FPGA Based Petascale Computing

Dan Poznanovic
SRC Computers, Inc.

The Salishan Conference on High-Speed Computing
April 20, 2005

poz@srccomp.com

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Dan Poznanovic
SRC Computers, Inc.

The Salishan Conference on High-Speed Computing
April 20, 2005

poz@srccomp.com

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

The Inevitable Petascale Computer

Dan Poznanovic
SRC Computers, Inc.

The Salishan Conference on High-Speed Computing
April 20, 2005

poz@srccomp.com

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Agenda

- Background on SRC's use of FPGAs
- What are FPGAs anyway?
- Explicit Computing Model
- Implicit+Explicit System Architecture
- The future for FPGA & FPGA-like chips
- Let's build up a Petascale System
- Comments & Conclusions

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Some Background on SRC

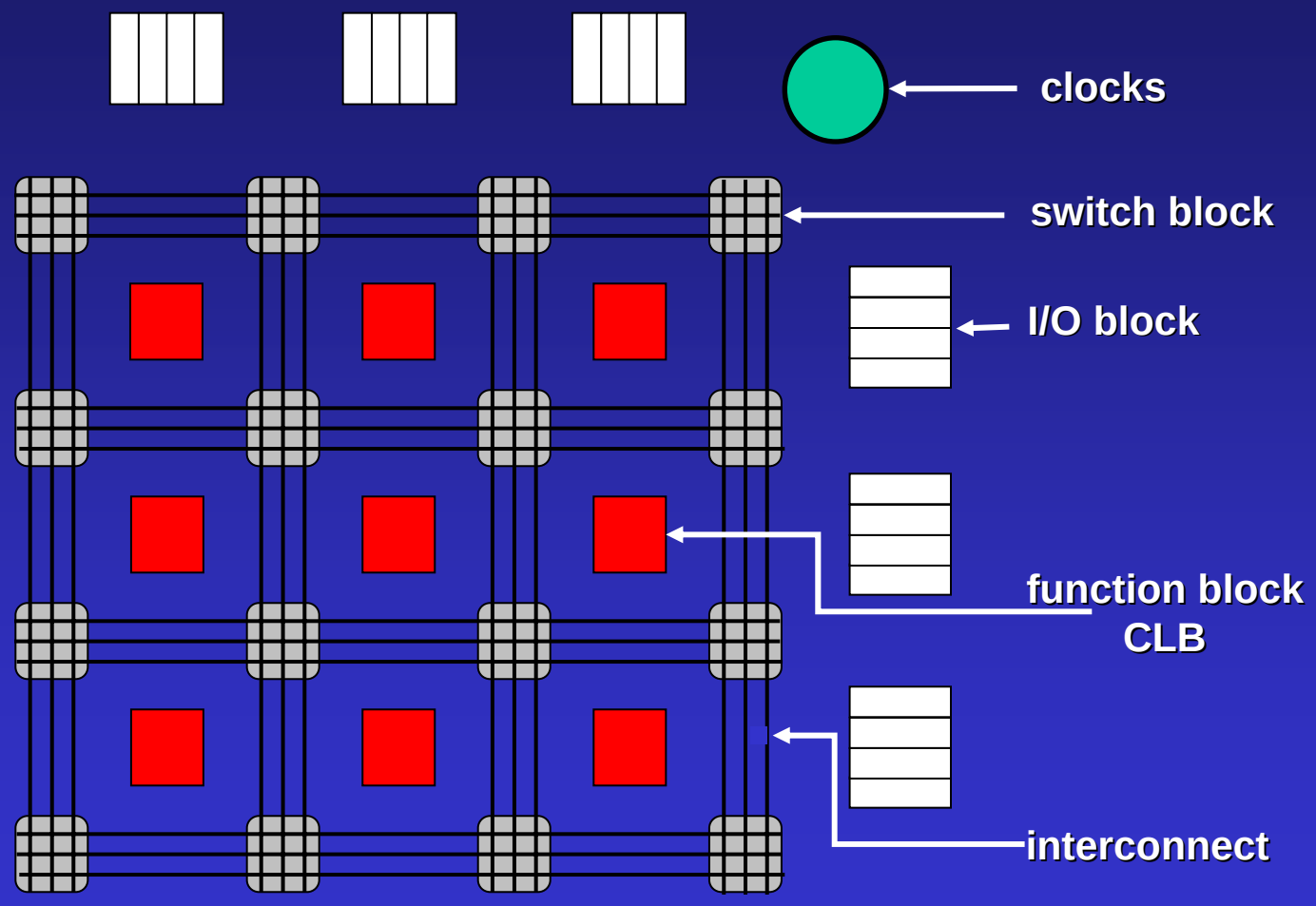
- Founded in 1996 by Seymour R. Cray (SRC)
- Goal:
  Cray class systems & Commodity parts
- Performance, Balance & No custom ASICs
- Yield: Systems of Microprocessors + FPGAs
- Cray-1,Cray-2, Cray-3, Cray-4, Cray-5, SRC-6

