Design and Implementation of Low-Power Embedded 3D Graphics Rendering Engine for Mobile Applications using the Embedded Memory Logic Technology

Ramchan Woo

2000. 12. 19

Semiconductor System Laboratory
Department of Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology (KAIST)

Outline

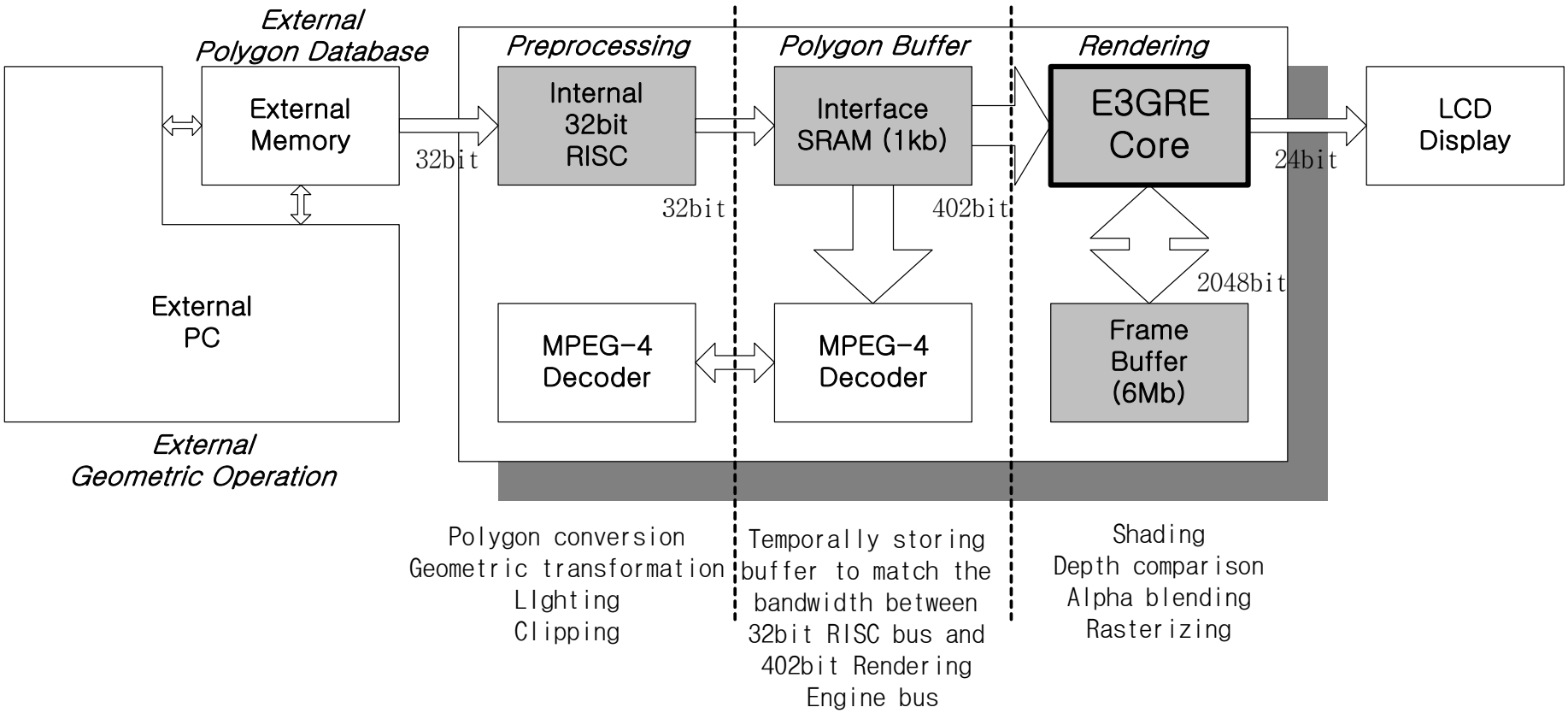
- Introduction
- Architecture
- Memory Access
- Circuit Implementation
- Performance Comparison
- Conclusion and Further Works

Introduction

- **Multimedia PDA-Chip**
  - Personal Information Management
  - MP3 Playback
  - Realtime Video Decoding
  - **3D Graphics Rendering**

World's First
One-chip Implementation
for PDA

3D-CG in PDA-Chip



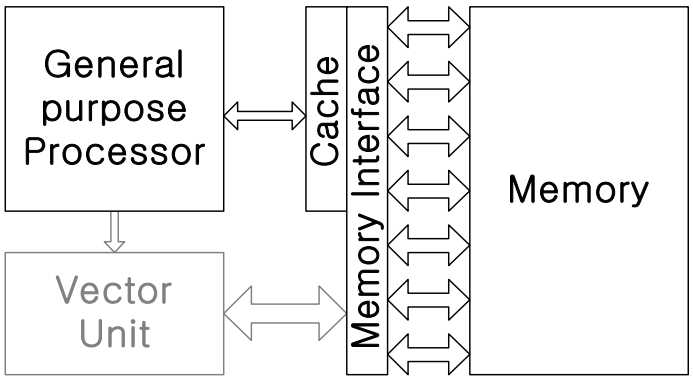
E3GRE Features

- Gouraud Shading
- Hidden Surface Removal with 16bit Z-buffer
- Alpha-Blending for Transparency
- Double-Buffering for Flicker-Free Animation
- Direct Video Transfer through SAM
- 24bit True color on 256 x 256 screen
- 2.22Mtris/sec @ 20MHz
- 3.2GByte/sec Memory BW @ 20MHz
- 6Mb Embedded DRAM Frame-Buffer
- Low Power consumption

Outline

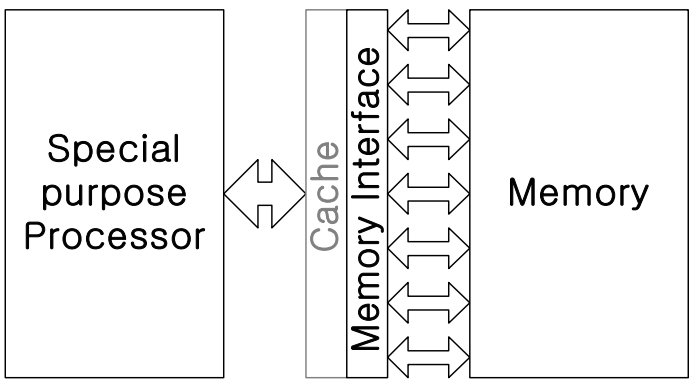
- Introduction
- **Architecture**
- Memory Access
- Circuit Implementation
- Performance Comparison
- Conclusion and Further Works

Previous EML Architecures



GPP + Single DRAM

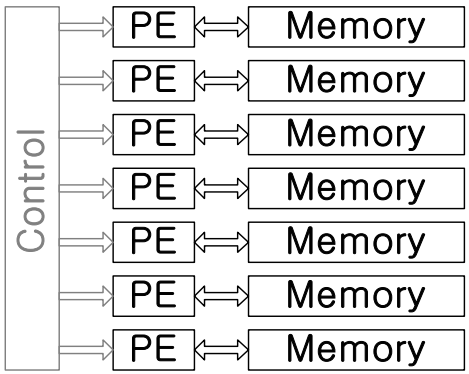
- IRAM (Berkeley), M32RD (Mitsubishi)
- General-Purpose Scalar Processor
- Sequential Memory Access with Bank Interleaving
- Small-Width of Datapath (32b or 64b)
- Optional Vector Unit (1024b -> 32b)
- Huge Area, Huge Power



SPP + Single DRAM

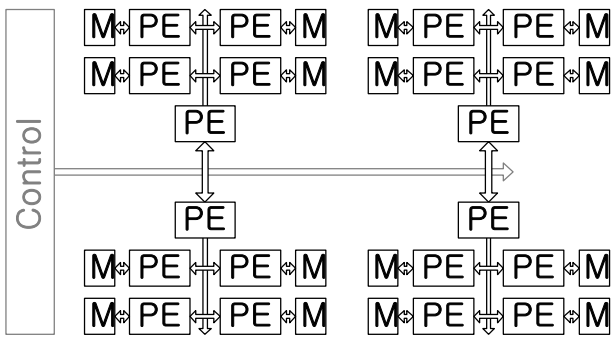
- 3D-RAM (Mitsubishi), GS@PS2 (SONY)
- Special-Purpose Scalar Processor
- Large-Width of Datapath (128b)
- Sequential Memory Access with Bank Interleaving
- High Power due to High-CLK @ SPP

Previous EML Architecures (Cont'd)



1D Array (PEs+DRAMs)

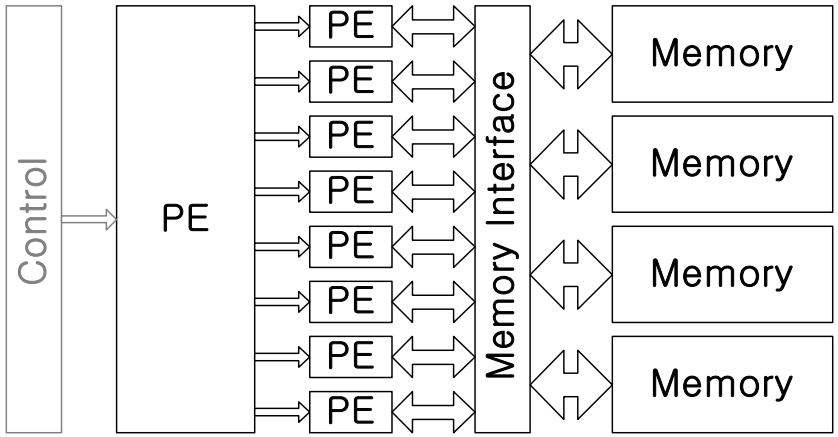
- Media-Chip (Hitachi), PIP-RAM (NEC)
- Processing Elements
- Independent Memory Access
- Extra Controller is Required for 3D-CG
- Low-Power
- Area Panelty due to Bad DRAM Cell-Efficiency
- Layout Difficulty



