Principle and Architecture of Ultra Low Power SRAM Device

Chih Cheng Hsiao 2024/03/26

Content

1. **Features**
2. **Compare with current technology**
3. **Circuits**
4. **Place and Louting**
5. **Power comparison**
6. **Patent family**

Features 1

- 1. The power of the new SRAM device is 10%~20% compare to current technology.
- 2. The area of the new SRAM device is around the same compare to current technology.
- 3. The implementation of new SRAM device is simple and easy.
- 4. The circuit interface of new SRAM device is the same as current technology.
- 5. The operation voltage is the same as digital voltage.(1.0 V at 28nm process)
  => This invention is not base on voltage scaled down.
- 6. The new SRAM device requires 4 metal layers.
- 7. The new SRAM device is mainly applied to the L1/L2 cache of the CPU/GPU in the SOC chip.
- 8. The clock speed is the same as current technology.

Compare with current technology 1

	New Invention	Dolphin/Surecore/Etc.
Method	SRAM array power almost goes to 0.	Reduce 50% of bit line power.
Power	Only 10~20% compare to foundry solution.	80 ~ 90% compare to foundry solution.
Area	90%~110% compare to foundry solution.	110% compare to foundry solution

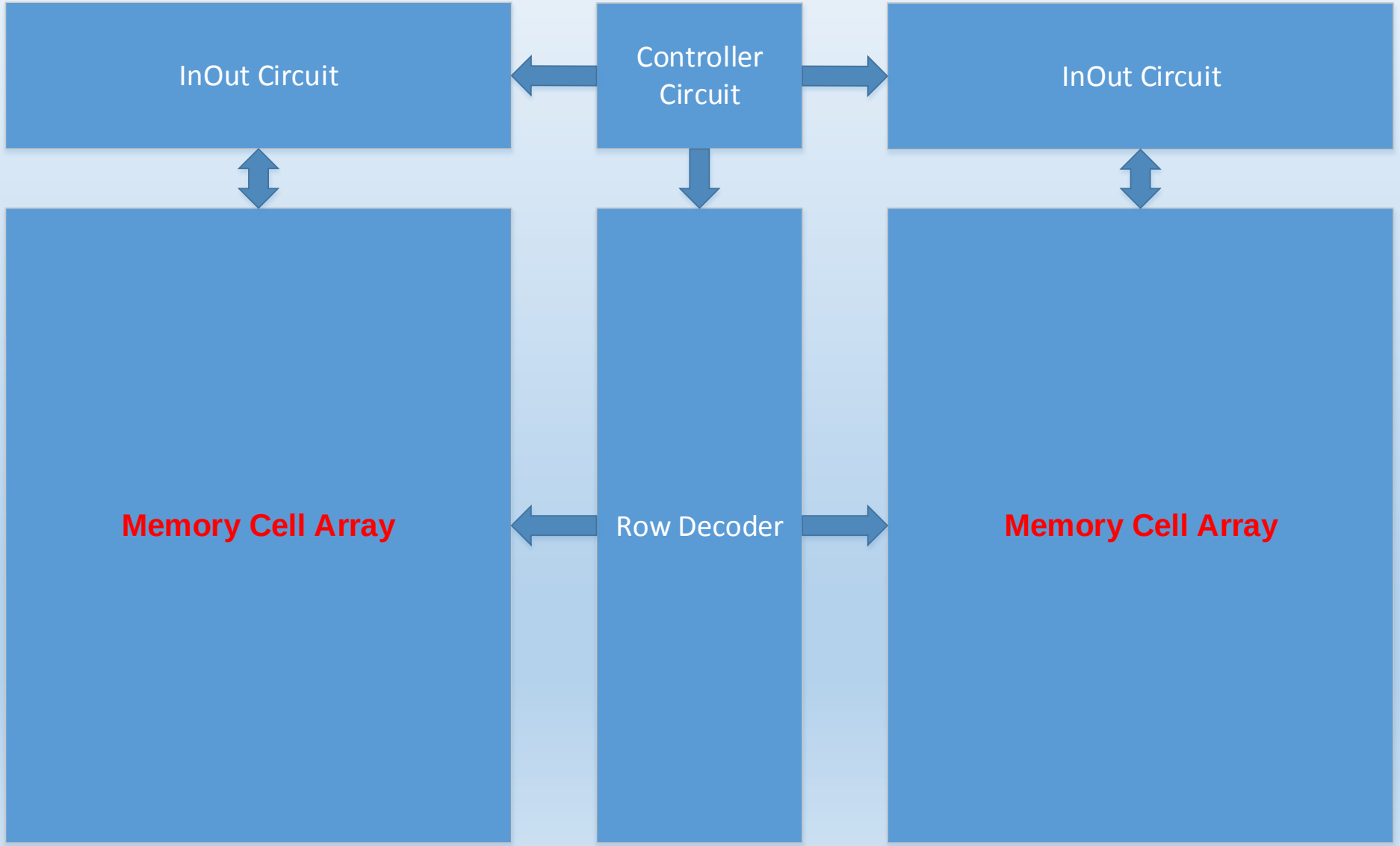
Compare with current technology 2

256w * 64b * 8 blocks	TSMC	SureCore/Dolphin	The New One
Power of bit line	**1**	**0.5**	**1**
Power of cell array	**30**	**30**	**0.5**

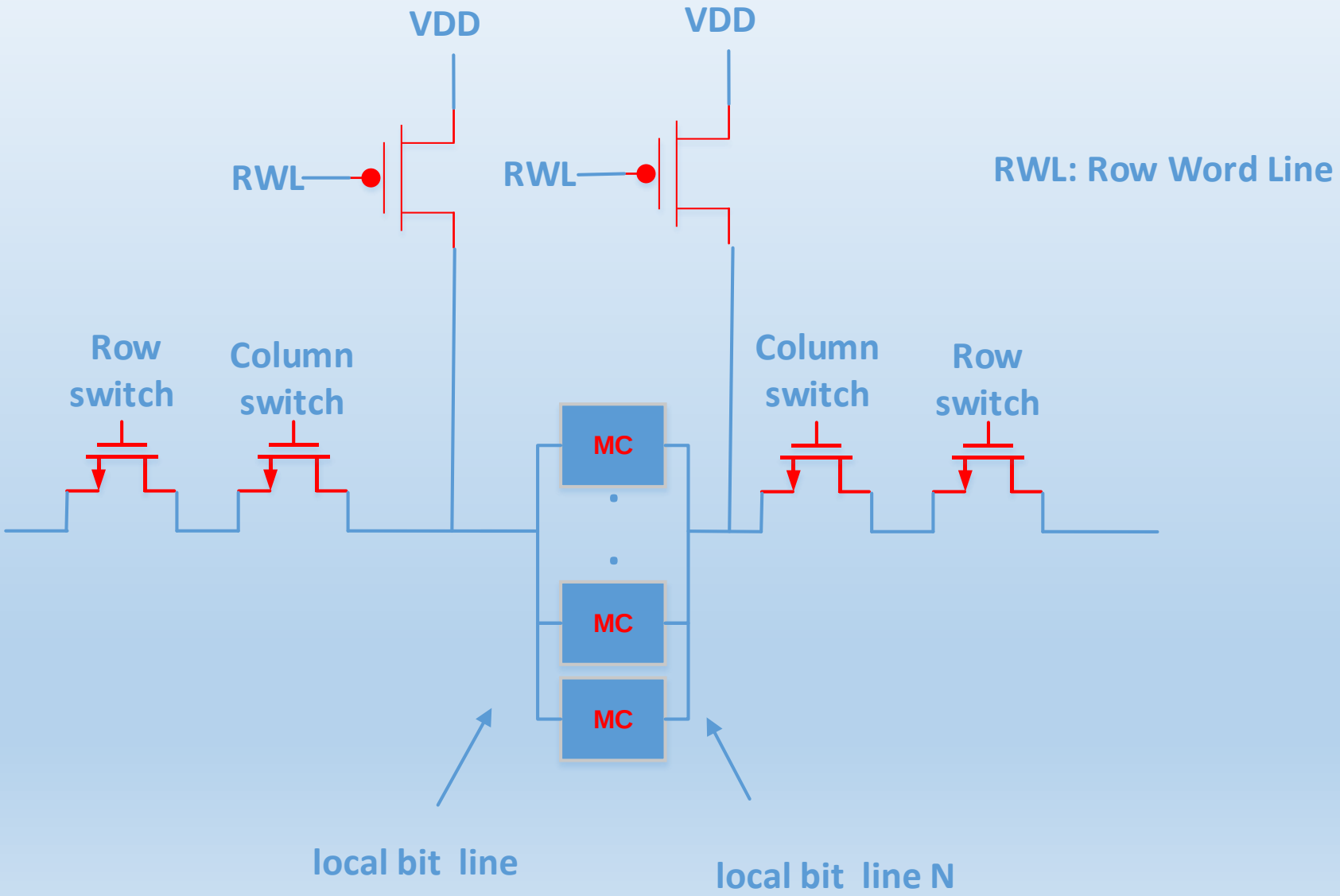
256w * 64b * 8 blocks : Each bit line with 256 SRAM cells, each data word has 64 bits.