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

What is an FPGA anyway?

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

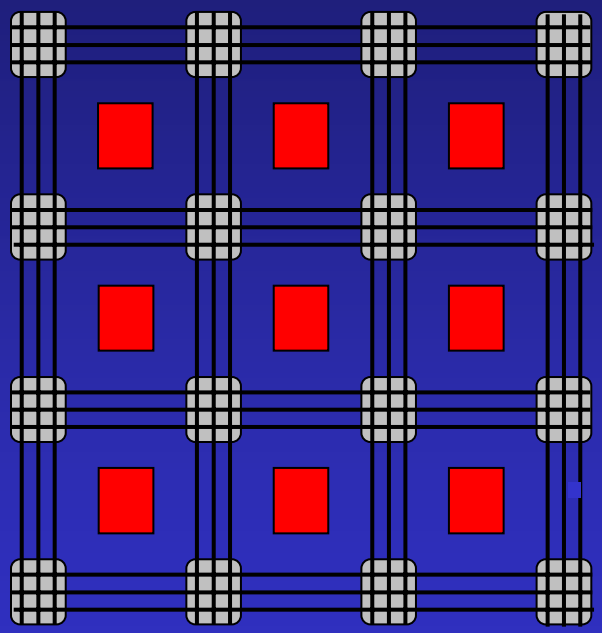
Generic FPGA Structure



Field Programmable Gate Array

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

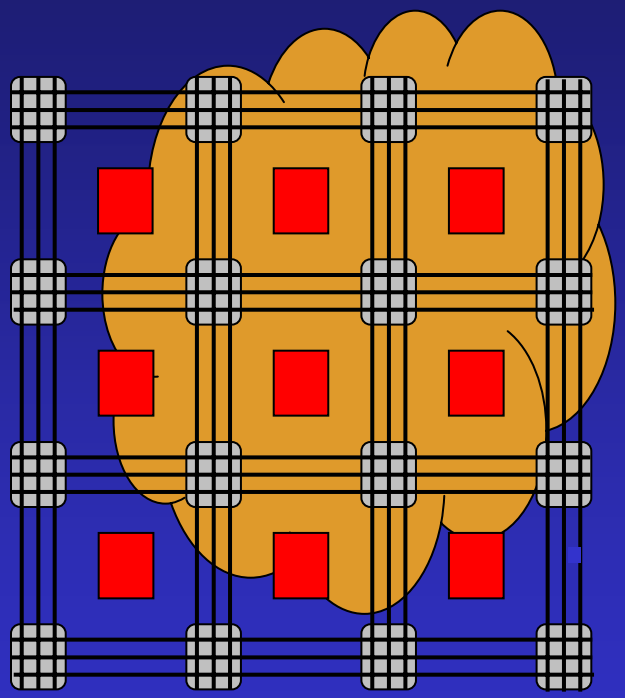
Generic FPGA Structure



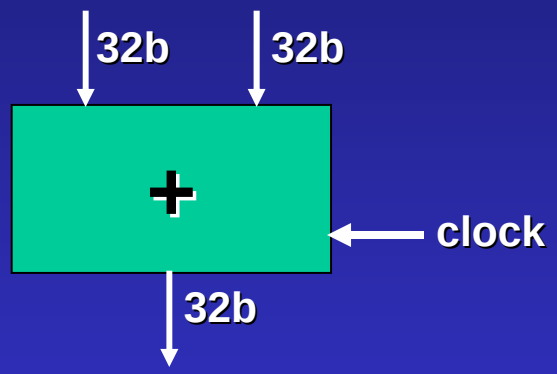
- Simple view
  - Little “islands” of logic
  - Grid of interconnecting wires
- Logic functions and connections can be reconfigured
- “Bitstream” is a complete configuration for the chip
- Configuration takes ~50 msec (Xilinx v6000)