2D Array (PEs+DRAMs)

- RamP-1 (KAIST)
- Processing Elements
- Independent Memory Access
- Well matched to 3D-CG
- Low-Power, High-Performance
- Large Area Panelty due to Bad DRAM Cell-Efficiency
- Layout Difficulty

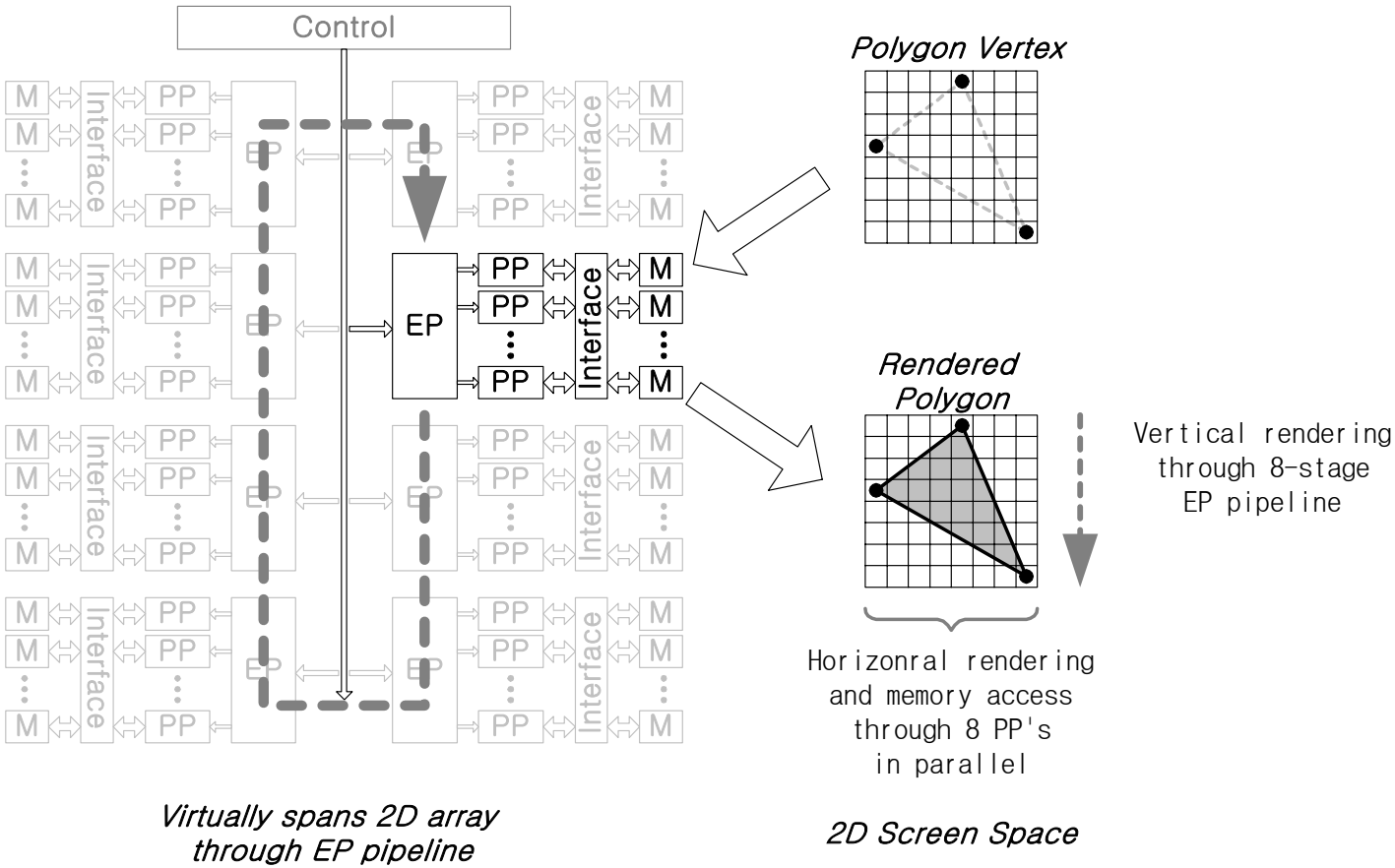
Proposed ViSTA Architecture



Virtually Spanning 2D Array (ViSTA)

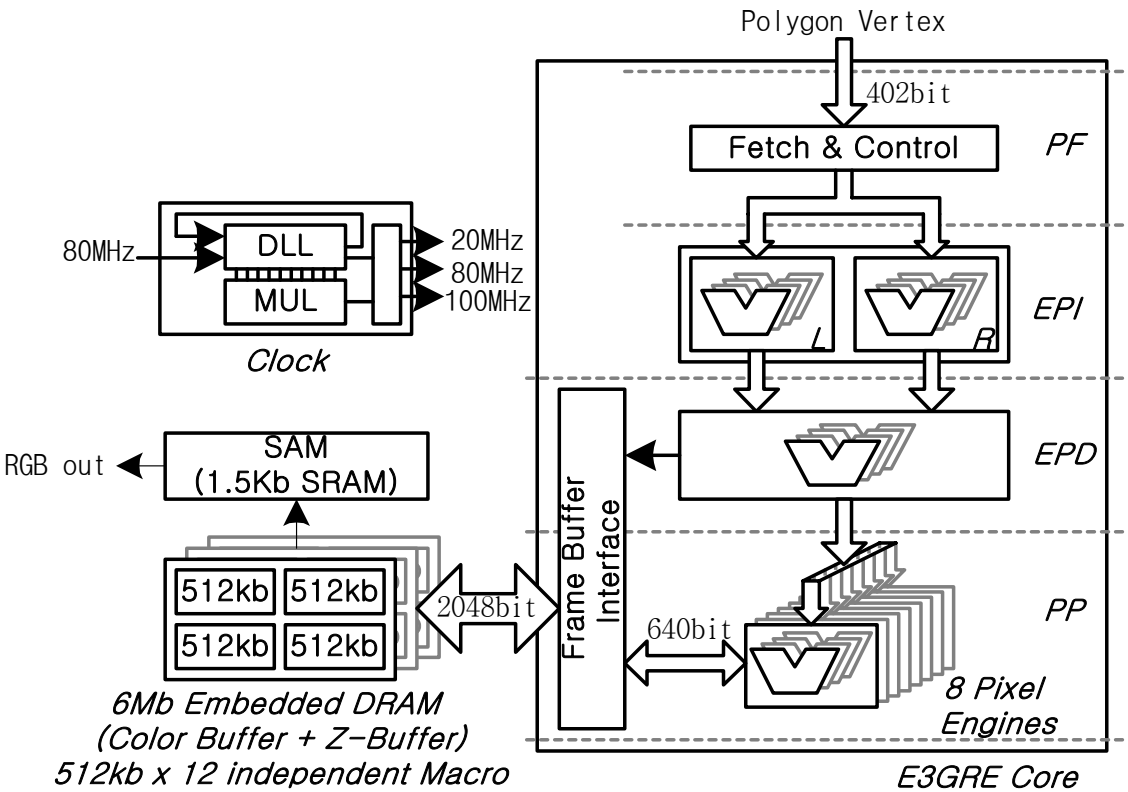
- Hierarchical 1+8 Processing Elements
- Independent-controlled DRAM's
- "Logically Local" / "Physically Global" Frame Buffer
- Intellegent Memory Interface Circuit
- Low-Power, High-Performance, Small-Area
- Applicable to PDA-3D

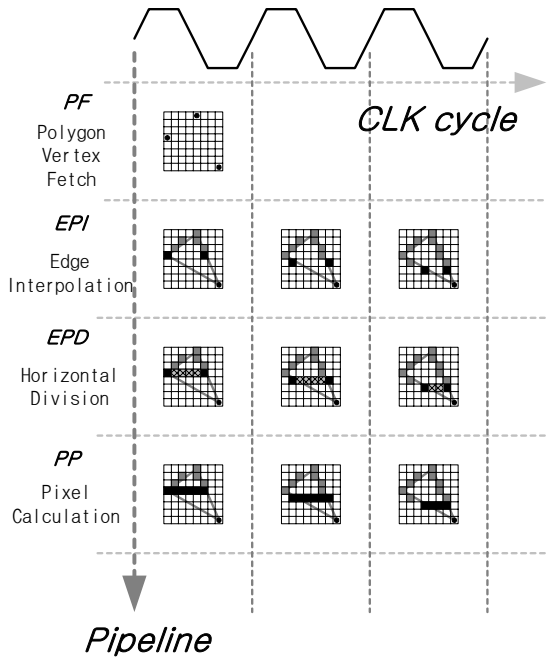
ViSTA Operation



E3GRE Block Diagram

- **Core and eDRAM are separated by FBI**

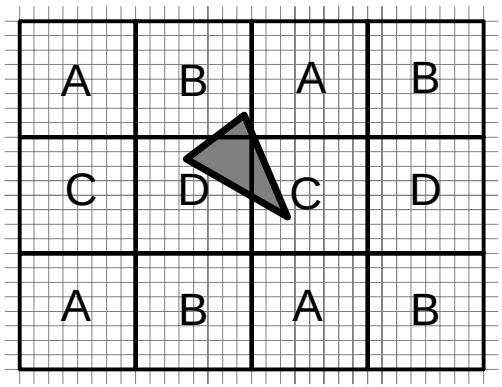


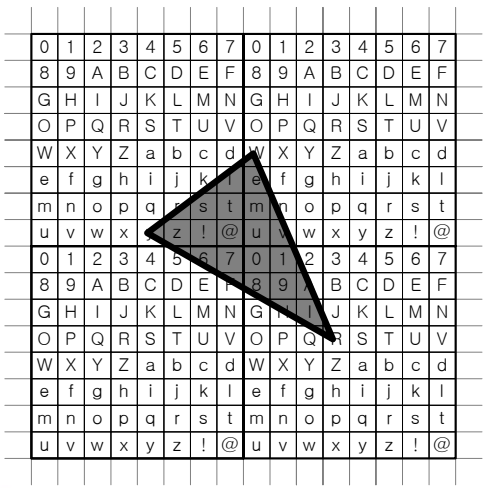


Outline

- Introduction
- Architecture
- **Memory Access**
- Circuit Implementation
- Performance Comparison
- Conclusion and Further Works

