Quantitative Modeling of Racetrack Memory

A Tradeoff among Area, Performance, and Power

[1]Chao Zhang, [1]Guangyu Sun, [1]Weiqi Zhang,
[2]Fan Mi, [2]Hai Li, [3]Weisheng Zhao

[1] Peking University, China

[2] University of Pittsburgh, U.S.A.

[3] Beihang University, China

Outline

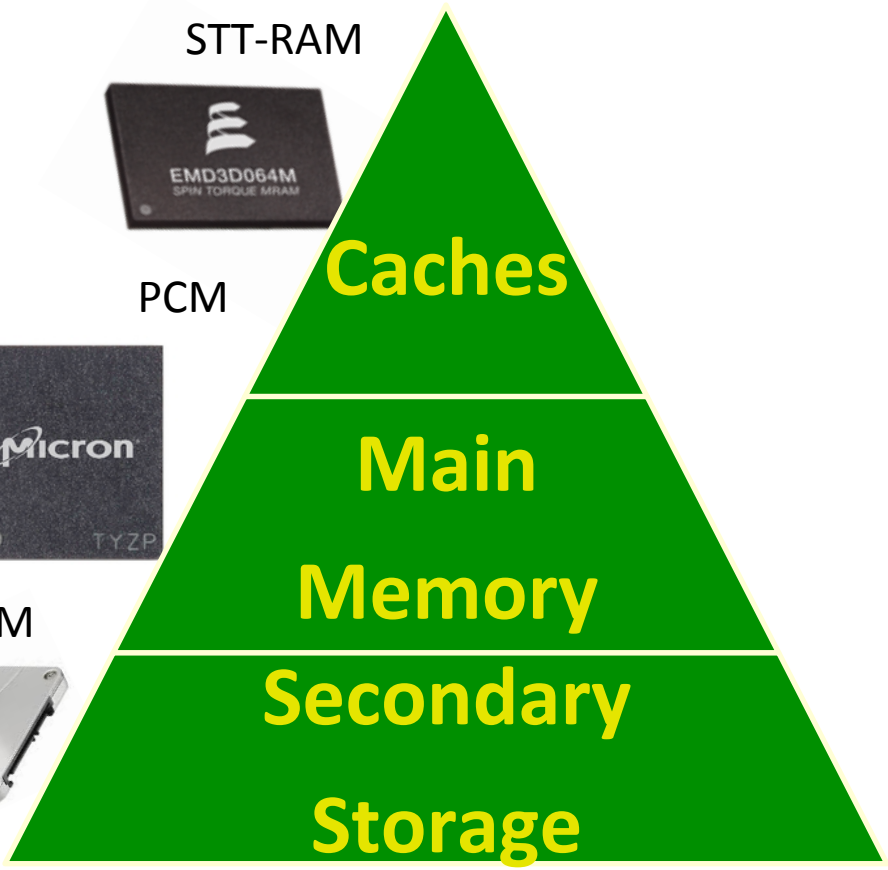
- **Background**
- **Layout Modeling**
- **Case Study**
- **Conclusions**

Non-Volatile Memory is Attractive

- High storage density
- Low standby power
- Good scalability
- Immunity to soft error

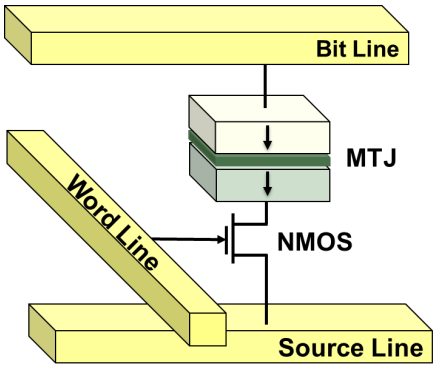
…

Universal Memory for all these layers?



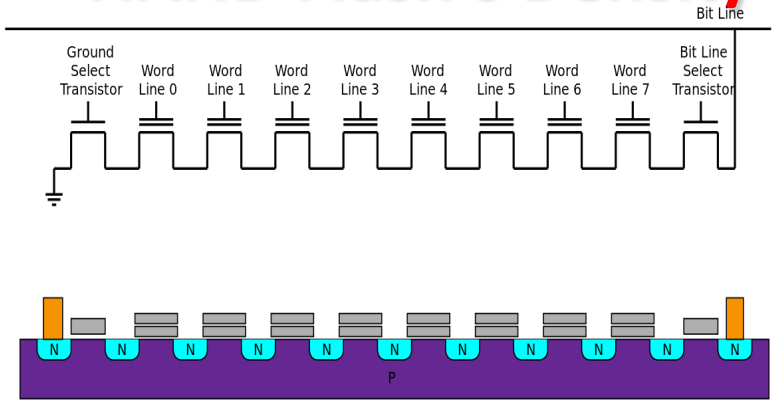
Racetrack is a Potential Universal Memory

STT-RAM's Speed



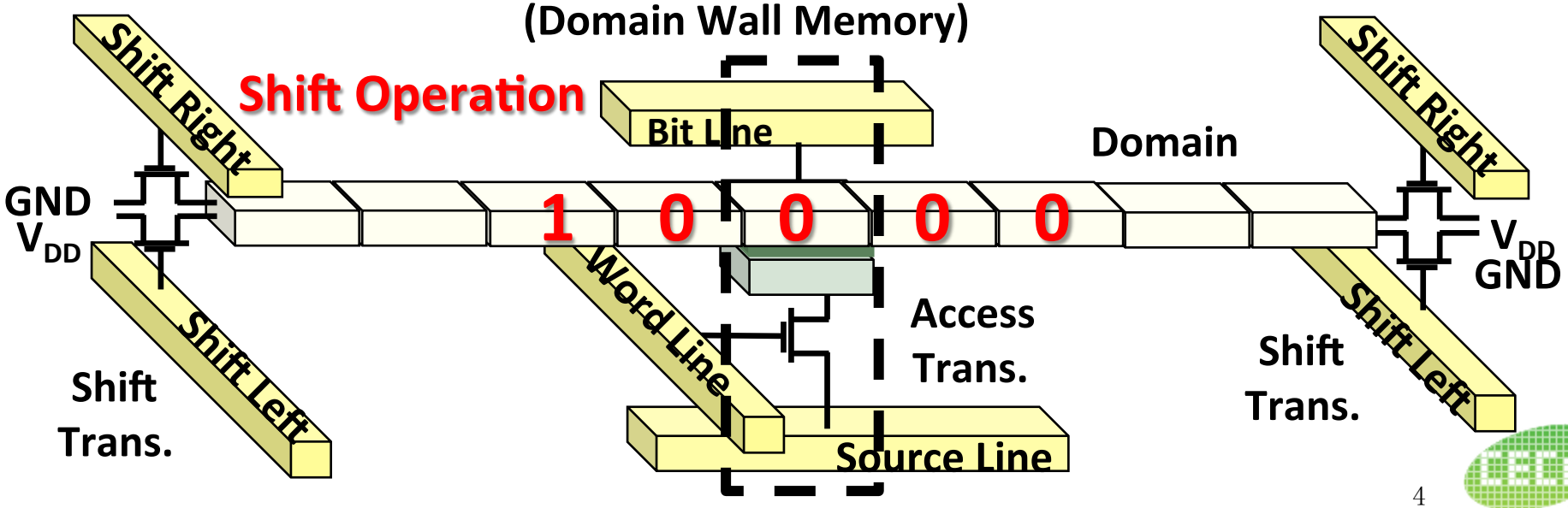
STT-RAM

NAND Flash's Density



NAND Flash

Racetrack Memory (Domain Wall Memory)