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Logic in FPGA



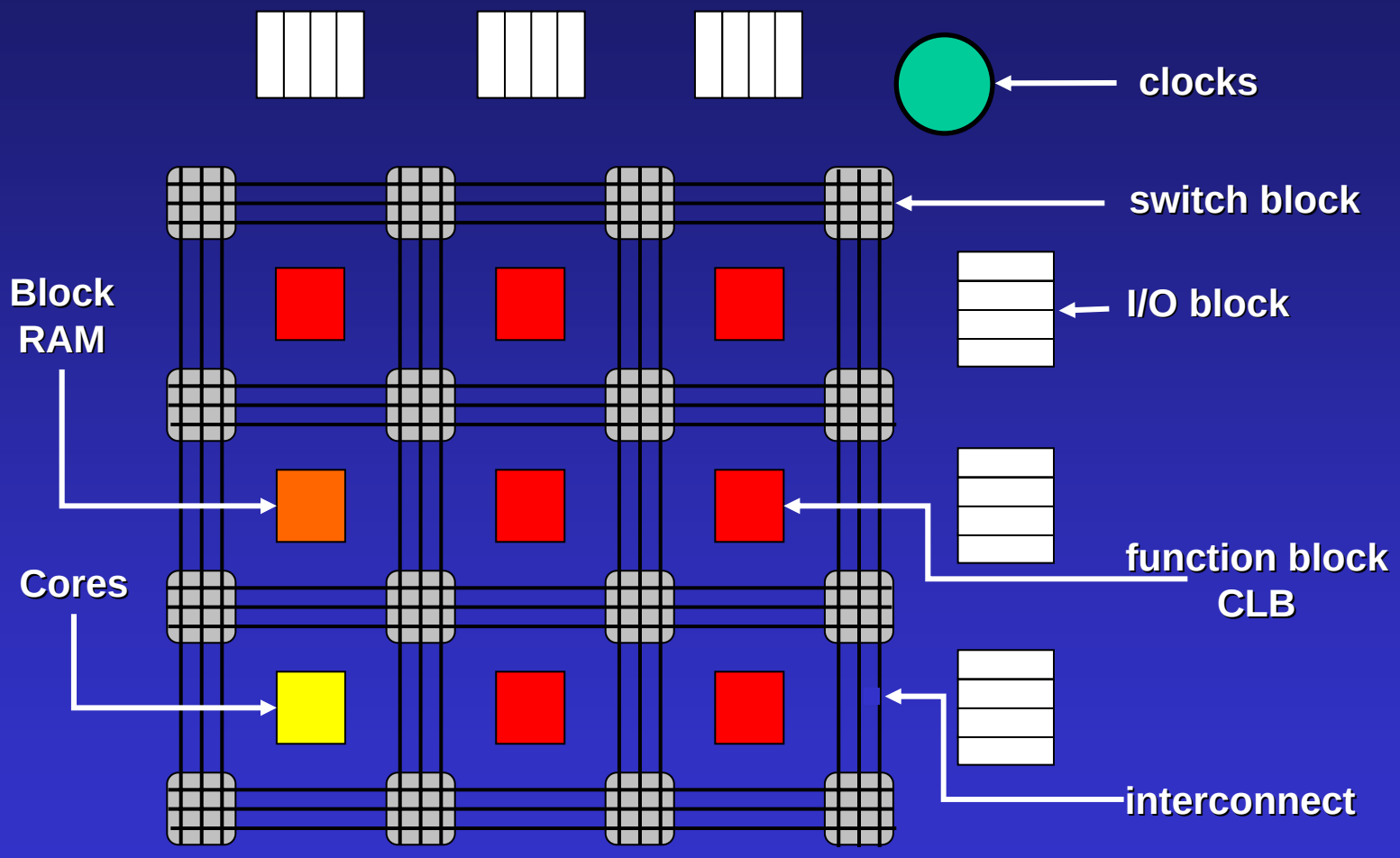
- Example – 32b Integer Add



- Described by VHDL, Verilog, Schematic Capture
- "Synthesized" to the chip's resources

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Generic FPGA Structure



Field Programmable Gate Array

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

