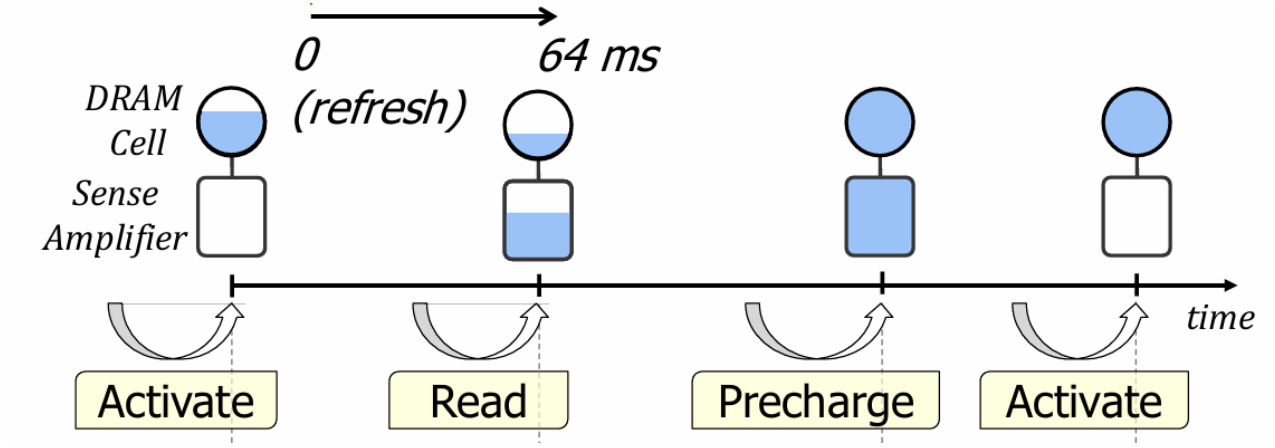
Ramulator 2.0 Summary

Intelligent System

Laboratory

DRAM Operations & States

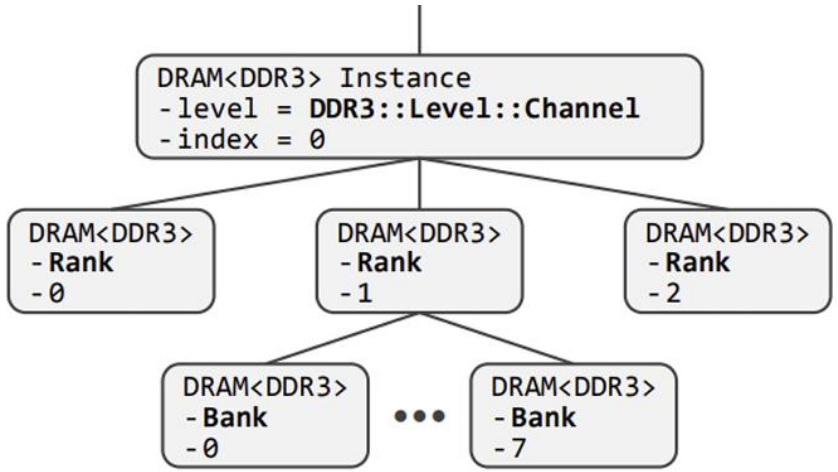
❑ DRAM Operations & States



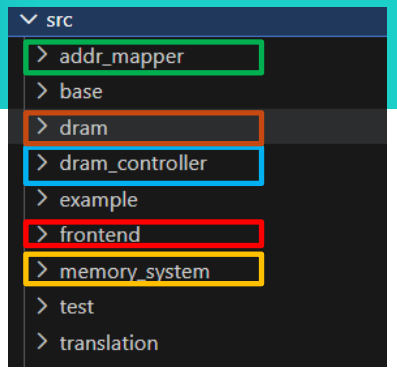
- **Main DRAM states**
  - **Activate**
  - **Read/Write**
  - **Precharge**

```
1  // DRAM.h
2  template <typename T>
3  class DRAM {
4    DRAM<T>* parent;
5    vector<DRAM<T>*>
6      children;
7    T::Level level;
8    int index;
9    // more code...
10 };
```

```
1  // DDR3.h/cpp
2  class DDR3 {
3    enum class Level {
4      Channel, Rank, Bank,
5      Row, Column, MAX
6    };
7
8
9    // more code...
10 };
```