Unit: uA/Mhz

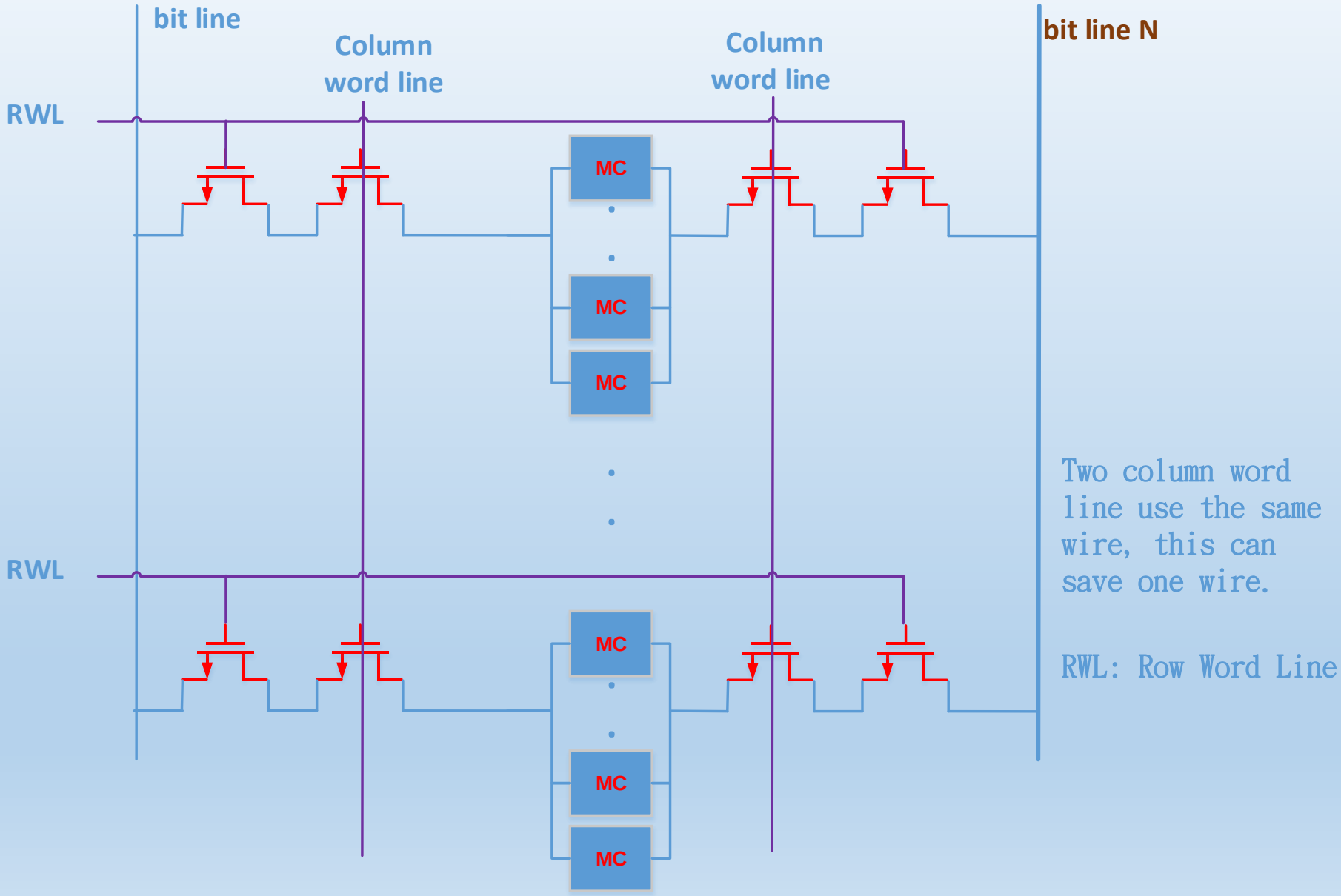
Circuits : Memory Overview



Circuits : Memory Cell Group