Related Work on RM

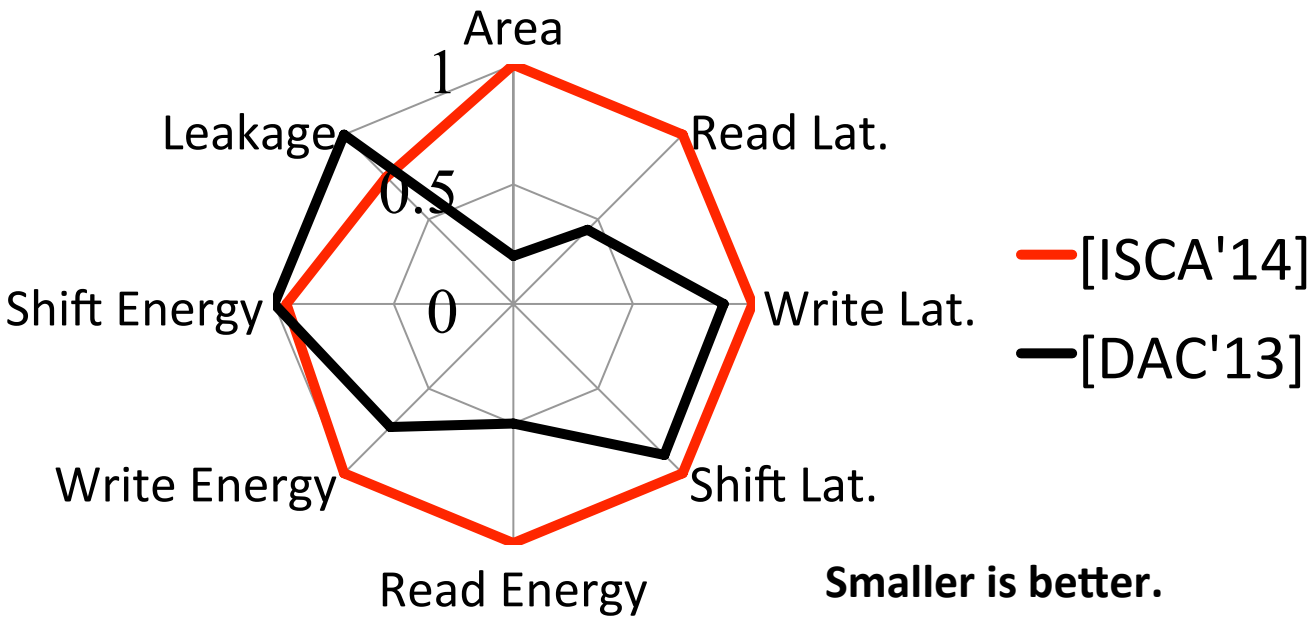
- **Physical level**
  - [Science'08], introduction of racetrack memory.
  - [IEDM'11], a 256 cells demo die.
  - [VLSI'09], [JAP'11], [JAP'12], Physical modeling.

RM design is not fully exploited due to lack of circuit-level modeling

- **Architecture level**
  - Density: 2X~6X, Speed up: 60%~140% (over STT-RAM cache)
  - [ISLPED'12, ISCA'14], large GPGPU caches
  - [DAC'13, TC'14], cross-layer exploration and optimization.

RM Design Space is Huge!

- **Metrics used to describe a design**

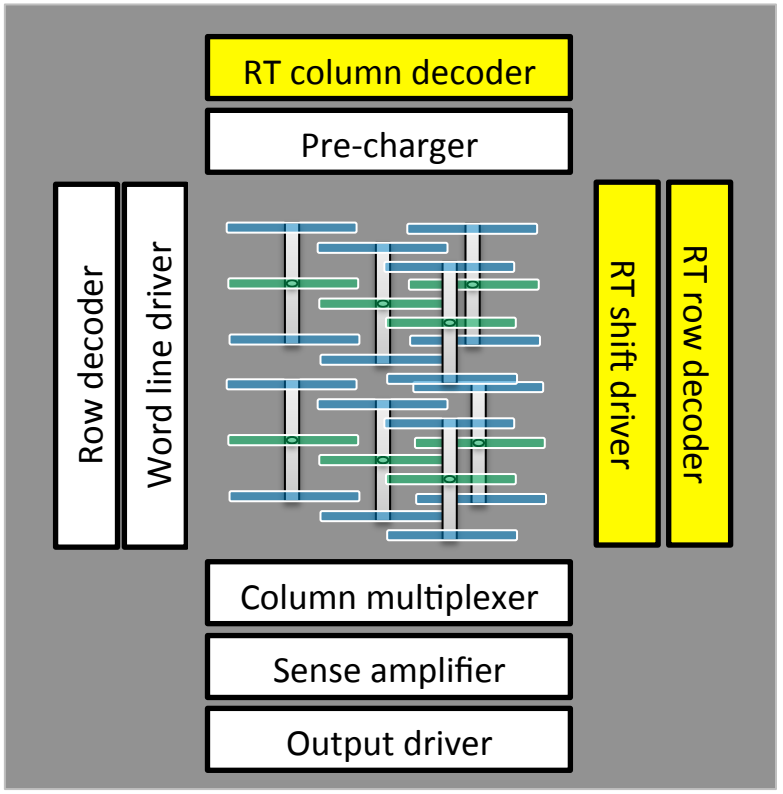


Modeling of RM is challenging

Challenges of Modeling RM Array

- **Decoding is more complicated**
  - Access port , domain
- **Support for shift operation**
  - Extra peripheral circuitry
  - Position registers
- **Layout modeling is complicated**
  - Cell has many parameters
  - Cells can be overlapped
  - Cell-Circuit co-optimization

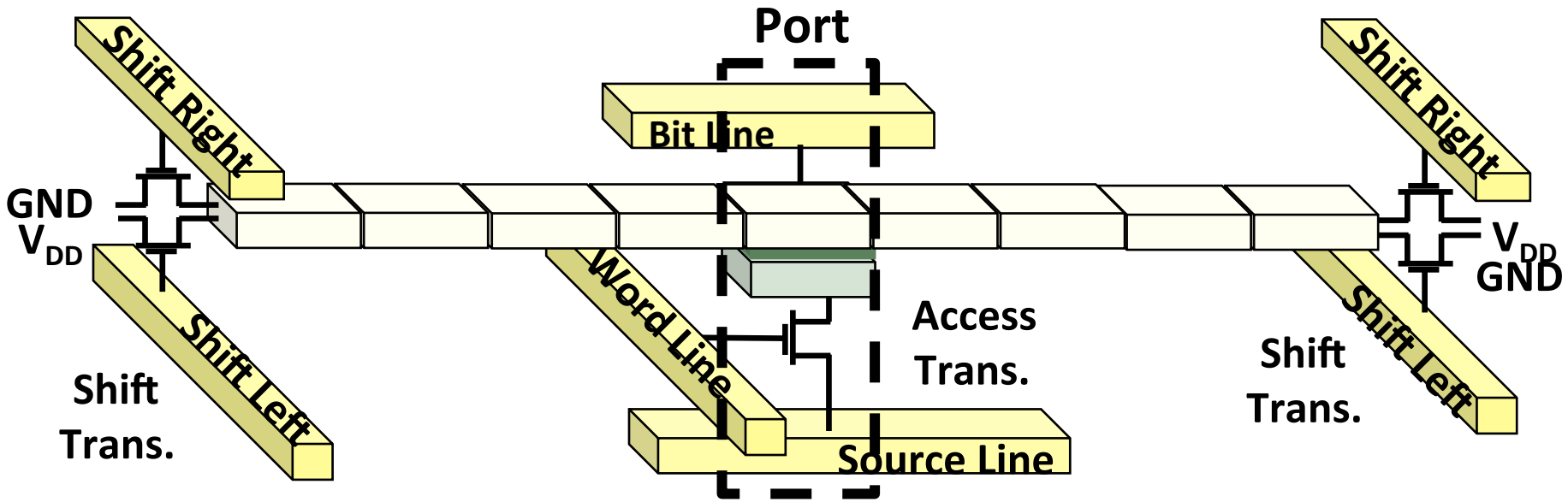
Racetrack Memory



Outline

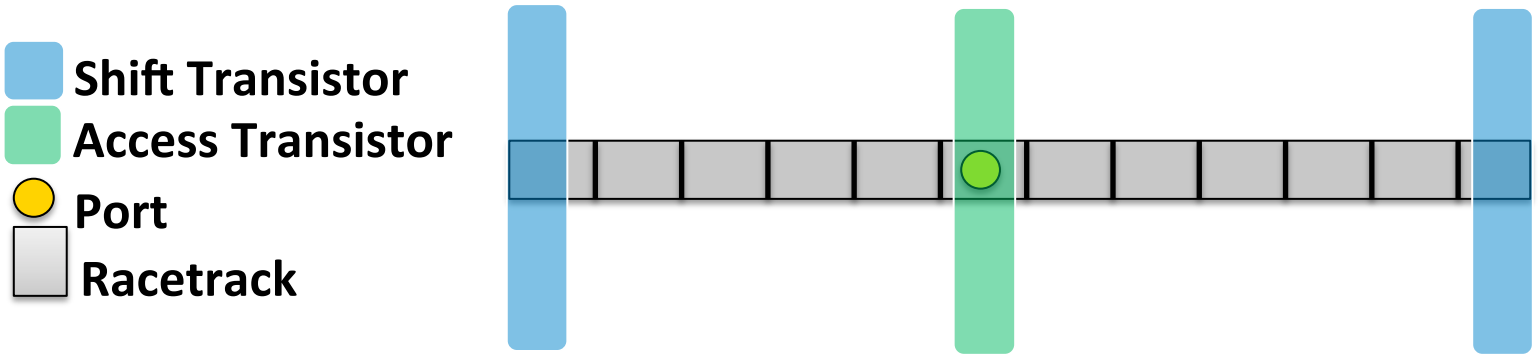
- Background
- **Layout Modeling**
- **Case Study**
- **Conclusions**

