



Programming Tips

Ing-June Lu

Skymizer, TAIWAN

2020/5/6

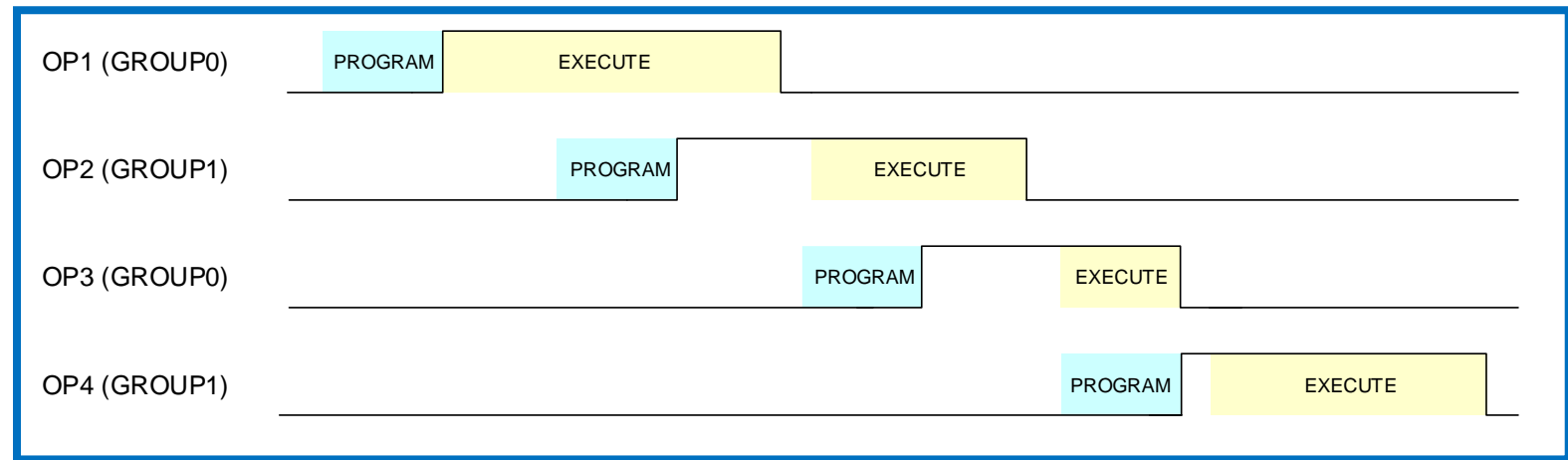