Conventional Memory Mappings





2D Blocking

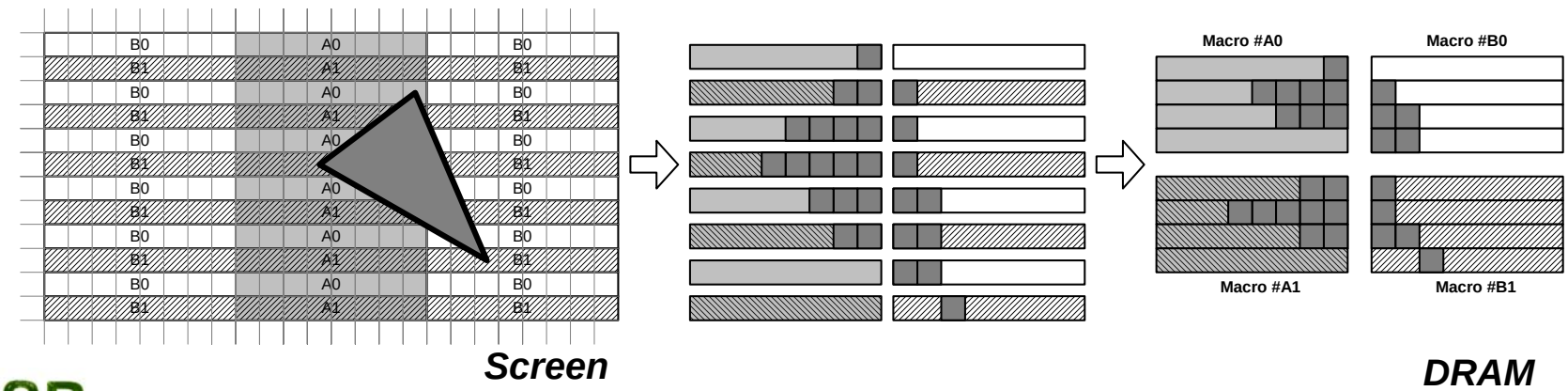
- PC-Graphics
- 4-way Bank Interleaving
- 1 Block maps into 1 DRAM Row
- Unnecessary Power Consumption
- Serial Pixel Access
- Doesn't fit well to Parallel Rasterization

Independent Assignment

- RamP-1, PixelFlow, InfiniteReality
- Independent Memory Control
- Reduced Power Consumption
- Parallel Pixel Access
- Well matched to Parallel Rasterization
- Increased Die Area due to DRAM cell efficiency
- Layout difficulty

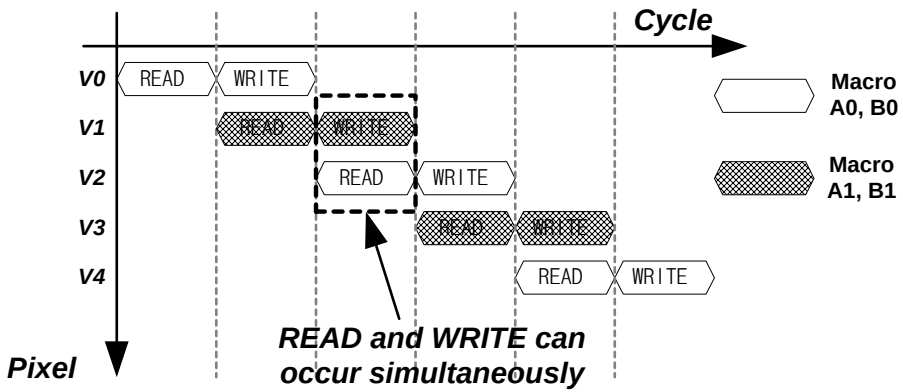
Proposed Memory Mapping

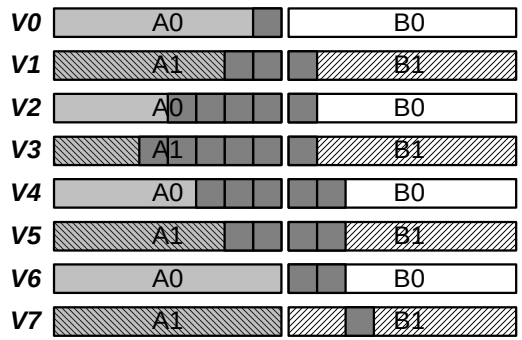
- **Selective and Alternative Line-Block Activation (SALBA)**
  - 4 Independent Memory
  - Line-Block consists of 8x1 Screen Pixels
  - 1 Line-Block maps into 1 DRAM SWD
  - Line-Block Read/Write with embedded-DRAM (x320/block)

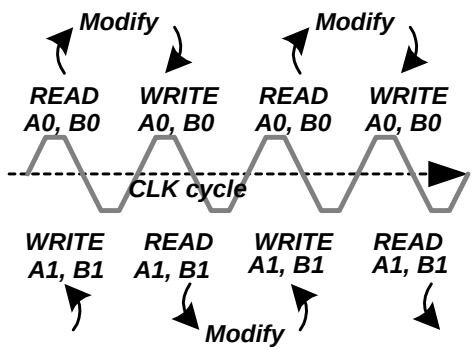


Memory Mapping (cont'd)

- **Simultaneous and Continuous RMW**

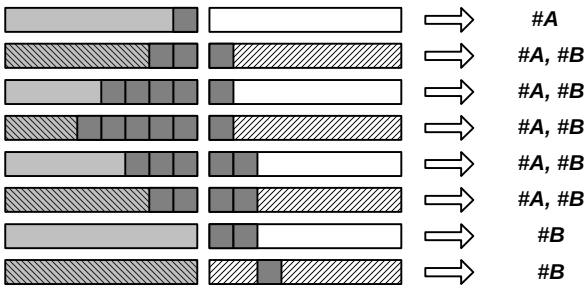






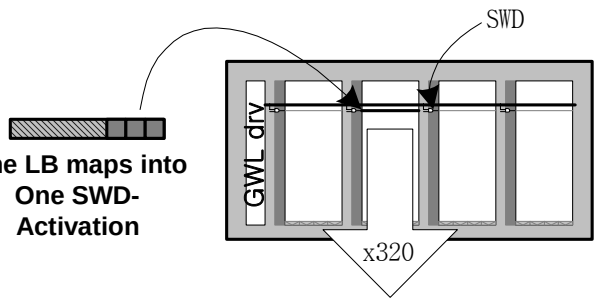
Low-Power DRAM Access

- **Selective Macro Activation (SMA)**



Active only necessary Macro(s)

- **Partial Wordline Activation (PWA)**

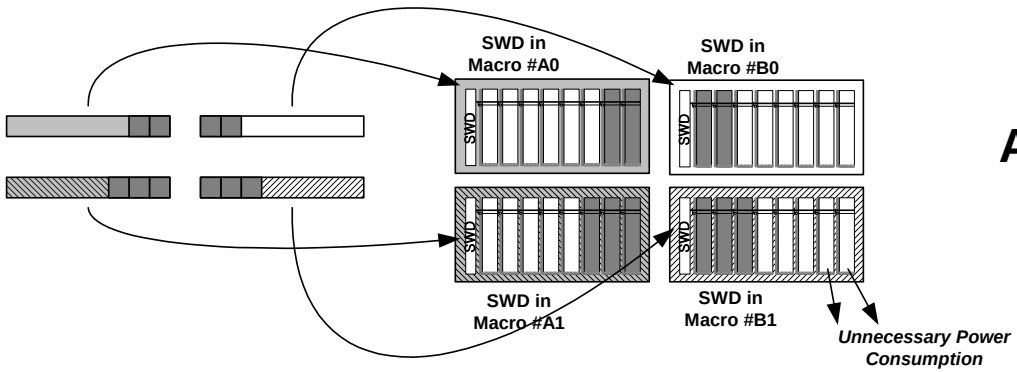


Simultaneously read or write
8pixel x (24bit RGB + 16bit Z)