Layout of a Single Cell



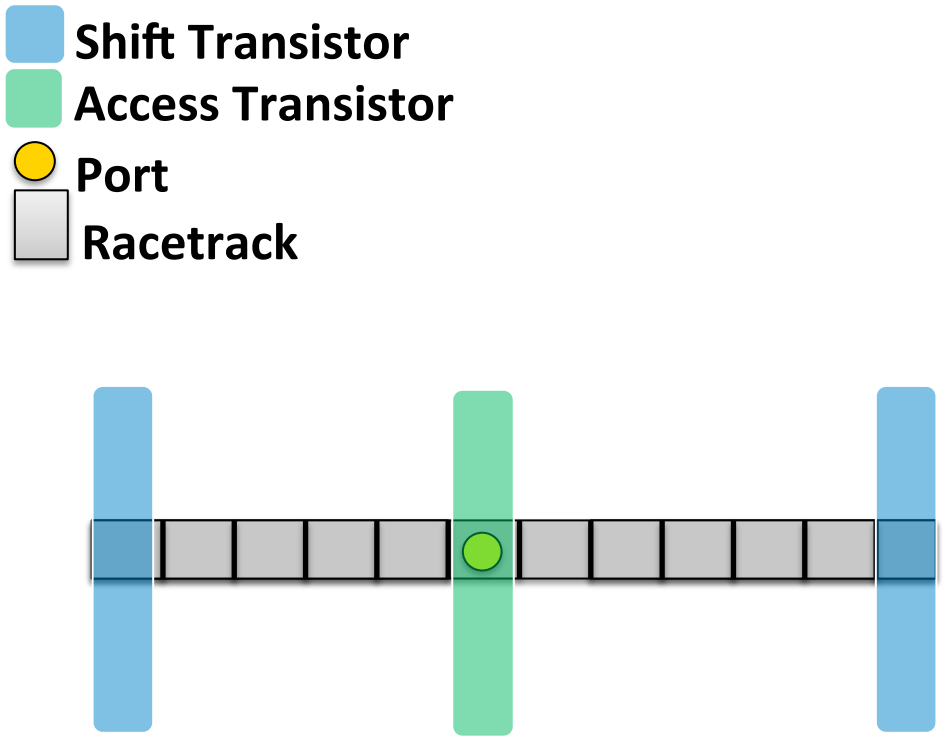
Detailed view

Simple view



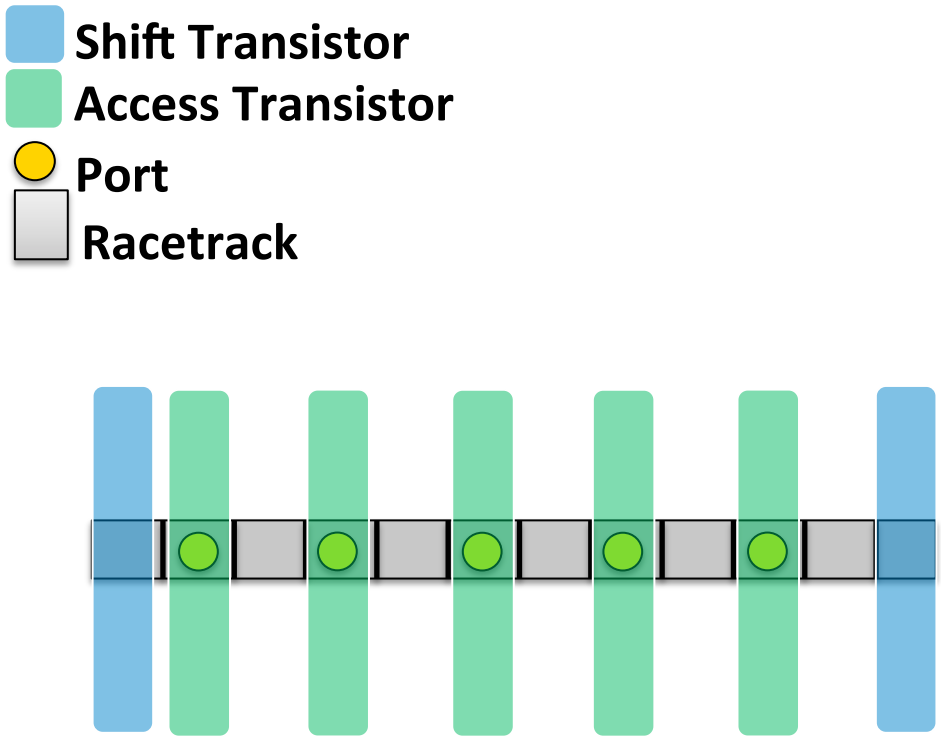
Cell Layout Modeling

- Cell parameters
  - Racetrack length
  - Number of ports
  - Width of transistor
  - Domain size
  - Overhead domains
- Cell overlapping



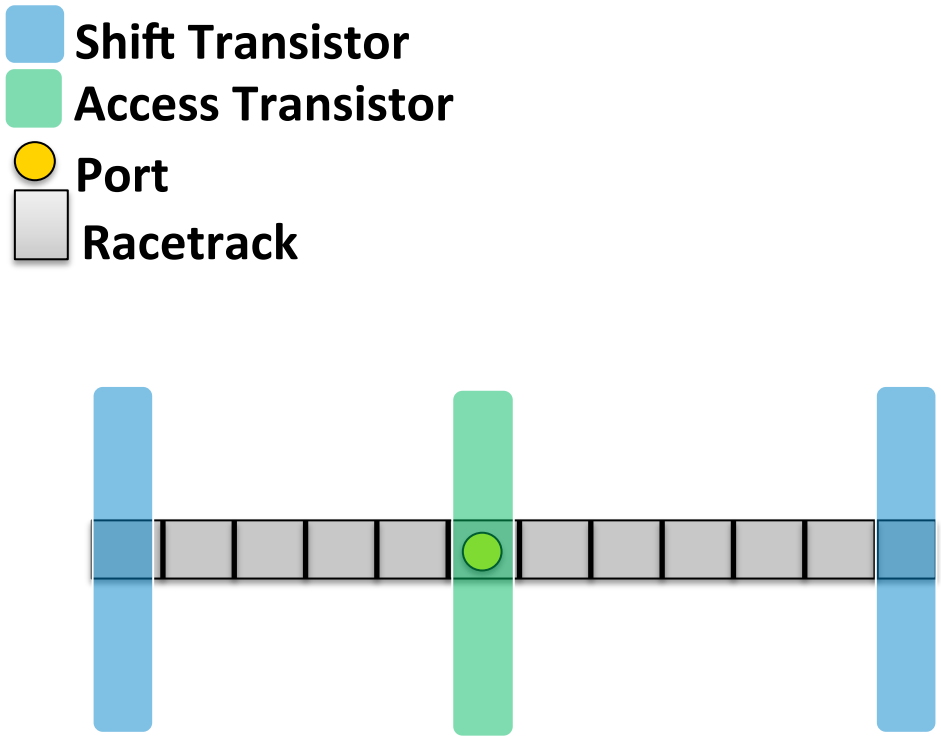
Cell Layout Modeling

- Cell parameters
  - Racetrack length
  - Number of ports
  - Width of transistor
  - Domain size
  - Overhead domains
- Cell overlapping