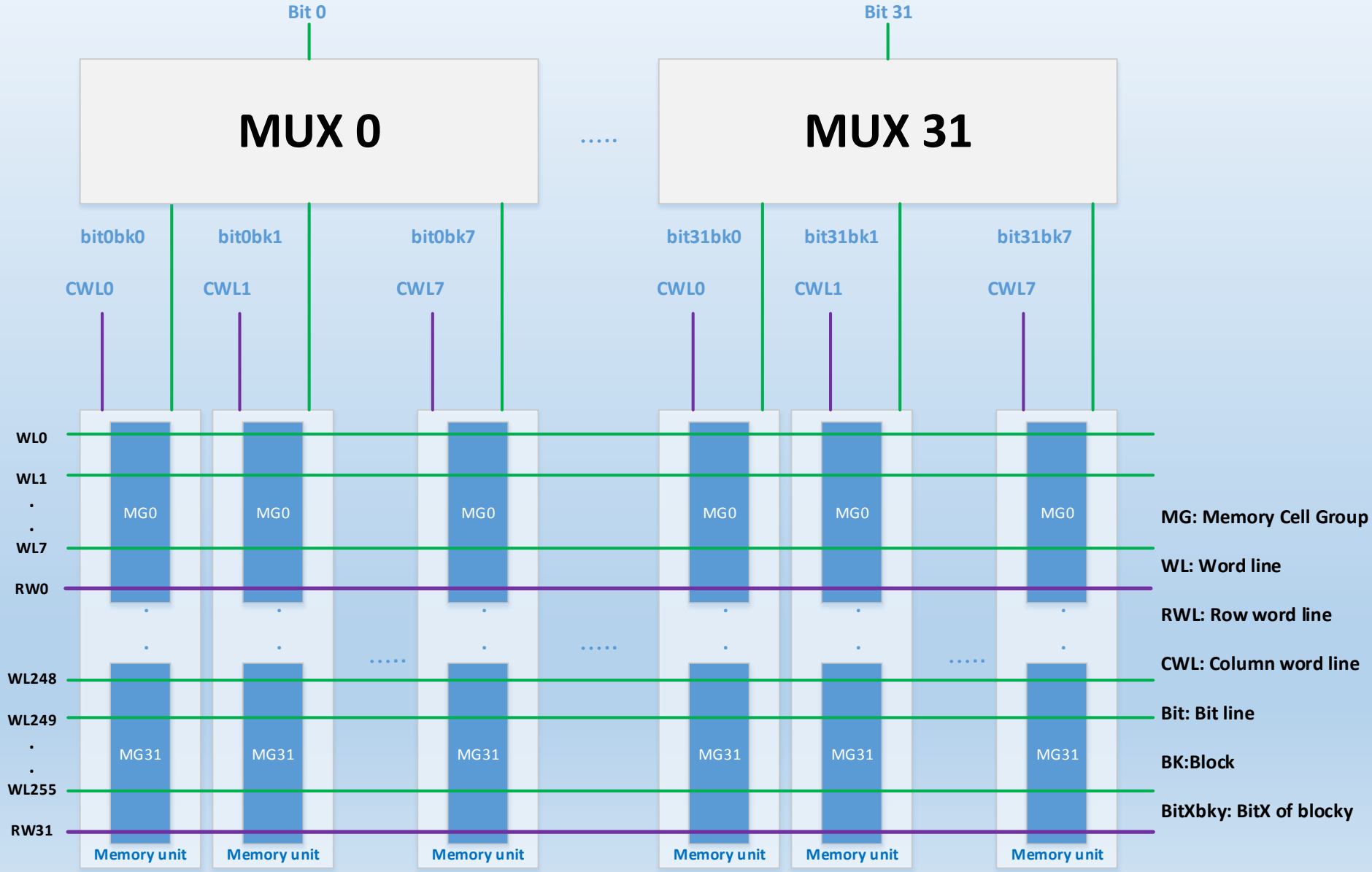
Circuits : Memory unit



Two column word line use the same wire, this can save one wire.

