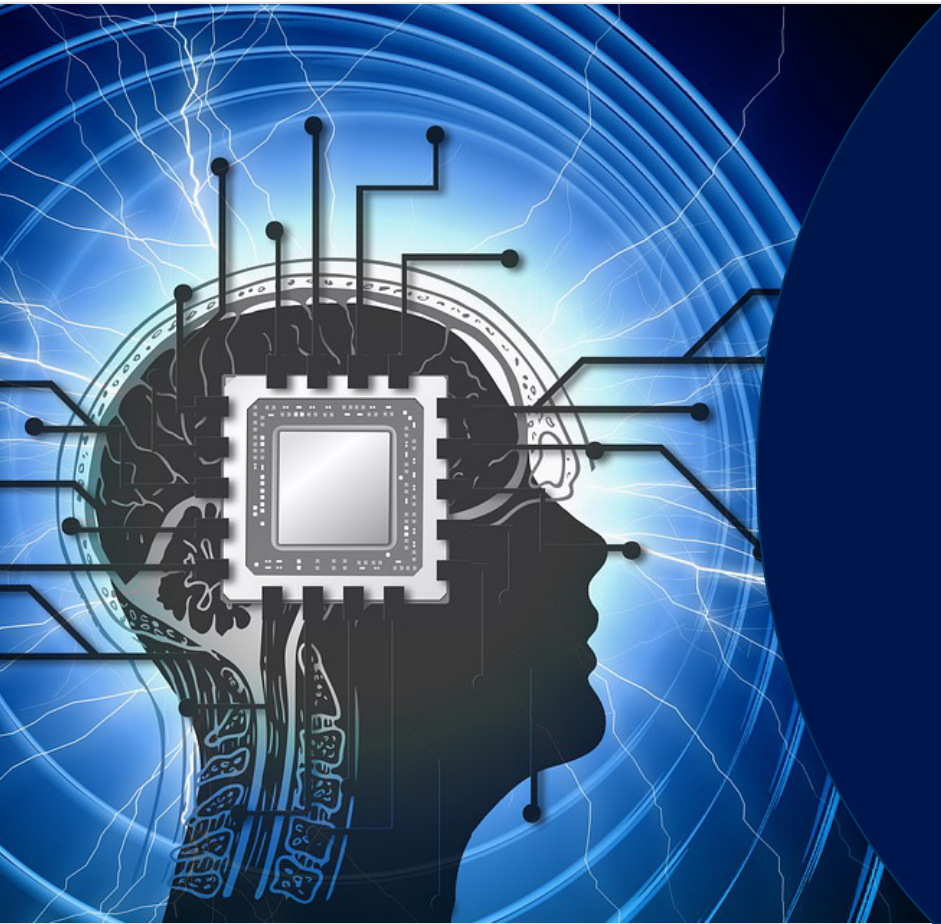
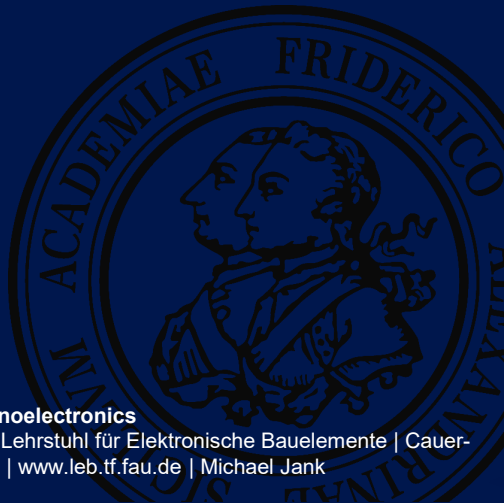
Friedrich-Alexander-Universität
Technische Fakultät



Halbleitertechnik IV- Nanoelectronics

Friedrich-Alexander-Universität Erlangen-Nürnberg | Lehrstuhl für Elektronische Bauelemente | Cauerstraße 6 | 91058 Erlangen | www.leb.tf.fau.de | Michael Jank

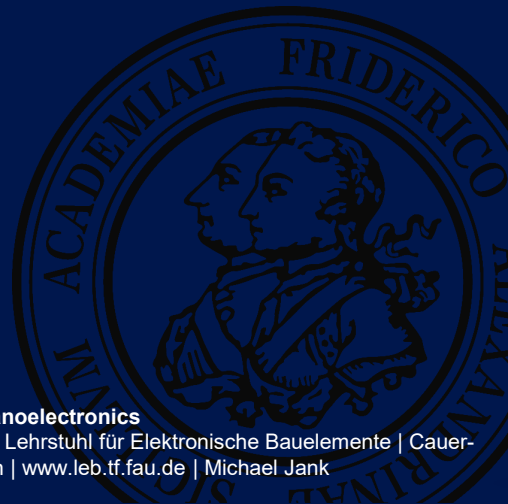
Metal-Oxide-Semiconductor Memory Devices



Halbleitertechnik IV - Nanoelectronics
FAU Erlangen-Nürnberg | Lehrstuhl für Elektronische Bauelemente | Cauerstraße 6 | 91058 Erlangen | www.leb.tf.fau.de | Michael Jank

Objectives of the lecture

Mission and goals?



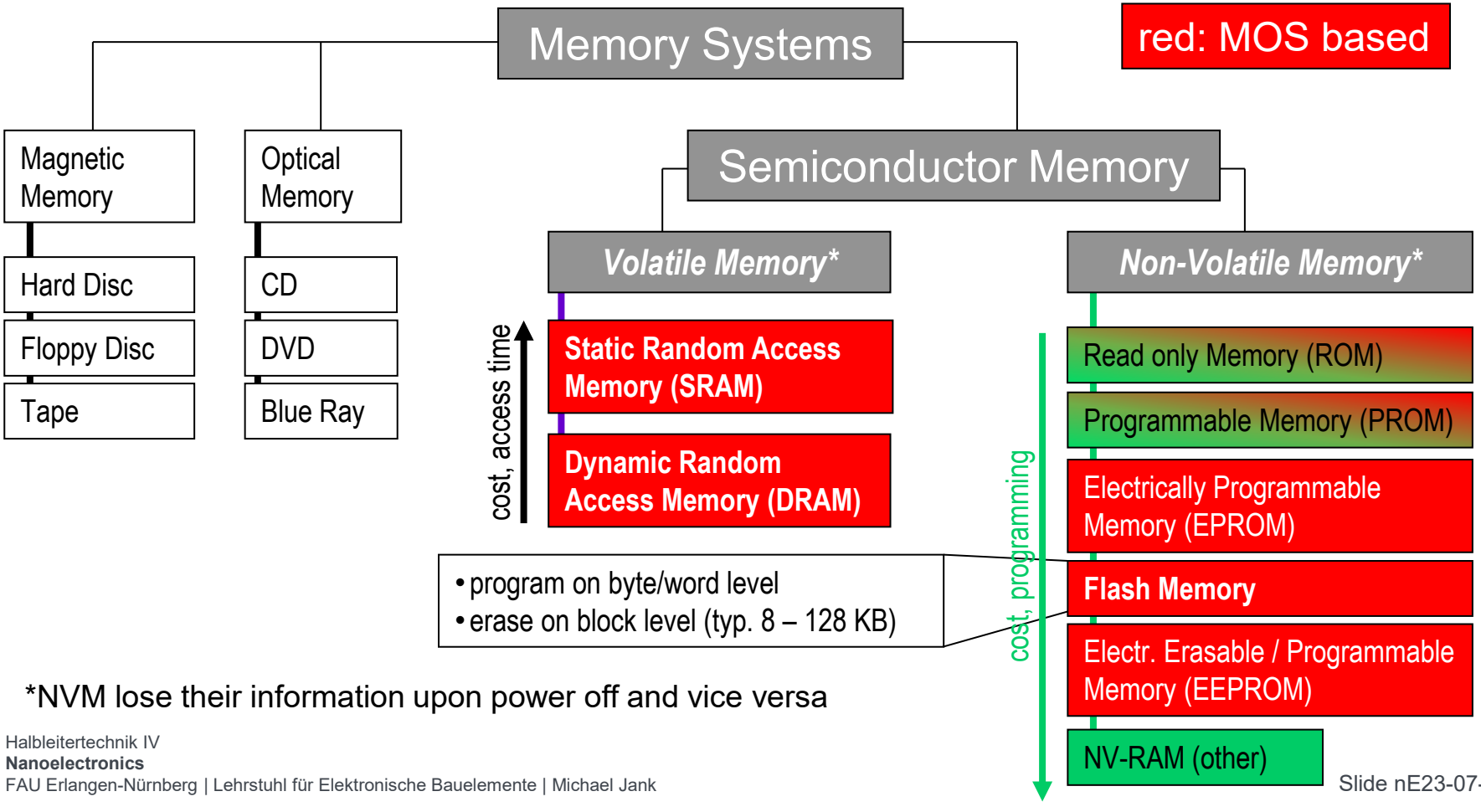
MOS Memory Devices

Objectives

- This lecture introduces the knowledge on the three classes of most-adopted semiconductor-based memory devices
  - *SRAM*
  - *DRAM*
  - *Flash*
- You should be able to
  - *describe the memory architectures and principles of operation*
  - *know about the critical parameters in scaling of DRAMs*
  - *explain problems and solutions in the operation and scaling of Flash memories*
  - *consider measures for maximizing the memory density in Flash- and SONOS memories*