src files <=> DRAM Operation



❑ Simulation Configuration

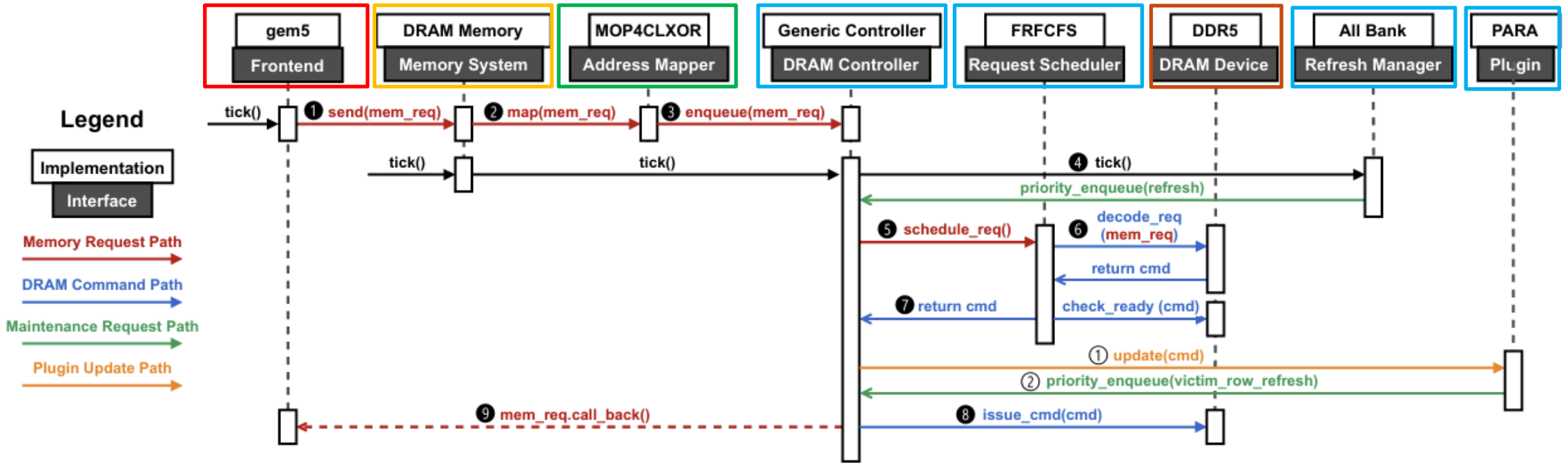


Fig. 1: High-level software architecture of Ramulator 2.0 using an example DDR5 system configuration

src files <=> DRAM Operation

❑ Simulation Configuration

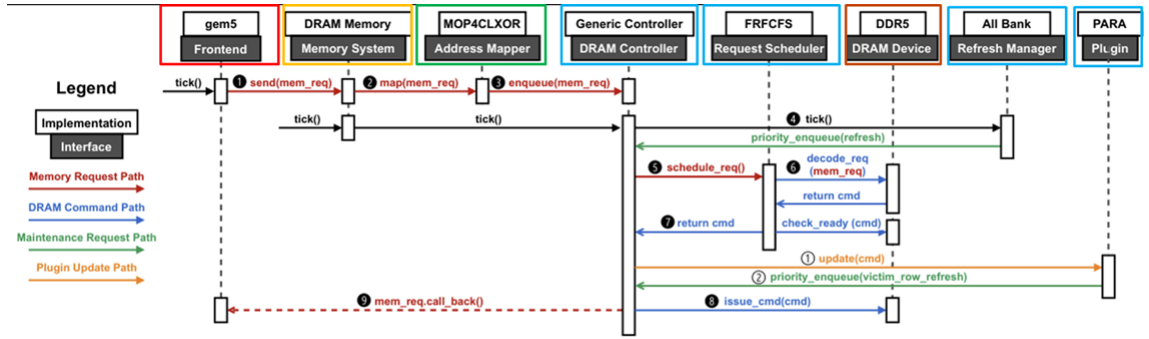
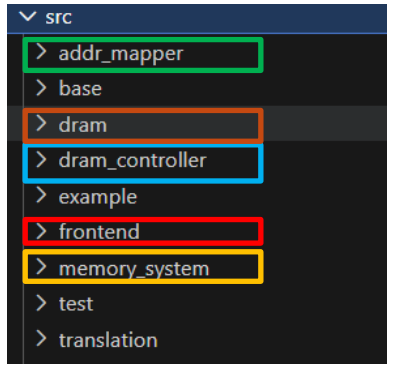