RWL: Row Word Line

Circuits : Cell Array



Circuits : Reliability and stability

- The SRAM cell use the foundry’s. So the reliability and stability is the same as foundry’s.

Circuits : Read Margin

- Because there insert more one transistor between SRAM cell and bit line, the read margin is no issue.

- For read access speed, two solution:

  (a) Use independent double gate transistor to replace the two access switch transistor.

  (b) Double the size of two access switch transistor. Due to read operation is operate at liner region, so the speed is the same as current technology

Circuits : Write Margin

- Write margin can use negative bit line to improve it.
- The write access speed can improve as read access speed.

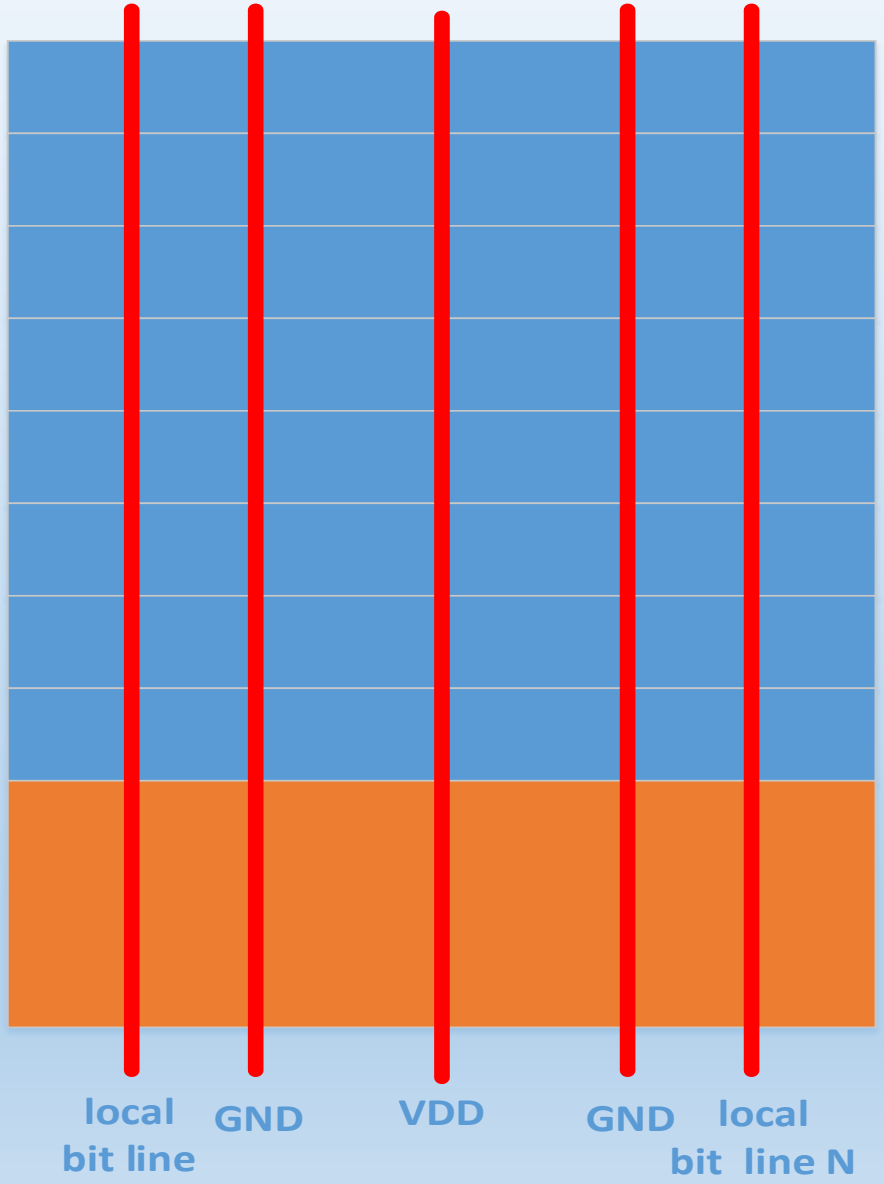
Circuits : Memory Array Explanation 1

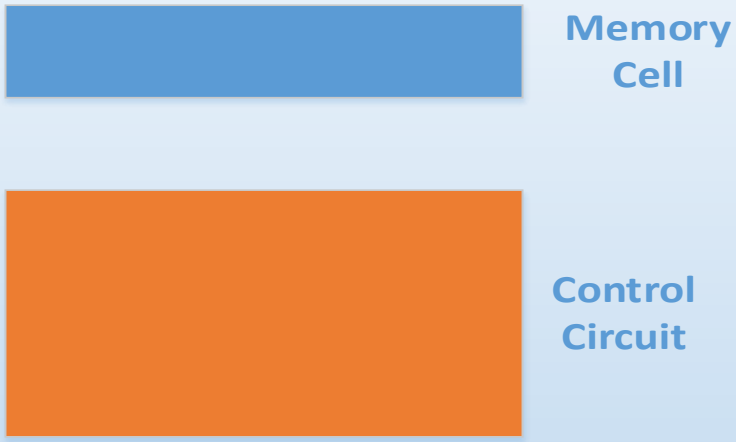
- There has 8 blocks, one block has 32 bit line set. One bit line set has bit line and bit line N(not shown).
- The CWLX (CWL0 ~CWL7) is for turn on the gate switch of corresponding blocks.
- There has row word lines(RWL)(not shown). The selected row word line will also for turn on the gate switch. For example WL0 – WL7 are with RWL0, WL248 – WL255 are with RWL31. So when SRAM cell is selected by WL0 – WL7, RWL0 will active, etc.
- Both CWLX and corresponding row word line are active, the two gate switch will turn on.

Circuits : Memory Array Explanation 2

- Thus only the bit lines of selected block will charge or discharge when read or write.
- And the un-select blocks will only charge or discharge the local bit lines. And the local bit line is very short, the connected SRAM cell is very few, so the this power is almost can ignore.
- Due to only the bit lines of selected block will charge/discharge, the pre-charge can only active in read cycle.