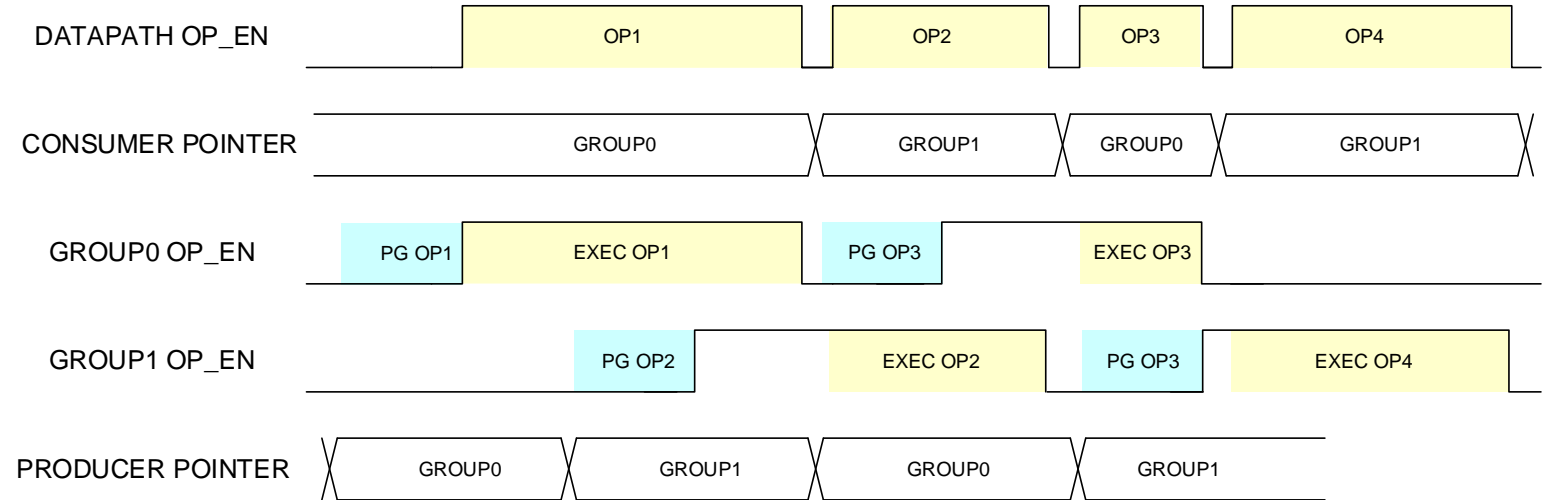
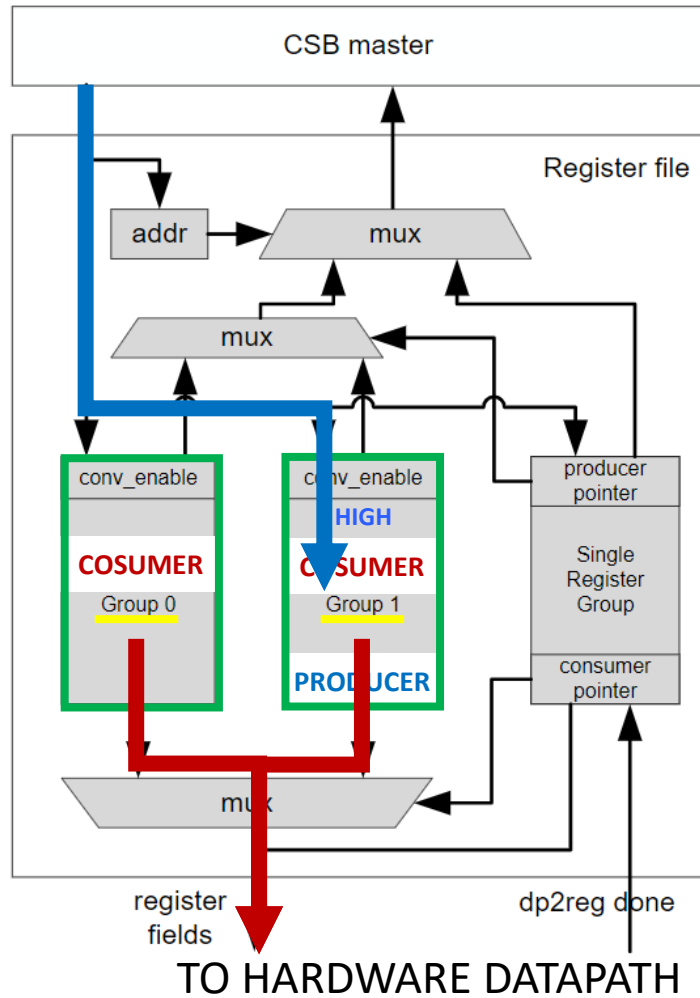


