

A 3bit/Cell 32Gb NAND Flash Memory at 34nm with 6MB/s Program Throughput and with Dynamic 2b/Cell Blocks Configuration Mode for a Program Throughput Increase up to 13MB/s

Speaker :

Naso Giovanni – Micron Flash Design Center Avezzano Italy

ISSCC 2010 paper 24.7

Design team

Micron Flash Design Center Avezzano (Italy) :

L. Botticchio, C. Cerafogli, P. Conenna, A. D'Alessandro,
D. Di Cicco, L. De Santis, W. Di Francesco, M.L. Gallese,
G. Imondi, M. Incarnati, C. Lattaro, A. Macerola, G. G. Marotta,
V. Moschiano, C. Musilli, G. Naso, D. Orlandi, F. Paolini,
S. Perugini, L. Pilolli, G. Santin, F. Rori, M. Rossini, E. Sirizotti,
M. Tiburzi, A. Torsi, T. Vali

Micron Flash Design Center S. Jose' (California) :

R. Ghodsi, F. Roohparvar

Micron Product Engineering Boise (Idaho) : D. Rivers

Intel Flash Design Center Folsom (California) :

M. Goldman, C. Haid

Agenda

- Device features
- Dynamic bits per cell configuration
- Device architecture and organization
- Sensing
- Data flow
- VT placement
- Throughput
- Source bias
- Erase ramp
- Analog system
- Temperature compensation

Device features

Technology	34nm triple-well 3 metals
Cell size	0.0054 μm^2 (select gates included) 64 WLs per string
Chip size	126 mm^2
Organization	4096 Bytes per page x 384 pages x 684 blocks x 4 planes x 8 IO
ECC per 4KB page	224 Bytes
Array read time	60 μs typ – 100 μs max
Program time	2 ms typ – 10 ms max
Erase time	10 ms typ – 30 ms max
Clock cycle time	12 ns
Supported multiple bit/cell	1 bc – 2 bc – 3 bc (dynamic configuration)

Dynamic bits per cell configuration

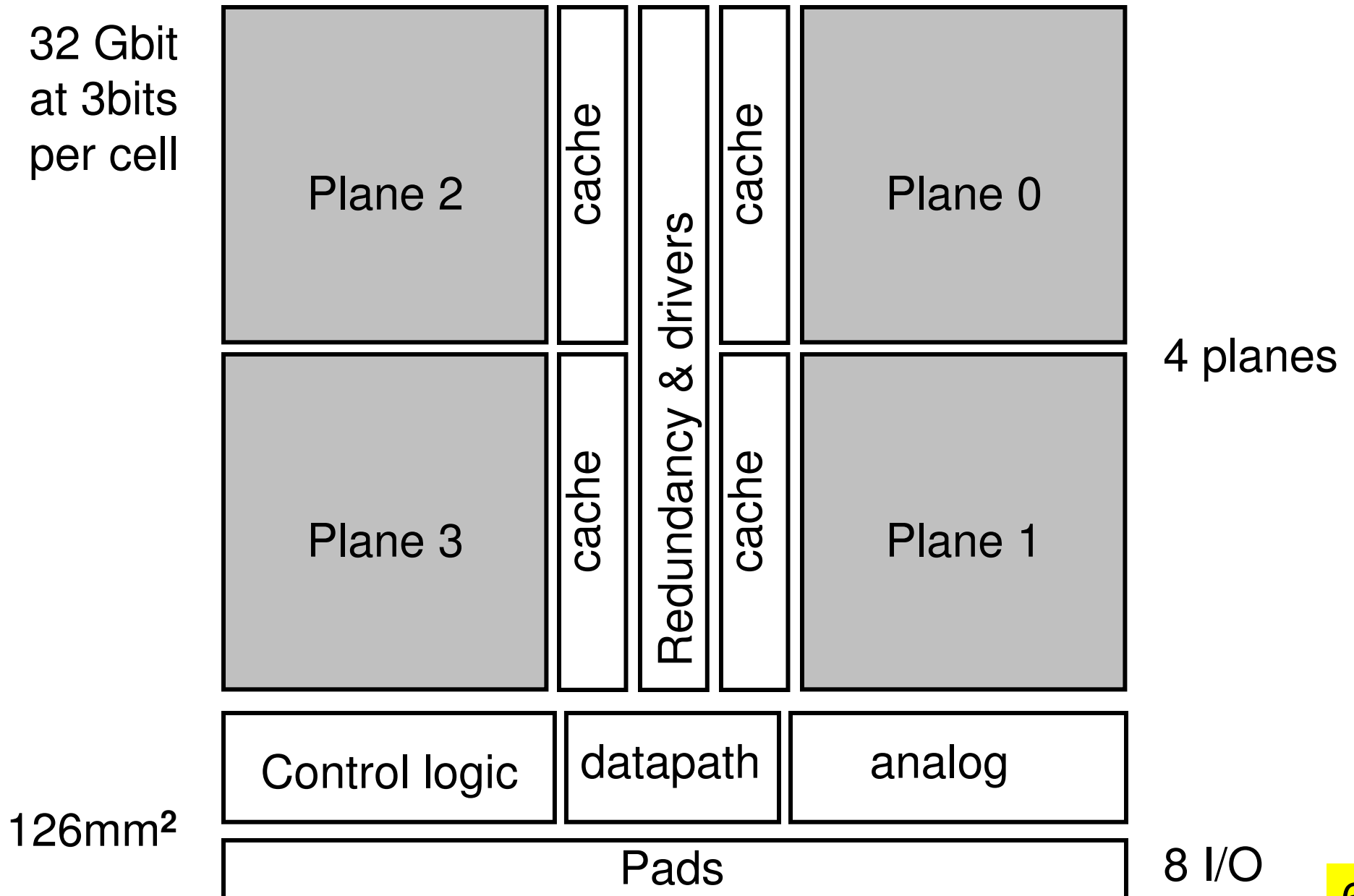
Number of bits per cell can be dynamically set by the user to 1, 2, 3 through the set feature command.

Program and read algo are optimized for the different bit per cell configurations to achieve maximum performance and margins.

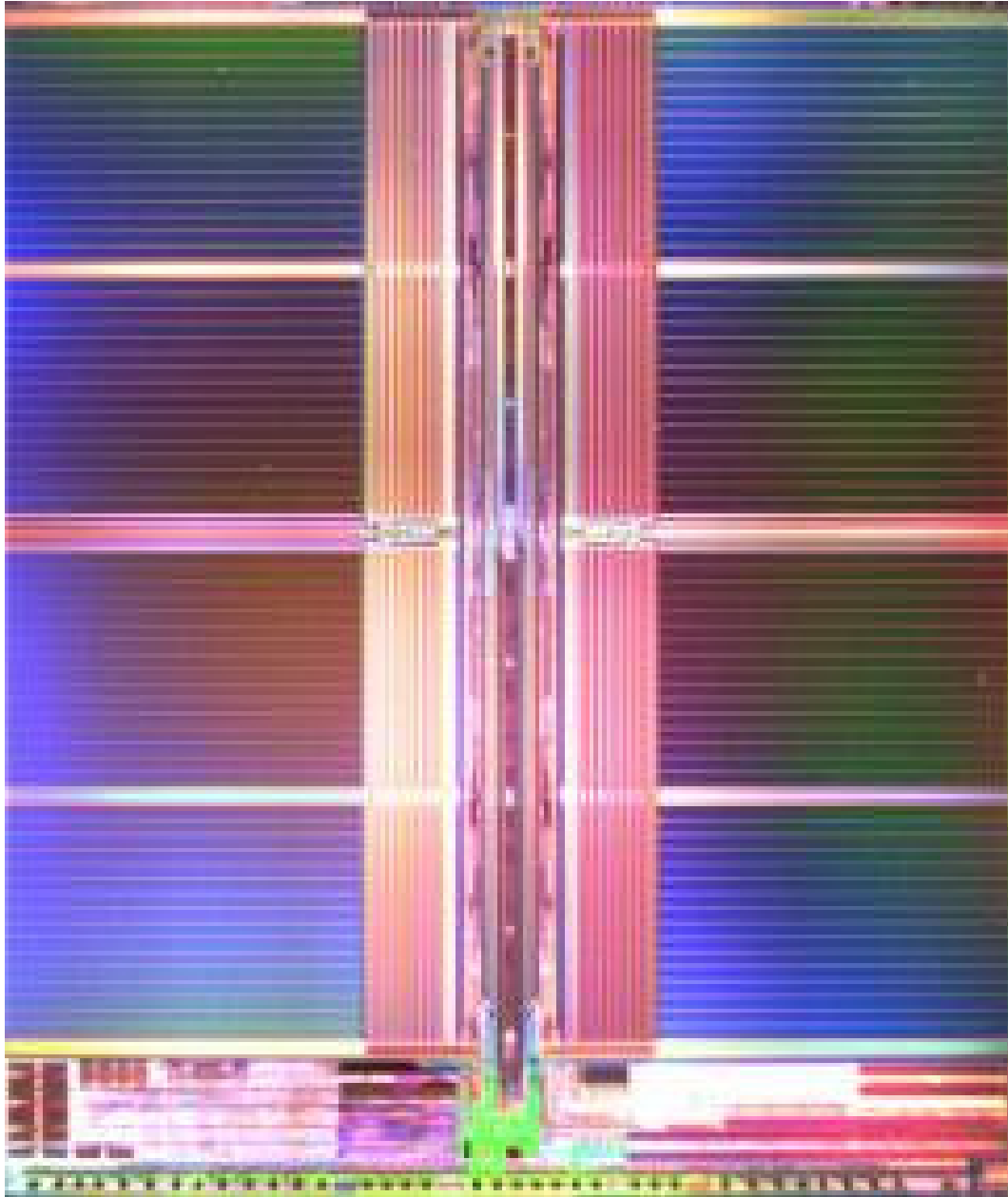
	3 bits per cell	2 bits per cell	1 bit per cell
tBERS (ms)	10 - 30	10 - 30	10 - 30
tPROG (ms)	2 - 10	1 - 2.2	0.25 - 0.9
tR (us)	60 - 100	50	30

Pages addressing is automatically adjusted based on bits per cell configuration.