MOS Memory Devices

Overview on information storage systems



*NVM lose their information upon power off and vice versa

MOS Memory Devices

Overview on information storage systems - Abbreviations

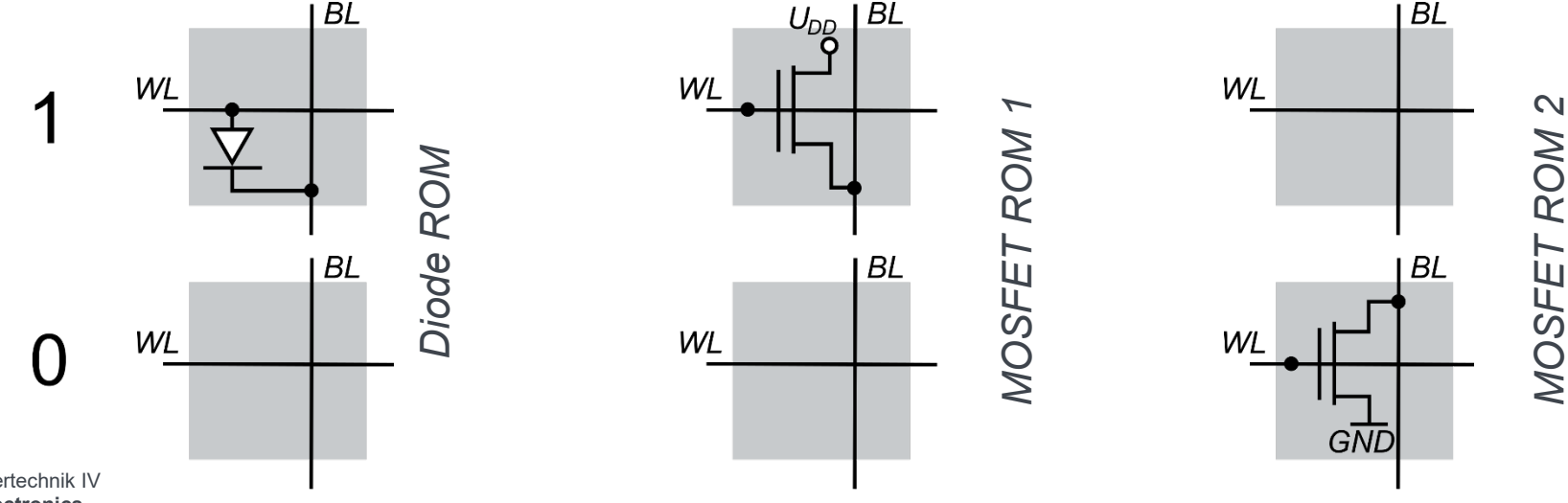
- **ROM** Read Only Memory
- EPROM Electrically Programmable ROM
- OTP-EPROM One-Time Programmable EPROM
- E^2PROM Electrically Erasable Programmable ROM

- **RAM** Random Access Memory
- DRAM Dynamic RAM
- SRAM Static RAM
- FRAM Ferroelectric RAM (auch: FeRAM)
- MRAM Magnetic RAM
- CBRAM Conductive Bridging RAM
- RRAM Resistive RAM
- PCRAM Phase Change RAM

MOS Memory Devices

Read-only memory (principle is not MOS based)

- Only one-time programming
  - *e.g. by formation or opening of wiring at a via (i.e. cross-point connection between two metal layers) between word and bit line in an array*
  - *during processing (photo mask) or eletrically in the chip (burning)*
- Concepts for realization of ROM memories
  - *rectifiying/switching elements avoid cross-talk between the cells*

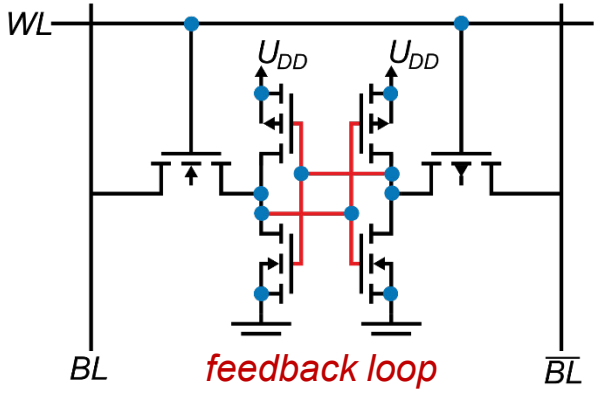


MOS Memory Devices

MOS-based Semiconductor Memories

SRAM

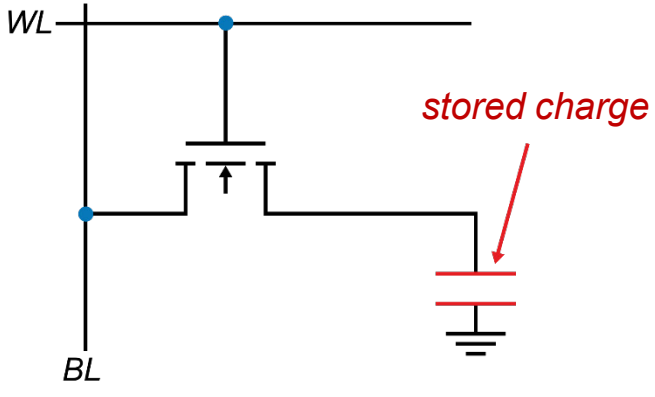
6-transistor cell (6T)



+ *very fast*
+ *CMOS*
+ *no refresh necessary*
- *transient*
- *very large cell*

DRAM

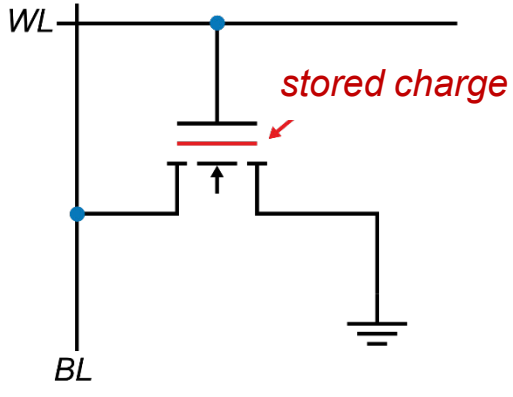
1-transistor/1 capacitor cell (1T1C)



+ *fast*
+ *small cell*
- *volatile*
- *refresh*
- *elaborate technology*

Flash

1T



+ *non volatile*
+ *small cell*
+ *multibit capability*
- *slow*
- *erase by block*
- *limited cycling capability*

MOS Memory Devices

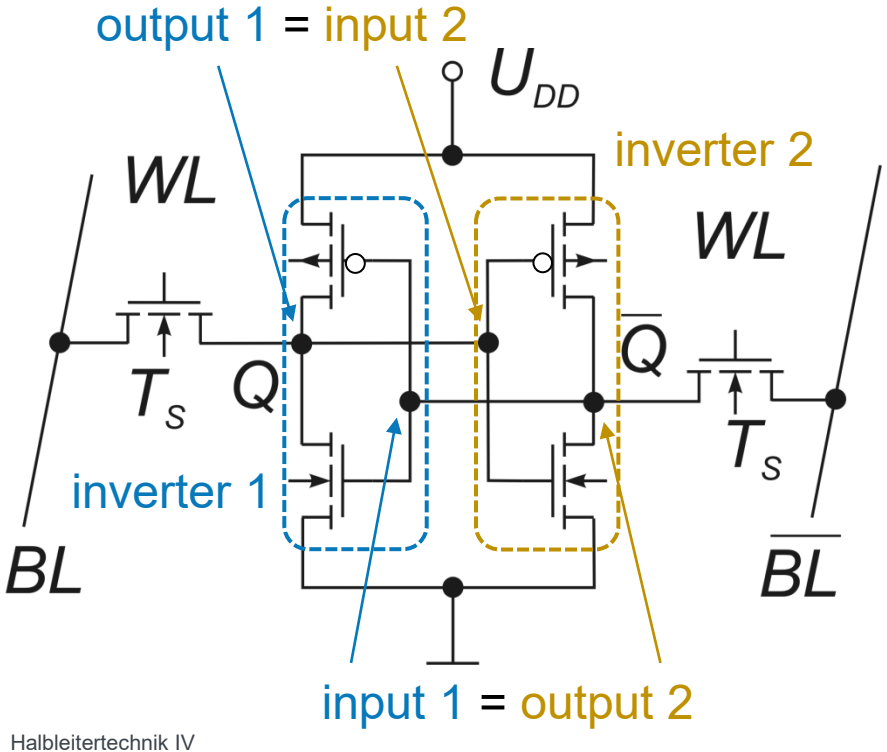
Outline

- Overview on highly-integrated information storage
- Volatile semiconductor memory
  - *Static memory with random access (SRAM)*
  - *Dynamic memory with random access (DRAM)*
- Non-volatile semiconductor memory
  - *Flash memory transistors*
  - *Flash architectures*
  - *SONOS storage stacks*
  - *Nitride-ROM memory (NROM)*

MOS Memory Devices

Volatile Semiconductor Memory

- SRAM (6T): more a circuit concept than a device
  - *two fed-back CMOS inverters that conservate their state*



SRAM: write (high/"1")

- set inverted bitline $\overline{BL}$ to low/L and bitline BL to high/H
- activate wordline WL, nodes Q / $\overline{Q}$ of the flip-flop will be set to H / L
- deactivate wordline => information is stored statically

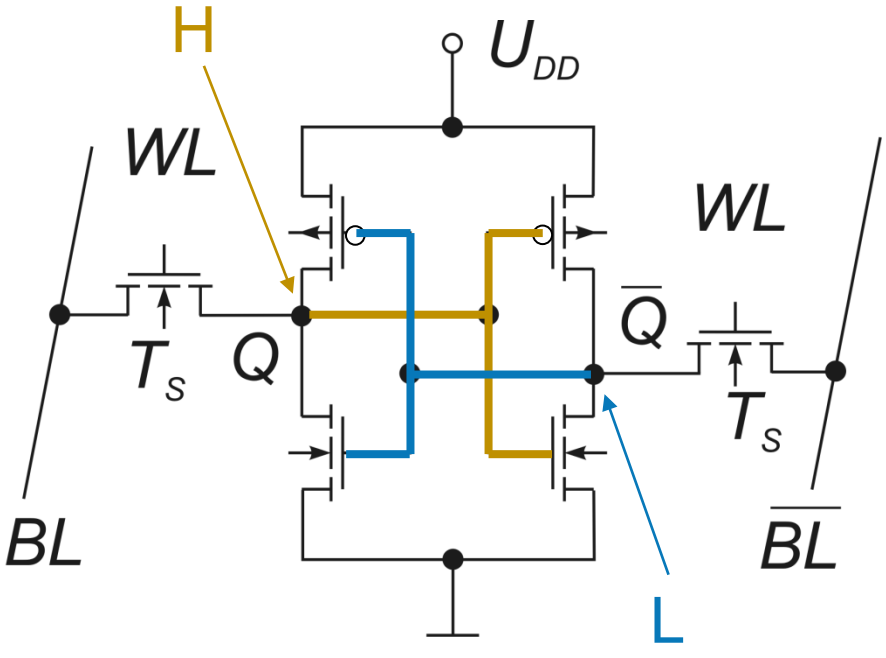
SRAM: read

- activate wordline
- bitline is charged by flip-flop, read result

MOS Memory Devices

Volatile Semiconductor Memory

- SRAM (6T): more a circuit concept than a device
  - *two fed-back CMOS inverters that conservate their state*



SRAM: write (high/"1")

- set inverted bitline $\overline{BL}$ to low/L and bitline BL to high/H
- activate wordline WL, nodes Q / $\overline{Q}$ of the flip-flop will be set to H / L
- deactivate wordline => information is stored statically