Active only necessary Sub-Wordline

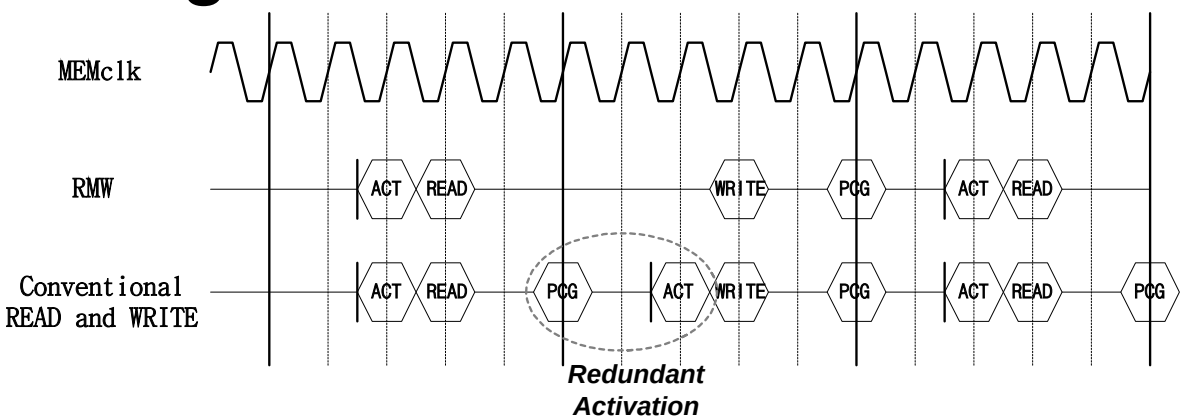
Low-Power DRAM Access

- **Partial I/O Activation (PIA)**



Active only necessary I/O

- **Eliminating Redundant Commands**

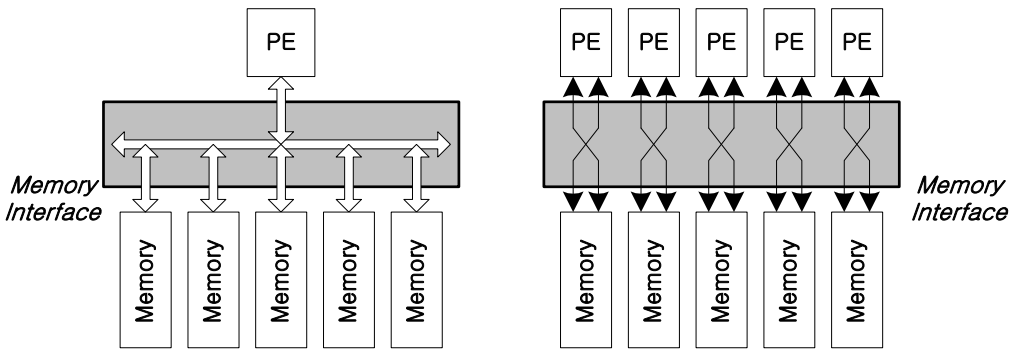


Memory Interface in ViSTA/SALBA

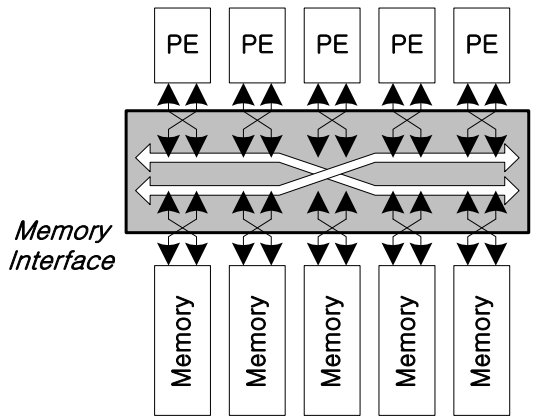
- **Comand Generation for 12 Memories**
- **Run-time bus reconfiguration**
  - Optimized for RMW data transaction
  - Data transfer between (2048bit @ 100MHz 2.5V MEM) and (640bit @ 20MHz 1.5V PP)
- **Macro-based Design**
  - Any kind of DRAM macro can be attached
  - No wait for DRAM macro in RE core design
  - No more pitch-matched design of PE's
  - More layout freedom to designers

Run-Time Bus Reconfiguration

- **Conventional**

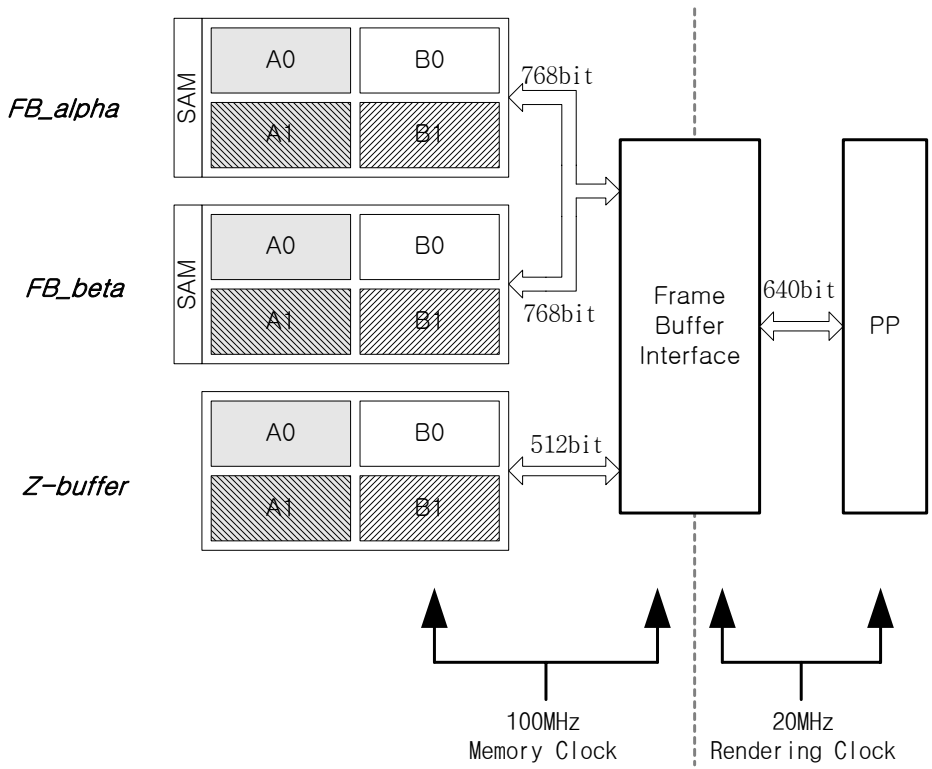


- **Proposed**



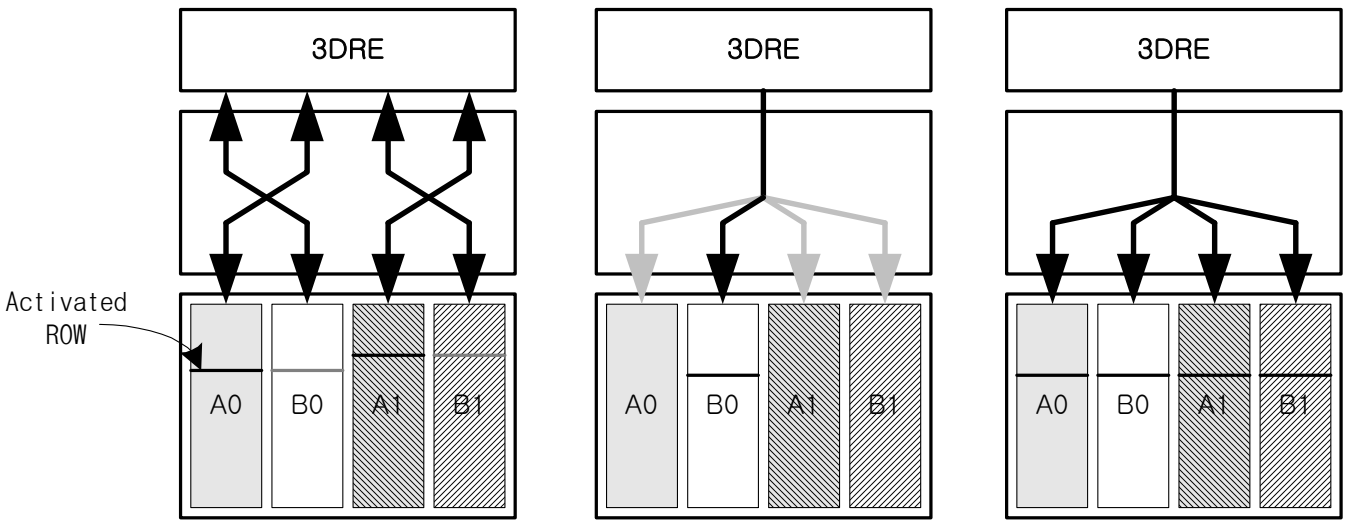
Memory Configuration

- **3.2GByte/sec Memory BW @ 20MHz**
  - 20MHz x 2rmw x 8PPs x 40RGBZ x 2ABmem



Memory Access Modes

- (a) : Normal Read-Modify-Write
- (b) : Memory Test / FB Initilize
- (c) : Refresh