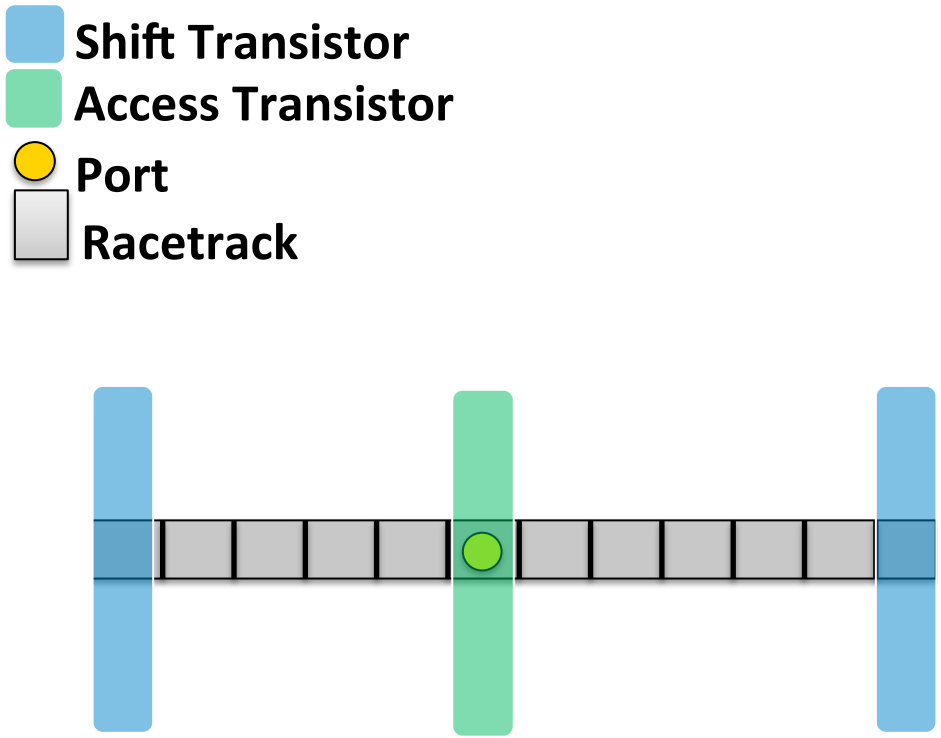
Cell Layout Modeling

- Cell parameters
  - Racetrack length
  - Number of ports
  - Width of transistor
  - Domain size
  - Overhead domains
- Cell overlapping



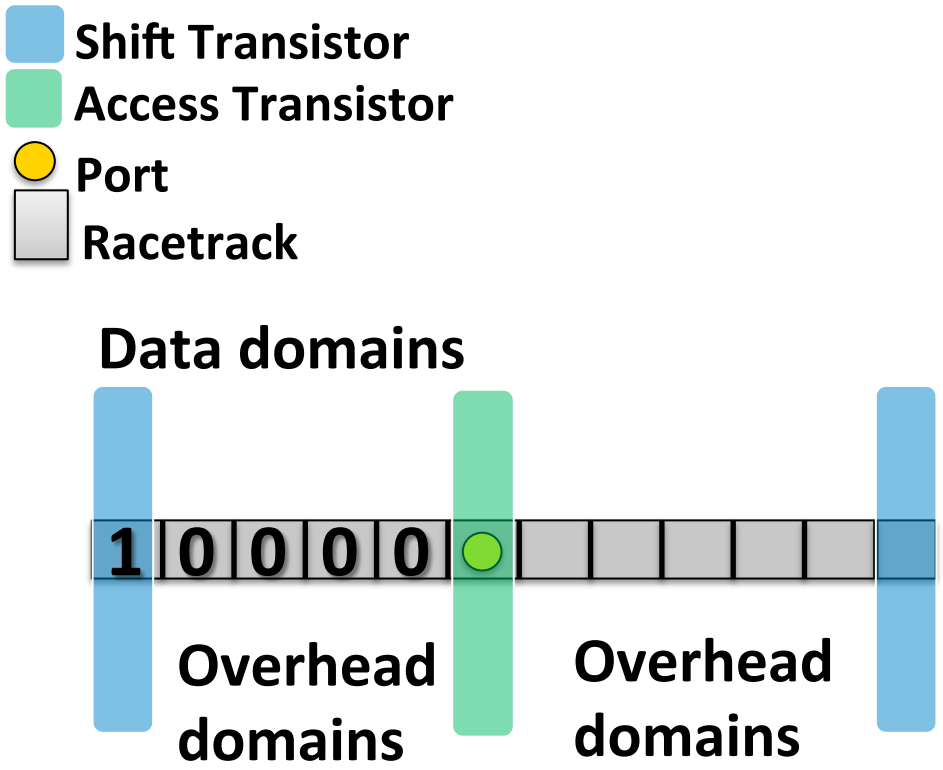
Cell Layout Modeling

- Cell parameters
  - Racetrack length
  - Number of ports
  - Width of transistor
  - Domain size
  - Overhead domains
- Cell overlapping



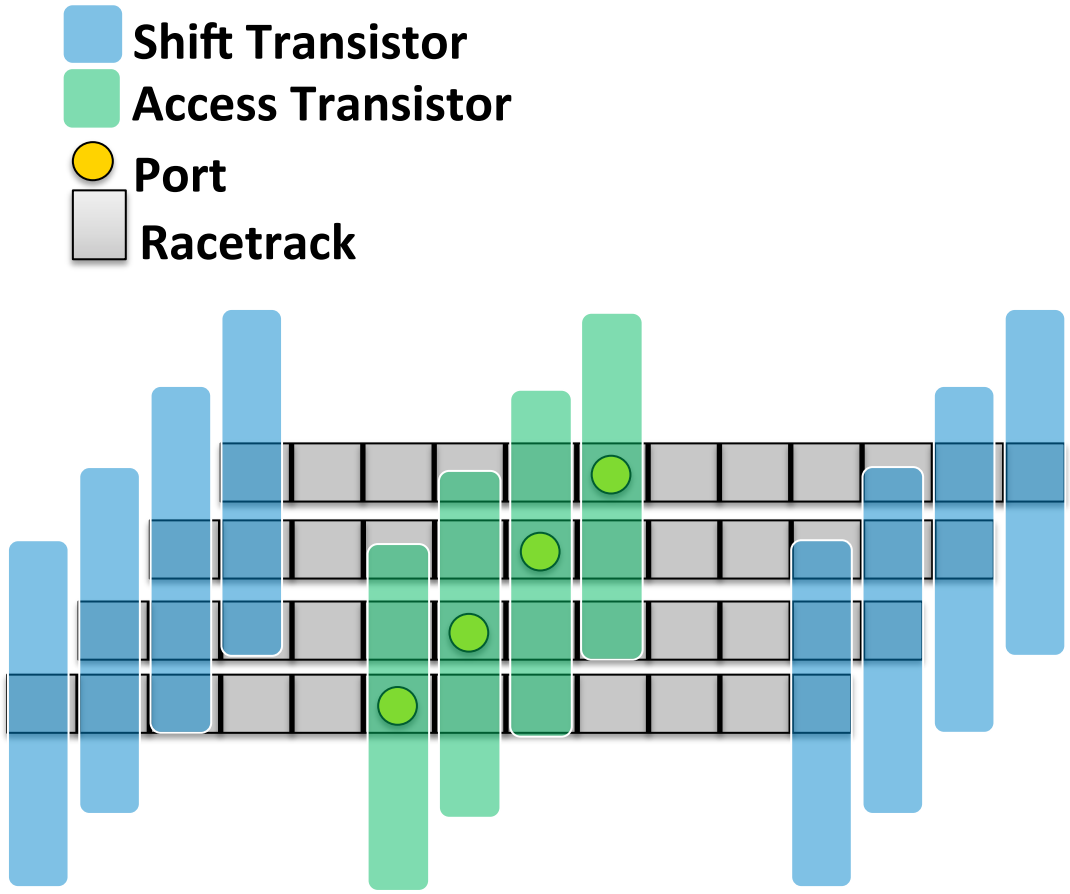
Cell Layout Modeling

- Cell parameters
  - Racetrack length
  - Number of ports
  - Width of transistor
  - Domain size
  - Overhead domains
- Cell overlapping