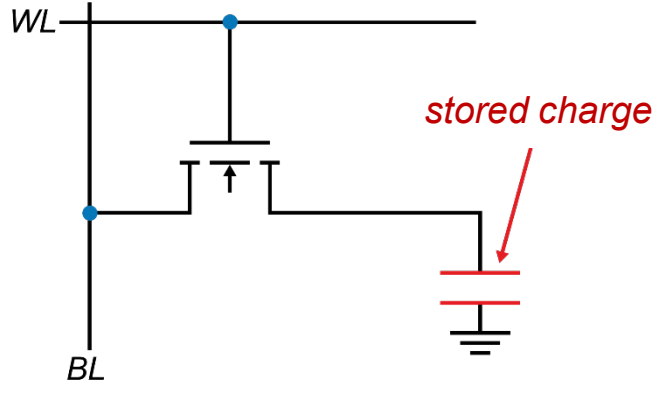
SRAM: read

- activate wordline
- bitline is charged by flip-flop, read result

MOS Memory Devices

Volatile Semiconductor Memory

- DRAM (1T1C): two-device assembly
  - *"0" no charge on capacitor*
  - *"1" charge on capacitor*
  - *can be realized in CMOS*
  - *refresh necessary due to leakage*
- Concepts:
  - *planar cell (formerly)*
  - *"Trench" cell*
  - *"Stacked" cell*
- Definition of the smallest structure size (F):
  - *Diameter of the capacitor/gate electrode*
  - *"Pitch": Distance between 2 conductor tracks (word/bit lines)*

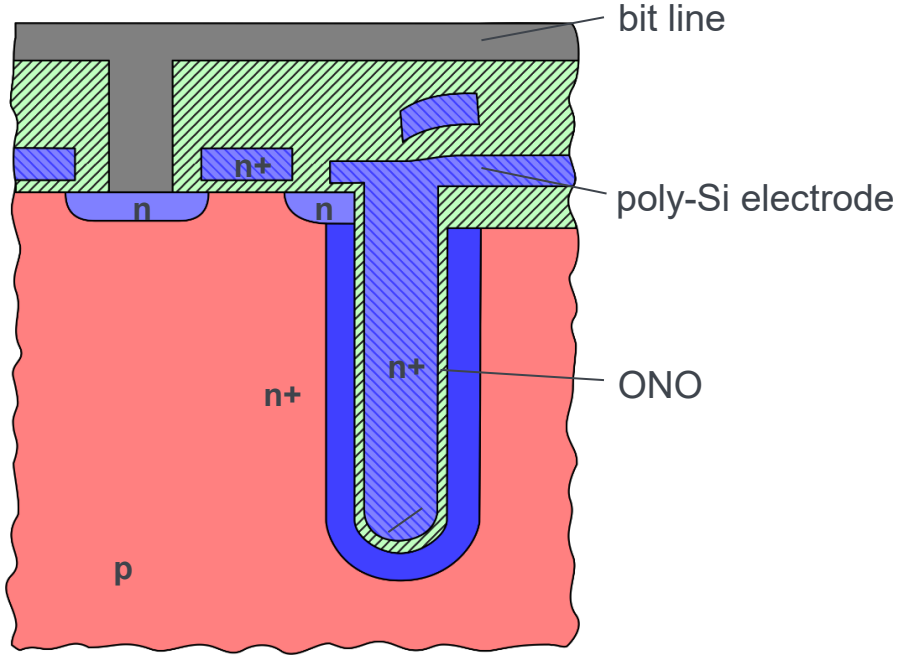


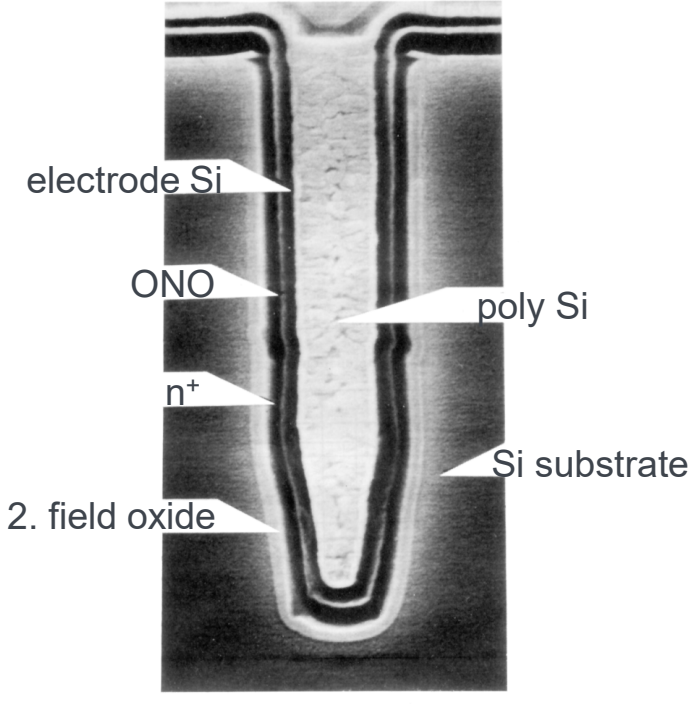
Cell Size $4F^2$

MOS Memory Devices

Volatile Semiconductor Memory

- DRAM Cell Concepts: Trench Cell
  - *The aim is to maximize the area of the storage capacitor to store as much charge as possible*
  - *Burying creates large capacitor gate area and consumes only a small amount of chip area*

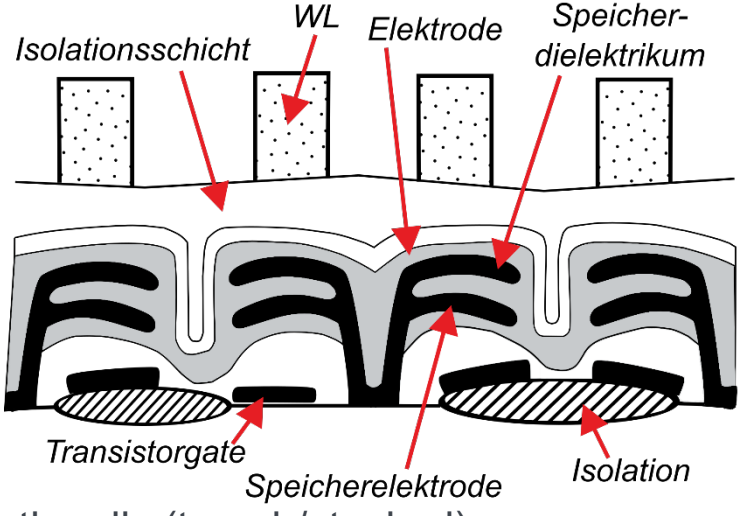


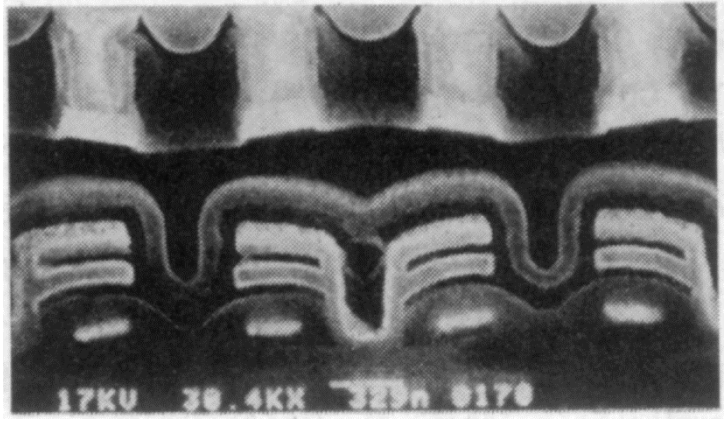


MOS Memory Devices

Volatile Semiconductor Memory

- DRAM Cell Concepts: Stacked Capacitor
  - *here the capacitor is mounted above the transistors*
  - *generation of a large effective capacitor area with low consumption of chip area*



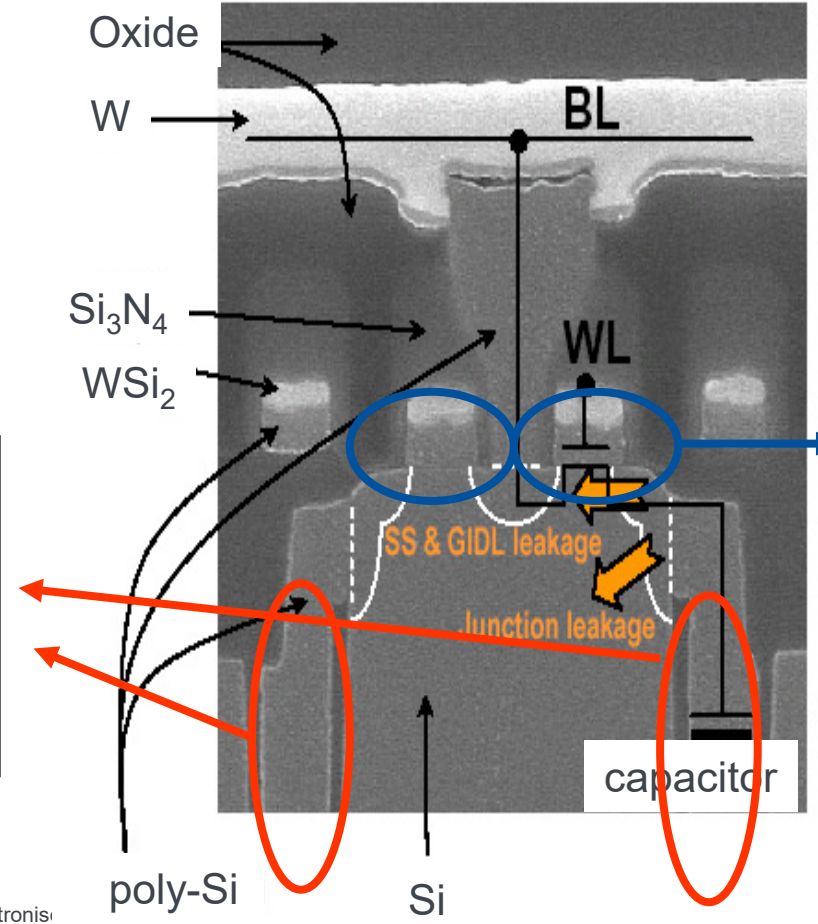


- Both cells (trench/stacked) are very complex to produce
  - *practically, deep trench etching is more difficult to scale (the narrower, the deeper you have to etch and refill!) → Stack capacitor has prevailed*