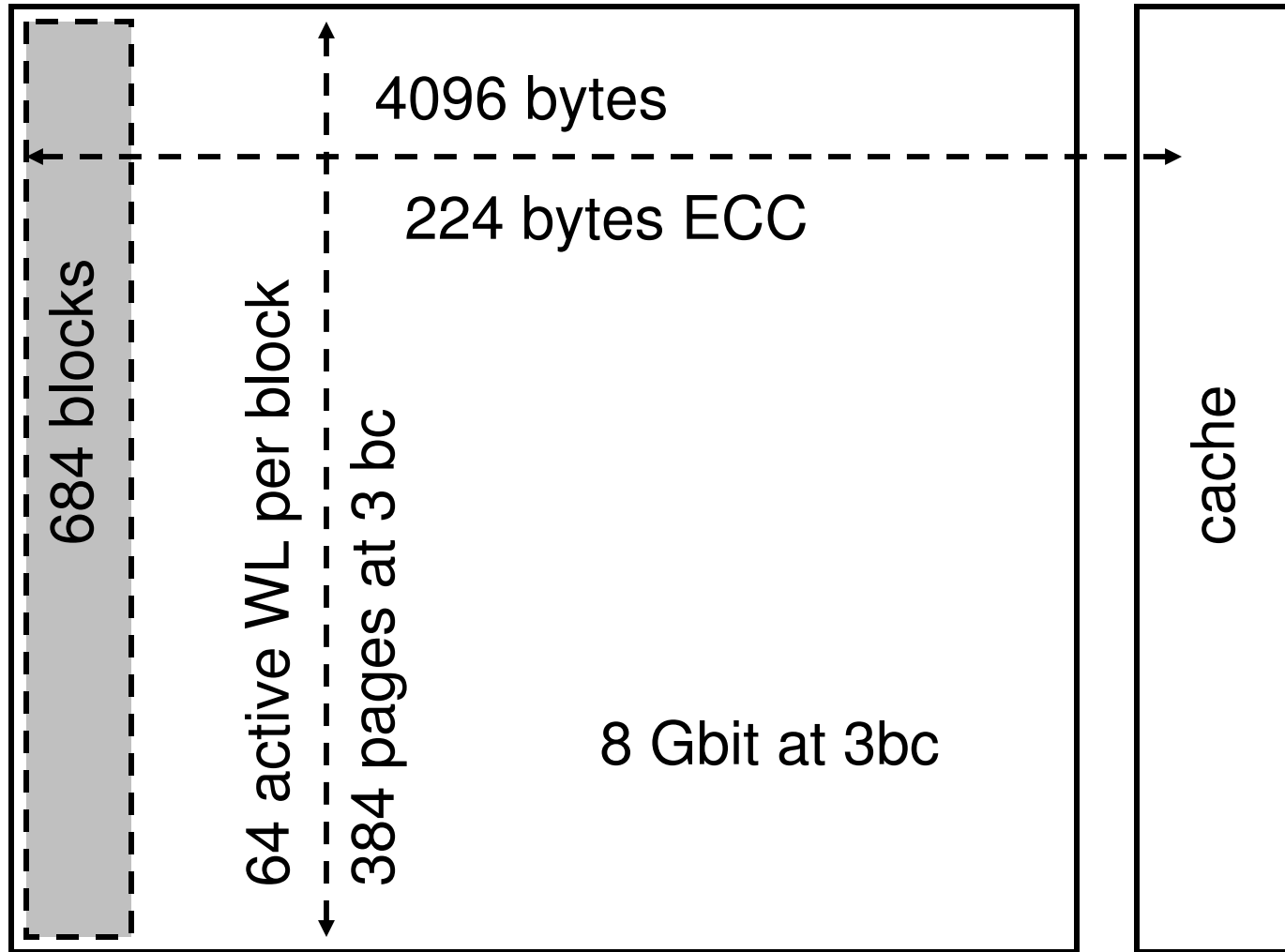
Device architecture



Device photo

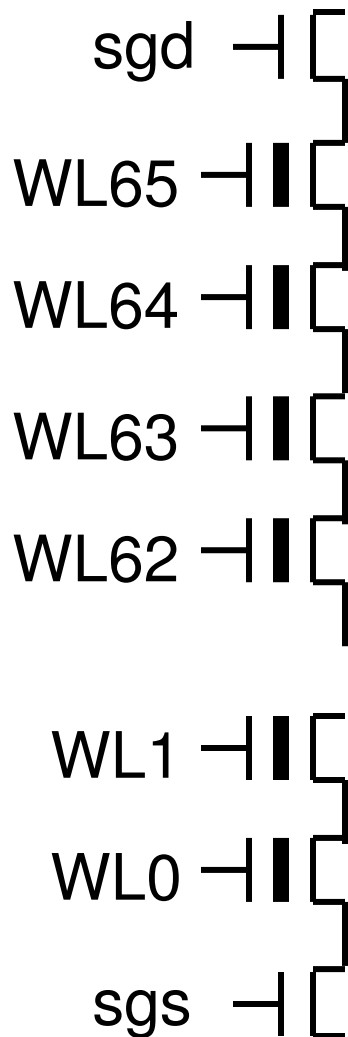


Plane organization



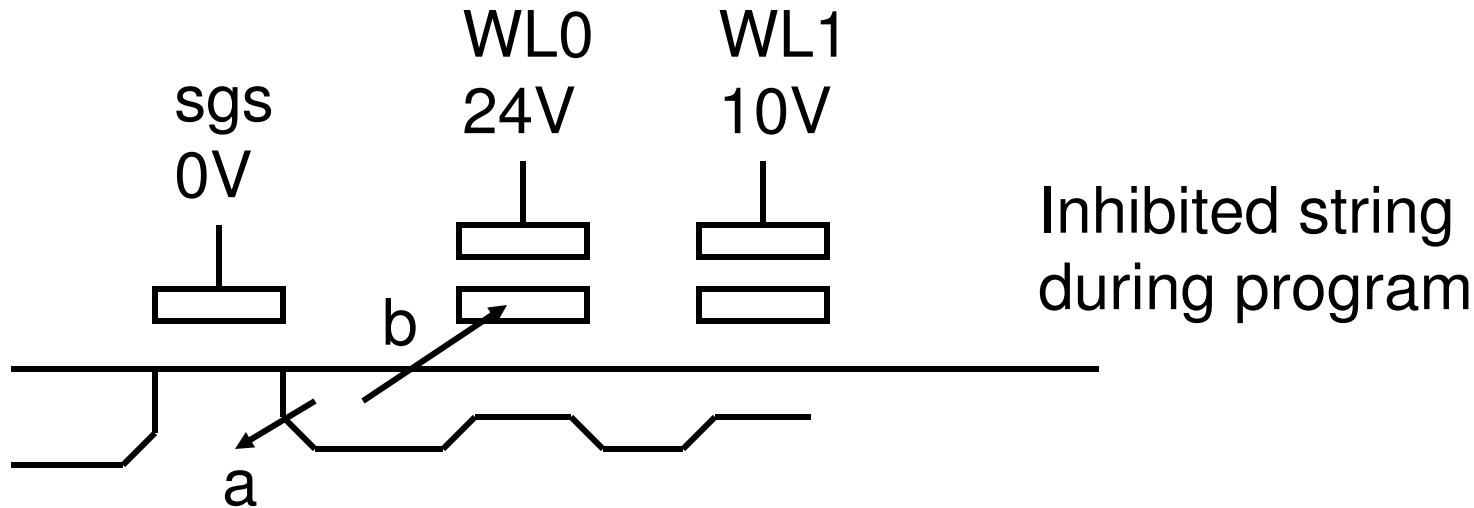
Array string structure

66 physical WLs with different options to manage 'hot carrier disturb' at the edge



WL	option1	option2	option3
65	1 bc	dummy	1 bc
64	2 bc	3 bc	1 bc
63	2 bc	3 bc	3 bc
62	3 bc	3 bc	3 bc
2	3 bc	3 bc	3 bc
1	3 bc	3 bc	3 bc
0	1 bc	dummy	1 bc
even/odd pages	384	384	384