Cell Layout Modeling

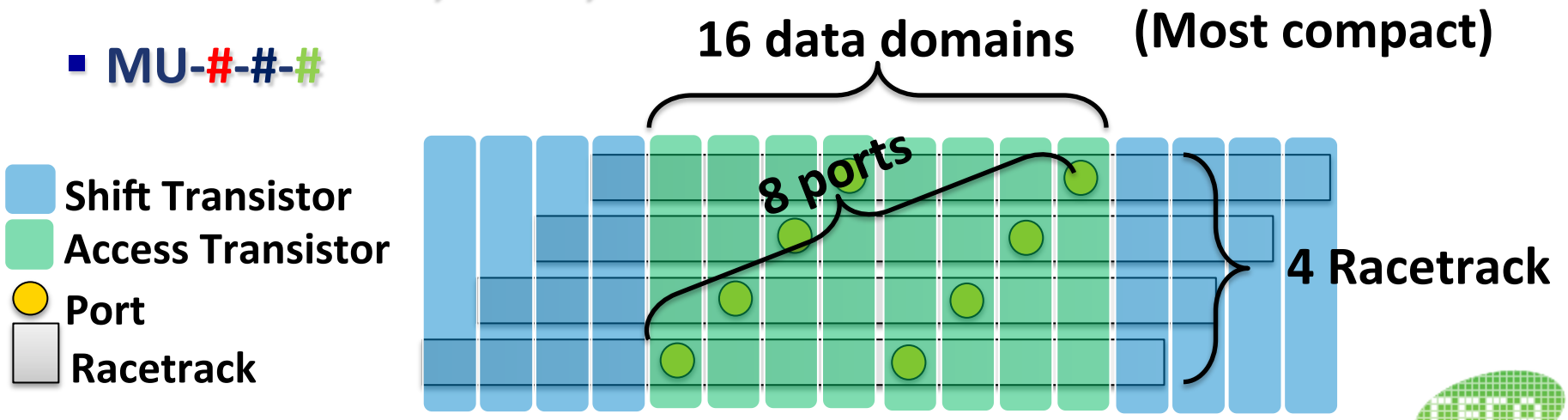
- **Cell parameters**
  - Racetrack length
  - Number of ports
  - Width of transistor
  - Domain size
  - Overhead domains
- **Cell overlapping**



Macro Unit

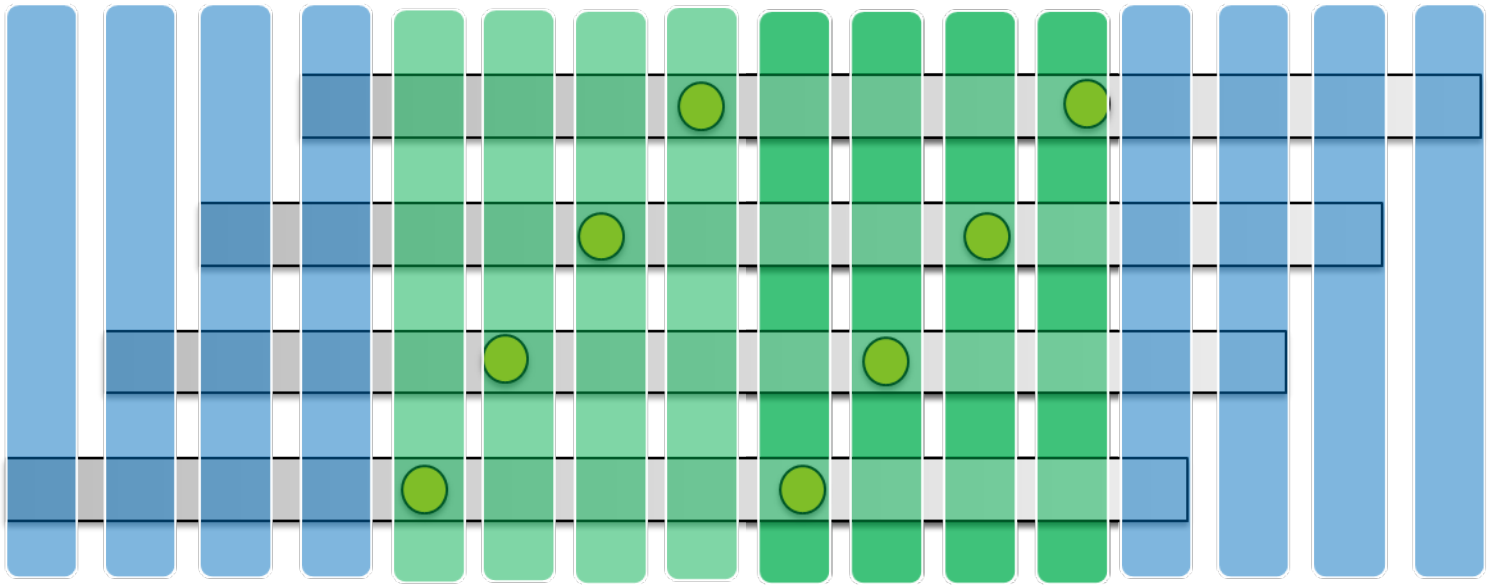
- Use macro unit to build an array.
- Definition
  - Several cells that are overlapped with each other
- Configuration
  - #Data Domain, #Port, #Racetrack
  - MU-#-#-#

MU-16-8-4
(Most compact)



Interaction: #Port vs. Racetrack Length

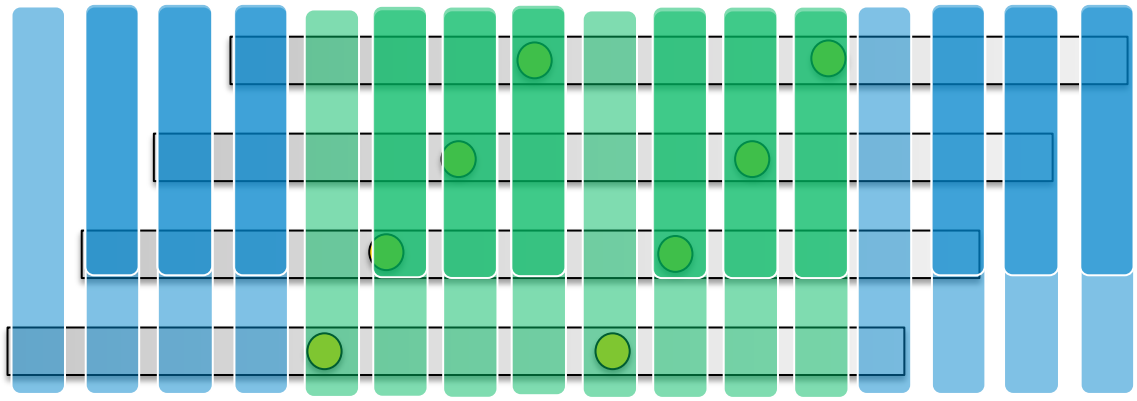
Max. port: 12



$$N_P \leq N_{Pmax} = \frac{L_{RT} - N_D L_D 2^{1-N_{PPR}}}{L_{NMOS} + G_{MOS}}$$