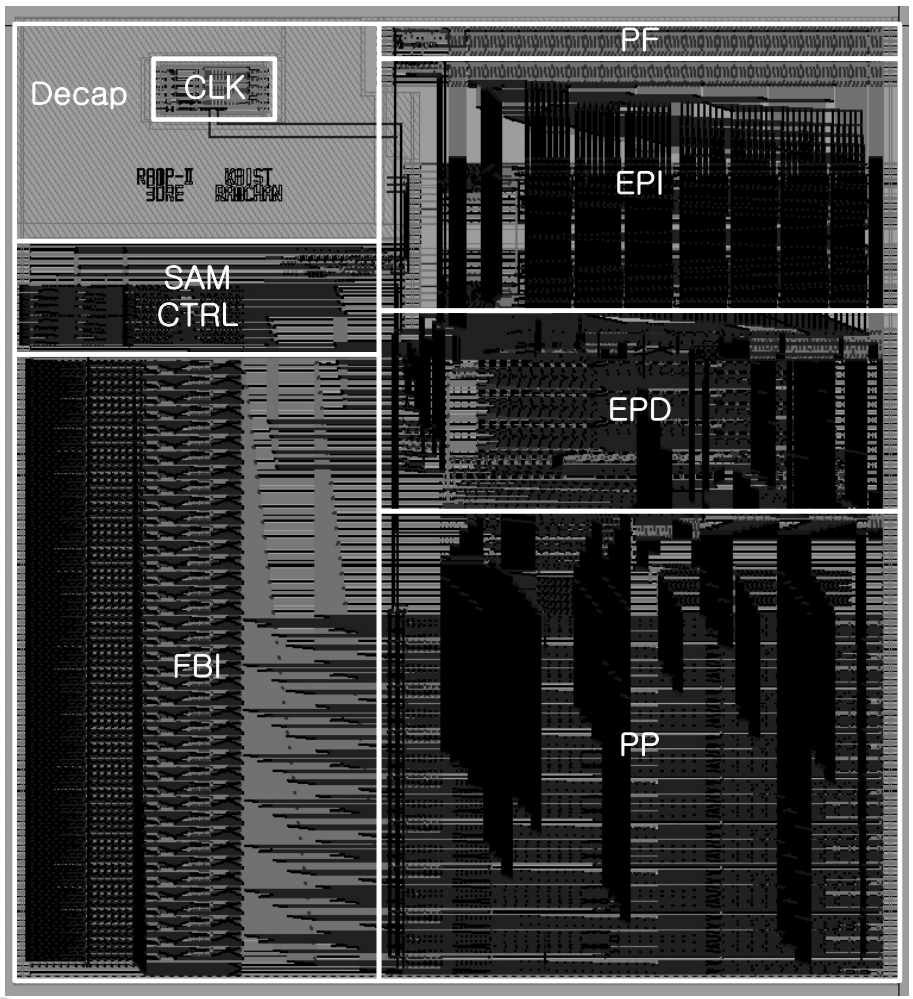


(a) Communication (b) Individual Activation (c) Broadcasting

Outline

- Introduction
- Architecture
- Memory Access
- **Circuit Implementation**
- Performance Comparison
- Conclusion and Further Works

E3GRE Core Floorplan



Low-Power Datapath

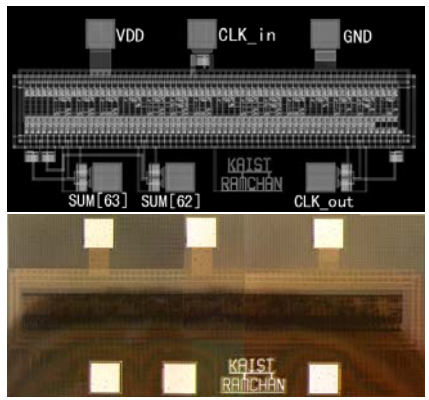
- PowerPC 603 Latch
- Adder
- Divider
- Alpha-Blend Unit

Special Functional Unit

- Bus Interface Circuit
- Clock Multipler/Generator

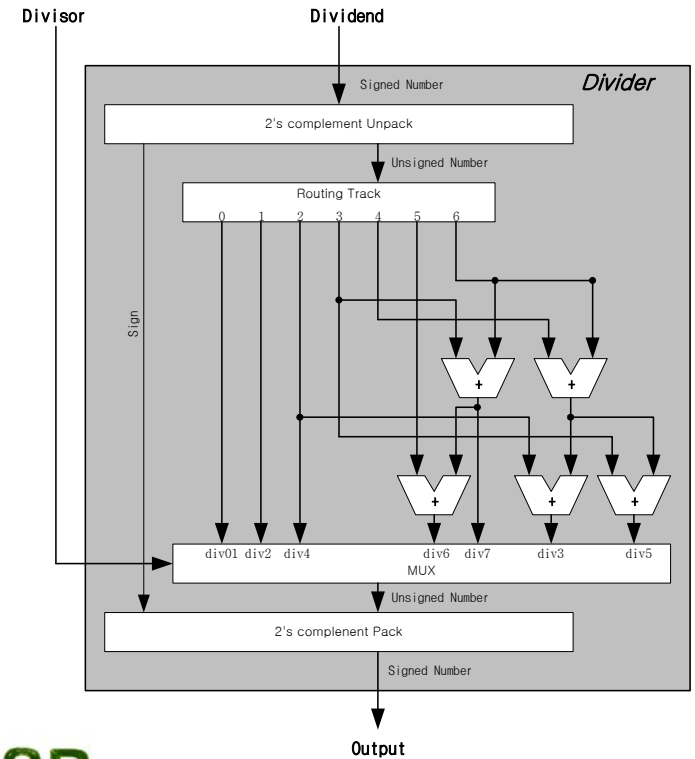
Adder

- **Low-Power and Fast Addition**
  - More than 100 adders
  - Static version of “A 670ps, 64bit Dynamic Low-Power Adder @ 0.25um, 2.5V”, [ISCAS 2000]
    - DORP instead of XORP
    - Separated Carry Propagation Path
    - Efficient Carry Grouping
  - 11bit fixed-point for 8bit R,G,B,X,Y
  - 19bit fixed-point for 16bit Z
  - Less than 0.01mW power consumption
  - Less than 1ns 19bit addition time



SAS Divider

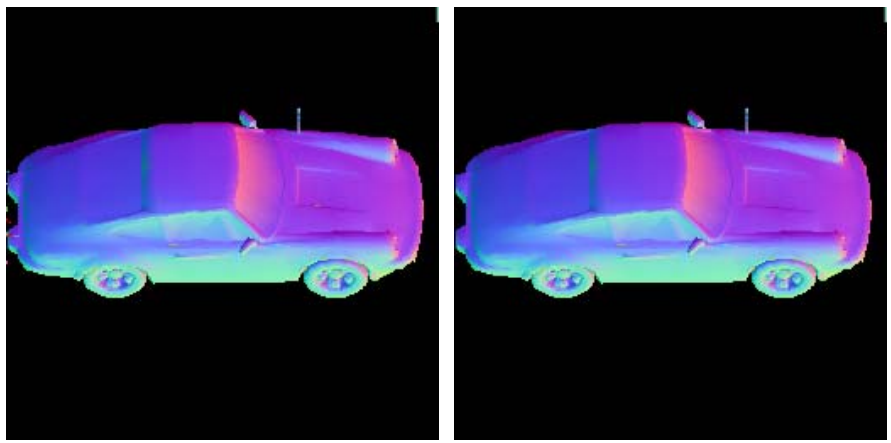
- **Shift-Add-Select Approximation for Low-Power and 1-cycle Division**



D	1/D (exact)	1/D (approx.)	Error (% of X)
0 (000)	Not Defined		
1 (001)	1	1 $=(1/2)^0$	0
2 (010)	0.5	0.5 $=(1/2)^1$	0
3 (011)	0.333333…	0.328125 $=(1/2)^2+(1/2)^4+(1/2)^6$	0.52%
4 (100)	0.25	0.25 $=(1/2)^2$	0
5 (101)	0.2	0.203125 $=(1/2)^3+(1/2)^4+(1/2)^6$	0.31%
6 (110)	0.166666…	0.171875 $=(1/2)^3+(1/2)^5+(1/2)^6$	0.52%
7 (111)	0.142857	0.140625 $=(1/2)^3+(1/2)^6$	0.22%

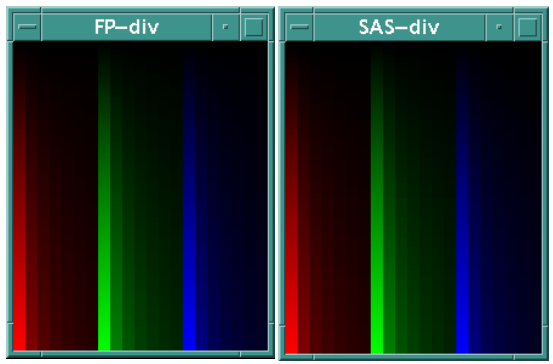
SAS Divider (Cont'd)

- **Image Comparison between FP-div and SAS-div**



Conventional Floating-Point Divider

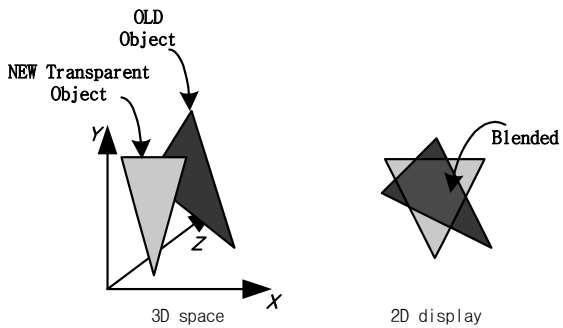
Proposed SAS-Approximation Divider



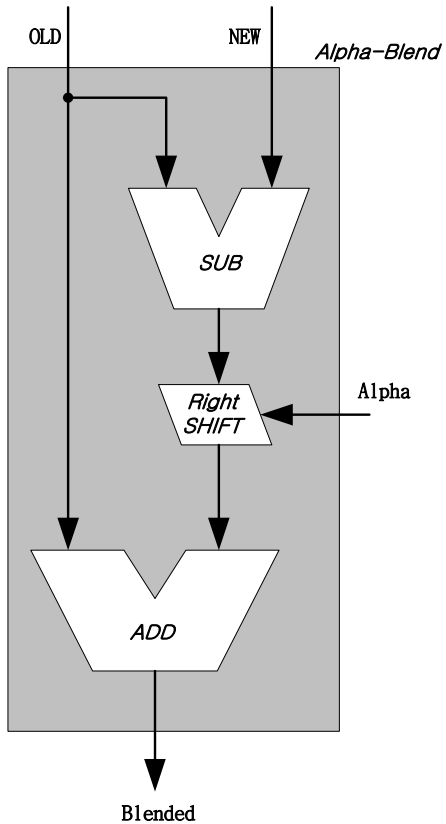
FP-div SAS-div

Alpha Blend Unit

- **Restrict Alpha Value for Low-Power Operation -> Mul-Free**
  - A=100%, 50%, 25%, 12.5%, 0%
  - 24 Alpha-Blend Unit