Outline

- Register Interface
- Memory Layout of Feature Data
- Memory Layout of Weights
- Memory Layout of Elementwise Operands
- Operator Descriptor
- Memory Descriptor
- Data Cube Structure



Register Interface

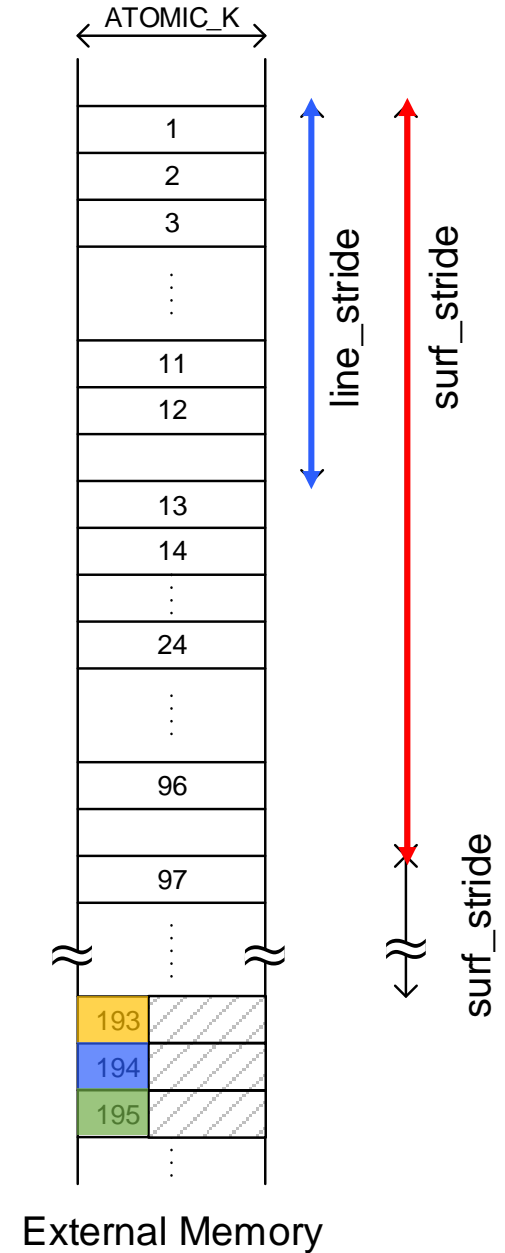
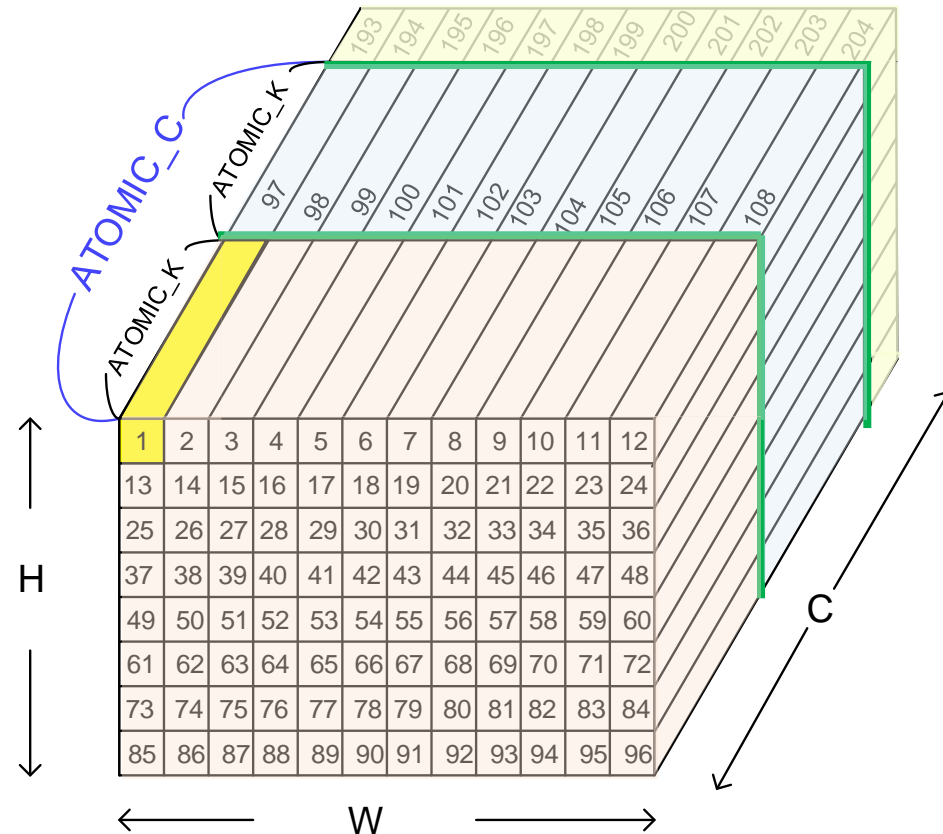