MOS Memory Devices

Volatile Semiconductor Memory

- DRAM: challenges



capacitor:

minimum capacity of 30fF must be maintained

select transistor:

leakage < 1fA

I_{ON} > few µA

$I_{ON}/I_{OFF} > 10^9$

MOS Memory Devices

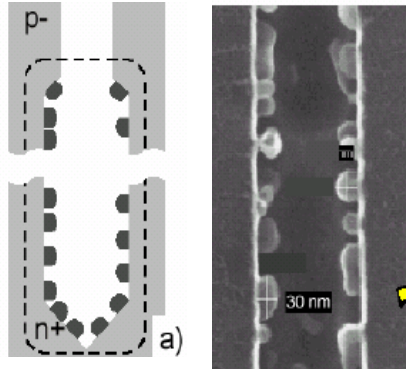
Volatile Semiconductor Memory

- DRAM: increasing the capacitance
- besides the area enlargement by trench or stacked capacitors there are mainly materials-driven approaches

$$C = A \cdot \frac{\varepsilon}{d}$$

increase of capacity by high-k dielectrics
e.g. Al_2O_3, Hf_2O, Zr_2O, Ta_2O_5, BST

reduction of thickness



Increase of surface by hemispherical Si grains (HSG)

Nucleation and ripening at etch defects (crystal defects)

MOS Memory Devices

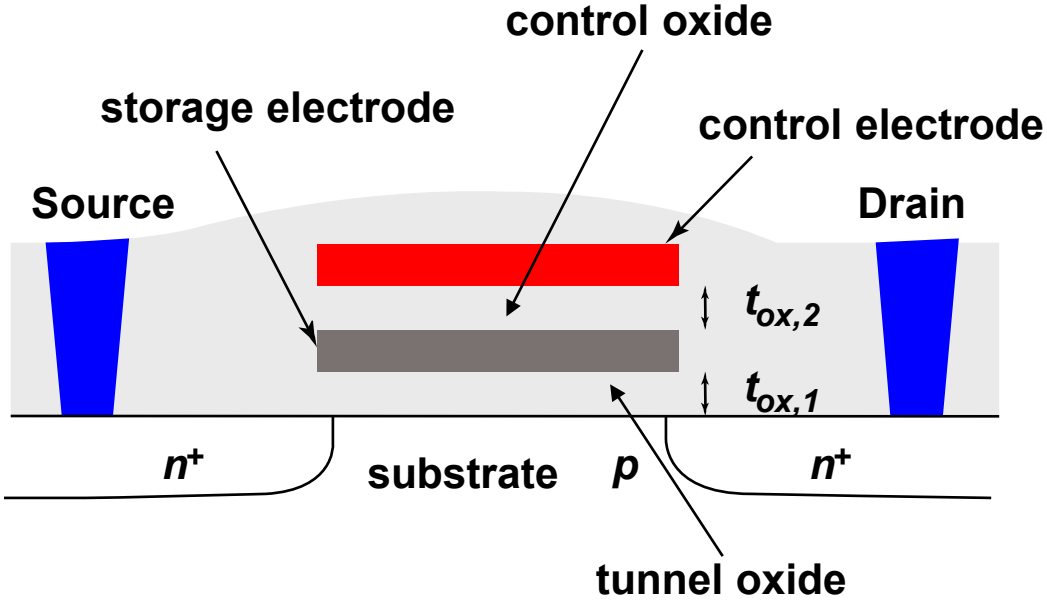
Outline

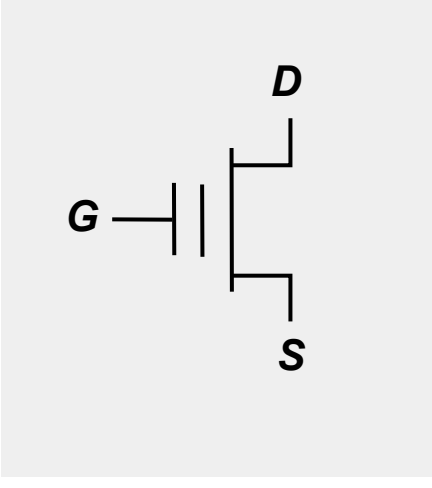
- Overview on highly-integrated information storage
- Volatile semiconductor memory
  - *Static memory with random access (SRAM)*
  - *Dynamic memory with random access (DRAM)*
- Non-volatile semiconductor memory
  - *Flash memory transistors*
  - *Flash architectures*
  - *SONOS storage stacks*
  - *Nitride-ROM memory (NROM)*

MOS Memory Devices

Flash Memory Transistors

- Combination of storage and control electrode in a single gate stack
  - *electrically programmable and erasable*
  - *1T architecture*
  - *storage electrode is not contacted: floating gate*

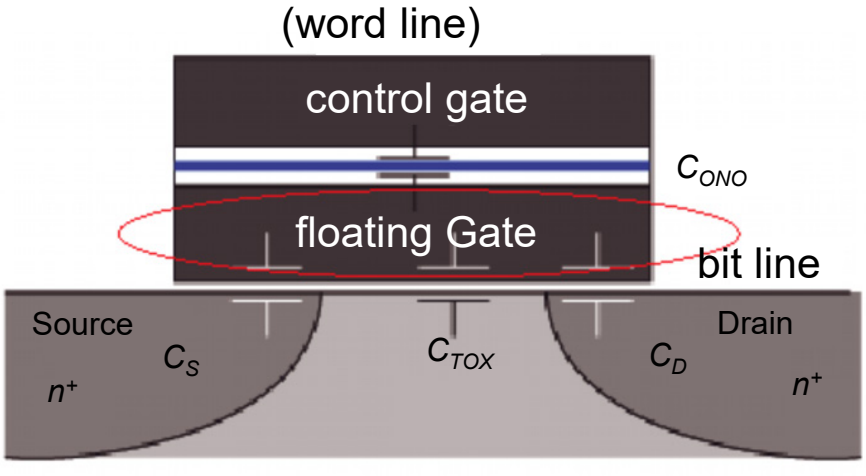


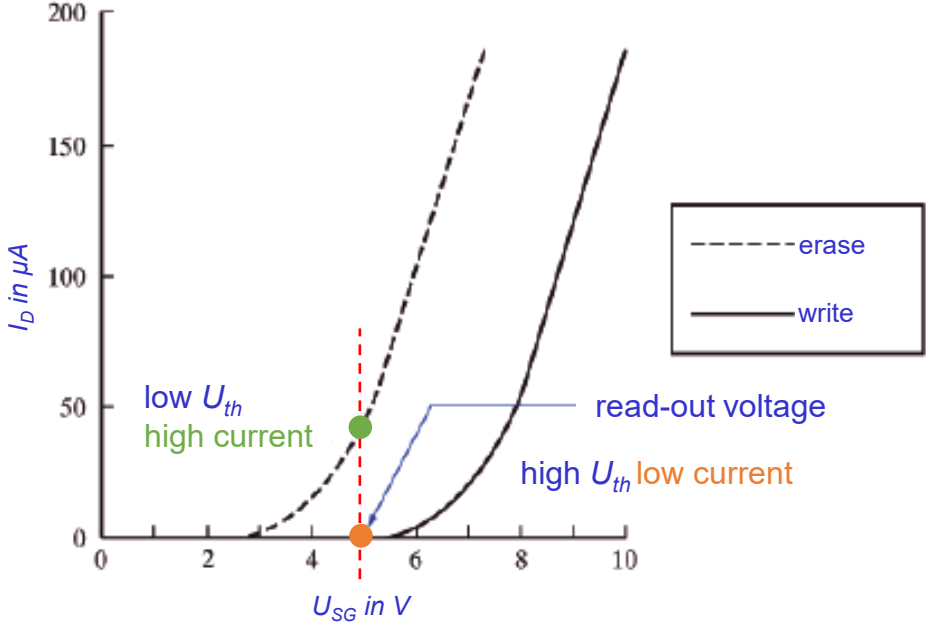


MOS Memory Devices

Flash Memory Transistors

- Readout of Flash memory
  - *stored charges (mostly electrons) increase U_{Th} (see lectures on basics or on hot electrons)*
  - *read-out voltage is selected between lower and upper U_{th}, the associated current is detected*

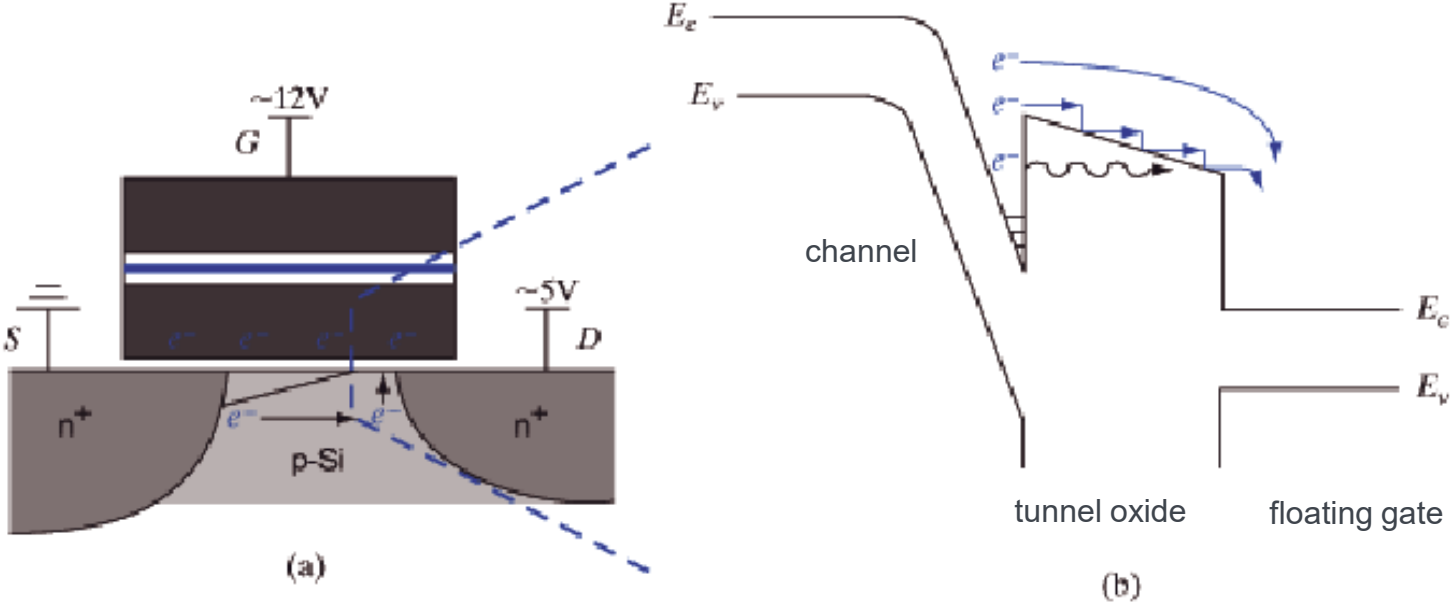




MOS Memory Devices

Flash Memory Transistors

- Programming (writing) of Flash memory
  - *hot electrons generate electron-hole pairs at the drain. Electrons with sufficient energy reach FG directly (if energy is sufficient) or through FOWLER NORDHEIM tunneling*
  - *programming by FOWLER NORDHEIM tunneling*