Interaction: #RT vs. Transistor Width

Max. RT: 3

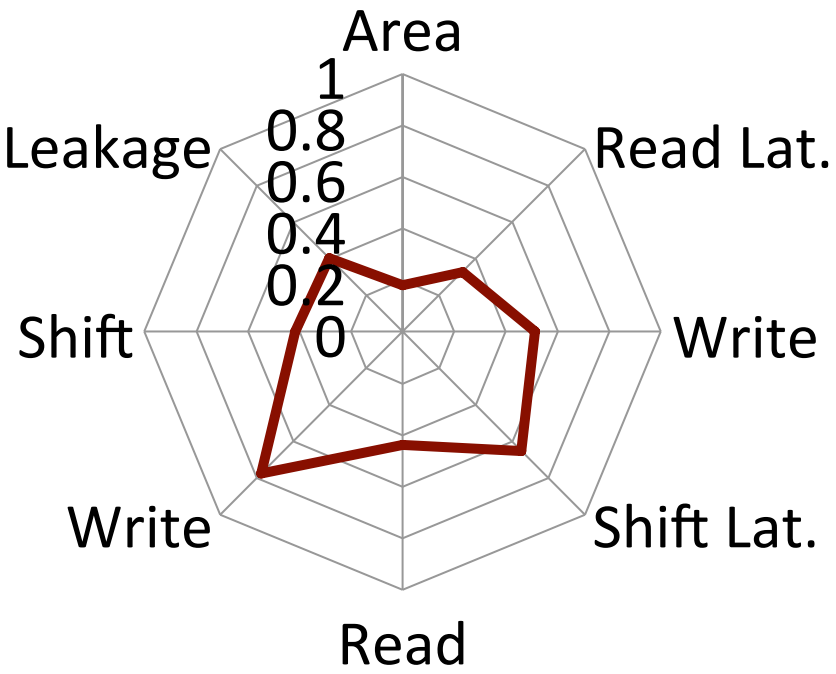


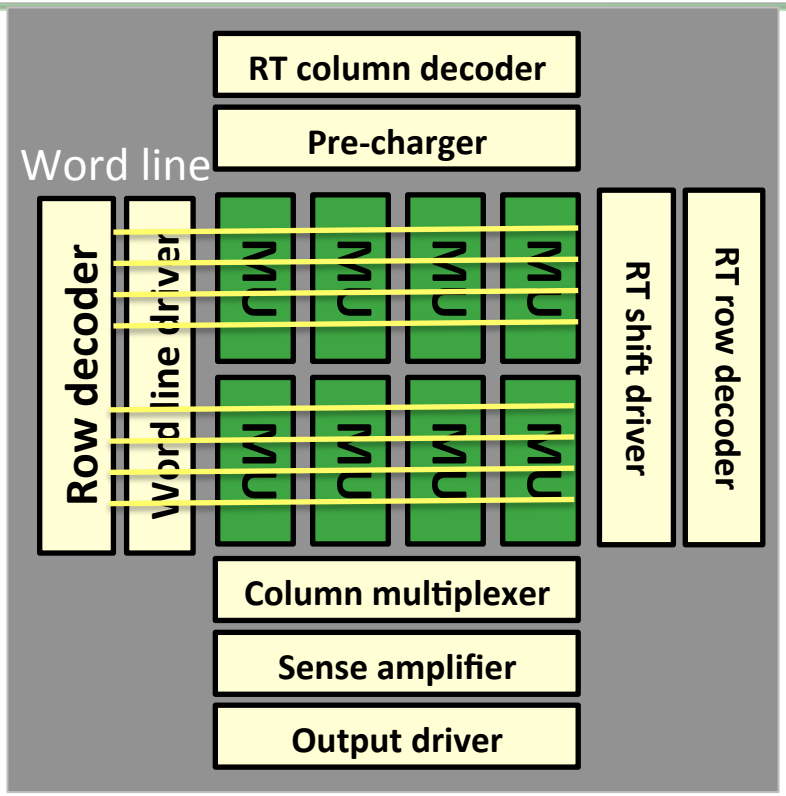
$$N_{RT} \leq N_{RTmax} = \lfloor \frac{W_{MOS} - W_{via}}{W_{RT} + G_{RT}} \rfloor$$

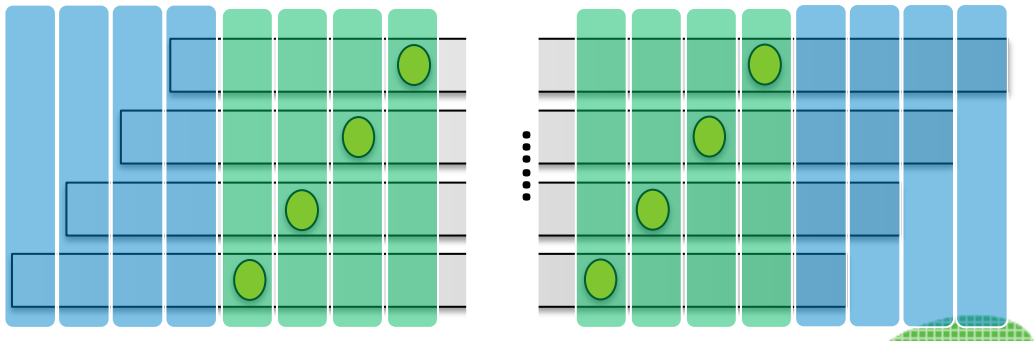
Cell (MU)-Circuit co-optimization is also important