Edge WL 'hot carrier disturb'



a = channel de-boost due to GIDL

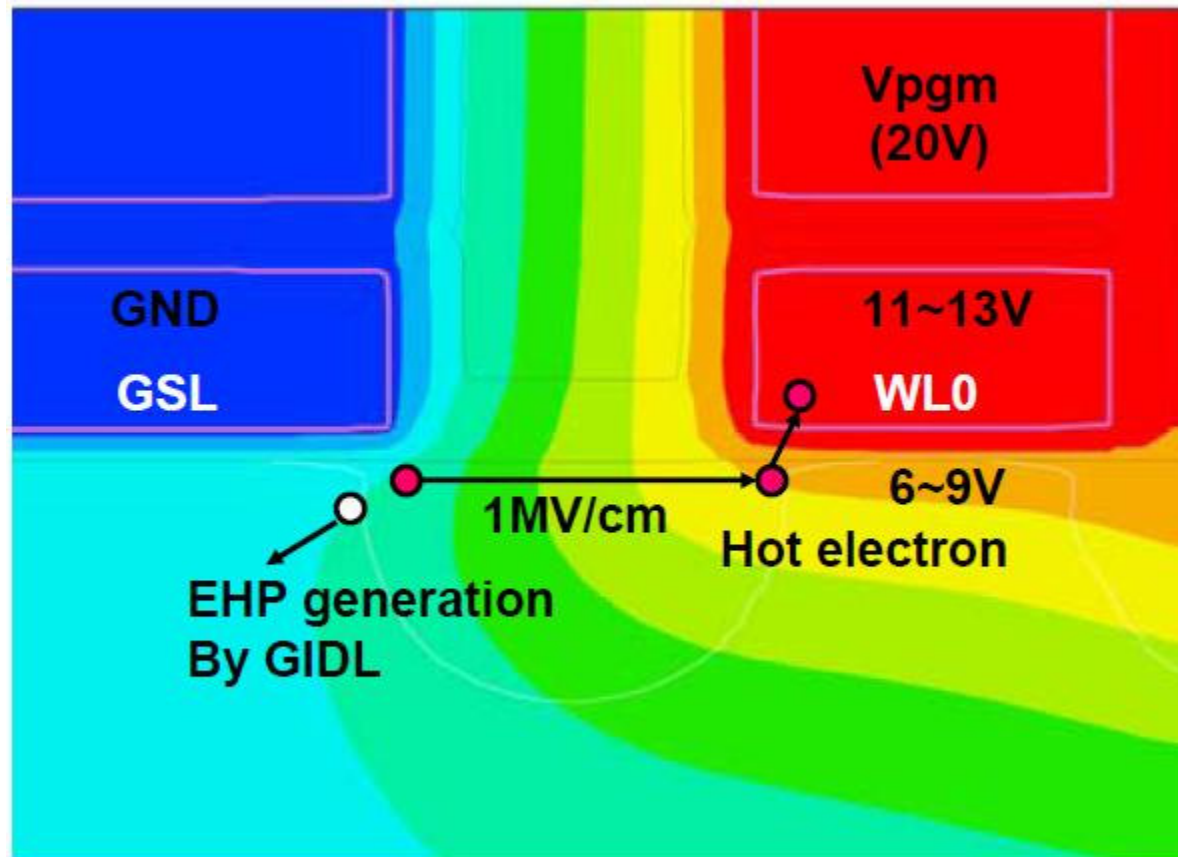
b = hot electrons due to WL0-sgs E field

Hot carrier disturb depends on :

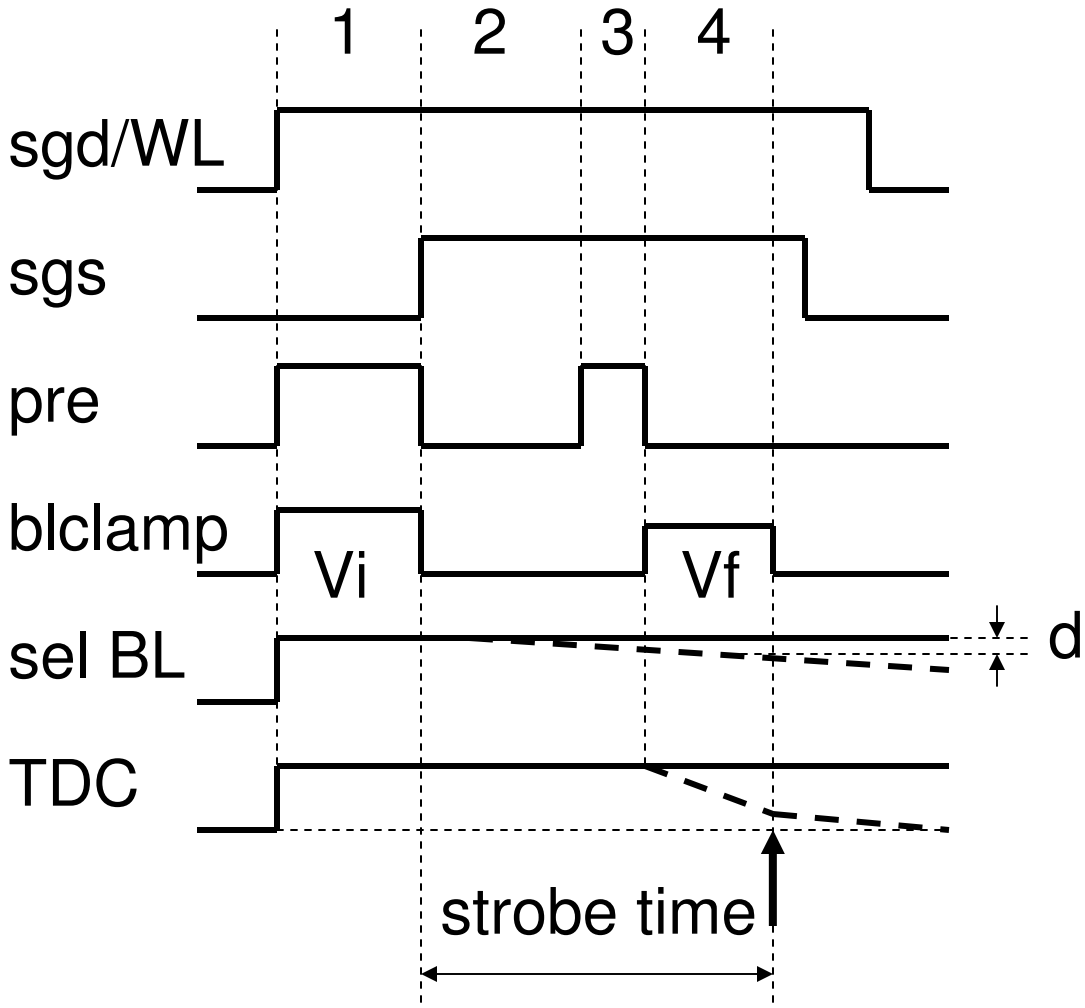
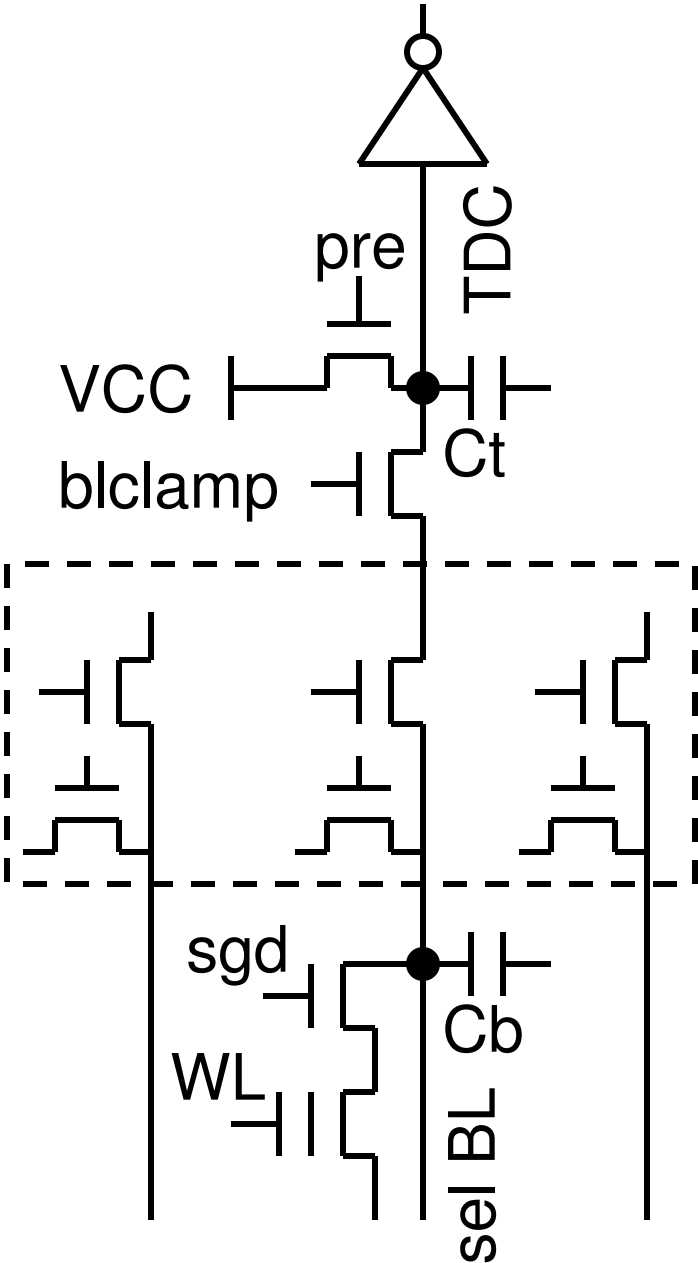
- sgs/WL0 dimension
- SGS bias
- WL0 voltage during program. In case 3bc WL0 is high (24V) in case 1bc WL0 is lower (15V).
- Hot carrier disturb applies also to WL65

Edge WL 'hot carrier disturb'

Hot carrier disturb has been widely investigated in IEEE NVSMW 2006 pag. 31



Sensing (erased and programmed cells)