MOS Memory Devices

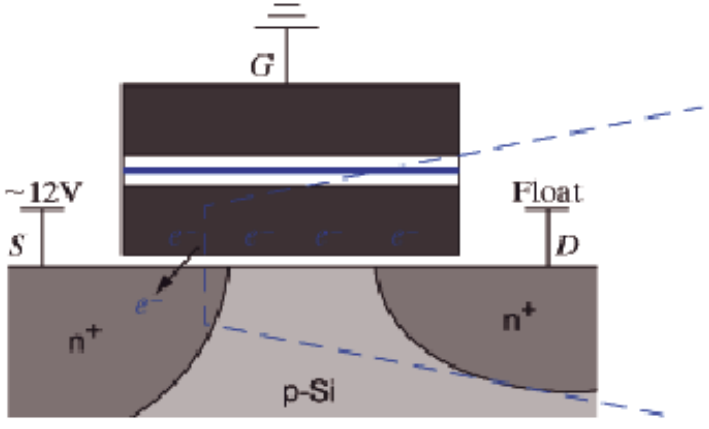
Flash Memory Transistors

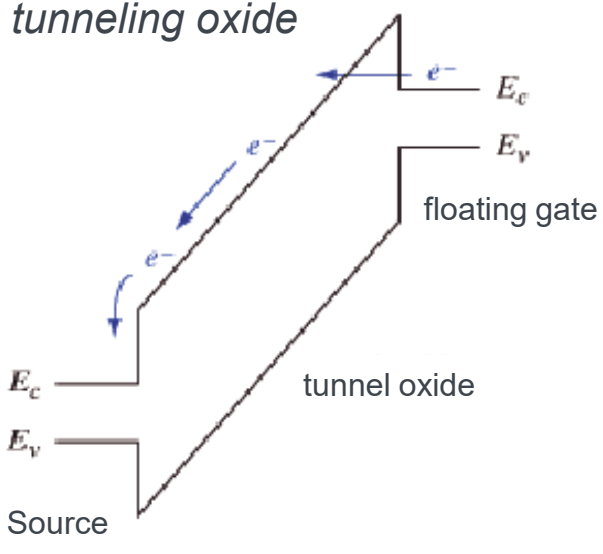
- Erase programming of Flash memory
  - *application of a high voltage between control gate and Source*
  - *FOWLER NORDHEIM tunneling of carriers*

$$J_{FN} \sim AE_{ox}^2 \exp(-B / E_{ox})$$

A, B *constants*

E_{ox} *electric field in tunneling oxide*

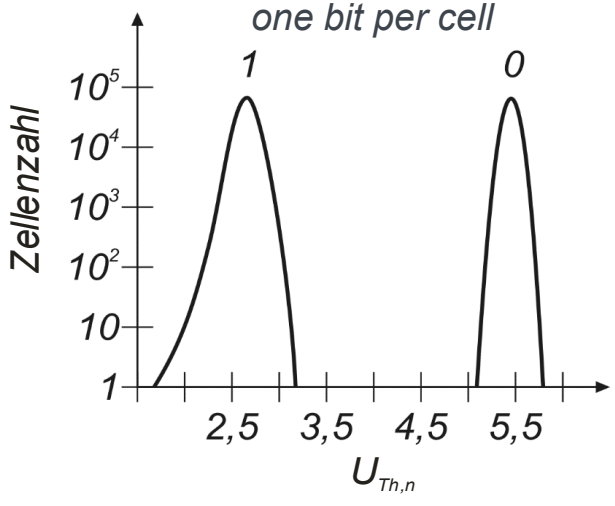


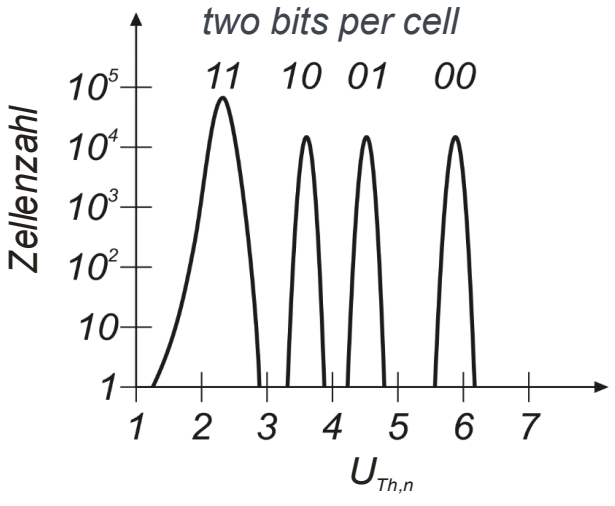


MOS Memory Devices

Flash Memory Transistors

- Increase of memory capacity by multilevel programming (MLC: MultiLevel Cell)
  - *differentiation of multiple U_{Th} levels in a cell*



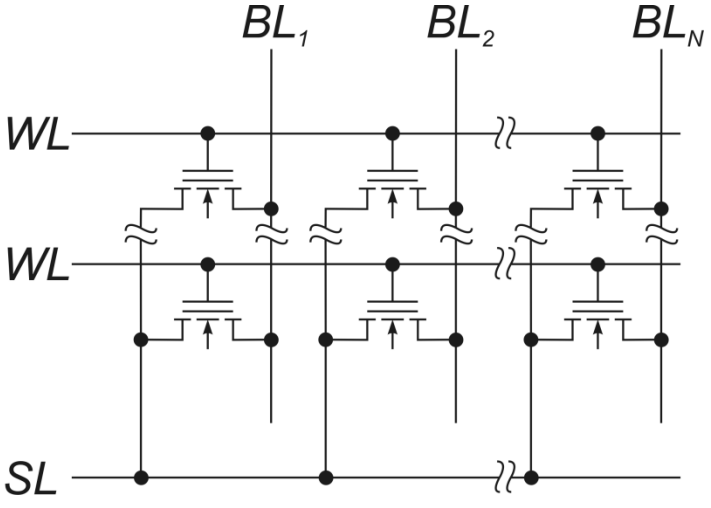


  - *precise control of programmed U_{th}s is mandatory*
    - active „post-writing“ / „-erasing“
  - *high requirements for retention time*

MOS Memory Devices

Flash Architectures

- NOR architecture (e.g. for code memory in microcontrollers)



cell size ≈10F²

due to individual addressing capability

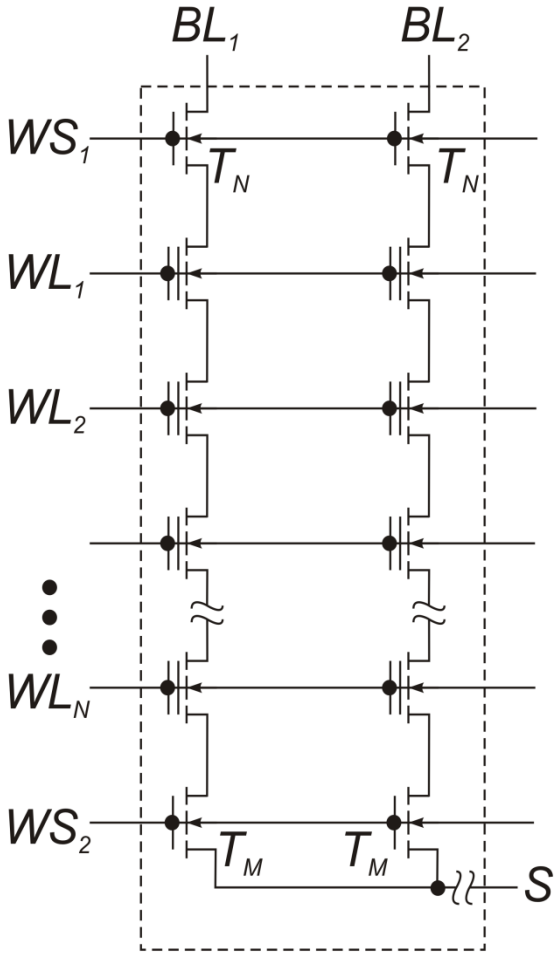
- *fast random access to single bits (<100ns)*
- *fast programming (1-10 µs) by hot electron injection (high currents limit parallel programming)*
- *erase by FN tunneling (Flash) per block*

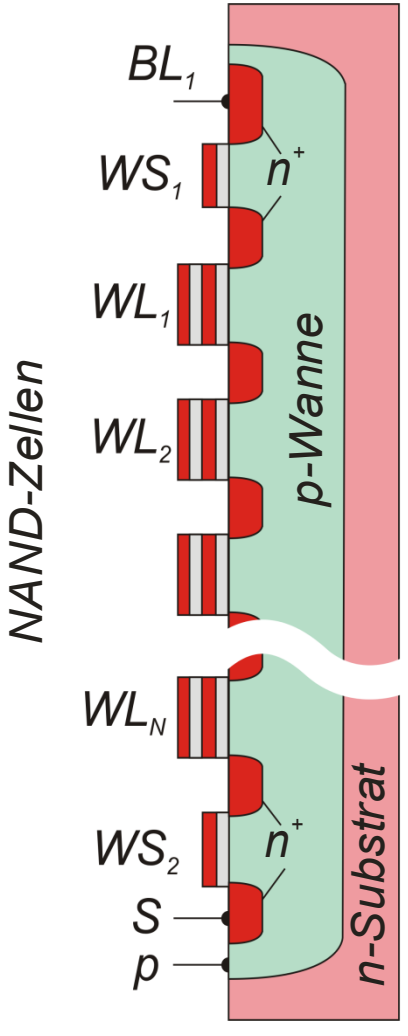
MOS Memory Devices

Flash Architectures

- NAND architecture
  (e.g. for high volume data storage)
  - *Pros*
    - space-saving design (shared S/D)
    - massively parallel programming (FN)
  - *Cons*
    - random access is very slow (>10 µm)
    - programming is very slow (>100 µm)