What Is the Compute Model For FPGAs?

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

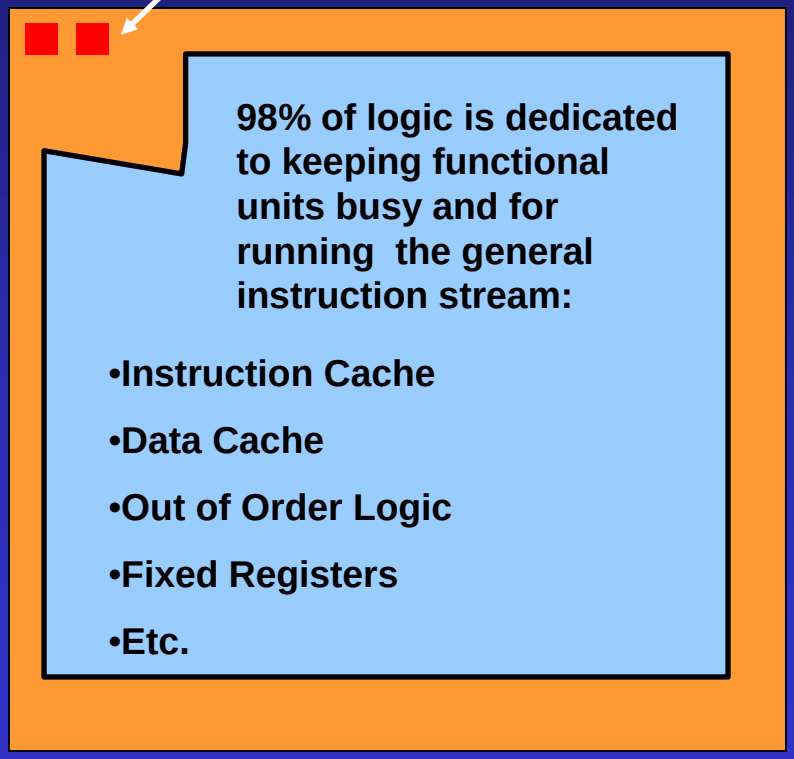
Explicit Computing

- FPGA Chips means to define logic & connections
- Functional units can be defined
- Expressions created from functional units
- Programs created from expressions
- Direct Execution Logic
- No instructions