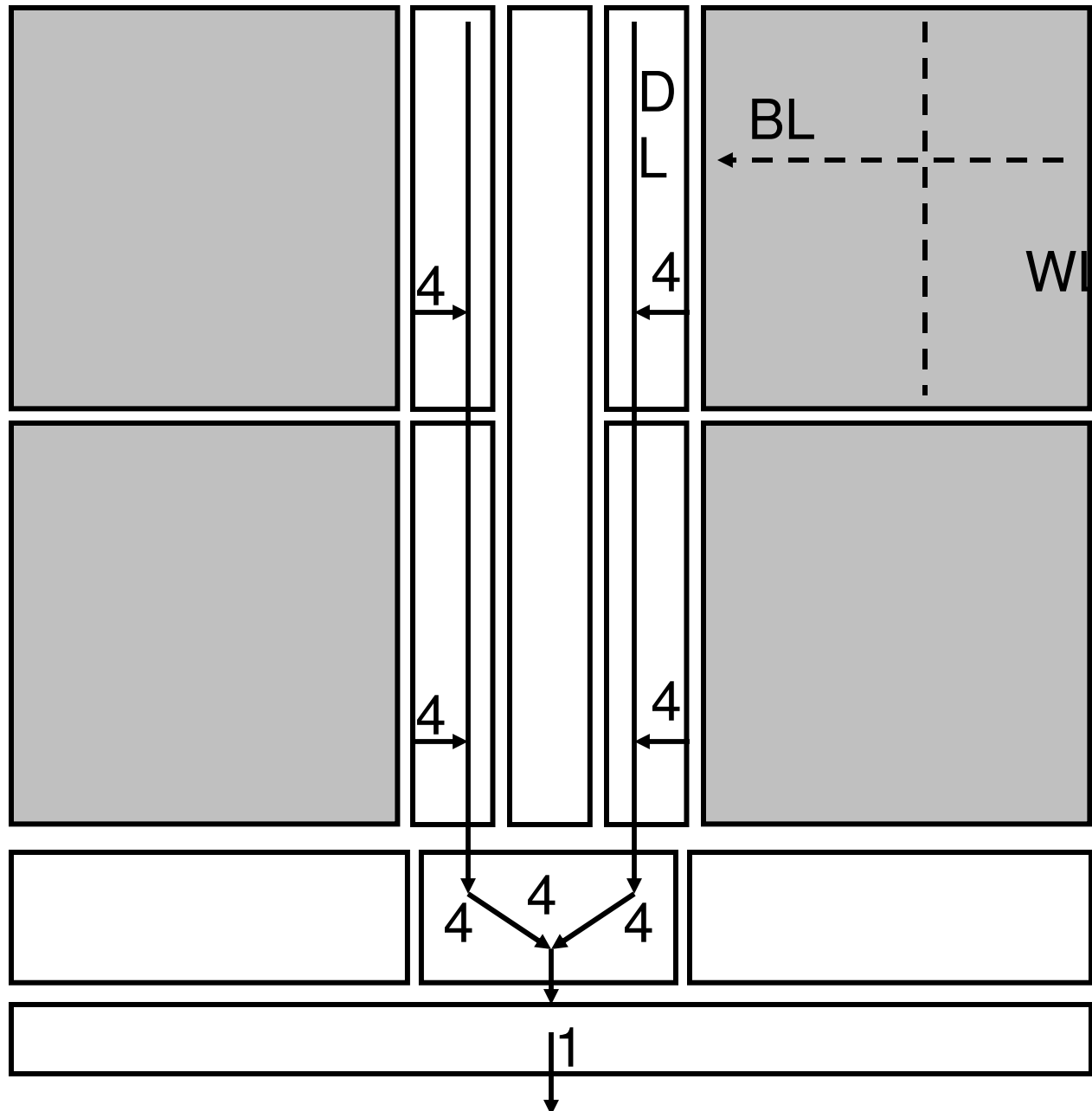


Vi – Vf accounts for BL leakage
 Unsel BLs are used as gnd-shield

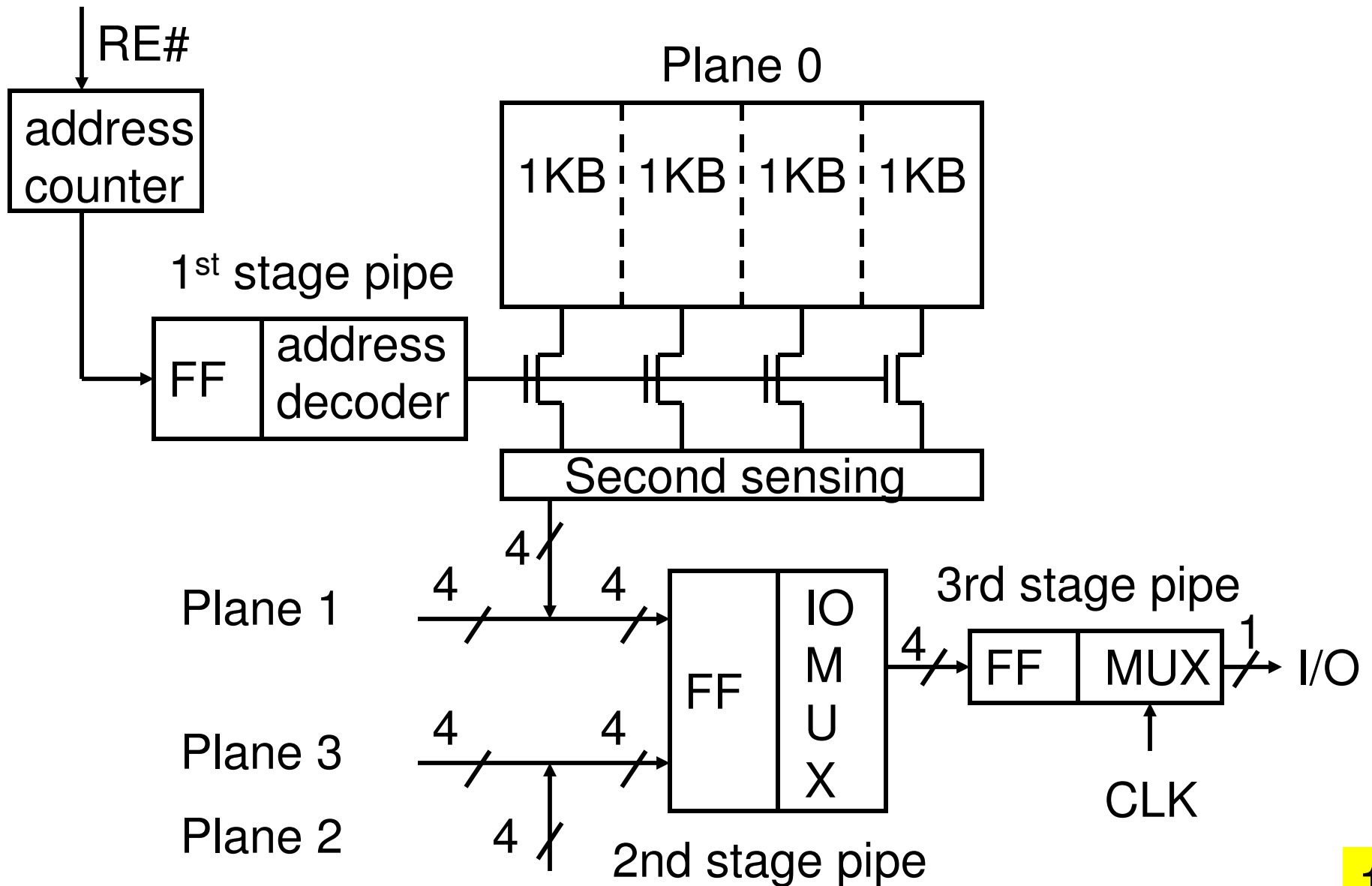
Quad planes

Quad planes architecture with reduced bitlines RC and alternate bitlines program with reduced bitlines interference (the unselected are used as shield) allow fast program operation and high program sustainable throughput.