Fig. 1: High-level software architecture of Ramulator 2.0 using an example DDR5 system configuration



1. **Requests are sent**: Front-end(trace file)에서 Memory Request를 보냄
2. **Memory Addresses are Mapped**: Address Mapper가 Request Address를 DRAM 구조에 맞게 변환
3. **Enqueue**: DRAM Ctrlr의 Buffer에 Request를 넣음
4. **DRAM Ctrlr -** Ticking Refresh Manager: Ctrlr가 Refresh Manager를 호출해 high-priority maintenance request(ex. Refresh)을 추가
5. **DRAM Ctrlr - Request Scheduling**: Request Scheduler에게 최적의 Request을 선택하라고 요청
6. **DRAM Device가 Request 확인**: Scheduler가 DRAM Device Model을 참조해 적합한 Command를 Decode
7. **Issue Command**: DRAM Ctrlr가 DRAM Command를 보냄
8. **Updates the behavior and timing information:** DRAM Command Issue시 State & Timing이 Update
9. **Notify the frontend**: Memory Request가 끝나면 callback으로 frontend에 알림

main function

❑ main.cpp

```cpp
int main(int argc, char* argv[]) {
  // Parse command line arguments
  argparse::ArgumentParser program("Ramulator", "2.0");
  program.add_argument("-c", "--config").metavar("\"dumped YAML configuration\"")
    .help("String dump of the yaml configuration.");
  program.add_argument("-f", "--config_file").metavar("path-to-configuration-file")
    .help("Path to a YAML configuration file.");
  program.add_argument("-p", "--param").metavar("KEY=VALUE")
    .append()
    .help("Specify parameter to override in the configuration file. Repeat this option to change multiple parameters.");
```

•
•
•

```cpp
  // Connect the frontend and the memory system together,
  // this recursively calls the "setup" function in all instaniated components
  // so that they can get each other's parameters (if needed) after their initialization
  frontend->connect_memory_system(memory_system);
  memory_system->connect_frontend(frontend);

  // Get the relative clock ratio between the frontend and memory system
  int frontend_tick = frontend->get_clock_ratio();
  int mem_tick = memory_system->get_clock_ratio();

  int tick_mult = frontend_tick * mem_tick;

  for (uint64_t i = 0;; i++) {
    if (((i % tick_mult) % mem_tick) == 0) {
      frontend->tick();
    }

    if (frontend->is_finished()) {
      break;
    }

    if ((i % tick_mult) % frontend_tick == 0) {
      memory_system->tick();
    }
  }

  // Finalize the simulation. Recursively print all statistics from all components
  frontend->finalize();
  memory_system->finalize();

  return 0;
}
```

main.cpp

1. **Argument 받는 부분**
   - **Options**
   1. -c: command line dump
   2. -f: YAML document
   3. -p: overriding parameters in a YAML document