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

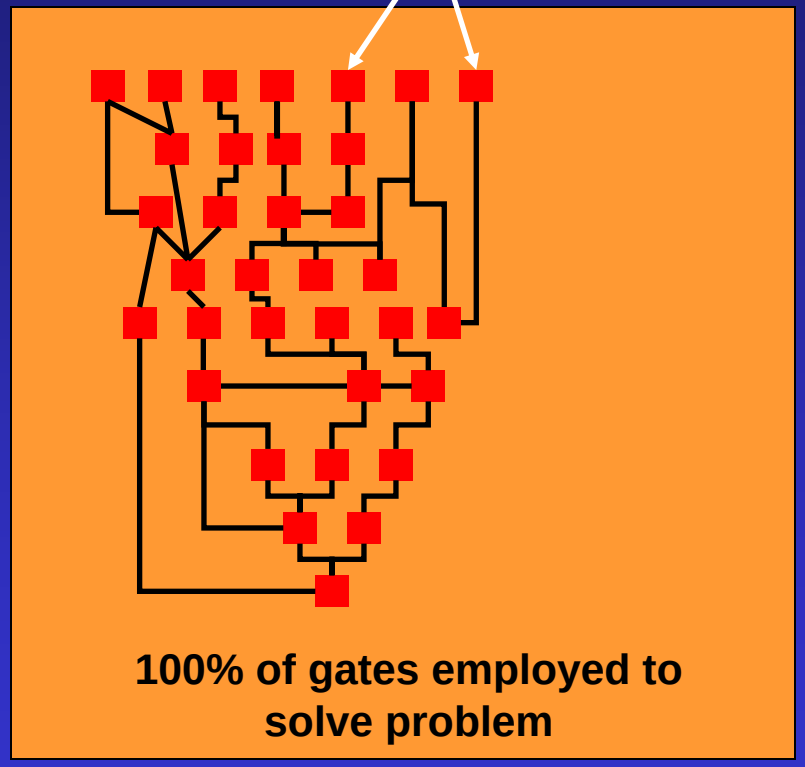
von Neumann vs. Direct Execution Logic

Functional Units



Intel Pentium 4 Processor

100s of Functional Units

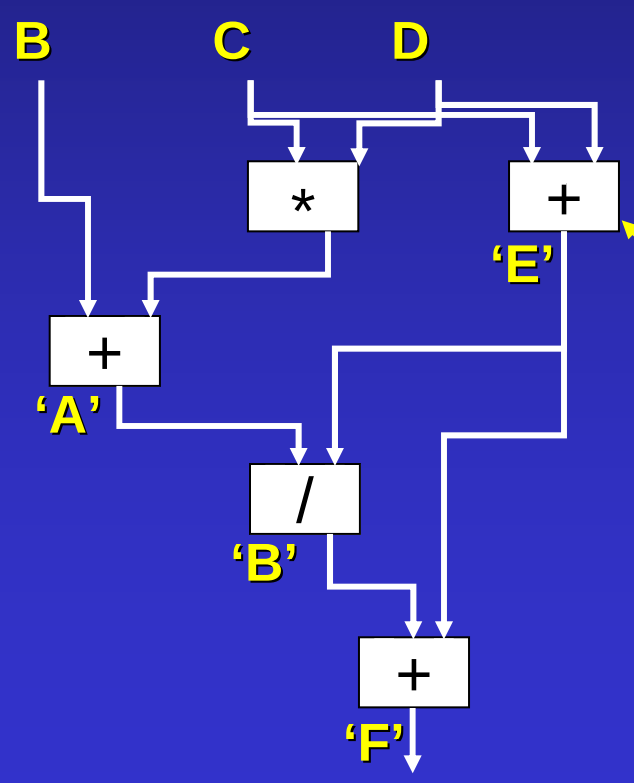


Direct Execution Logic

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Dataflow

A = B + C * D;
E = C + D;
B = A / E;
F = B + E;