cell size ≈5F²

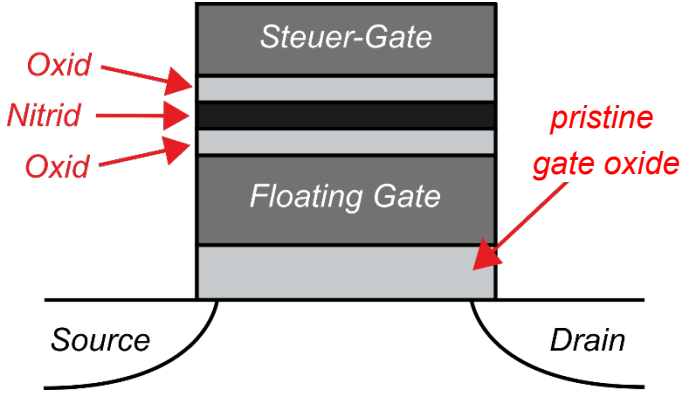




MOS Memory Devices

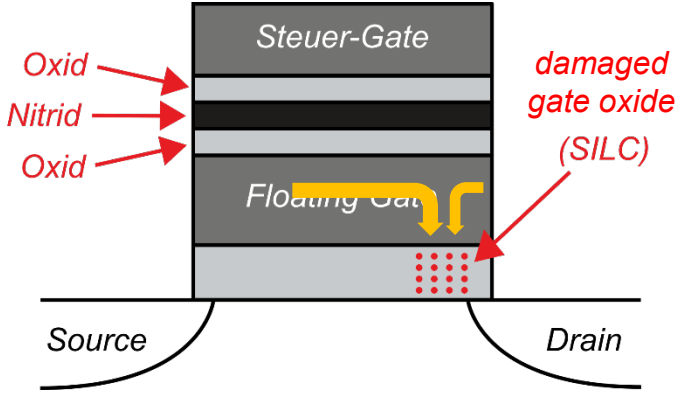
SONOS Memory stacks

- Flash memory: problems
  - *Retention: charge loss*
    - write/erase cycling degrades gate and tunnel oxide
    - SILC: Stress Induced Leakage Currents over defect cascades
  - *Endurance: allowed number of cycles*
  - *U_{Th} shift by defects (traps) in tunnel oxide*



before P/E cycling

retention approx. 100 yrs.

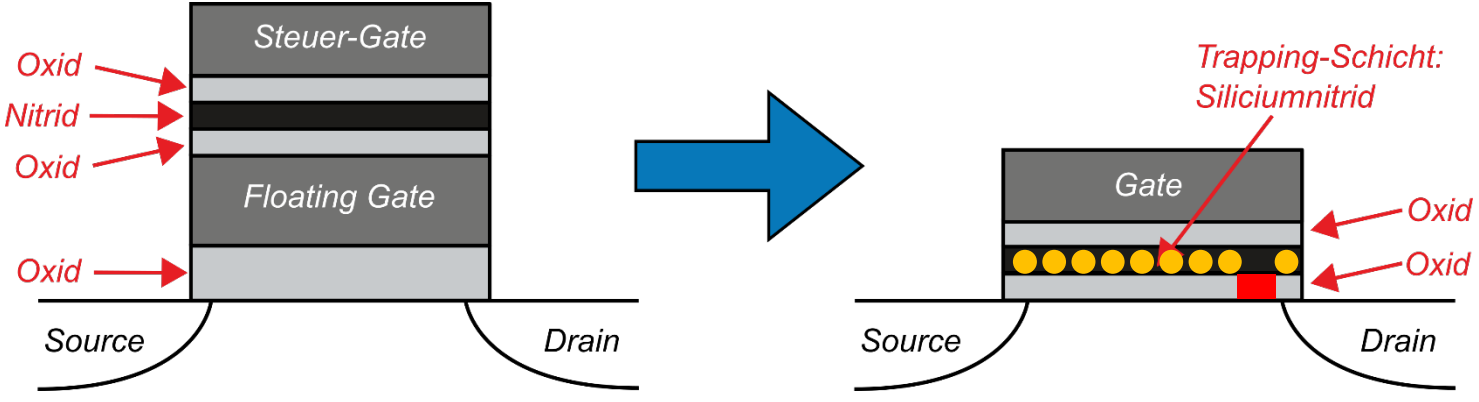


after 10^5 P/E cycles

retention approx. 10 yrs.

MOS Memory Devices
SONOS Storage Stacks

- Solution: Flash memories with storage layer from silicon nitride
  - *dielectric ONO stack (already known from reinforcement of control oxide)*

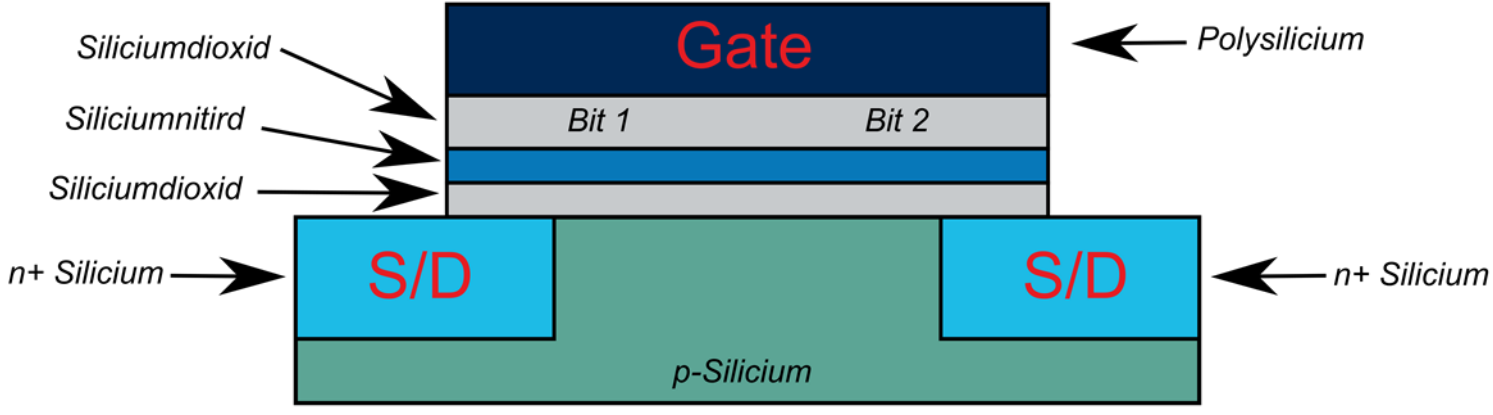


  - *Stationary charge storage in adhesion points of silicon nitride*
    - Better integration (lower vertical layer thickness)
    - Only local, no full charge loss if tunnel oxide is locally damaged

MOS Memory Devices

Nitride ROM Memory

- NROM memory based on ONO layer stack
  - *dedicated operation mode of SONOS memories (preferrably in NAND architecture)*
  - *device is symmetrical, as any „normal“ MOSFET*

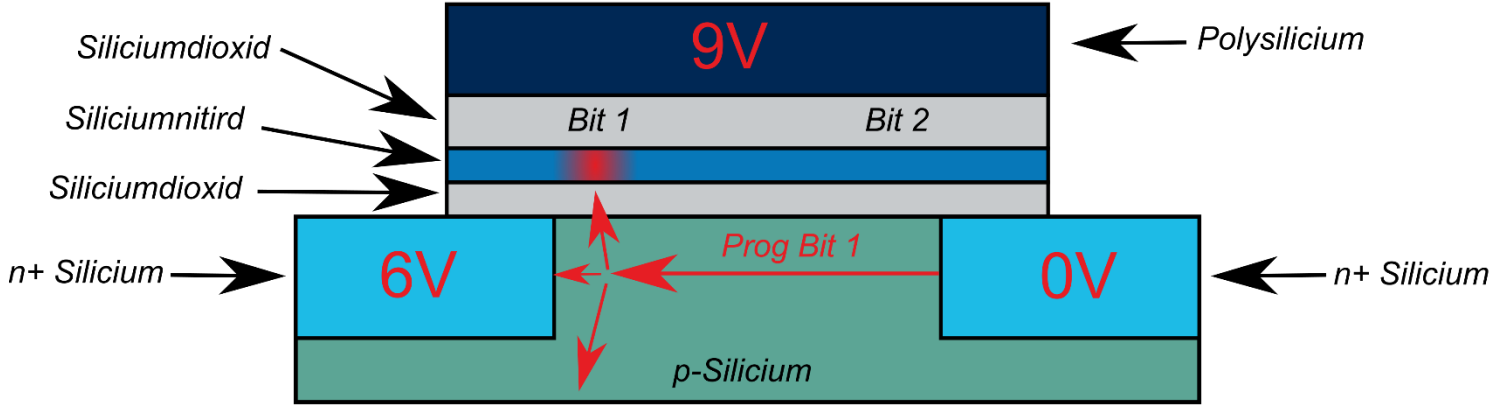


- *Brand names*
  - TWIN Flash, Mirror Bit,

MOS Memory Devices

Nitride ROM Memory

- NROM memory based on ONO layer stack
  - *programming*
    - injection of hot channel electrons (high current)
    - local storage of electrons in silicon nitride close to drain area

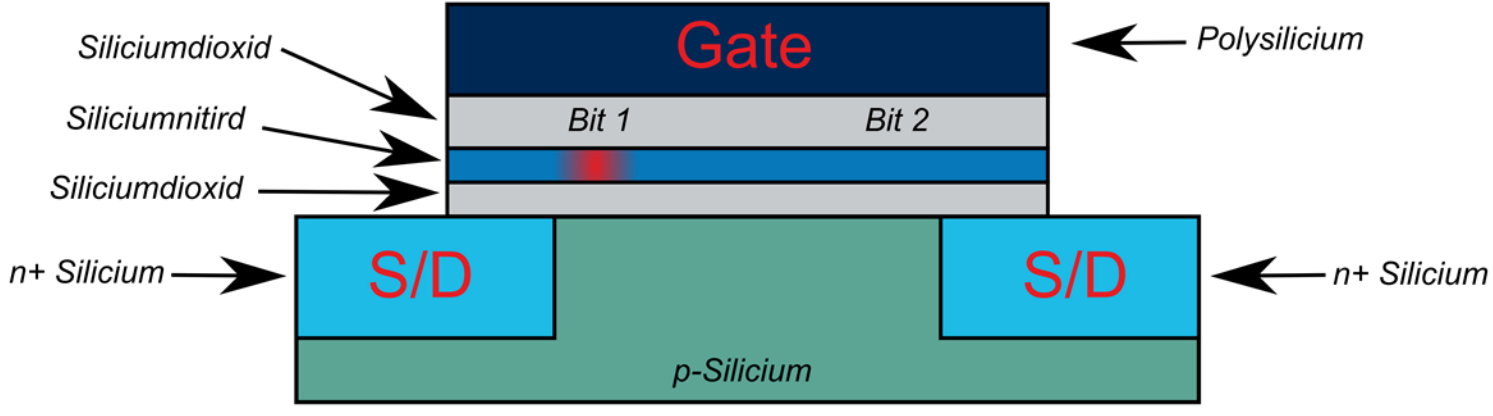


  - *the individual assignment of the contacts as source or drain is done by voltage polarity*