2. **Long for loop를 통한 tick() 기반 simul**
   1. frontend(core)가 발행한 예상 instructions들을 모두 처리시 is_finished()가 true가 됨

yaml file

❑ example_config.yaml

```yaml
Frontend:
  impl: SimpleO3
  clock_ratio: 8
  num_expected_insts: 500000
  traces:
    - example_inst.trace

  Translation:
    impl: RandomTranslation
    max_addr: 2147483648

MemorySystem:
  impl: GenericDRAM
  clock_ratio: 3

  DRAM:
    impl: DDR4
    org:
      preset: DDR4_8Gb_x8
      channel: 1
      rank: 2
    timing:
      preset: DDR4_2400R

  Controller:
    impl: Generic
    Scheduler:
      impl: FRFCFS
    RefreshManager:
      impl: AllBank
    RowPolicy:
      impl: ClosedRowPolicy
      cap: 4
    plugins:

  AddrMapper:
    impl: RoBaRaCoCh
```

1. Frontend Interface(IFrontEnd) 부분

- **trace file에서 Instruction읽고, Memory Request 생성**
- *impl: SimpleO3*
  ⇒ Simple Out-of-Order (O3) CPU
- *clock_ratio: 8*
  ⇒ global CLK 대비 Frontend CLK 속도
- *num_expected_insts: 500000*
  ⇒ Simulation이 해당 instruction 수에 도달 시 종료
- *traces*
  ⇒ Instruction trace file(include memory access Inst)
- *impl: RandomTranslation*
  ⇒ Physical Memory ↔ Virtual Memory 변환
  ⇒ System의 Page Table 등을 간단히 Modeling
- *max_addr: 2147483648*
  ⇒ Translation 시 address overflow 방지

yaml file

❑ example_config.yaml

```yaml
Frontend:
  impl: SimpleO3
  clock_ratio: 8
  num_expected_insts: 500000
  traces:
    - example_inst.trace

  Translation:
    impl: RandomTranslation
    max_addr: 2147483648

MemorySystem:
  impl: GenericDRAM
  clock_ratio: 3

  DRAM:
    impl: DDR4
    org:
      preset: DDR4_8Gb_x8
      channel: 1
      rank: 2
    timing:
      preset: DDR4_2400R

  Controller:
    impl: Generic
    Scheduler:
      impl: FRFCFS
    RefreshManager:
      impl: AllBank
    RowPolicy:
      impl: ClosedRowPolicy
      cap: 4
    plugins:

  AddrMapper:
    impl: RoBaRaCoCh
```

2. MemorySystem Interface 부분

- **Frontend의 Request를 받아 DRAM Ctrlr을 통해 처리**
- **Latency, en/dequeue, Timing Constraints 처리**

- *impl: GenericDRAM*
  ⇒ 기본 DRAM 기반 System, Ctrlr와 DRAM을 통합

- *clock_ratio: 3*
  ⇒ global CLK 대비 MemorySystem CLK 속도
  ⇒ 현재: DRAM이 CPU보다 느린 System (= 3:8)

yaml file

❑example_config.yaml

```yaml
Frontend:
  impl: SimpleO3
  clock_ratio: 8
  num_expected_insts: 500000
  traces:
    - example_inst.trace

Translation:
  impl: RandomTranslation
  max_addr: 2147483648

MemorySystem:
  impl: GenericDRAM
  clock_ratio: 3

  DRAM:
    impl: DDR4
    org:
      preset: DDR4_8Gb_x8
      channel: 1
      rank: 2
    timing:
      preset: DDR4_2400R

  Controller:
    impl: Generic
    Scheduler:
      impl: FRFCFS
    RefreshManager:
      impl: AllBank
    RowPolicy:
      impl: ClosedRowPolicy
      cap: 4
    plugins:

  AddrMapper:
    impl: RoBaRaCoCh
```

