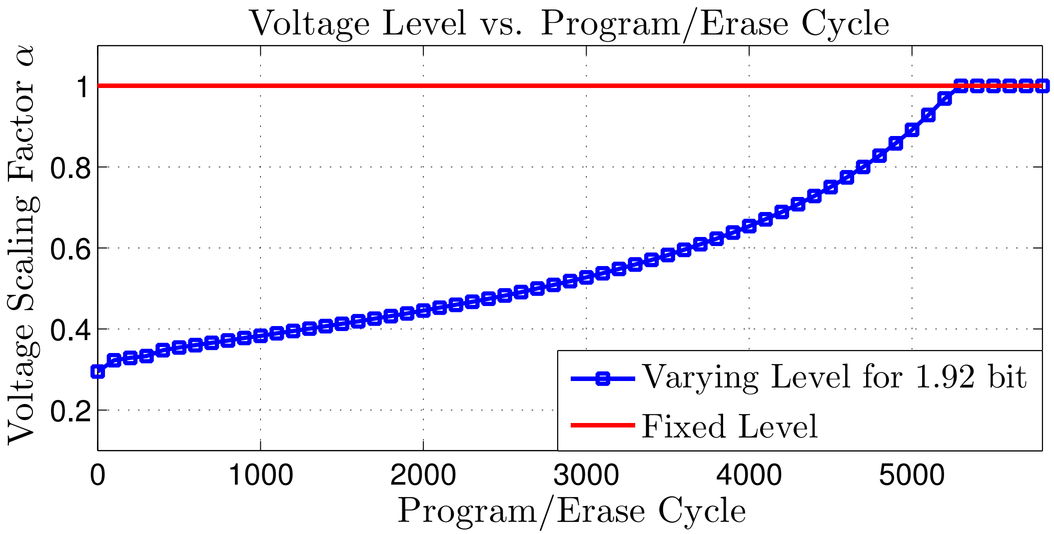
Setting Voltage Levels to Achieve Needed Capacity

- We can use information theory to **set the code rate** to match the capacity provided by the current channel conditions and the standard voltage levels.

- A much better idea is to **set the voltage levels** to provide the capacity needed to met the specified storage levels.

Dynamic Voltage Allocation





Conclusion

- Channel degradation follows total **number of electrons** pushed through the floating gate rather than the number of P/E cycles.
- Channel can be learned by occasionally using increased precision, which is done anyway to improve decoding
- This can mitigate the mean-shift portion of retention loss.
- Voltage levels can be **dynamically increased** to track the degrading channel.