Adapted from <http://nvdla.org/hw/v1/hwarch.html#register-interface>

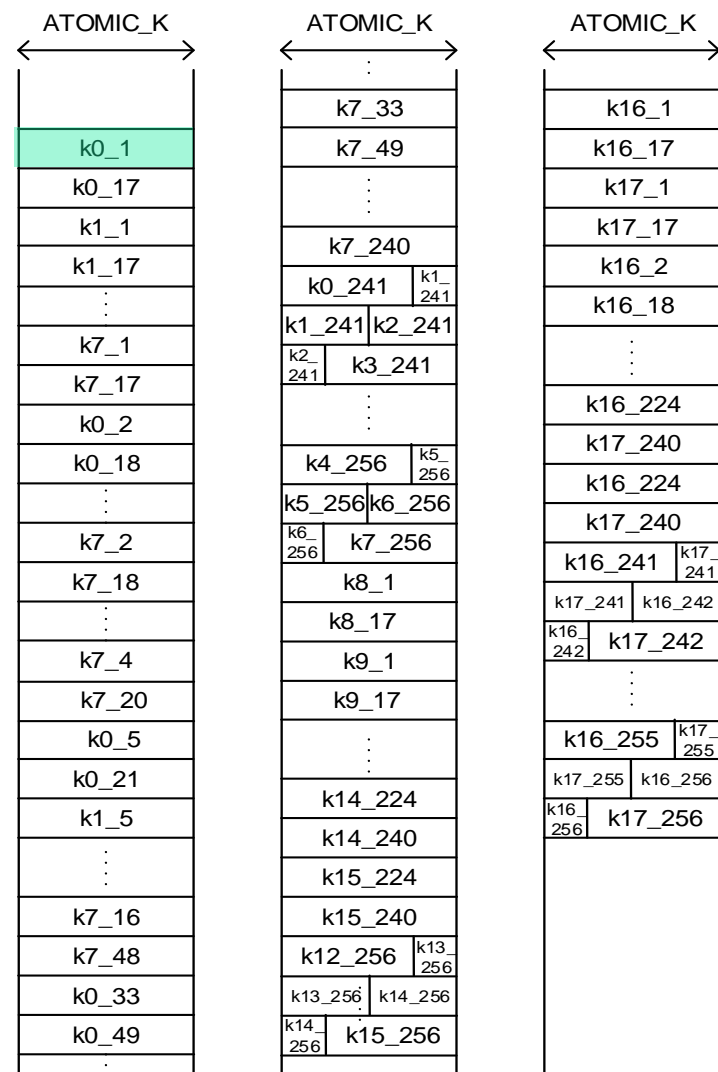
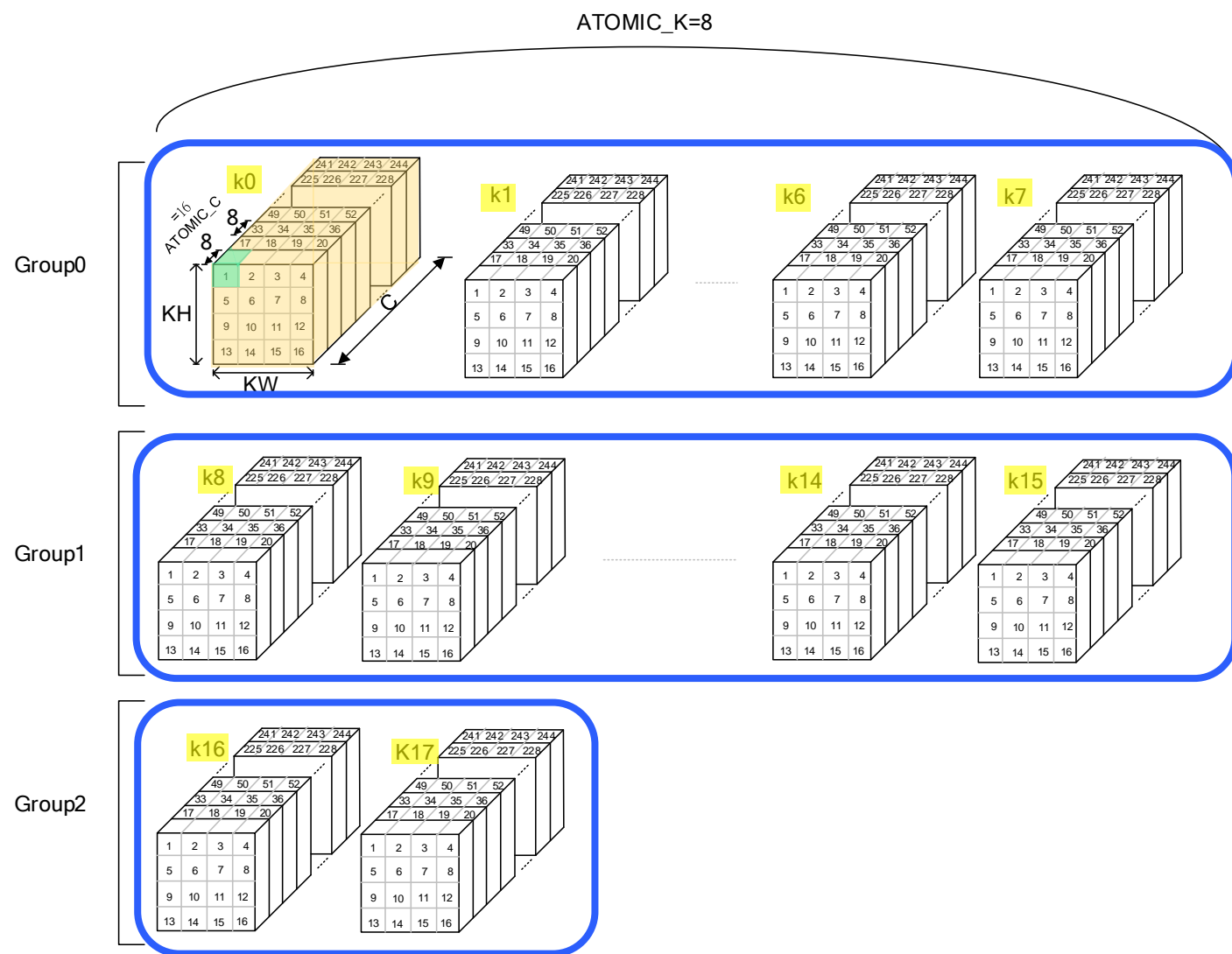