2. MemorySystem Interface 부분

- **DRAM Section**
- *impl: DDR4*
  ⇒ tick() 시 Timing Check / Command Issue 실행.
- *org: DDR4_8Gb_x8*
  ⇒ 현재 DRAM preset - 8Gb 용량, x8bit data bus
  ⇒ Channel/Rank 설정 시 기본 Preset설정을 Override 함
- *timing: DDR4_2400R*
  ⇒ Timing preset - nRCD등의 Timing Contraint 정의
  ⇒ 이를 이용해 tick() 시 Latency 계산

```
inline static const std::map<std::string, std::vector<int>> timing_presets = {
  //   name       rate   nBL  nCL  nRCD  nRP   nRAS  nRC   nWR  nRTP nCWL nCCDS nCCDL nRRDS nRRDL nWTRS nW
  {"DDR4_1600J", {1600,  4,  10,  10,  10,   28,   38,   12,   6,   9,   4,   5,   -1,   -1,   2,
  {"DDR4_1600K", {1600,  4,  11,  11,  11,   28,   39,   12,   6,   9,   4,   5,   -1,   -1,   2,
  {"DDR4_1600L", {1600,  4,  12,  12,  12,   28,   40,   12,   6,   9,   4,   5,   -1,   -1,   2,
  {"DDR4_1866L", {1866,  4,  12,  12,  12,   32,   44,   14,   7,   10,  4,   5,   -1,   -1,   3,
  {"DDR4_1866M", {1866,  4,  13,  13,  13,   32,   45,   14,   7,   10,  4,   5,   -1,   -1,   3,
  {"DDR4_1866N", {1866,  4,  14,  14,  14,   32,   46,   14,   7,   10,  4,   5,   -1,   -1,   3,
  {"DDR4_2133N", {2133,  4,  14,  14,  14,   36,   50,   16,   8,   11,  4,   6,   -1,   -1,   3,
  {"DDR4_2133P", {2133,  4,  15,  15,  15,   36,   51,   16,   8,   11,  4,   6,   -1,   -1,   3,
  {"DDR4_2133R", {2133,  4,  16,  16,  16,   36,   52,   16,   8,   11,  4,   6,   -1,   -1,   3,
  {"DDR4_2400P", {2400,  4,  15,  15,  15,   39,   54,   18,   9,   12,  4,   6,   -1,   -1,   3,
  {"DDR4_2400R", {2400,  4,  16,  16,  16,   39,   55,   18,   9,   12,  4,   6,   -1,   -1,   3,
  {"DDR4_2400U", {2400,  4,  17,  17,  17,   39,   56,   18,   9,   12,  4,   6,   -1,   -1,   3,
  {"DDR4_2400T", {2400,  4,  18,  18,  18,   39,   57,   18,   9,   12,  4,   6,   -1,   -1,   3,
```

```
class DDR4 : public IDRAM, public Implementation {
  RAMULATOR_REGISTER_IMPLEMENTATION(IDRAM, DDR4, "DDR4", "DDR4 Device Model")

public:
  inline static const std::map<std::string, Organization> org_presets = {
    //   name         density   DQ   Ch Ra Bg Ba   Ro     Co
    {"DDR4_2Gb_x4",   {2<<10,   4,  {1, 1, 4, 4, 1<<15, 1<<10}}},
    {"DDR4_2Gb_x8",   {2<<10,   8,  {1, 1, 4, 4, 1<<14, 1<<10}}},
    {"DDR4_2Gb_x16",  {2<<10,   16, {1, 1, 2, 4, 1<<14, 1<<10}}},
    {"DDR4_4Gb_x4",   {4<<10,   4,  {1, 1, 4, 4, 1<<16, 1<<10}}},
    {"DDR4_4Gb_x8",   {4<<10,   8,  {1, 1, 4, 4, 1<<15, 1<<10}}},
    {"DDR4_4Gb_x16",  {4<<10,   16, {1, 1, 2, 4, 1<<15, 1<<10}}},
    {"DDR4_8Gb_x4",   {8<<10,   4,  {1, 1, 4, 4, 1<<17, 1<<10}}},
    {"DDR4_8Gb_x8",   {8<<10,   8,  {1, 1, 4, 4, 1<<16, 1<<10}}},
    {"DDR4_8Gb_x16",  {8<<10,   16, {1, 1, 2, 4, 1<<16, 1<<10}}},
    {"DDR4_16Gb_x4",  {16<<10,  4,  {1, 1, 4, 4, 1<<18, 1<<10}}},
    {"DDR4_16Gb_x8",  {16<<10,  8,  {1, 1, 4, 4, 1<<17, 1<<10}}},
    {"DDR4_16Gb_x16", {16<<10,  16, {1, 1, 2, 4, 1<<17, 1<<10}}},
  };
```

```
class DDR4 : public IDRAM, public Implementation {
    void tick() override {

  void init() override {
    RAMULATOR_DECLARE_SPECS();
    set_organization();
    set_timing_vals();

    set_actions();
    set_preqs();
    set_rowhits();
    set_rowopens();
    set_powers();

    create_nodes();
  };
```