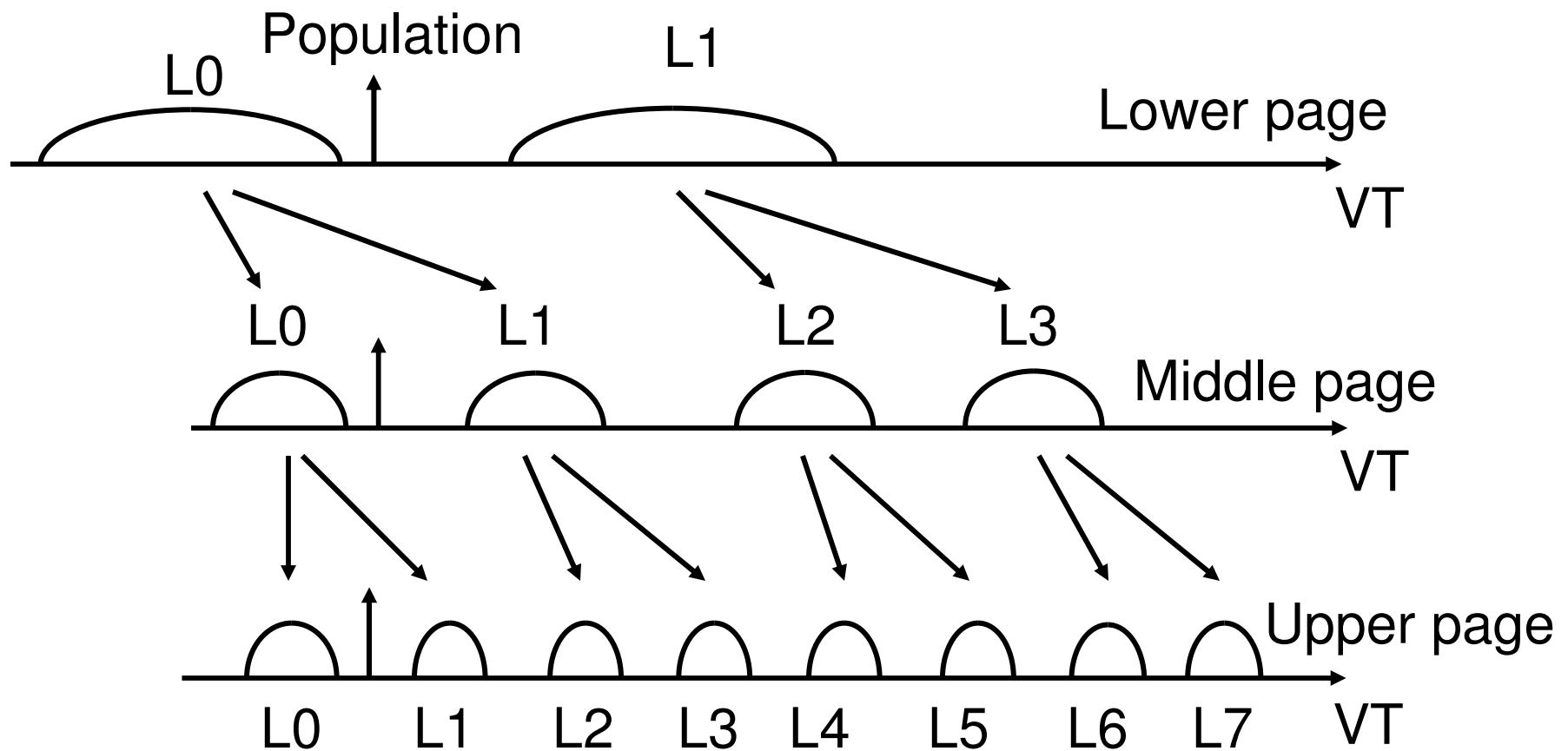
Byte data flow



Data path

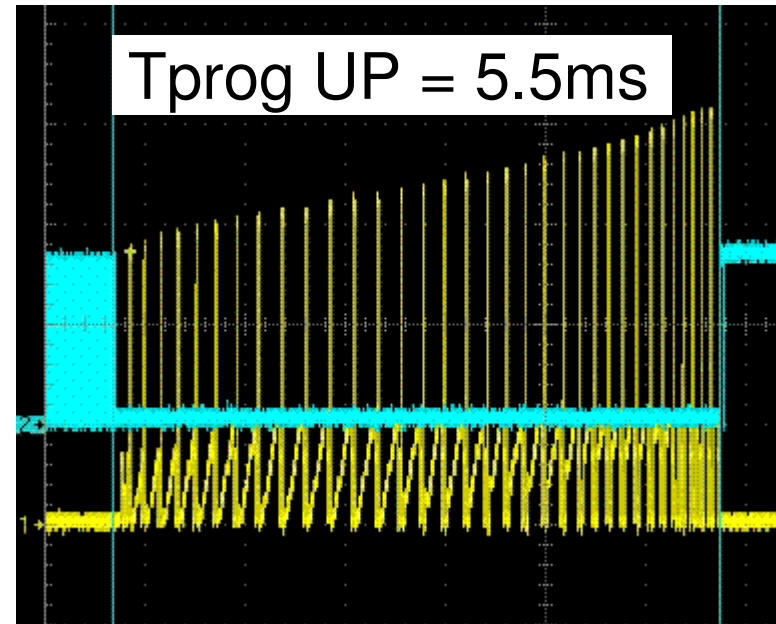
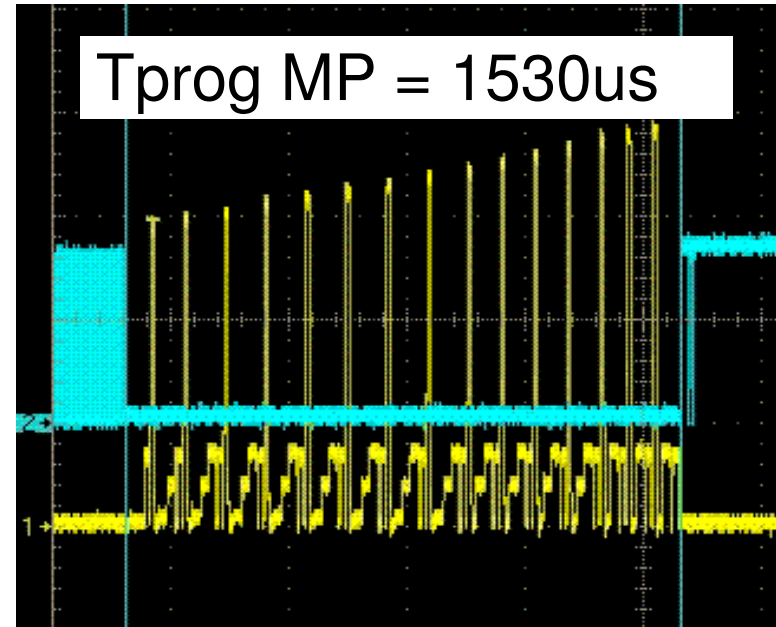
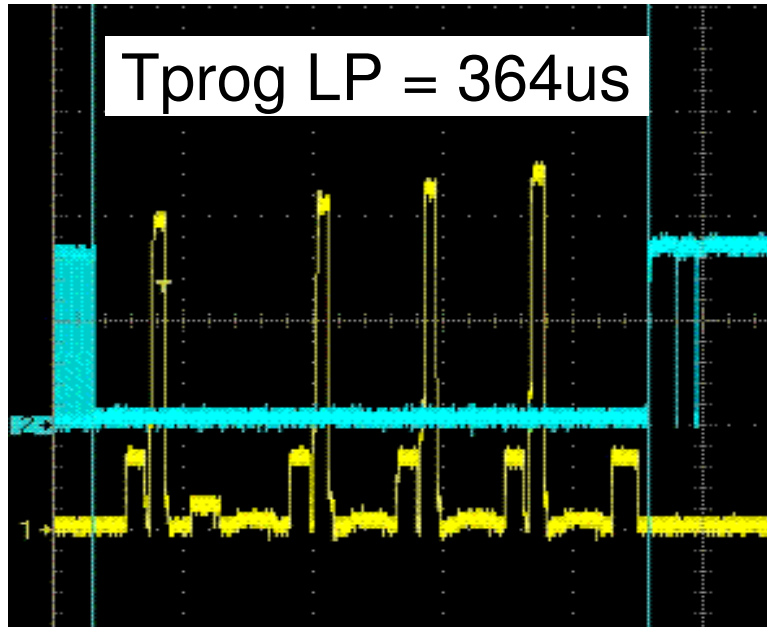


Low-Mid-Upper page VT placement (LMU) for FG-FG reduction



When user programs a page, all the non programmed surrounding lower order pages can be programmed by an internal algorithm to minimize the FG-FG interference.

Programming time



Mean
 $T_{\text{prog}} = 2.46 \text{ msec}$

3bits/cell throughput

$$\text{CKtr} = \text{Data DDR throughput} = \frac{2 \text{ Bytes}}{12 \text{ nsec}} = 166 \text{ MB/s}$$

$$\text{Sustain din throughput (quad)} = \min\left(\text{CKtr}; \frac{4096 \text{ Bytes} \times 4}{2.46 \text{ msec}}\right) = 6.6 \text{ MB/s}$$

$$\text{Sustain dout throughput (quad)} = \min\left(\text{CKtr}; \frac{4096 \text{ Bytes} \times 4}{70 \text{ usec}}\right) = 166 \text{ MB/s}$$

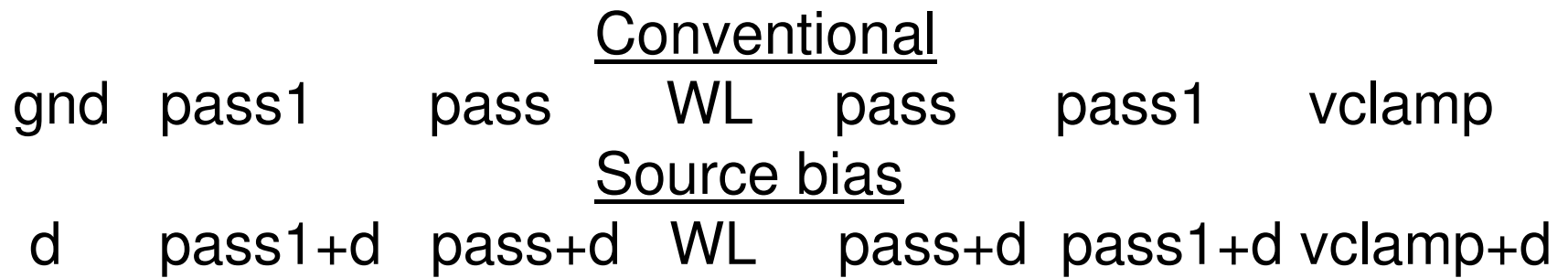
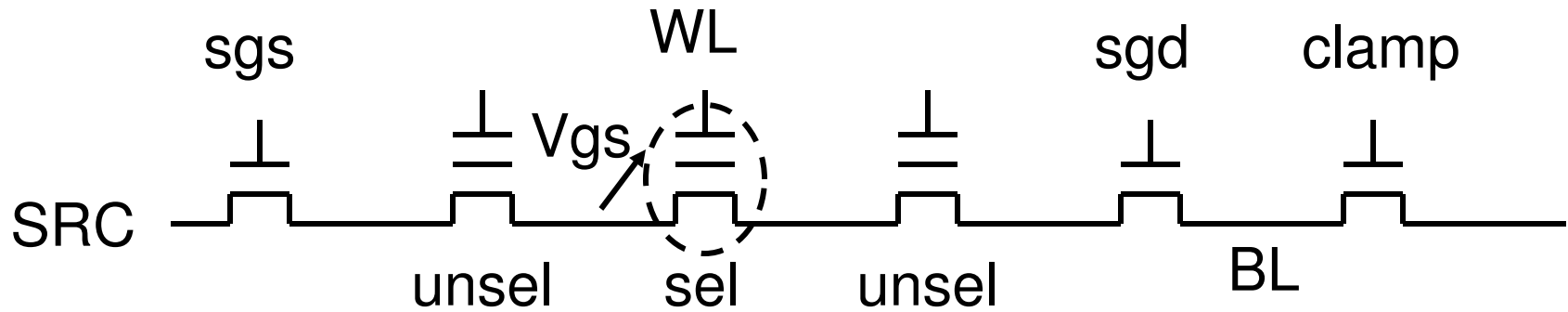
Source bias : total VT budget increase

Purpose of the Source bias technique is to 'apparently' move down all the distributions of about 1.5V and expand the negative VT region without the use of negative voltages on the word lines.