Memory Layout of Feature Data

- ATOMIC_K
- surface
- atomic_cube
- bubbles
- line_stride
- surf_stride



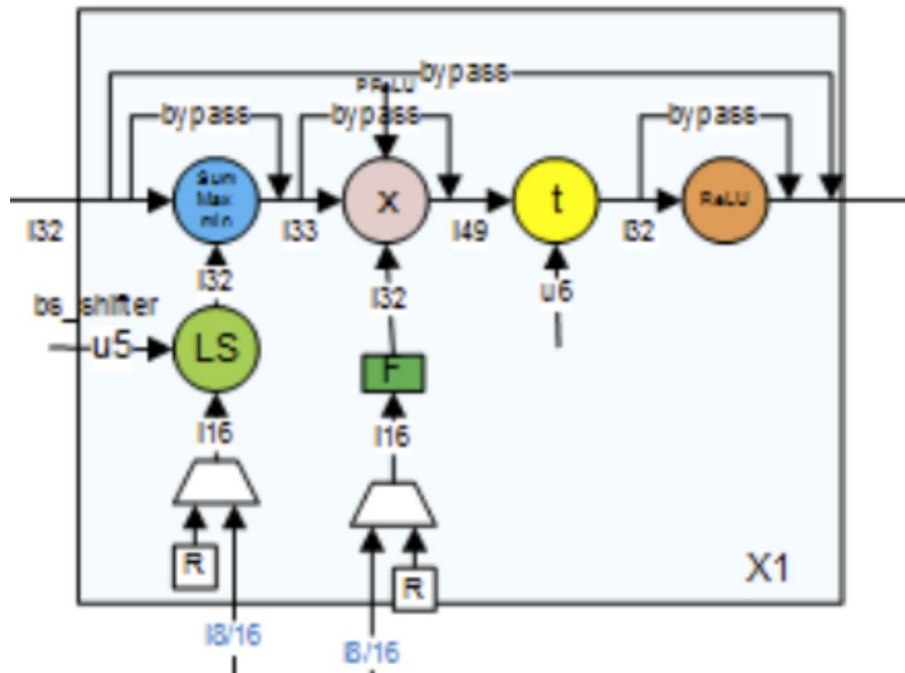
Memory Layout of Weight



External Memory



Memory Layout of Elementwise Operands



- Per-layer
- Per-channel
- Per-element

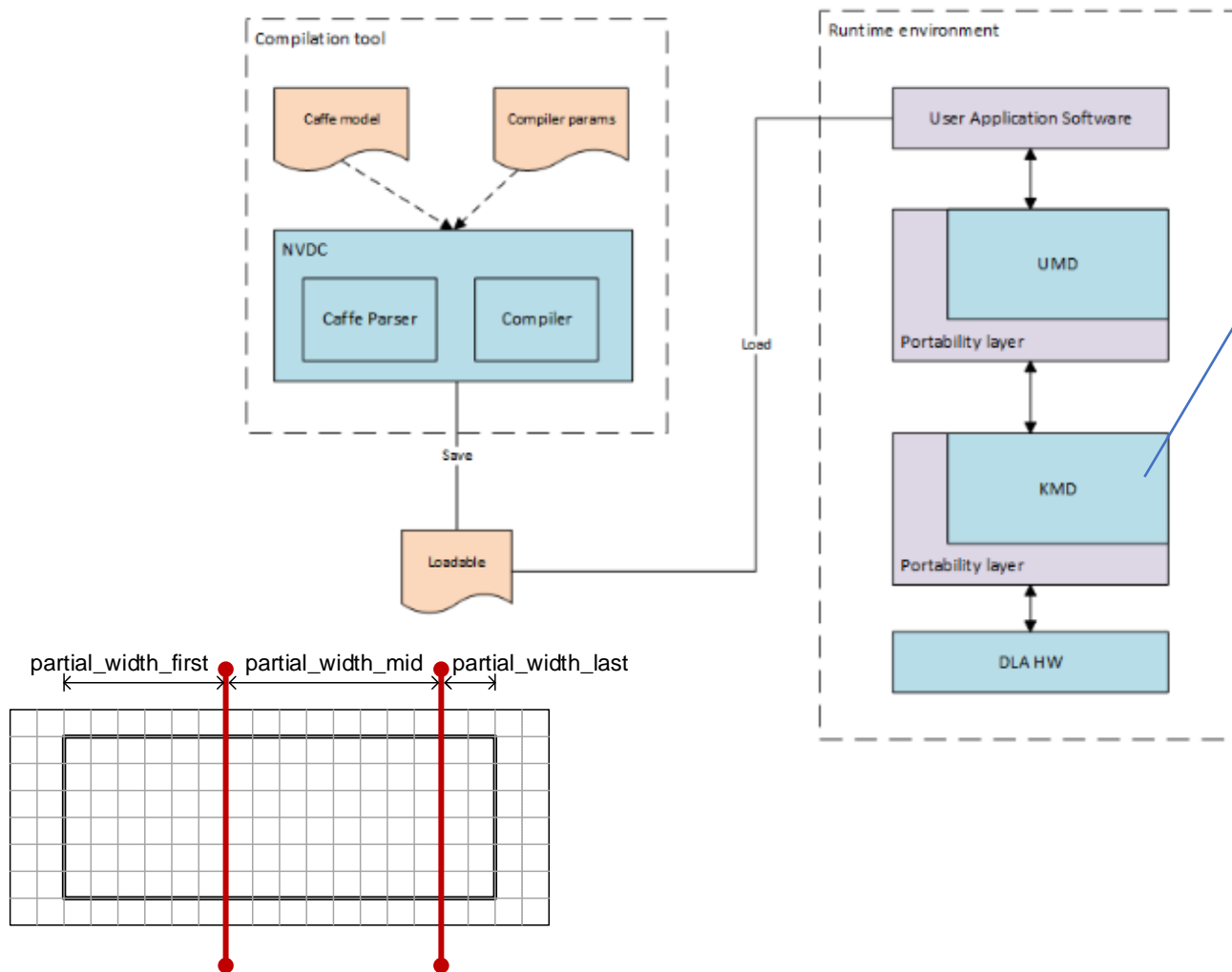
Address	0x0	0x1	0x2	0x3
ALU only	ALU0		ALU1	
MUL only	MUL0		MUL1	
ALU & MUL	ALU0		MUL0	