yaml file

❑example_config.yaml

```yaml
Frontend:
  impl: SimpleO3
  clock_ratio: 8
  num_expected_insts: 500000
  traces:
    - example_inst.trace

  Translation:
    impl: RandomTranslation
    max_addr: 2147483648

MemorySystem:
  impl: GenericDRAM
  clock_ratio: 3

  DRAM:
    impl: DDR4
    org:
      preset: DDR4_8Gb_x8
      channel: 1
      rank: 2
    timing:
      preset: DDR4_2400R

  Controller:
    impl: Generic
    Scheduler:
      impl: FRFCFS
    RefreshManager:
      impl: AllBank
    RowPolicy:
      impl: ClosedRowPolicy
      cap: 4
    plugins:

  AddrMapper:
    impl: RoBaRaCoCh
```

2. MemorySystem Interface 부분

- **Controller Section**
- *impl: Generic*
  ⇒ Generic - 기본 Ctrlr
  ⇒ Request Queue/Scheduling 등 관리
- *Scheduler - impl: FRFCFS*
  ⇒ FRFCFS - First-Ready First-Come-First-Serve
  ⇒ 준비된 Request를 Queue에서 꺼내 우선 처리
- *RefreshManager - impl: AllBank*
  ⇒ AllBank - 모든 Bank simultaneous Refresh
- *RowPolicy - impl: ClosedRowPolicy*
  ⇒ ClosedRowPolicy - 사용 후 Row 즉시 닫음(Precharge)
  ⇒ cap:4 - 열려있는 Row 최대 수 제한
- *plugins*
  ⇒ 현재 Ramulator에서는 Row Hammering 완화 기법을 plugin으로 제공해줌

yaml file

❑example_config.yaml

```yaml
Frontend:
  impl: SimpleO3
  clock_ratio: 8
  num_expected_insts: 500000
  traces:
    - example_inst.trace

  Translation:
    impl: RandomTranslation
    max_addr: 2147483648

MemorySystem:
  impl: GenericDRAM
  clock_ratio: 3

  DRAM:
    impl: DDR4
    org:
      preset: DDR4_8Gb_x8
      channel: 1
      rank: 2
    timing:
      preset: DDR4_2400R

  Controller:
    impl: Generic
    Scheduler:
      impl: FRFCFS
    RefreshManager:
      impl: AllBank
    RowPolicy:
      impl: ClosedRowPolicy
      cap: 4
    plugins:

  AddrMapper:
    impl: RoBaRaCoCh
```

2. MemorySystem Interface 부분

- **AddrMapper Section**

- *impl: RoBaRaCoCh*
  ⇒ Row-Bank-Rank-Column-Channel Mapping Scheme
  ⇒ Requested Addres 변환(Physical → DRAM Vector)

  [Physical → DRAM Vector Example]
  ⇒ Physical Address: 0x12345678
  ⇒ DRAM Vector:
  [Channel:0, Rank:1, Bank:2, Row:128, Column:512]

trace file

❑ example_inst.trace & trace.cpp & core.cpp

example_inst.trace

1	3	20734016	
2	1	20846400	
3	6	20734208	
4	1	20846400	
5	8	20841280	20841280
6	0	20734144	
7	2	20918976	20734016
8	1	20846400	
9			