MOS Memory Devices

Nitride ROM Memory

- NROM memory based on ONO layer stack
  - *programming*
    - injection of hot channel electrons (high current)
    - local storage of electrons in silicon nitride close to drain area

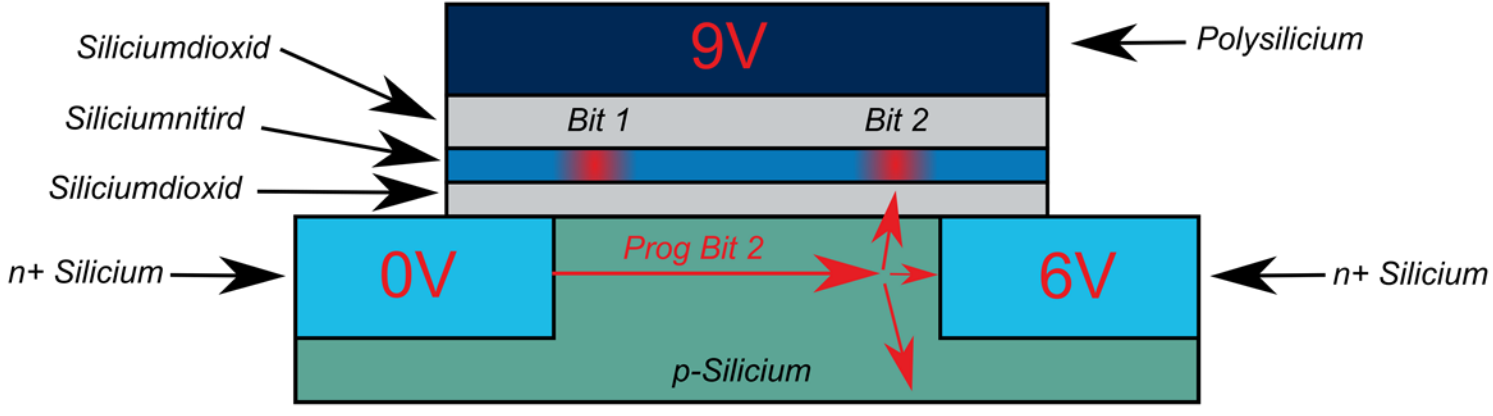


  - *the individual assignment of the contacts as source or drain is done by voltage polarity*

MOS Memory Devices

Nitride ROM Memory

- NROM memory based on ONO layer stack
  - *programming*
    - injection of hot channel electrons (high current)
    - local storage of electrons in silicon nitride close to drain area

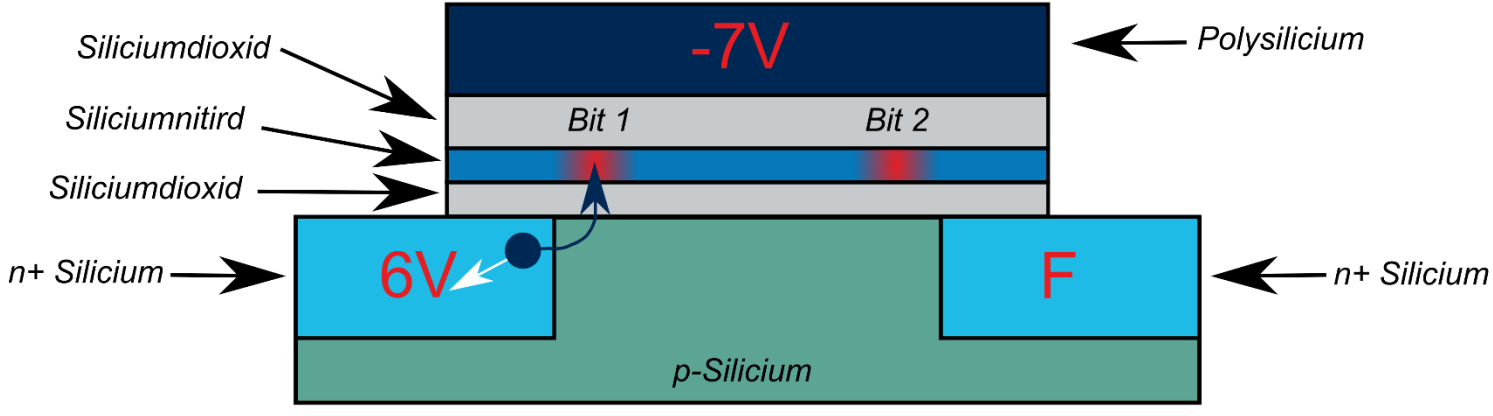


  - *a second bit can be addressed by changing source and drain*
    - 2-bit storage, doubling the memory capacity

MOS Memory Devices

Nitride ROM Memory

- NROM memory based on ONO layer stack
  - *erasing*
    - local injection of hot holes from drain into nitride layer
    - recombination of injected holes with trapped electrons

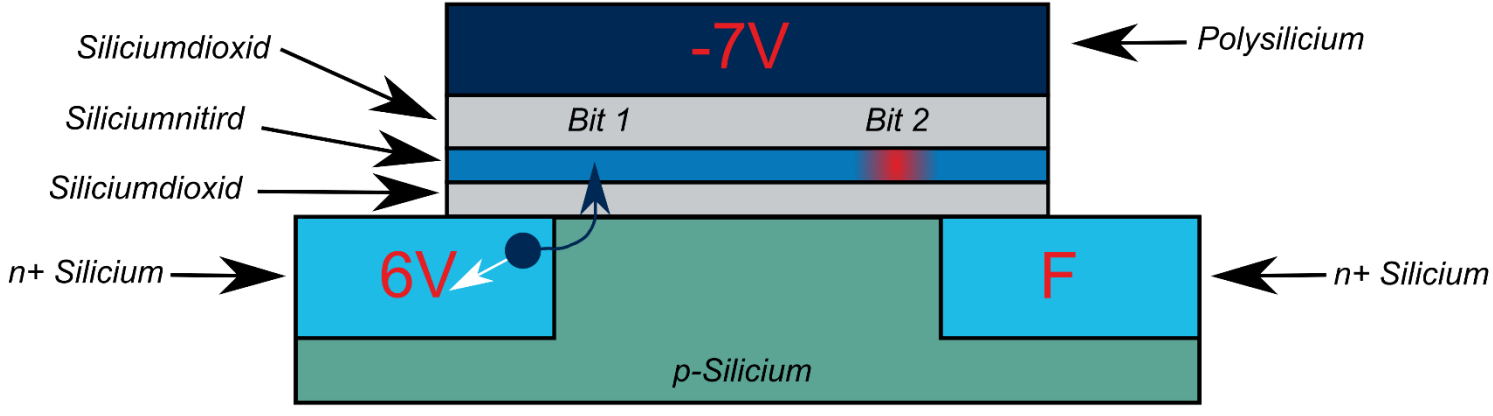


  - *disadvantages*
    - high current consumption due to low ratio of trapped vs. injected carriers
    - limitation of lateral scaling by charge packet extension and misalignment of electron and hole packets

MOS Memory Devices

Nitride ROM Memory

- NROM memory based on ONO layer stack
  - *erasing*
    - local injection of hot holes from drain into nitride layer
    - recombination of injected holes with trapped electrons

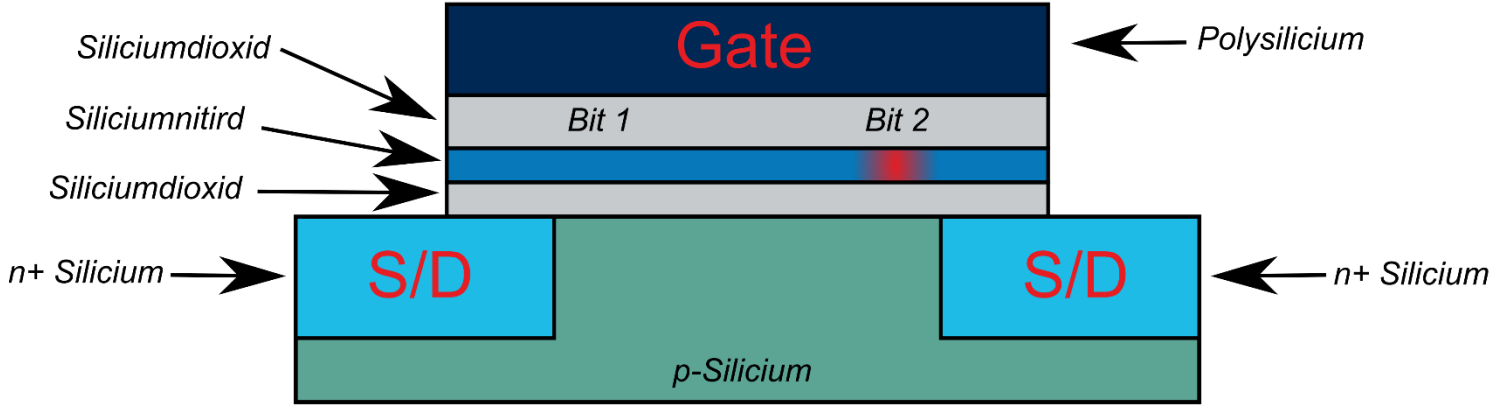


  - *disadvantages*
    - high current consumption due to low ratio of trapped vs. injected carriers
    - limitation of lateral scaling by charge packet extension and misalignment of electron and hole packets

MOS Memory Devices

Nitride ROM Memory

- NROM memory based on ONO layer stack
  - *erasing*
    - local injection of hot holes from drain into nitride layer
    - recombination of injected holes with trapped electrons