In this manner the total VT budget is expanded.

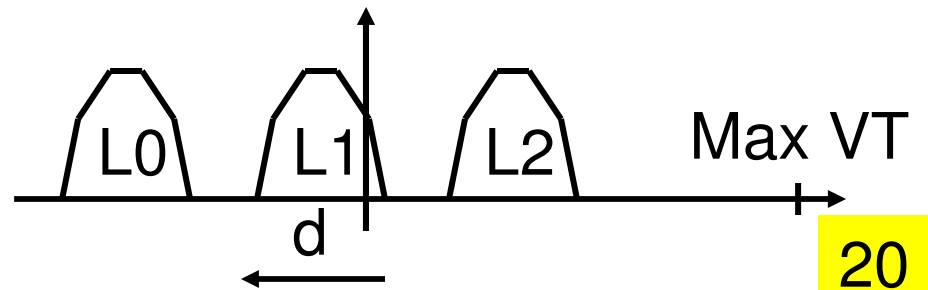
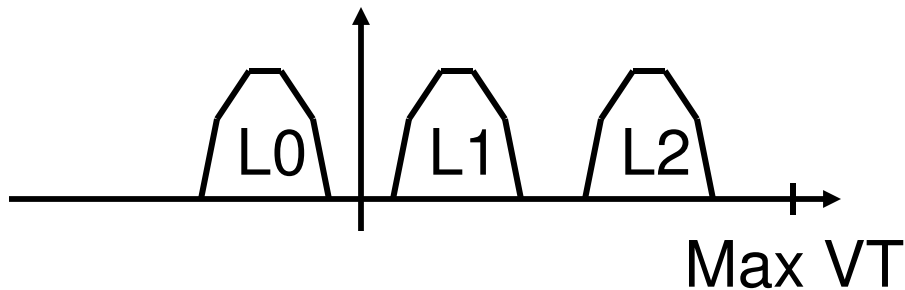
The source bias technique is performed during read and program verify.

Source bias

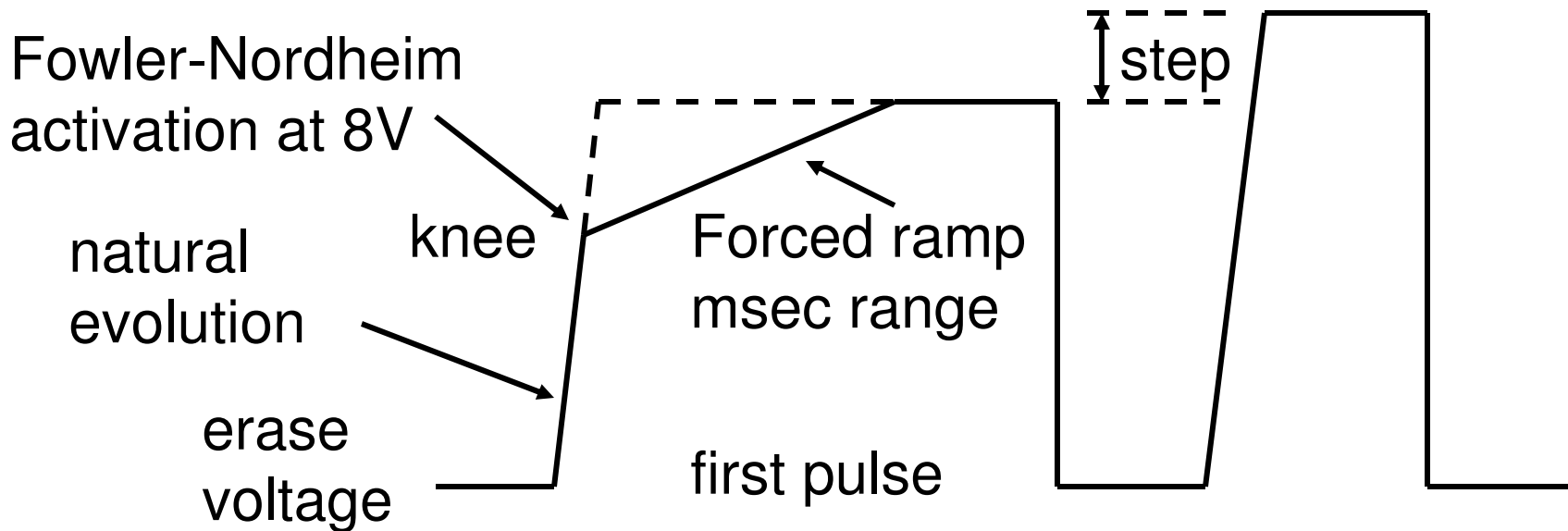


$$V_{gs} = (WL - SRC) = V_T$$

$$V_{gs} = WL - (SRC+d) = V_T - d$$



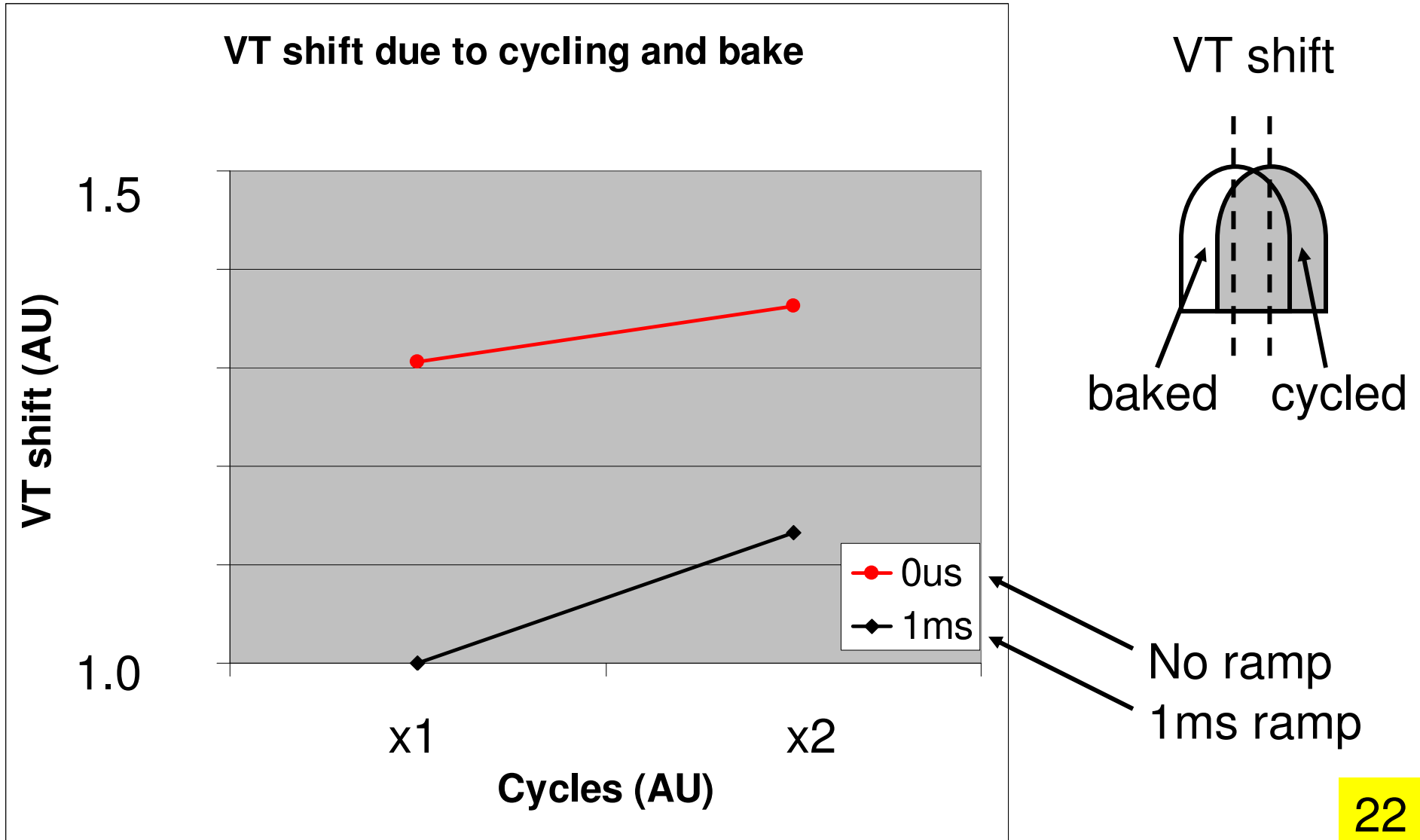
VT placement : erase ramp (charge loss reduction)