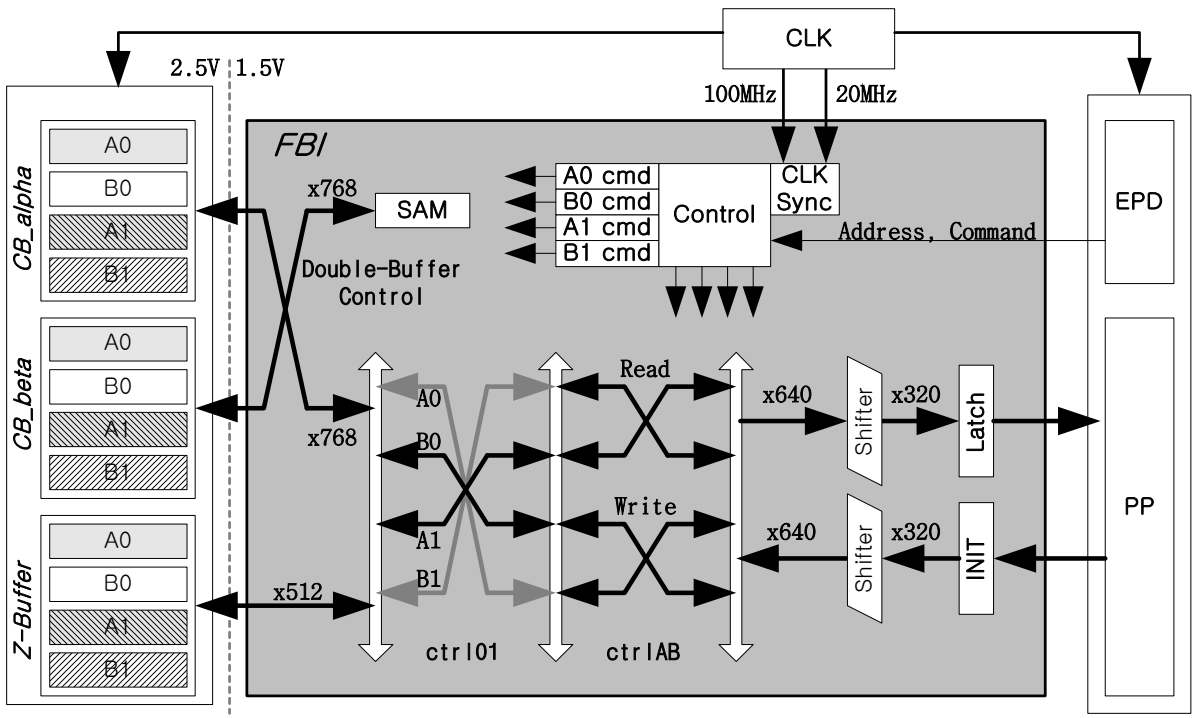


$$\mathrm{BLEND} = \mathrm{A} \times \mathrm{NEW} + (1 - \mathrm{A}) \times \mathrm{OLD}$$
$$= \mathrm{A} \times (\mathrm{NEW} - \mathrm{OLD}) + \mathrm{OLD}$$