Cell-Circuit Co-optimization

- **Optimization for Area**
  - **More cells and ports in a macro unit**



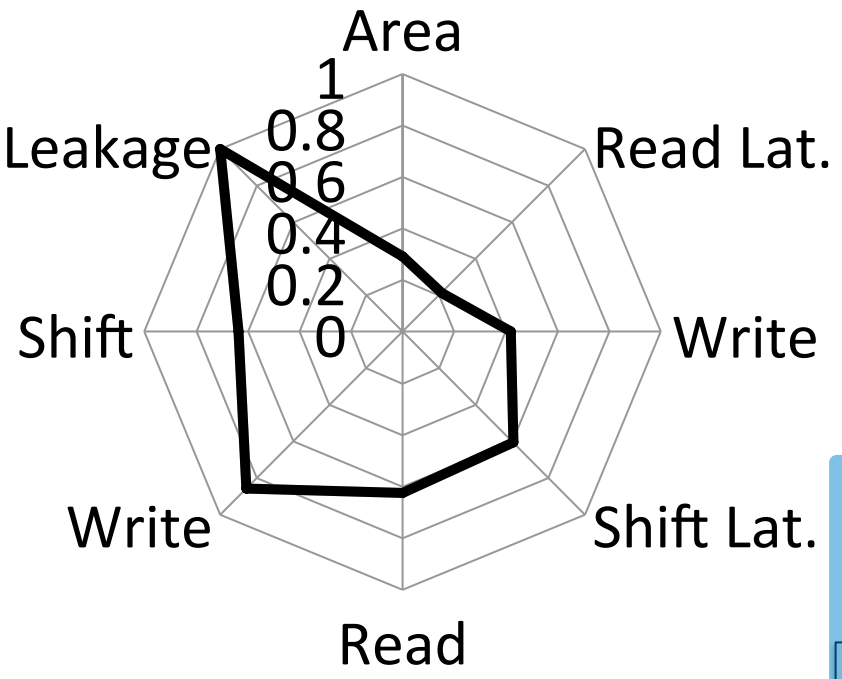


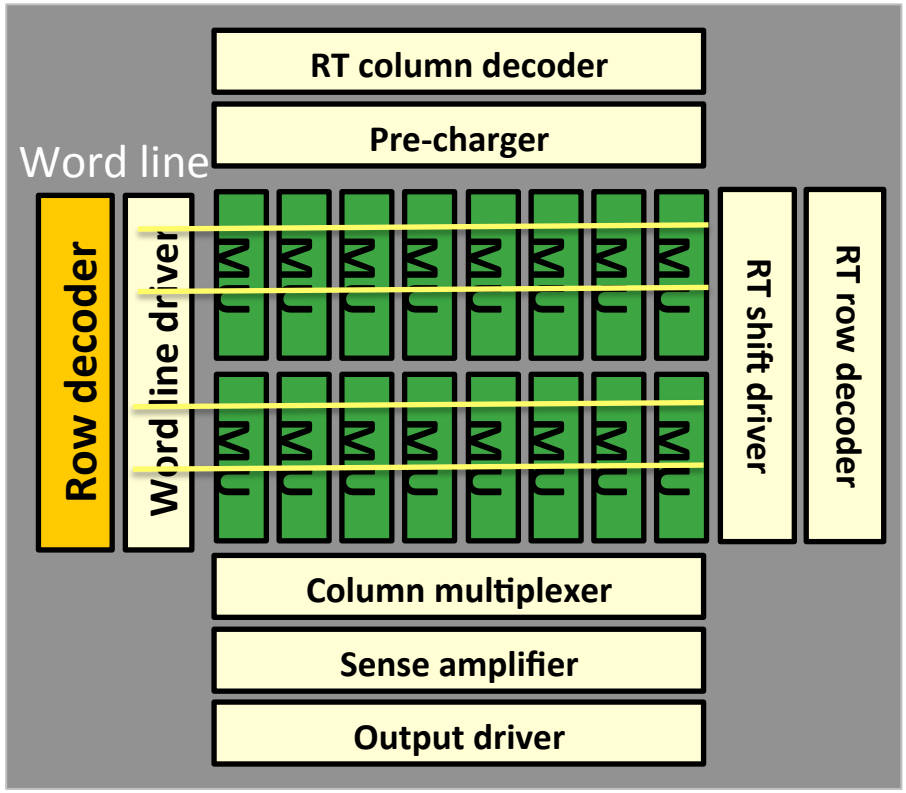


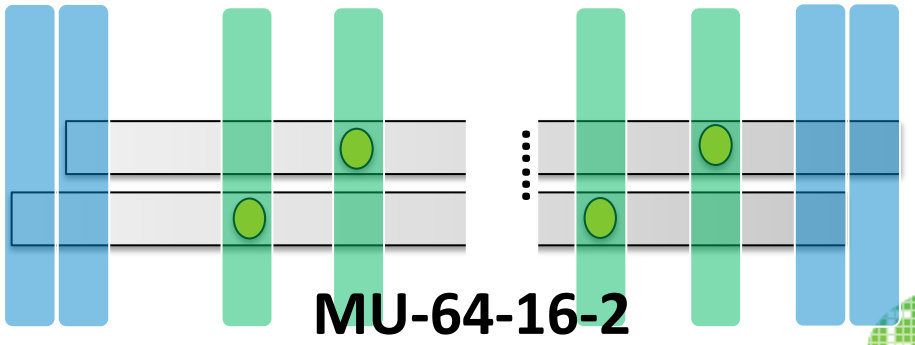
MU-64-32-4

Cell-Circuit Co-optimization

- **Read Latency**
  - **Trade density for speed**



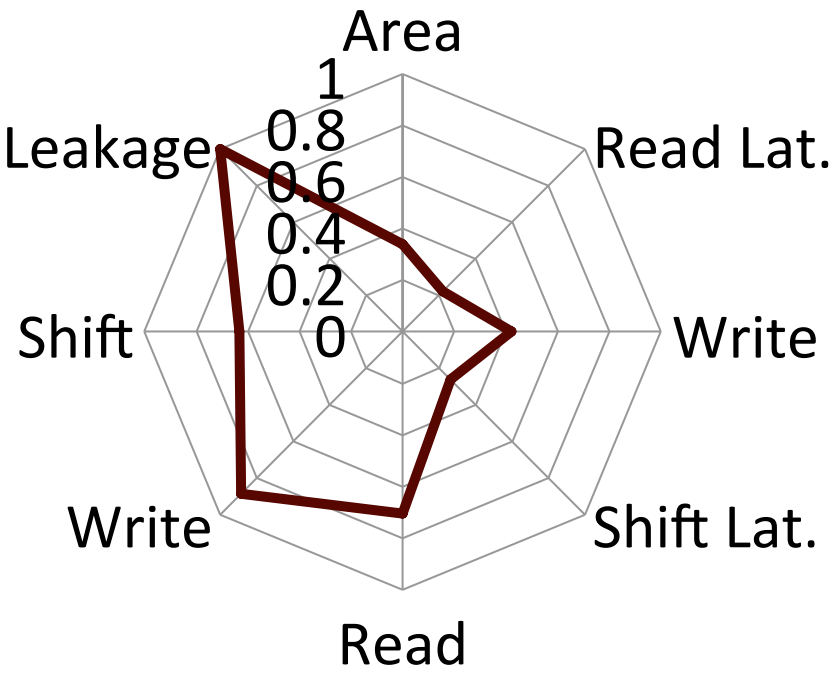


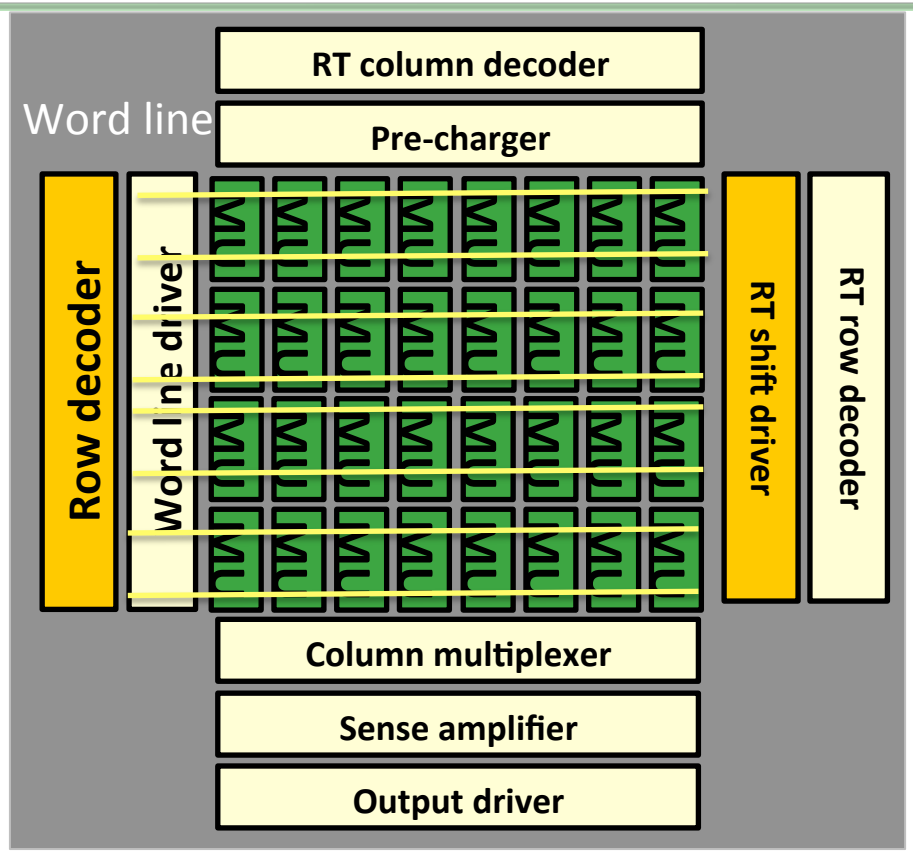


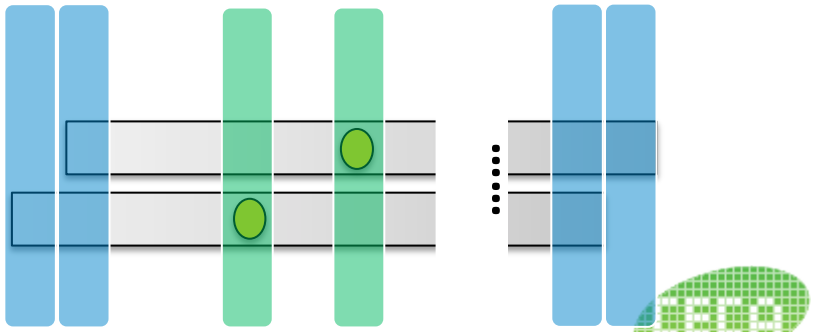
MU-64-16-2

Cell-Circuit Co-optimization

- **Shift Latency**
  - Short racetrack



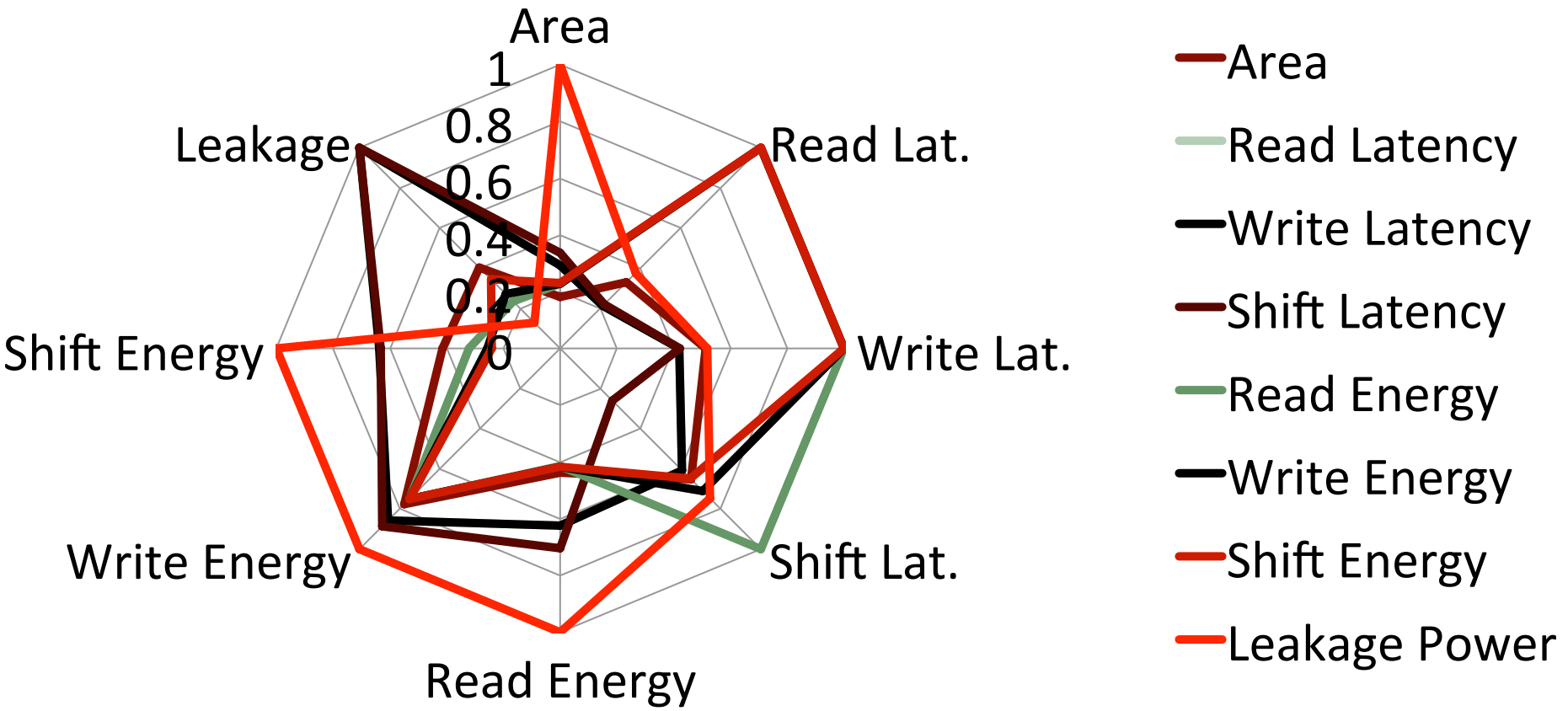




MU-16-8-2

Optimization Targets

- **Different target focuses on different metric.**



Outline

- Background
- Layout Modeling
- Case Study: RM Cache Design
- Conclusions

Experiment Setup

CPU: 4 simple cores