- simpleO3 CPU model기준
  - Ramulator는 “Memory” Simulator
  - Memory 명령어만 취급하기에, 3가지로만 Instruction을 분리한다.
    1. Not Memory Operation
    2. Load
    3. Store
- 각 line은 load(& store)동작을 나타낸다.
- 1st line : 3cycle동안 stall → load
- 5th line : 8cycle동안 stall → load → store
- 따라서, simpleO3 기반 trace파일:
  - 1st column은 Not Memory Operation Cycle 수 (or ticks 수)
  - 2nd column은 load operation address
  - 3rd column은 store operation address

trace file

❑ example_inst.trace & trace.cpp & core.cpp

- Trace file을 arg로 받아, Frontend (ex: simpleO3.cpp)에서 처리되어 메모리 접근 request를 생성
- simpleO3 CPU model기준: trace의 token[0]: bubble / token[1]: load address / token[2]: store address

trace file result

❑ ./ramulator2 -f ./example_config.yaml

```
root@947e591ed45c:/workspace# ./ramulator2 -f ./example_config.yaml
Frontend:
  impl: SimpleO3
  memory_access_cycles_recorded_core_0: 61
  cycles_recorded_core_0: 216815
  llc_mshr_unavailable: 0
  llc_read_misses: 37
  llc_read_access: 133336
  llc_write_misses: 8
  llc_write_access: 33334
  llc_eviction: 0
  num_expected_insts: 500000
  Translation:
    impl: RandomTranslation

MemorySystem:
  impl: GenericDRAM
  total_num_other_requests: 0
  total_num_write_requests: 0
  total_num_read_requests: 6
  memory_system_cycles: 81306
  DRAM:
    impl: DDR4
  AddrMapper:
    impl: RoBaRaCoCh

Controller:
  impl: Generic
  id: Channel 0
  avg_read_latency_0: 46.5
  read_queue_len_avg_0: 0.00232455181
  write_queue_len_0: 0
  queue_len_0: 245
  num_other_reqs_0: 0
  num_write_reqs_0: 0
  read_latency_0: 279
  priority_queue_len_avg_0: 0.000688756059
  row_hits_0: 2
  priority_queue_len_0: 56
  row_misses_0: 4
  row_conflicts_0: 0
  read_row_misses_0: 4
  queue_len_avg_0: 0.00301330769
  read_row_conflicts_core_0: 0
  read_row_hits_0: 2
  write_queue_len_avg_0: 0
  read_row_conflicts_0: 0
  write_row_misses_0: 0
  write_row_conflicts_0: 0
  read_queue_len_0: 189
  write_row_hits_0: 0
  read_row_hits_core_0: 2
  read_row_misses_core_0: 4
  num_read_reqs_0: 6
  Scheduler:
    impl: FRFCFS
  RefreshManager:
    impl: AllBank

  RowPolicy:
    impl: ClosedRowPolicy
    num_close_reqs: 0
```

- 기본 yaml file setting을 이용.
  => trace file의 Inst 실행결과가 분석됨
- 1. Frontend: CPU ↔ Memory 접근 분석
- 2. MemorySystem<DRAM> 관점 분석
  - Request 수
  - Cycle 수
  - Latency
  - Queue 길이
  - Row Hit / Miss