Adapted from http://nvdla.org/hw/v1/ias/unit_description.html#function-description



Operator Descriptors

operator
descriptor



[kmd/firmware/include/dla_interface.h](#)

```

1 struct dla_pdp_op_desc {
2     /* Performance parameters */
3     uint16_t partial_in_width_first;
4     uint16_t partial_in_width_mid;
5
6     uint16_t partial_in_width_last;
7     uint16_t partial_width_first;
8
9     uint16_t partial_width_mid;
10    uint16_t partial_width_last;
11
12    uint8_t split_num;
13
14    /* Algorithm parameters */
15    uint8_t pool_mode; /* dla_pool_mode */
16    uint8_t pool_width; /* dla_pool_width */
17    uint8_t pool_height; /* dla_pool_height */
18
19    uint8_t stride_x;
20    uint8_t stride_y;
21
22    /**
23     * The left/right padding size,
24     * pad_right might be less than pad_left
25     */
26    uint8_t pad_left;
27    uint8_t pad_right;
28
29    /* The top/bottom padding size */
30    uint8_t pad_top;
31    uint8_t pad_bottom;
32
33    /* Precision parameters */
34    uint8_t precision; /* dla_precision */
35    uint8_t reserved0;
36    /**
37     * if input has non-zero "offset", this value should be set
38     * There'll be 7 different padding values, the relationship between
39     * those versions are:
40     * padding_value[0] = -offset*scaling;
41     * padding_value[1] = 2*padding_value[0]
42     * padding_value[2] = 3*padding_value[0]
43     * ...
44     * The purpose is to avoid ucode implement FP16
45     * multiplier(for FP16 mode)
46     */
47    int32_t padding_value[PDP_PAD_VAL_NUM];
48 } __packed __aligned(4);