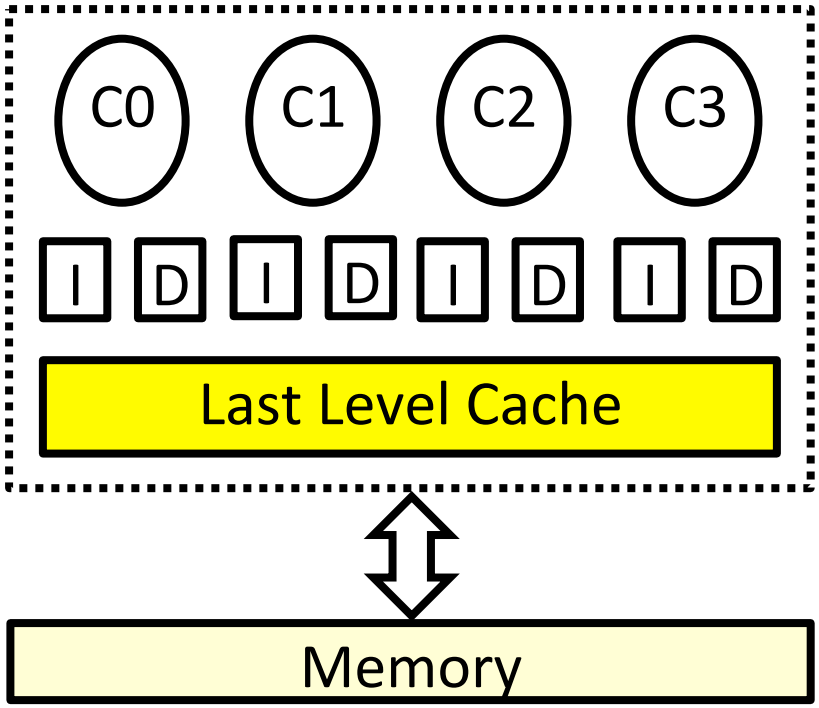
L1: 32KB I & D SRAM

LLC: 64MB RM

Benchmarks: SPEC CPU 2006

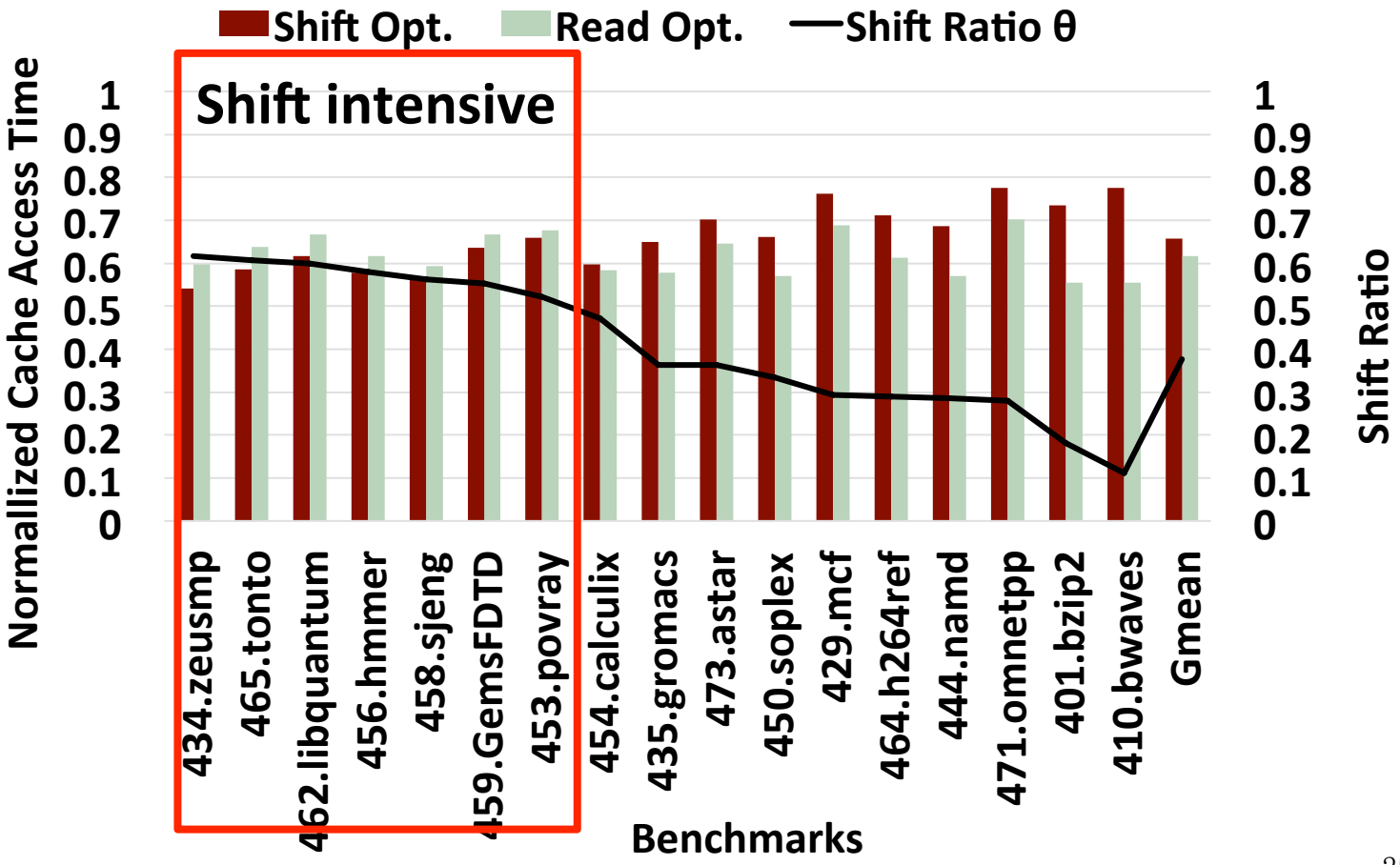
Simulator: gem5

Metric: System execution time



Optimize Read or Shift?

- Application with high shift intensity prefer shift optimized design.



Conclusions

- Racetrack Memory is attractive for ultra high storage density and fast access speed.
- The design space of Racetrack Memory is really large.
- Our model can facilitate cross-layer co-design.

About the Tool

- Codes will be released
  - http://ceca.pku.edu.cn/chaozhang
  - Suggestions and comments are welcome.
- Future work:
  - Access transistor folding
  - Exploration on racetrack width
  - 3D vertical cells support
  - Reliability simulation
  - …

Thank you!
Q & A