trace file result analysis

❑ ./ramulator2 -f ./example_config.yaml

```
root@947e591ed45c:/workspace# ./ramulator2 -f ./example_config.yaml
Frontend:
  impl: SimpleO3
  memory_access_cycles_recorded_core_0: 61
  cycles_recorded_core_0: 216815
  llc_mshr_unavailable: 0
  llc_read_misses: 37
  llc_read_access: 133336
  llc_write_misses: 8
  llc_write_access: 33334
  llc_eviction: 0
  num_expected_insts: 500000
  Translation:
    impl: RandomTranslation

MemorySystem:
  impl: GenericDRAM
  total_num_other_requests: 0
  total_num_write_requests: 0
  total_num_read_requests: 6
  memory_system_cycles: 81306
  DRAM:
    impl: DDR4
  AddrMapper:
    impl: RoBaRaCoCh
Controller:
  impl: Generic
  id: Channel 0
  avg_read_latency_0: 46.5
  read_queue_len_avg_0: 0.00232455181
  write_queue_len_0: 0
  queue_len_0: 245
  num_other_reqs_0: 0
  num_write_reqs_0: 0
  read_latency_0: 279
  priority_queue_len_avg_0: 0.000688756059
  row_hits_0: 2
  priority_queue_len_0: 56
  row_misses_0: 4
  row_conflicts_0: 0
  read_row_misses_0: 4
  queue_len_avg_0: 0.00301330769
  read_row_conflicts_core_0: 0
  read_row_hits_0: 2
  write_queue_len_avg_0: 0
  read_row_conflicts_0: 0
  write_row_misses_0: 0
  write_row_conflicts_0: 0
  read_queue_len_0: 189
  write_row_hits_0: 0
  read_row_hits_core_0: 2
  read_row_misses_core_0: 4
  num_read_reqs_0: 6
  Scheduler:
    impl: FRFCFS
  RefreshManager:
    impl: AllBank

RowPolicy:
  impl: ClosedRowPolicy
  num_close_reqs: 0
```

```
52  const SimpleO3Core::Trace::Inst& SimpleO3Core::Trace::get_next_inst(
53    const Inst& inst = m_trace[m_curr_trace_idx];
54    m_curr_trace_idx = (m_curr_trace_idx + 1) % m_trace_length;
55    return inst;
56  }
```

1. **Frontend**
   - Last Level Cache(LLC) 분석
     - Core 0의 Memory 접근 Cycles
     - LLC의 Read Request Cache Miss
     - LLC의 Read Access 수
     - Miss Status Handling Register(MSHR) unavailable cnt
       → MSHR가 가득 차서 Request 거부된 횟수

짧은 trace지만 반복 실행 되기에 많은 access 발생

trace file result

❑power는 어떻게 측정하느냐면

- yaml file에 아래 drampower_enable 옵션을 넣음
- dram.h -> DDR4.cpp 의 power함수 enable됨

```yaml
DRAM:
  impl: DDR4
  org:
    preset: DDR4_8Gb_x8
    channel: 1
    rank: 2
  timing:
    preset: DDR4_2400R
  drampower_enable: true
  # power_debug: true  # option: debug log
  voltage:
    preset: Default  # option: voltage preset
  current:
    preset: Default  # option: current preset

```

```yaml
DRAM:
  impl: DDR4
  active_cycles_rank1: 9361
  pre_background_energy_rank1: 3850.9652475000003
  total_background_energy_rank0: 4424.0171849999997
  idle_cycles_rank1: 68489
  total_cmd_energy: 3203.0342735999993
  total_cmd_energy_rank0: 1601.8769927999997
  total_energy_rank0: 6025.8941777999989
  act_background_energy_rank0: 572.15229750000003
  pre_background_energy_rank0: 3851.8648874999999
  active_cycles_rank0: 9345
  act_background_energy_rank1: 573.13190550000002
  total_energy: 12051.1486116
  idle_cycles_rank0: 68505
  total_energy_rank1: 6025.2544338000007
  total_background_energy_rank1: 4424.0971530000006
  total_background_energy: 8848.1143379999994
  total_cmd_energy_rank1: 1601.1572807999996
AddrMapper:
  impl: RoBaRaCoCh
```

src/frontend folder

- frontend

  - abcd

src/frontend folder

- ❑ frontend
  - abcd

src/frontend folder

- frontend

  - abcd

src/memory_system folder

- memory_system
  - abcd

src/memory_system folder

- memory_system
  - abcd

src/addr_mapper folder

- addr_mapper
  - abcd

src/addr_mapper folder

- addr_mapper
  - abcd

src/translation folder

- translation
  - abcd

src/translation folder

- translation
  - abcd

src/translation folder

- translation

  - abcd

src/dram folder

- dram
  - abcd

src/dram folder

- dram
  - abcd

src/dram folder

- dram
  - abcd

src/dram_controller folder

- dram_controller

  - abcd

src/dram_controller folder

- dram_controller

  - abcd

src/dram_controller folder

- dram_controller

  - abcd

src/dram_controller folder

- dram_controller

  - abcd

src/base folder

- base
  - abcd