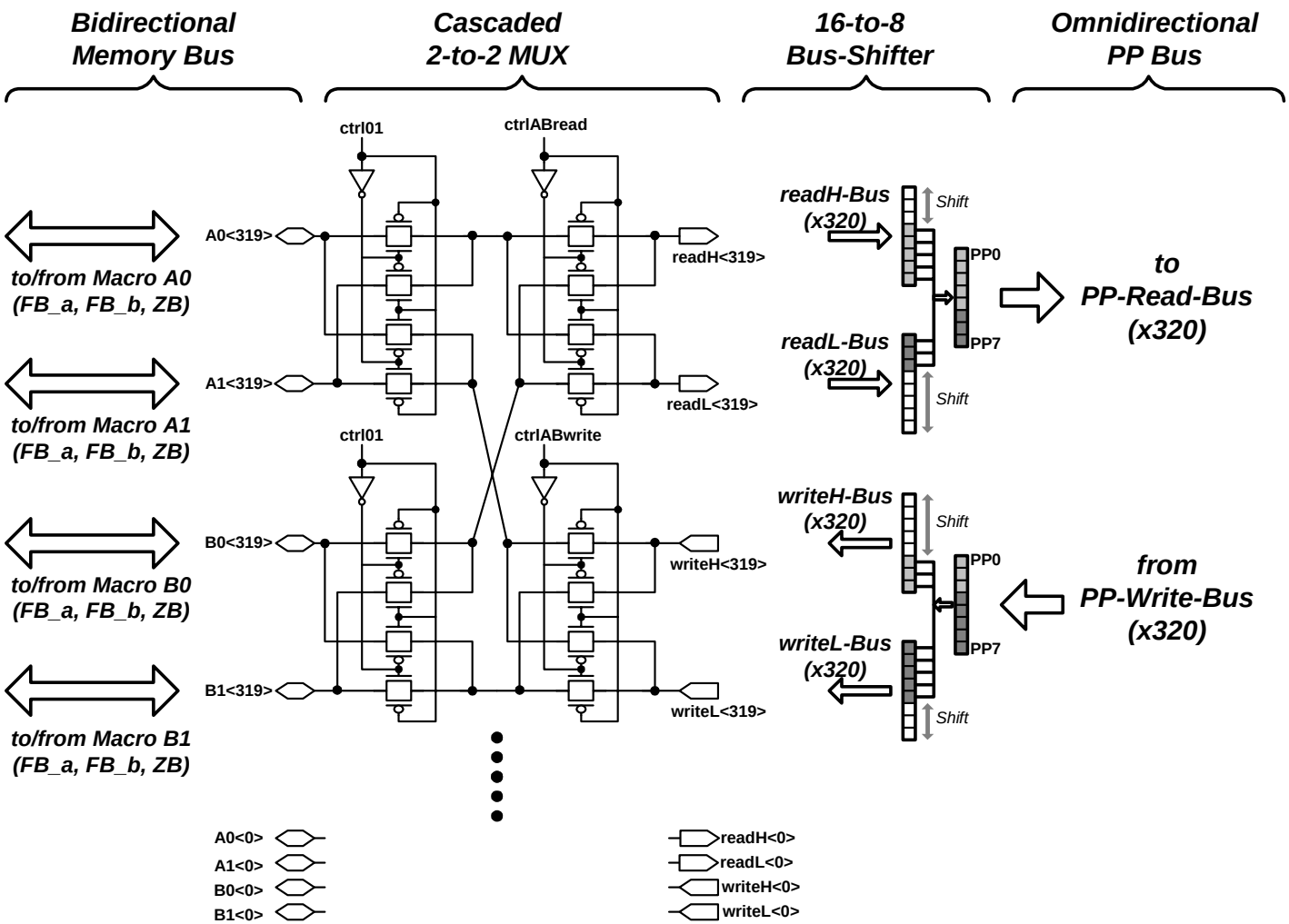


Frame Buffer Interface

- **Run-Time Bus Reconfiguration**
- **DRAM Control (12 independent Macros)**

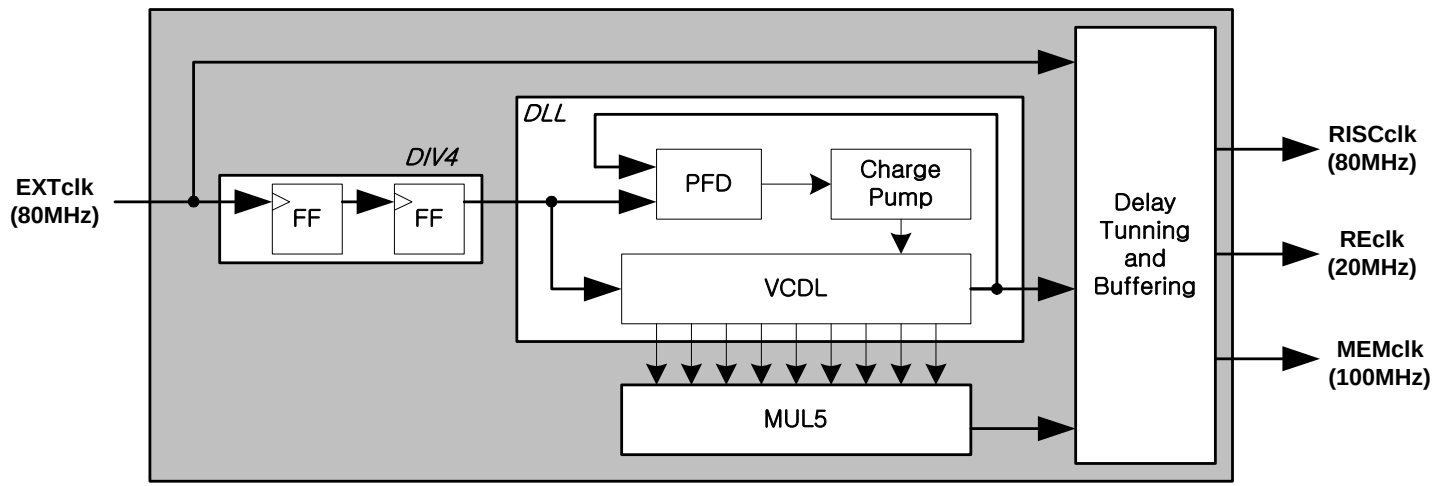


FBI Circuits

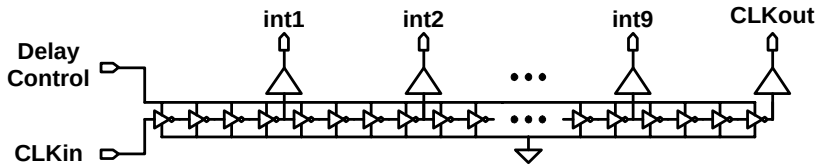


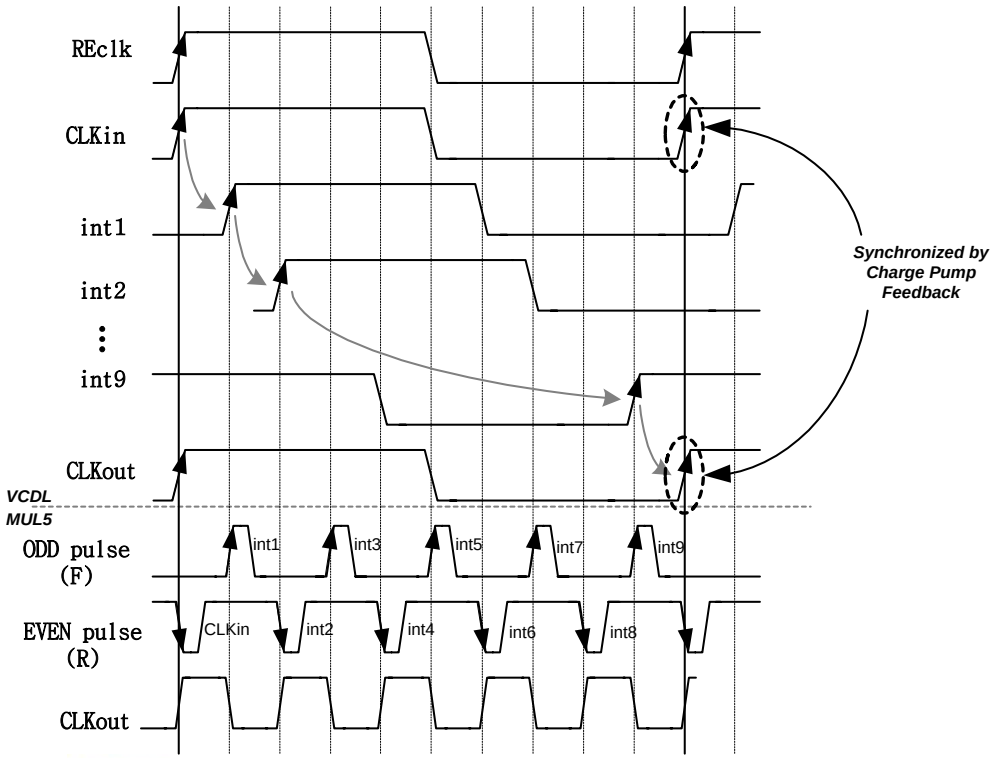
Clock Generation/Multiplication

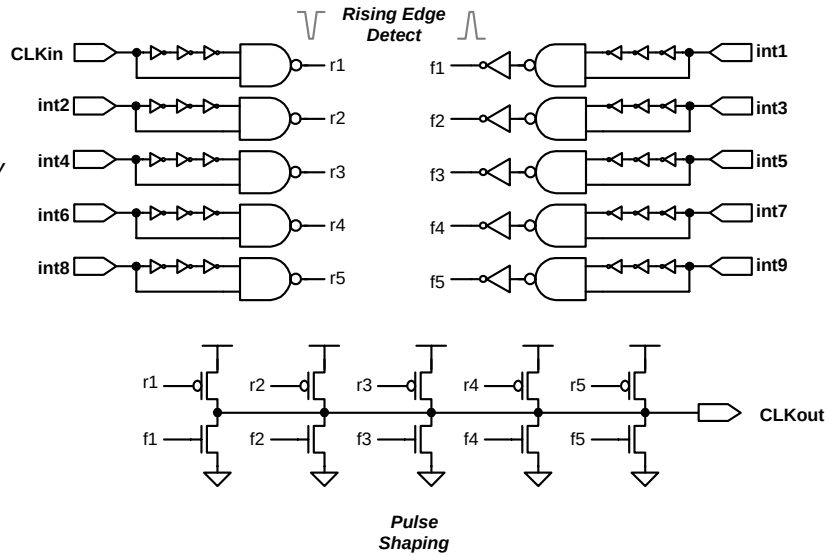
- **Multiple Clocks**
  - RISC : 80MHz / E3GRE : 20MHz / FB : 100MHz
- **DLL-based**



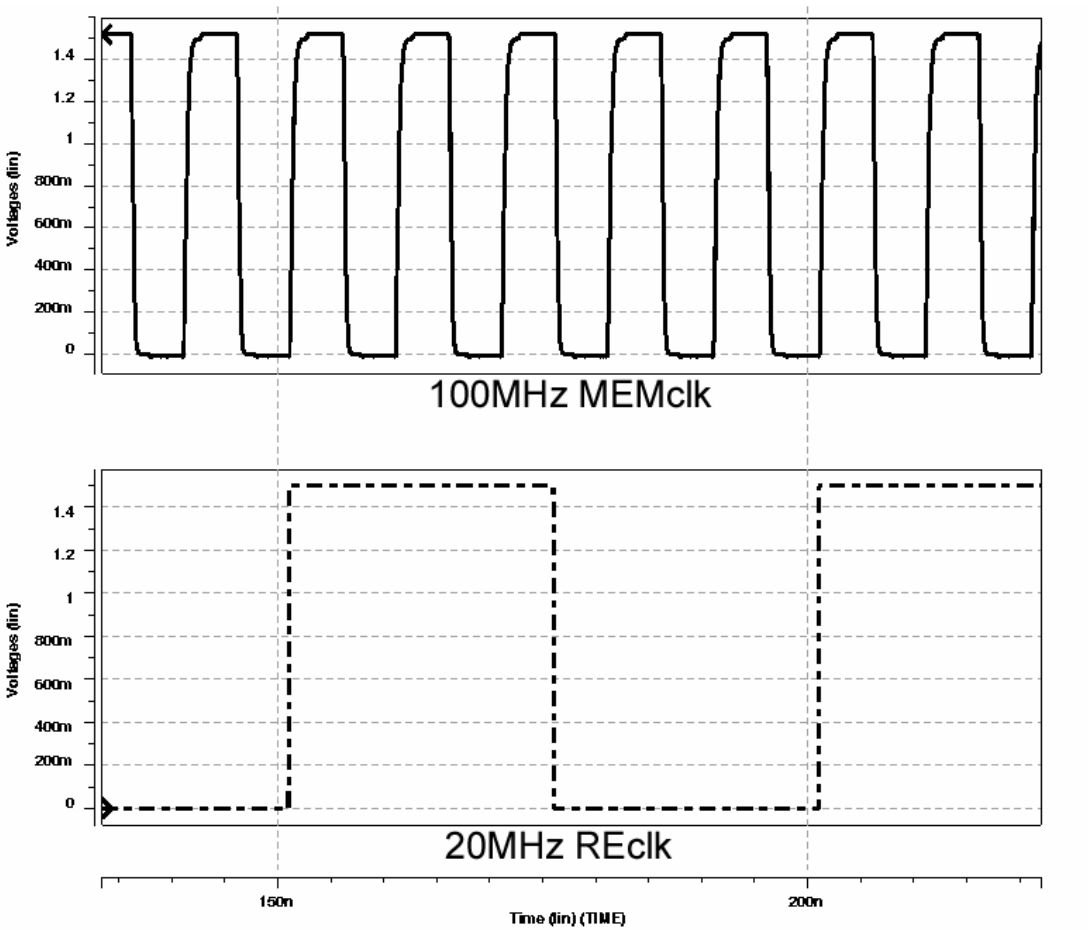
CLK Circuit

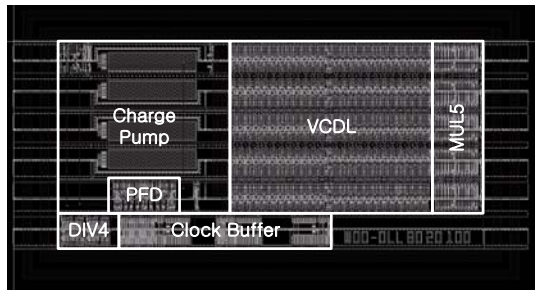

int1
int2
int9
CLKout
Delay Control
CLKin

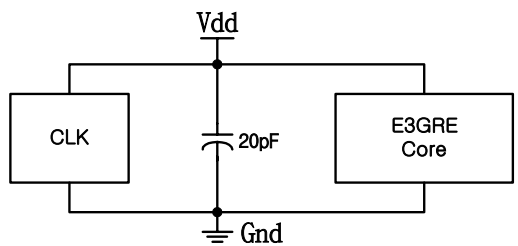

REclk
CLKin
int1
int2
int9
CLKout
VCDL
MUL5
ODD pulse (F)
EVEN pulse (R)
CLKout
int1
int3
int5
int7
int9
CLKin
int2
int4
int6
int8
Synchronized by Charge Pump Feedback


Rising Edge Detect
CLKin
int2
int4
int6
int8
r1
r2
r3
r4
r5
f1
f2
f3
f4
f5
int1
int3
int5
int7
int9
CLKout
Pulse Shaping