```

http://nvdla.org/hw/v1/ias/unit_description.html#planar-data-processor



Memory Descriptors

```

struct dla_sdp_surface_desc {
    /* Data cube */
    /* source input cube, available when SDP working on offline mode */
    struct dla_data_cube src_data;

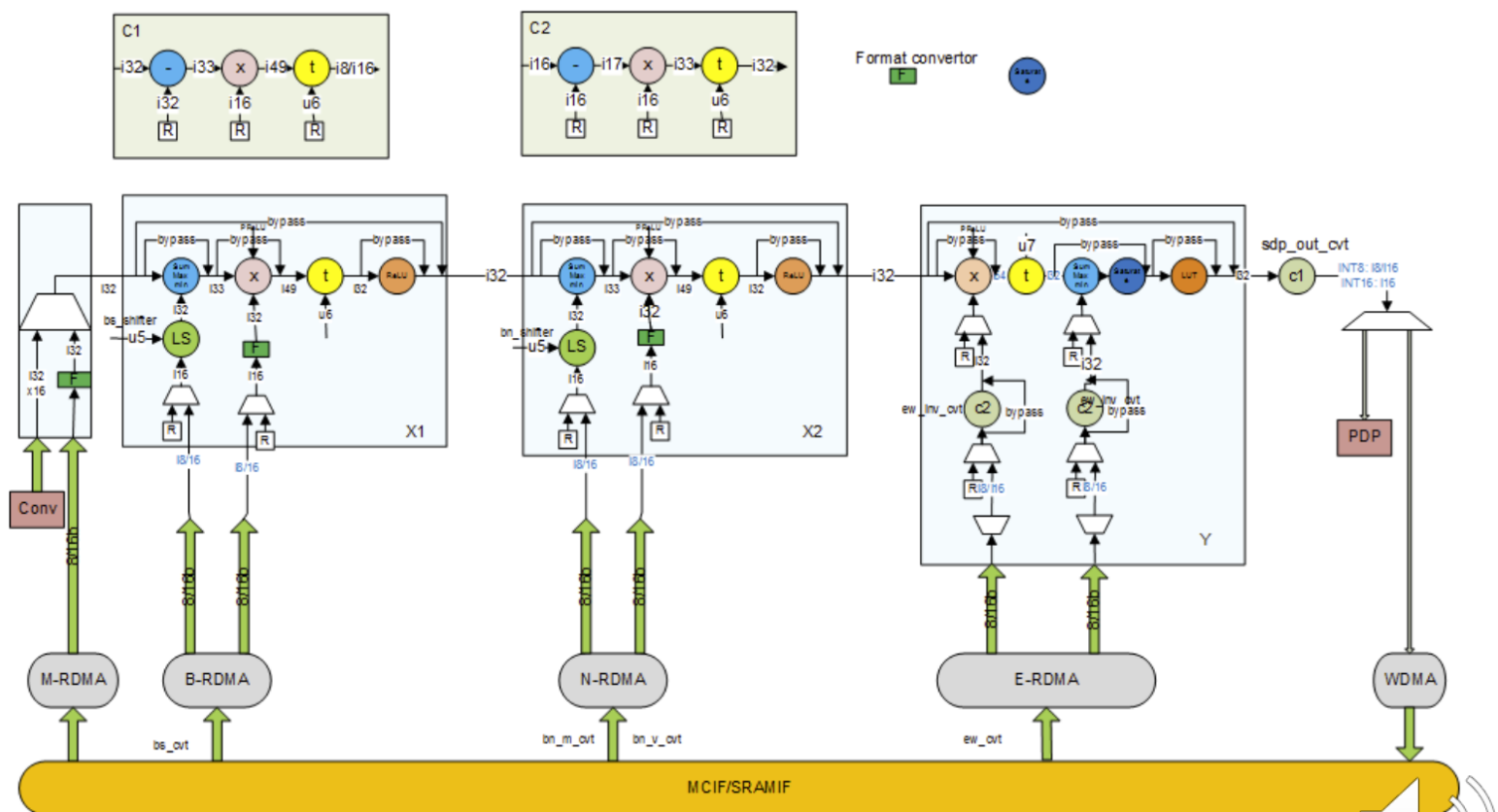
    /* X1 input cube */
    struct dla_data_cube x1_data;

    /* X2 input cube */
    struct dla_data_cube x2_data;

    /* Y input cube */
    struct dla_data_cube y_data;

    /* Output cube */
    struct dla_data_cube dst_data;
} __packed __aligned(4);
    
```

dla_sdp_surface_desc



Adapted from http://nvdla.org/hw/v1/ias/unit_description.html#function-description

Data Cube Structure

```
#define DLA_MEM_MC    0 /* External DRAM */  
#define DLA_MEM_CV    1 /* CV-SRAM */  
#define DLA_MEM_HW    2 /* DLA sub-module */
```

```
struct dla_data_cube {  
    uint16_t type; /* dla_mem_type */  
    int16_t address; /* offset to the actual IOVA in  
                     /* cube dimensions */  
    uint32_t offset; /* offset within address */  
    uint32_t size;  
  
    uint16_t width;  
    uint16_t height;  
  
    uint16_t channel;  
    uint16_t reserved0;  
  
    /* stride information */  
    uint32_t line_stride;  
    uint32_t surf_stride;  
  
    /* For Rubik only */  
    uint32_t plane_stride;  
} __packed __aligned(4);
```



References

■ NVDLA

- <http://nvdla.org/>
- <https://github.com/nvdla>

■ ONNC

- <https://github.com/ONNC/onnc>





Skymizer Taiwan Inc.

CONTACT US

E-mail sales@skymizer.com **Tel** +886 2 8797 8337

HQ 12F-2, No.408, Ruiguang Rd., Neihu Dist., Taipei City 11492, Taiwan

BR Center of Innovative Incubator, National Tsing Hua University, Hsinchu Taiwan



skymizer

Boost deep learning accelerator
with compiler technology



<https://skymizer.com>