Place and Louting : Memory Unit Metal Layer 1

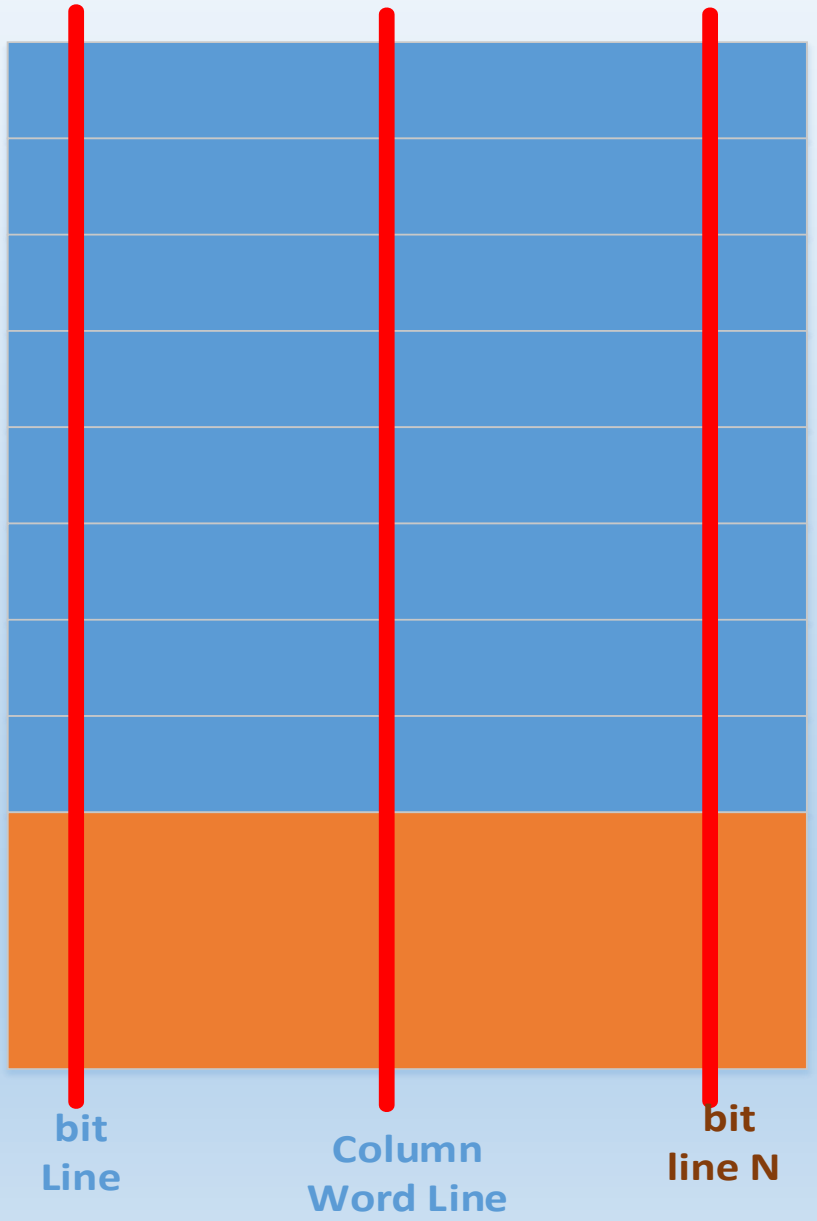




The placement is for memory unit. Metal layer 1.

The local bit line/local bit line N may place metal layer 0.