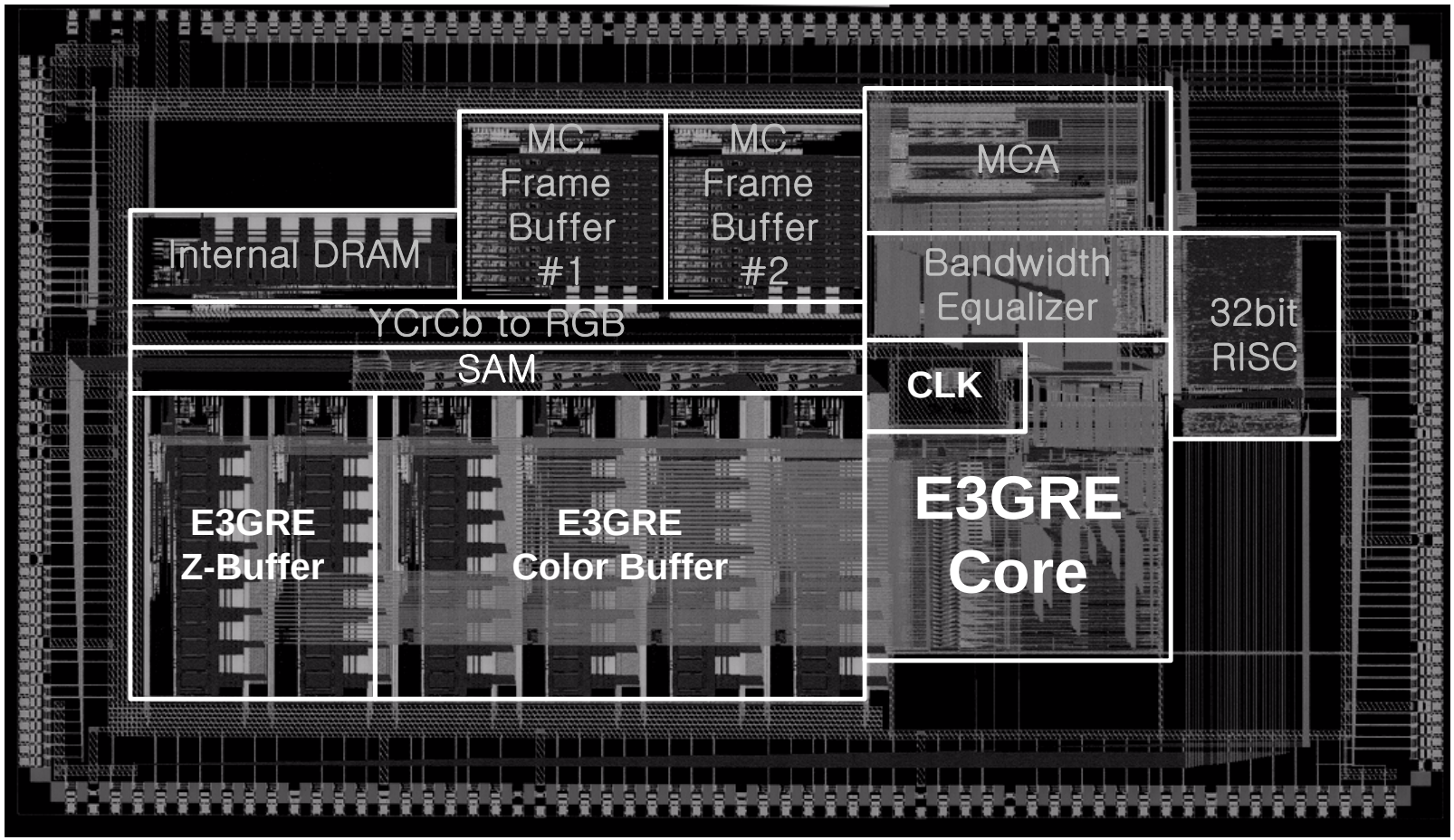
CLK Simulation Result



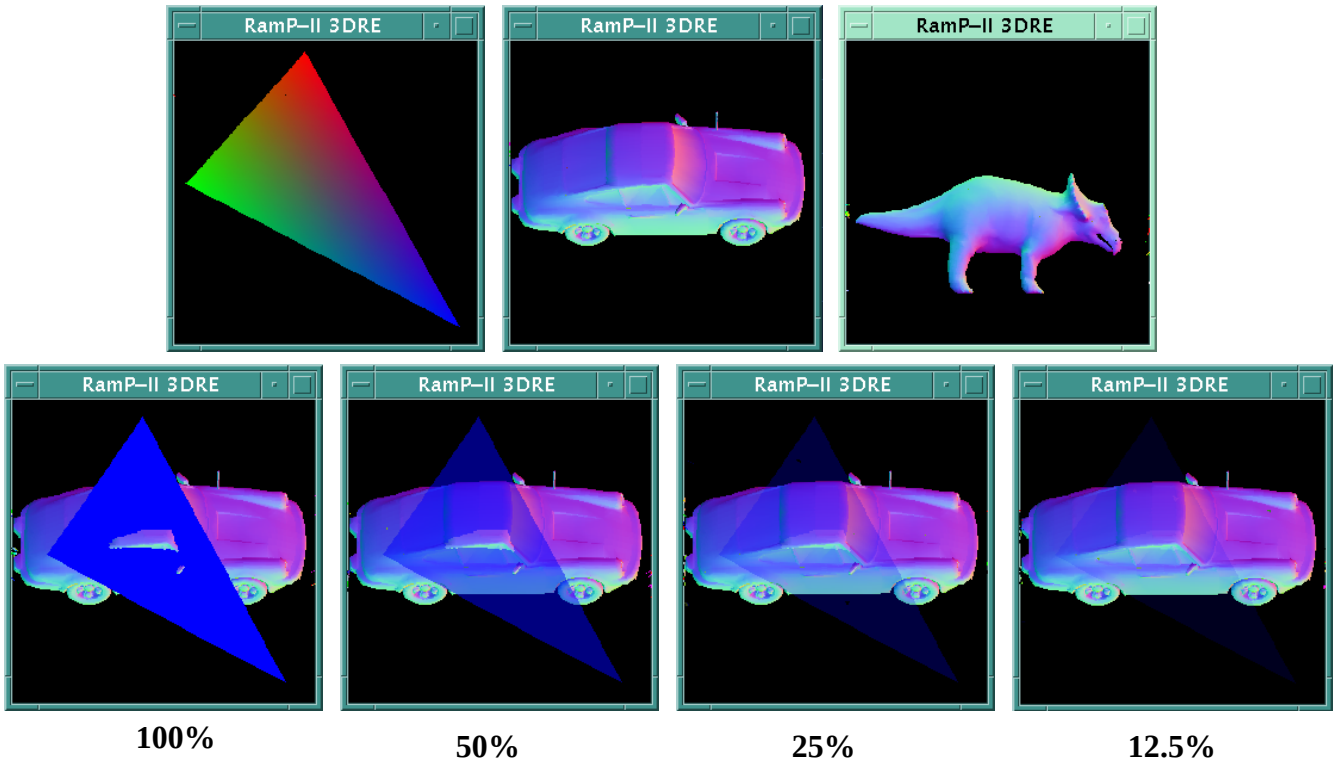




Chip Micrograph



Rendering Outputs



100% 50% 25% 12.5%

Name	Data Size	Number of Polygons
COLOR	100KByte	1,972
CAR	1.32MByte	26,700
DINO	352KByte	6,941

Circuit Summary

Process	0.18μm Hyundai CMOS EML 3P6M	
Power Consumption (mW)	Total	122
	E3GRE Core	36
	FB	84
	CLK	2
Area (mm^2)	Total	24
	E3GRE Core	5.75
	FB	18.2
	CLK	0.03
Number of Logic Transistors	Total	190,231
	E3GRE Core	188,188
	CLK	1,343
Embedded Memories	Color-Buffer	4Mb DRAM
	Z-Buffer	2Mb DRAM
	SAM	1.5Kb SRAM
Operating Frequency	E3GRE Core	20MHz
	Frame-Buffer	100MHz
Power Supply	E3GRE Core	1.5V
	Frame-Buffer	2.5V

Performance Comparison

	3D-RAM	Media-Chip	RamP-1	**E3GRE**
Year	JSSC1995	JSSC 1997	ISSCC 2000	**ISSCC 2001**
Number of Chips	12	1	8	**1**
Process	0.5μm	0.4μm	0.35μm	**0.18μm**
Main Clock	100MHz	100MHz	100MHz	**20MHz**
Maximum Pixel Fill Rate	400Mpixels (33Mpixels/chip)	400Mpixels for clearing	704Mpixels (88Mpixels/chip)	**71Mpixels (320Mpixels for clearing)**
Sustained Triangle Fill Rate	Less	Much Less	352Mpixels (44Mpixels/chip)	**71Mpixels**
Die Area	1680mm^2 (140mm^2/chip)	122mm^2	416mm^2 (52mm^2/chip)	**24mm^2**
Power Consumption	9600mW (for ALU and SAM)	640mW (for 8Mb DRAM)	4700mW	**122mW (36mW Logic 84mW 6Mb DRAM 2mW CLK)**
Functional Units	eDRAM, eSRAM SIMD-Shade SIMD-Blend Z-compare	eDRAM 4 ALUs	eDRAM, eSRAM SIMD-Shade, SIMD-Blend, Z-compare Horizontal Setup, SIMD-Divide	

Conclusion

- **E3GRE Design/Implementation for PDA-Chip**
  - ViSTA Architecture / SALBA Memory Mapping @ Low CLK
  - Low-Power Memory Access (SMA, PWA, PIA)
  - Low-Power Circuits
  - 100% Full-Custom Design (190,231 logic TRs with 6Mb DRAM Macro)
- **Features**
  - 2.22Mtris/sec (20MHz/9cycle)
  - 71Mpixel/sec Pixel Fill Rate (32pixel/triangle)
  - 122mW with Embedded Frame-Buffer and Clock Generation
  - 3.2GByte/sec Embedded-DRAM Bandwidth
  - 24bit True-Color on 256x256 screen
  - Hidden Surface Removal with 16bit Z-buffer
  - Double-Buffering for Flicker-Free Animation
  - Gouraud Shading, Depth Comparison, Alpha-Blending
  - Direct Video Transfer through SAM

Further Works

- **Efficient Texture Fetching using Unified DRAM Texture Cache**
- **Complete PDA-3D Pipeline with MobileGL**