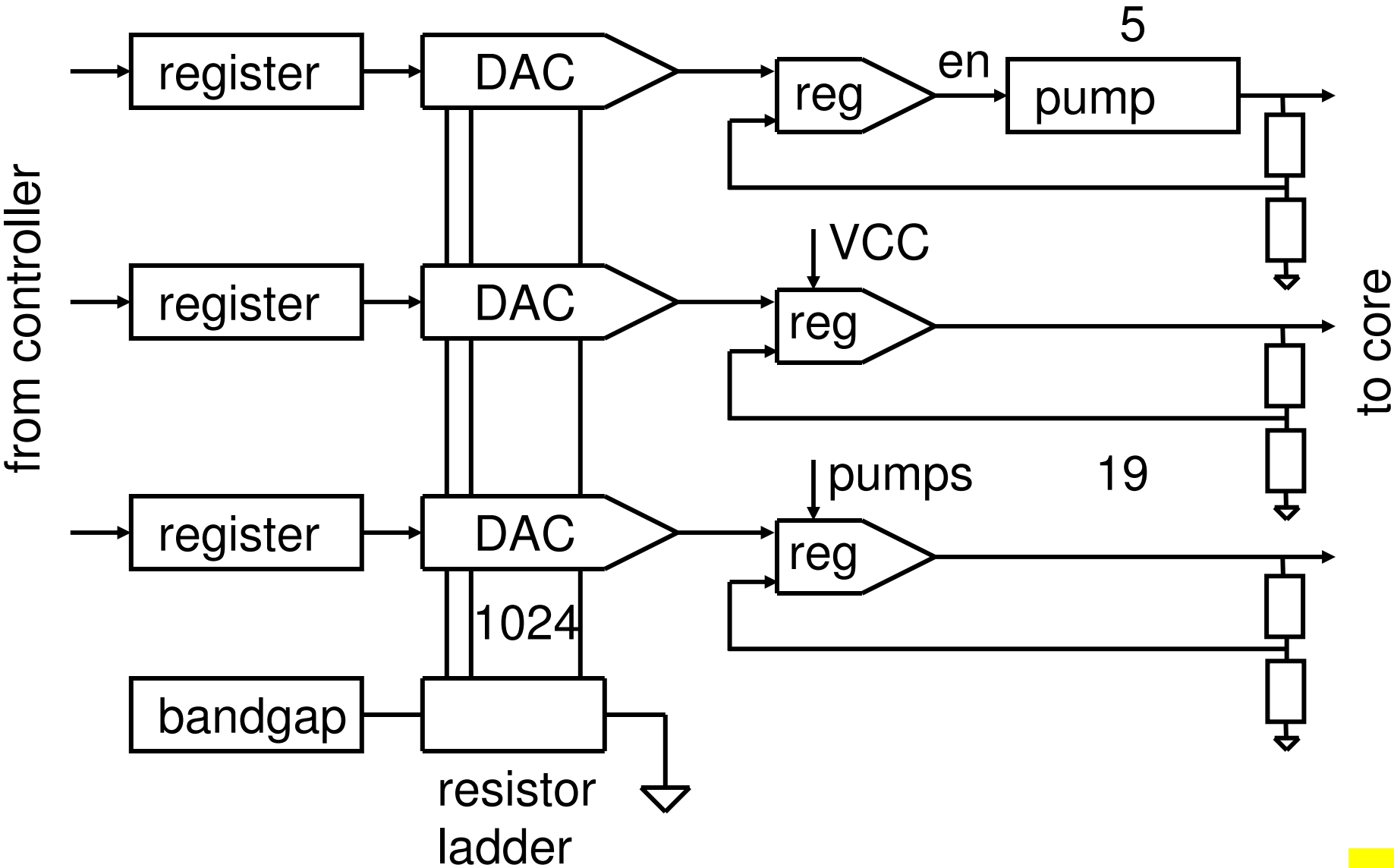
US pat 4628487

- Ramping first erase pulse decreases the Fowler-Nordheim peak current.
- Ramp applies only to the first pulse because charged floating gates prevent further peak currents.
- Pulse ramp technique is beneficial also in program but it is not realistic for performance reasons.

Effect of erase ramp on Threshold VT placement



Analog system



Analog system

Resistor ladder provides a maximum number of 1024 analog voltages to 24 DACs : each one of the DACs has a max resolution of 10 bits.

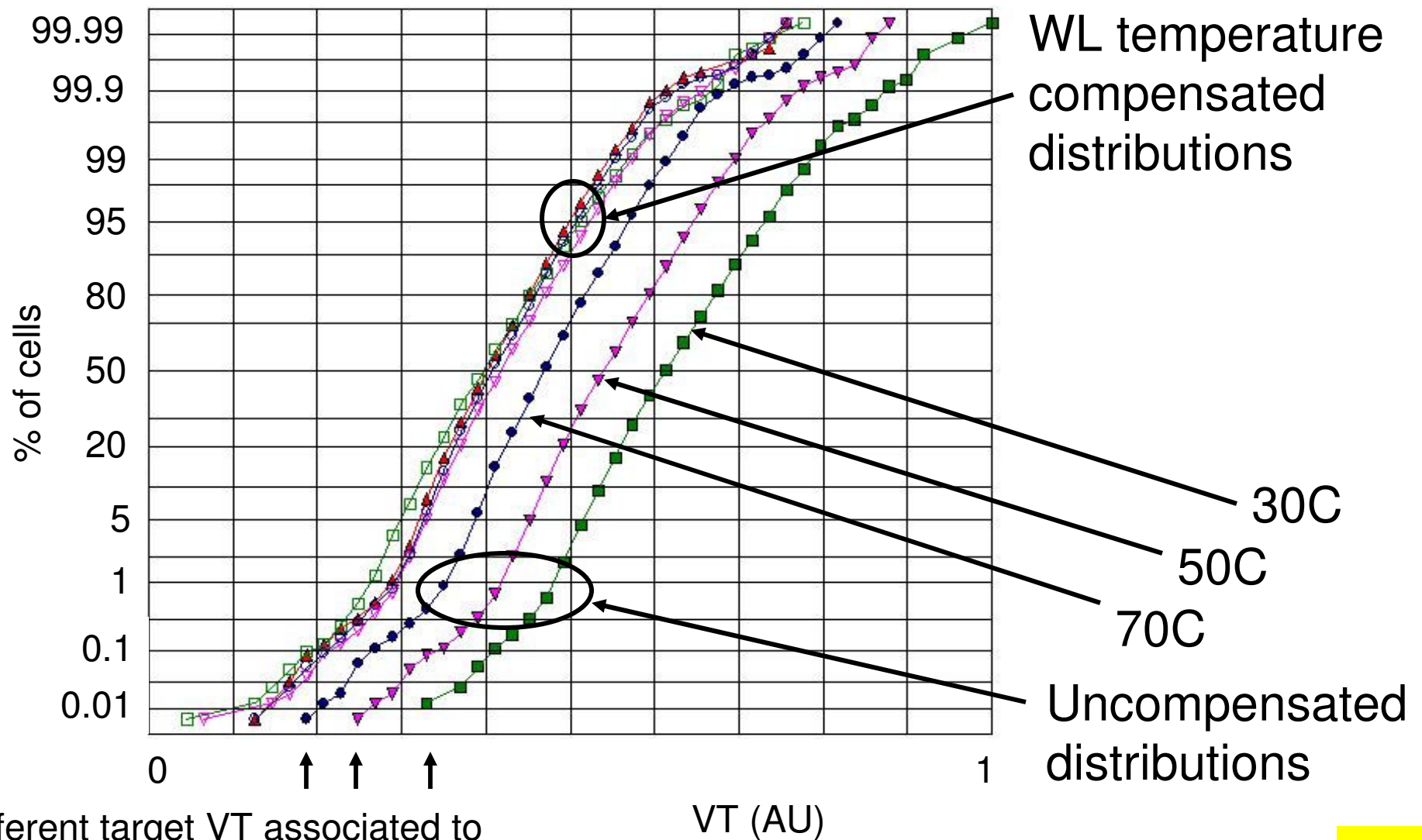
Regulators (reg) are used to generate voltage for the core.

Regulators are divided in two categories : 5 on-off regulators to enable pumps and 19 linear regulators to directly provide regulated voltage to the core.

Linear regulators can be powered by auxiliary pumps to provide medium range voltages to the core or can be powered directly by VCC to provide voltages to the core that are lower than VCC.