Place and Louting : Memory Unit Metal Layer 2

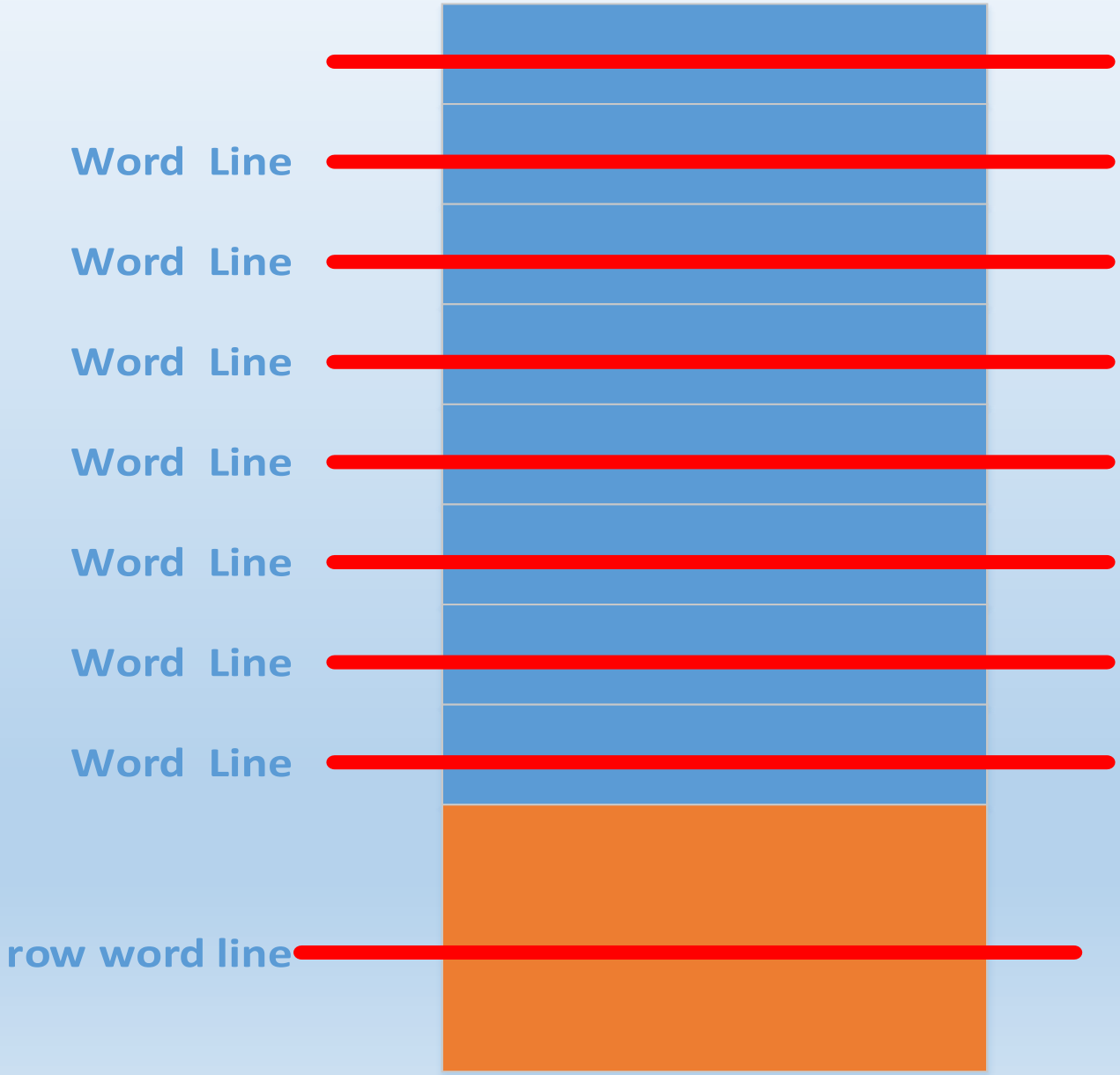


The placement is for memory unit.
Metal layer 2.

The two Column Word Line could be the same line for power saving.

Metal layer 2 can exchange with metal layer 3.

Place and Louting : Memory Unit Metal Layer 3



The placement is for memory unit.
Metal layer 3.

Power of Column Word Line

- Column switch should be parked before the selected block is change. So the column word line doesn't toggle before the selected block is change. This can save power.

- For example , a bit line is connected to 256 SRAM cells, due to the sequence access characteristic of memory, the column word line will require toggle only one cycle of 256 sequence access cycles. In cache memory, this effect will much more, so the column word line will require toggle only one cycle of 512/1024 access cycles. Then the column word line power is divided by 256, 512, 1024.

Power comparison 1

- Definition : Selected bit line power => **SBLP**
  Unselected Array power => **UAP**
  Decoder power => DP
  Cell read power => CRP
  Cell write power => CWP
  Sense amplifier power => SAP
  Column word line power => **CWLP**
  Row word line power => RWLP
  Word line power => WLP
  Pre-charge Power => PCP

Power comparison 2

- For example a memory device 256 bits/line X 32 bits/word X 8 blocks. In current design, one word line will enable 32bits X8 blocks = 256 bit cells. That is mean 256-32 = 224 bit cells waste active powers, because there only need 32 bit cells (**SBLP**) to output. For SRAM cells, one cell active current is around 66 uA, So 224 bit cells will waste 66 uA X 224 = 14784 uA (**UAP**). This is the main reason, that my invention makes the SRAM power to such low.