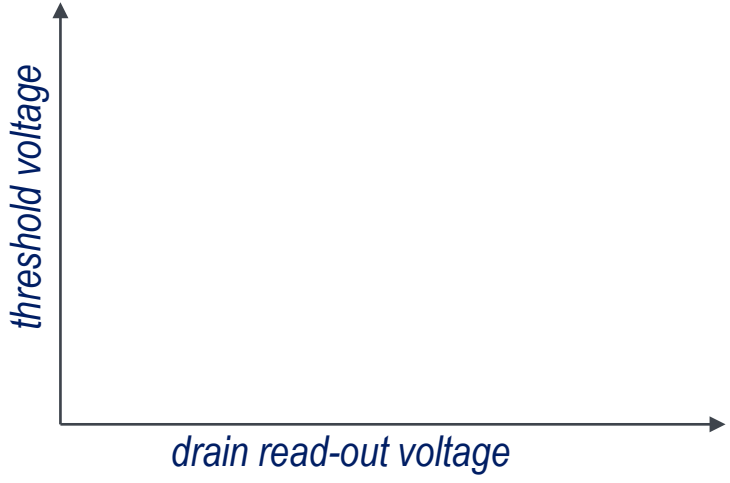


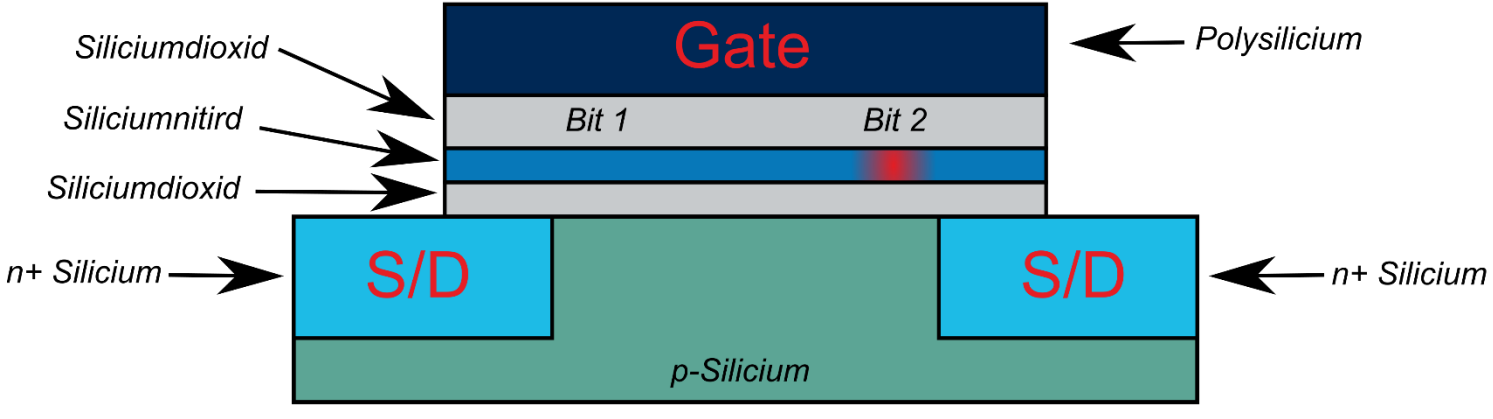
  - *disadvantages*
    - high current consumption due to low ratio of trapped vs. injected carriers
    - limitation of lateral scaling by charge packet extension and misalignment of electron and hole packets

MOS Memory Devices

Nitride ROM Memory

- NROM memory based on ONO layer stack
  - *reading*
    - operation in saturation mode
      $V_{DS} \geq V_{GS} - V_{TH}$
    - charge at source side defines V_{TH}
    - V_{TH} is increasing with (electron) charge

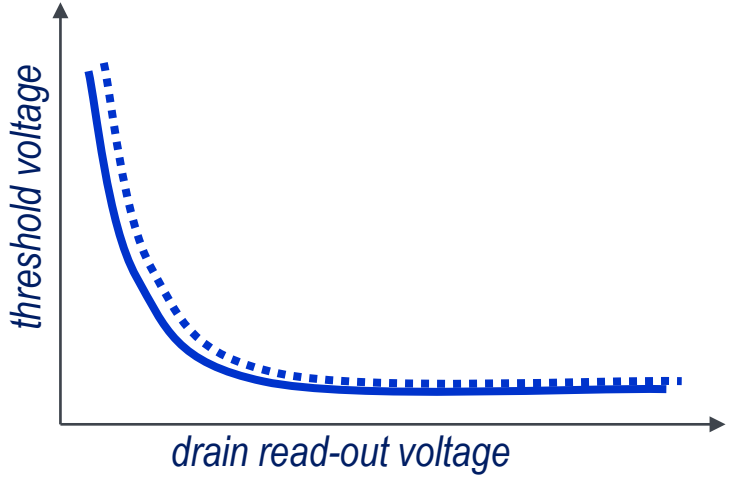


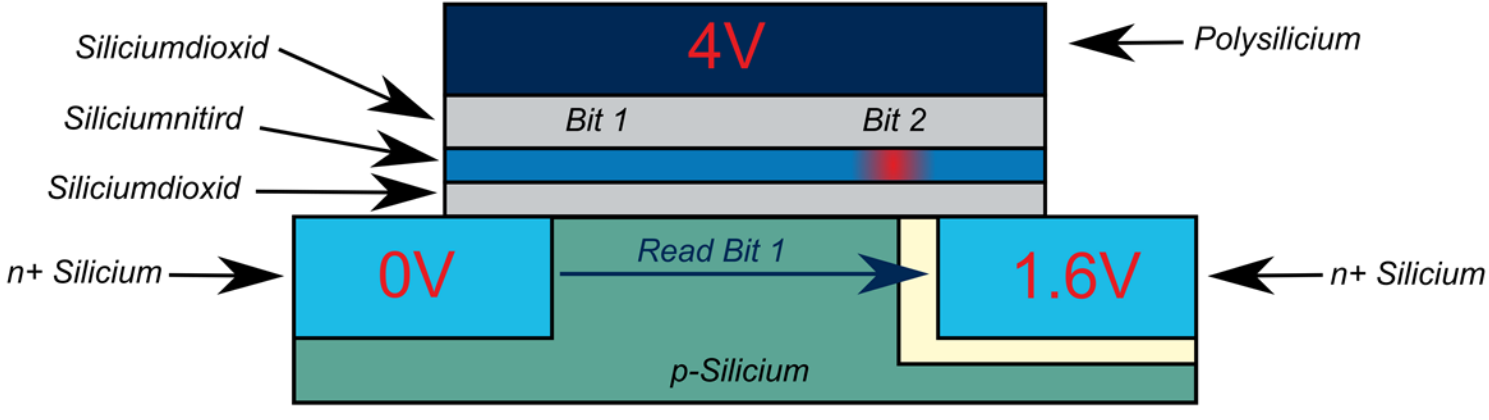


MOS Memory Devices

Nitride ROM Memory

- NROM memory based on ONO layer stack
  - *reading*
    - operation in saturation mode
      $V_{DS} \geq V_{GS} - V_{TH}$
    - charge at source side defines V_{TH}
    - V_{TH} is increasing with (electron) charge

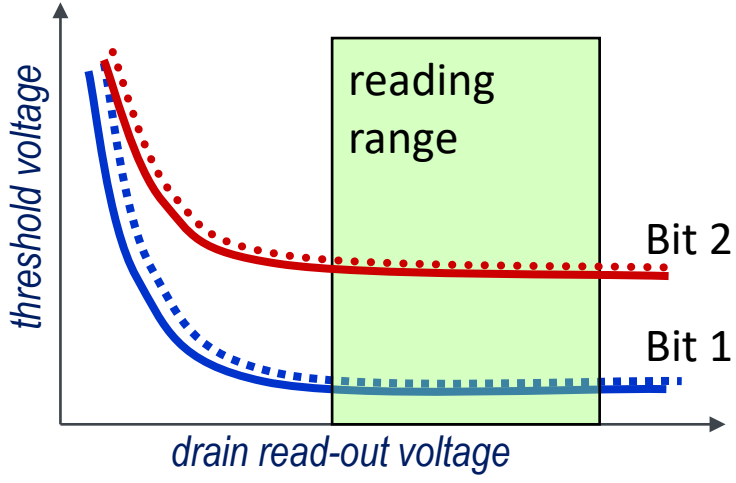


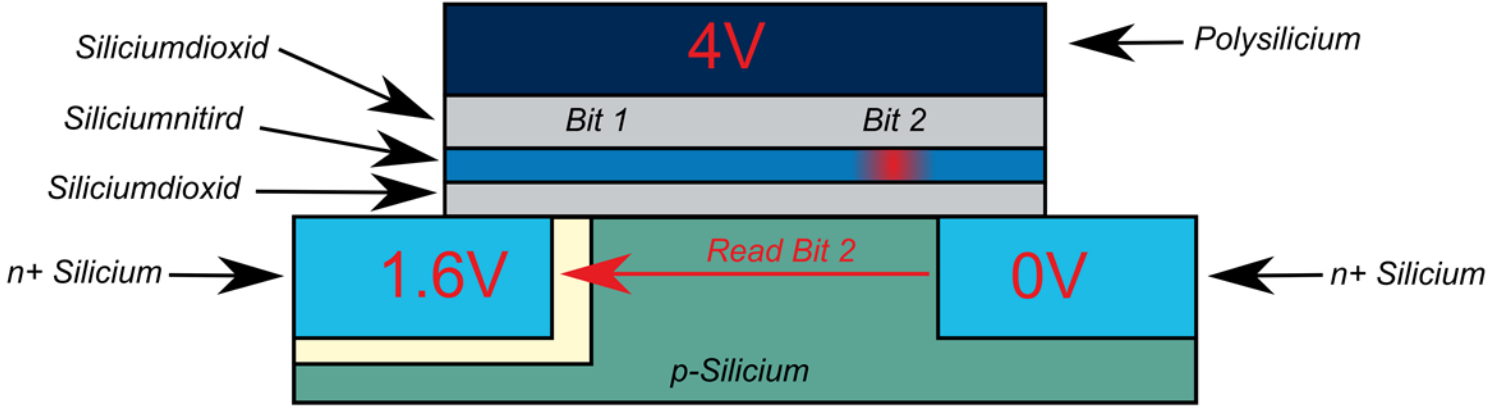


MOS Memory Devices

Nitride ROM Memory

- NROM memory based on ONO layer stack
  - *reading*
    - operation in saturation mode
      $V_{DS} \geq V_{GS} - V_{TH}$
    - charge at source side defines V_{TH}
    - V_{TH} is increasing with (electron) charge





MOS Memory Devices

Outline

- Overview on highly-integrated information storage
- Volatile semiconductor memory
  - *Static memory with random access (SRAM)*
  - *Dynamic memory with random access (DRAM)*
- Non-volatile semiconductor memory
  - *Flash memory transistors*
  - *Flash architectures*
  - *SONOS storage stacks*
  - *Nitride-ROM memory (NROM)*

Thanks for your attention!