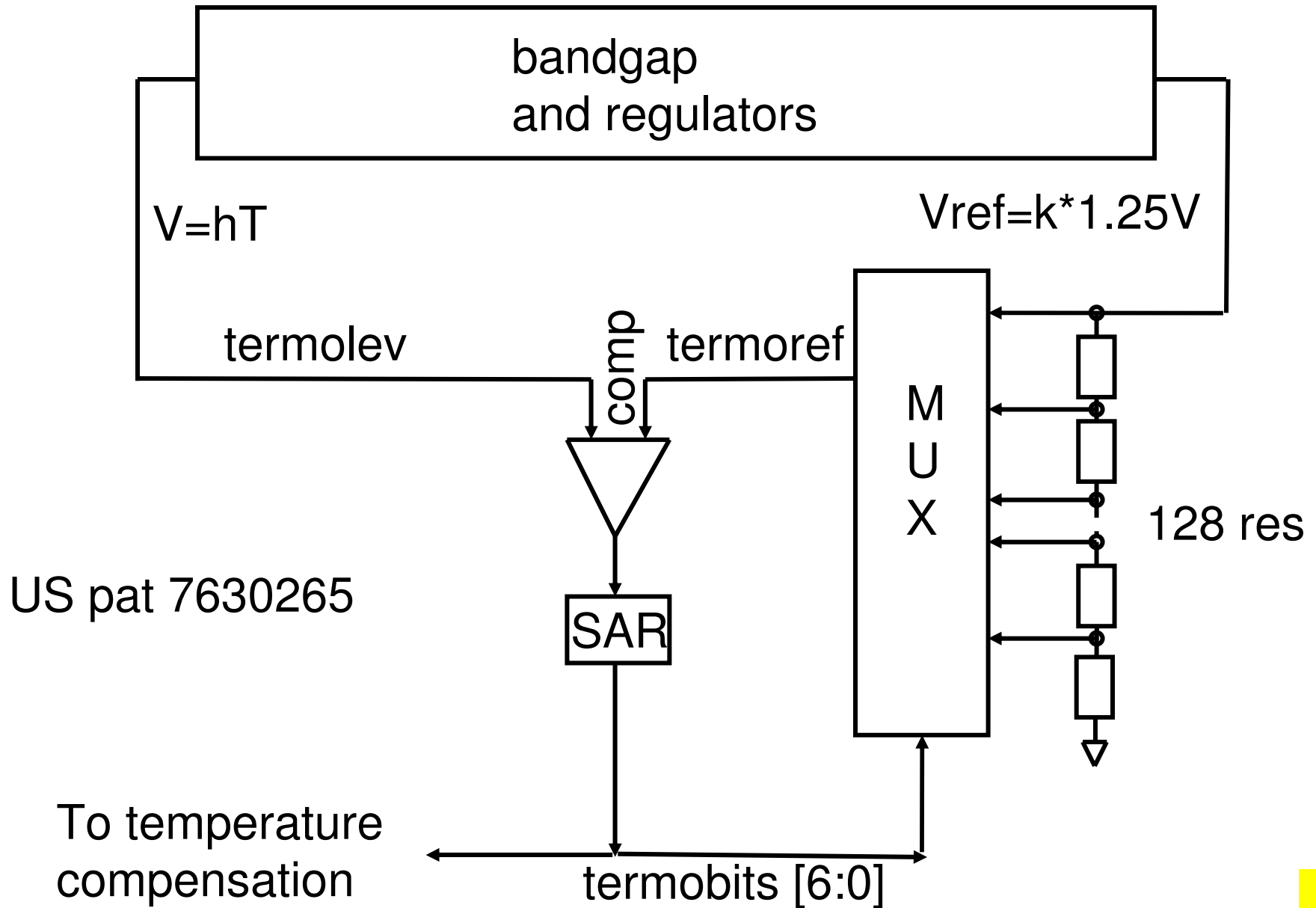
Temperature effect on VT

US pat 7630266

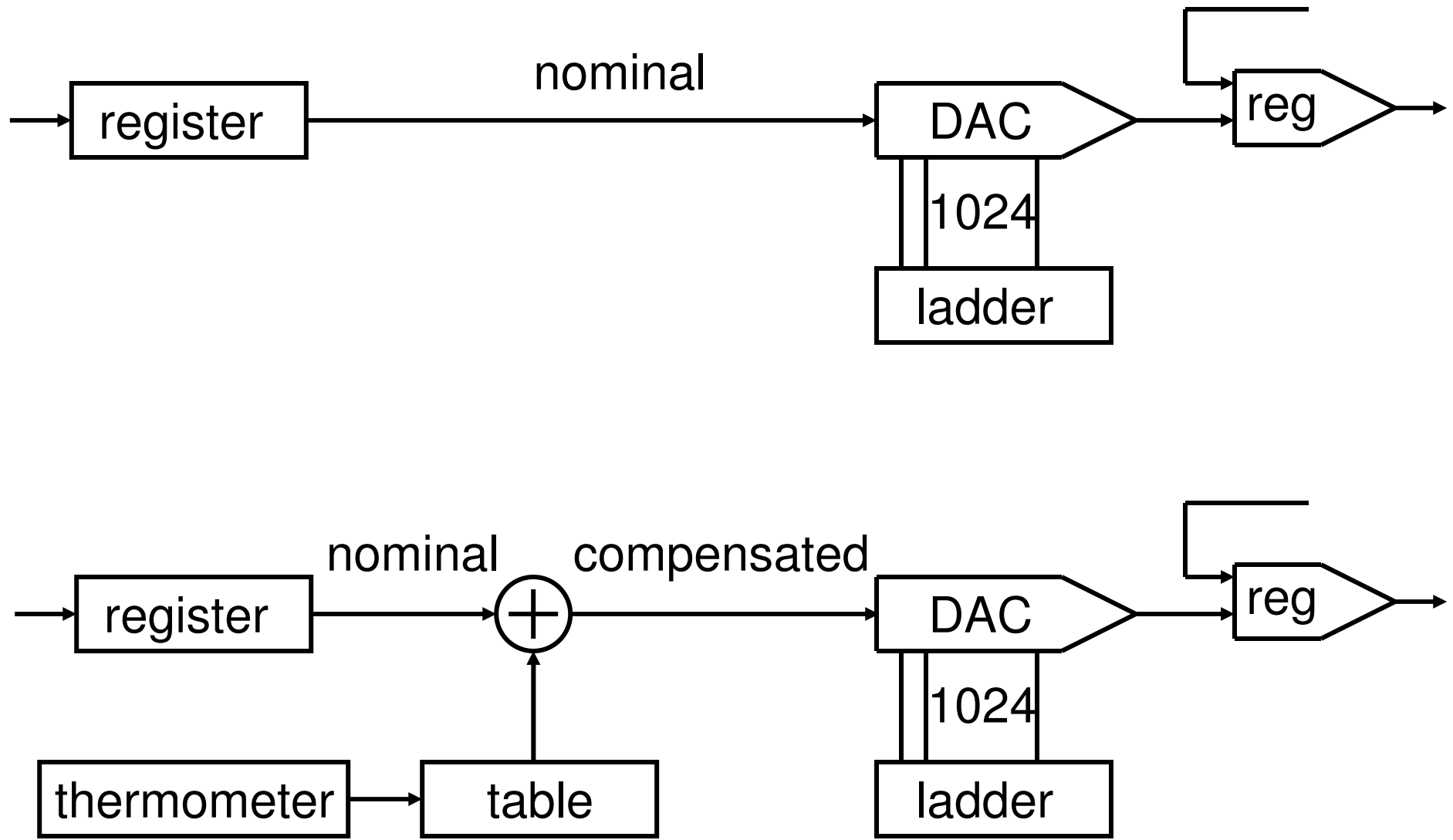


Different target VT associated to the same target WL in prog verify

Temperature detection



Temperature compensation



Temperature compensation features

7bits digital resolution

-40C to 90c temperature range

1C temperature resolution

Conversion time includes amplifier offset cancellation

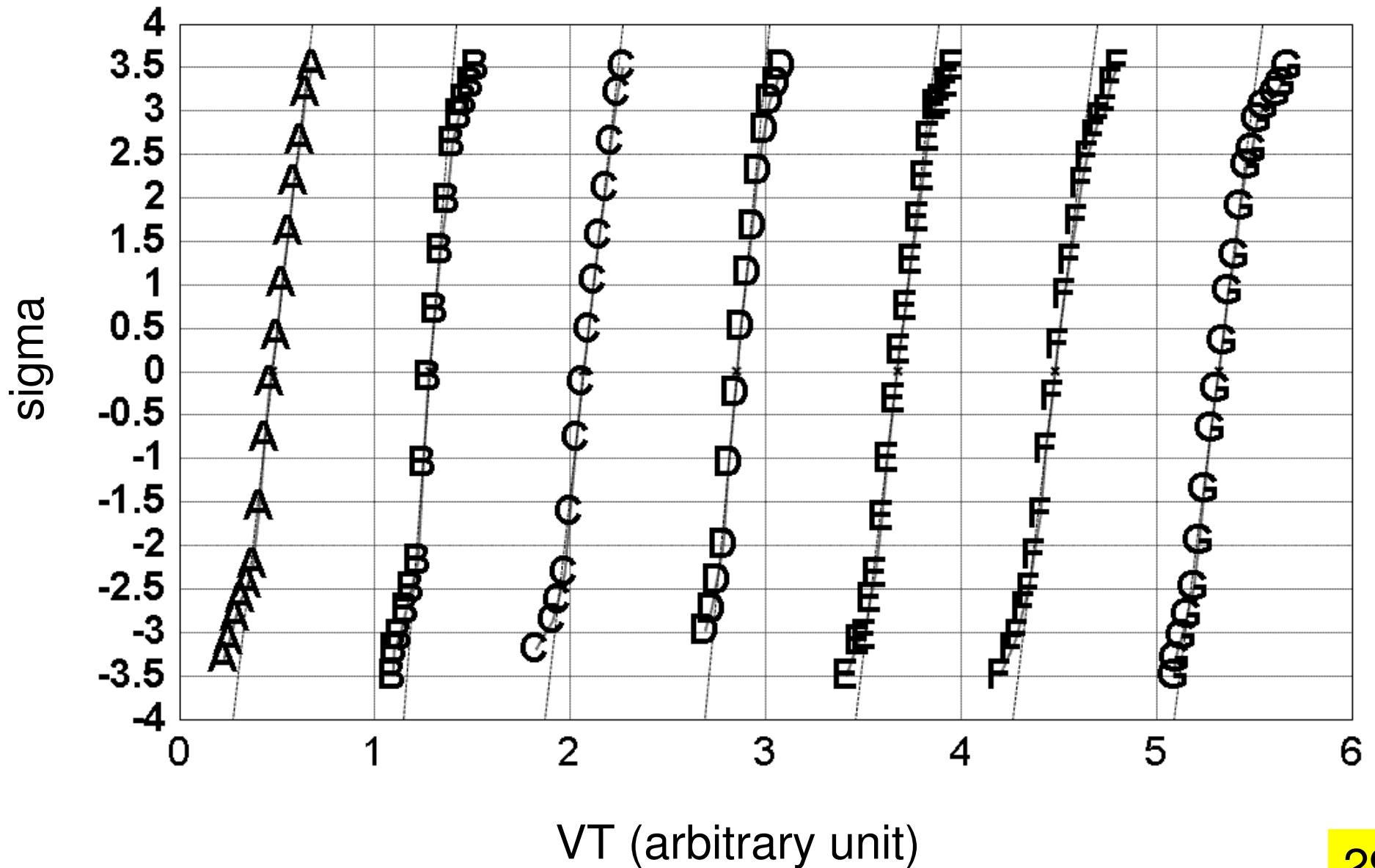
6 different regulated voltages can be compensated :

- wordline read/vfy

- vpass read/pgm

- SGS, SGD read/pgm

Typical uncycled distributions



Summary

A 3bit/cell 32Gb NAND Flash at 34nm was presented.

It has high sustainable write/read throughput obtained with the use of quad planes architecture and alternate BL shield in program and sensing operations.

Design techniques related to :

- LMU sequence

- Edge WL configuration

- Source bias

- Temperature compensation

- First erase pulse shaping

were used to increase VT distribution margins at time 0 and with age.