Power comparison 3

- Current technology read power :

  SBLP + DP + CRP + RWLP + WLP + PCP + UAP + SAP

- Current technology write power :

  SBLP + DP + CWP + RWLP + WLP + PCP + UAP

- New invention read power :

  SBLP/2 + DP + CRP + RWLP + WLP + PCP + CWLP/256 + UAP/32 + SAP

- New invention write power :

  SBLP/2 + DP + CWP + RWLP + WLP + CWLP/256 + UAP/32

Power comparison 4

- For 32cells/bit line X 32 bit line X 8 block (1K bytes) to 2048 cells/bit line X 64 bit line X 8 block (128K bytes), the UAP power is around 70% - 90% of SRAM power.
  And the other power is around 50 ~ 70 % of new invention, so the power of new invention is only 10 % ~ 20 % compare to current technology.

Power Consumption Information

✓28 nm process.

✓SRAM Device.

✓Speed : Up to 2.5 Ghz.

Bit Line Length X Data Width X Block Numbers	New invention Unit:(uA/MHz)
256 X 32 X 8 Read (8k bytes)	1.77
256 X 32 X 8 Write (8k bytes)	1.54
256 X 64 X 8 Read(16k bytes)	3.25
256 X 64 X 8 Write(16k bytes)	2.78

Patent Family

Country	Application number	Status
U.S.A	US15/950,183	Granted
U.S.A	US15/949,077	Granted
U.S.A	US18/182,382	Pending
China	CN201610483478.3A	Granted
China	CN201620653733.XU	Granted
China	CN202110789914.0A	Pending
Taiwan	TW105118704A	Granted
India	201614021619	Granted
India	202115041660	Pending

END