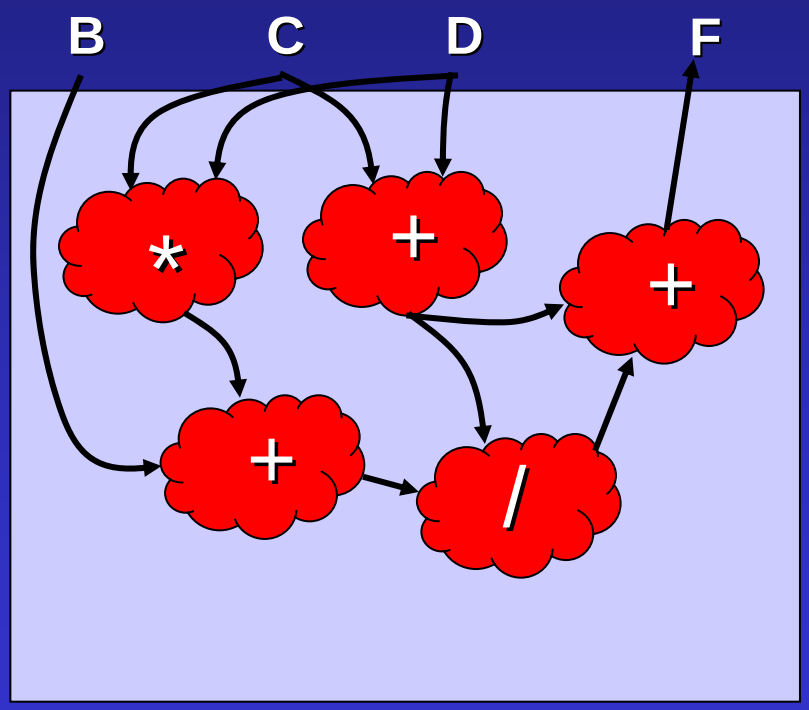
- Dataflow ideas have been around for decades
  - Ex: Manchester Dataflow
  - Ex: Monsoon Dataflow
- Basic idea: express computation with interconnected function units

Note concurrent operations

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Dataflow

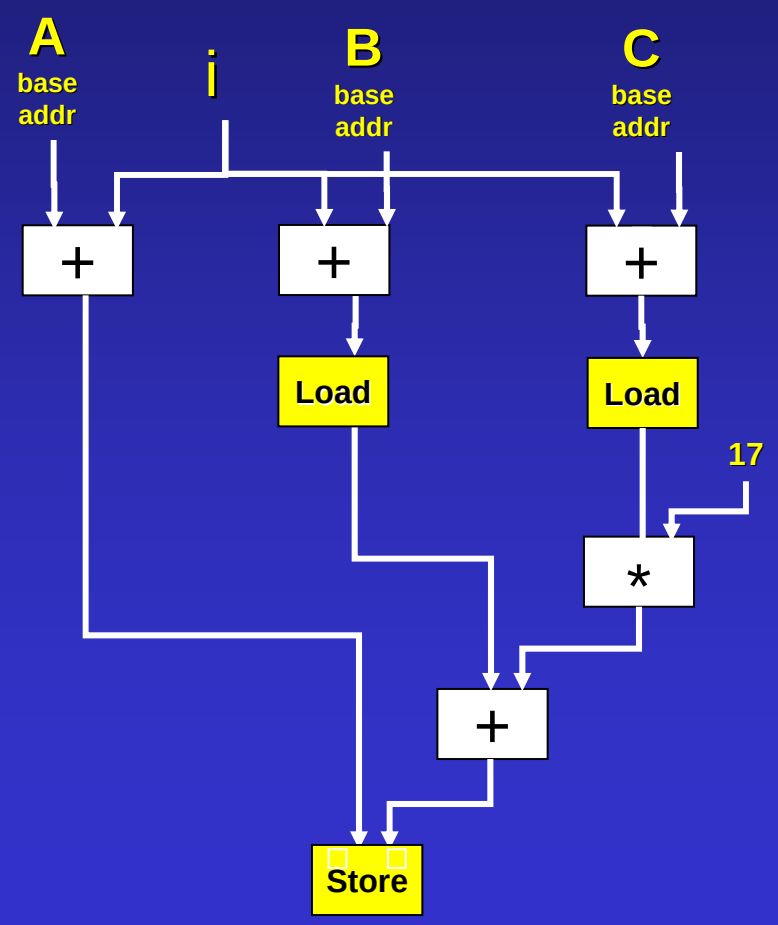
Dataflow can be implemented in FPGA Logic



Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Pipelined Loop

for (i=0; i<n; i++) A[i] = B[i] + C[i] * 17;



- DFG comes naturally from the loop body
- Compiler then inserts delays to balance the paths
- Compiler adds function units to:
  - Generate a stream of indices
  - Detect loop termination
- New result each clock

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

The “Other” Functional units

- Computational Functional Units
- Data Access Functional Units
- Data Pre-Fetch Functional Units

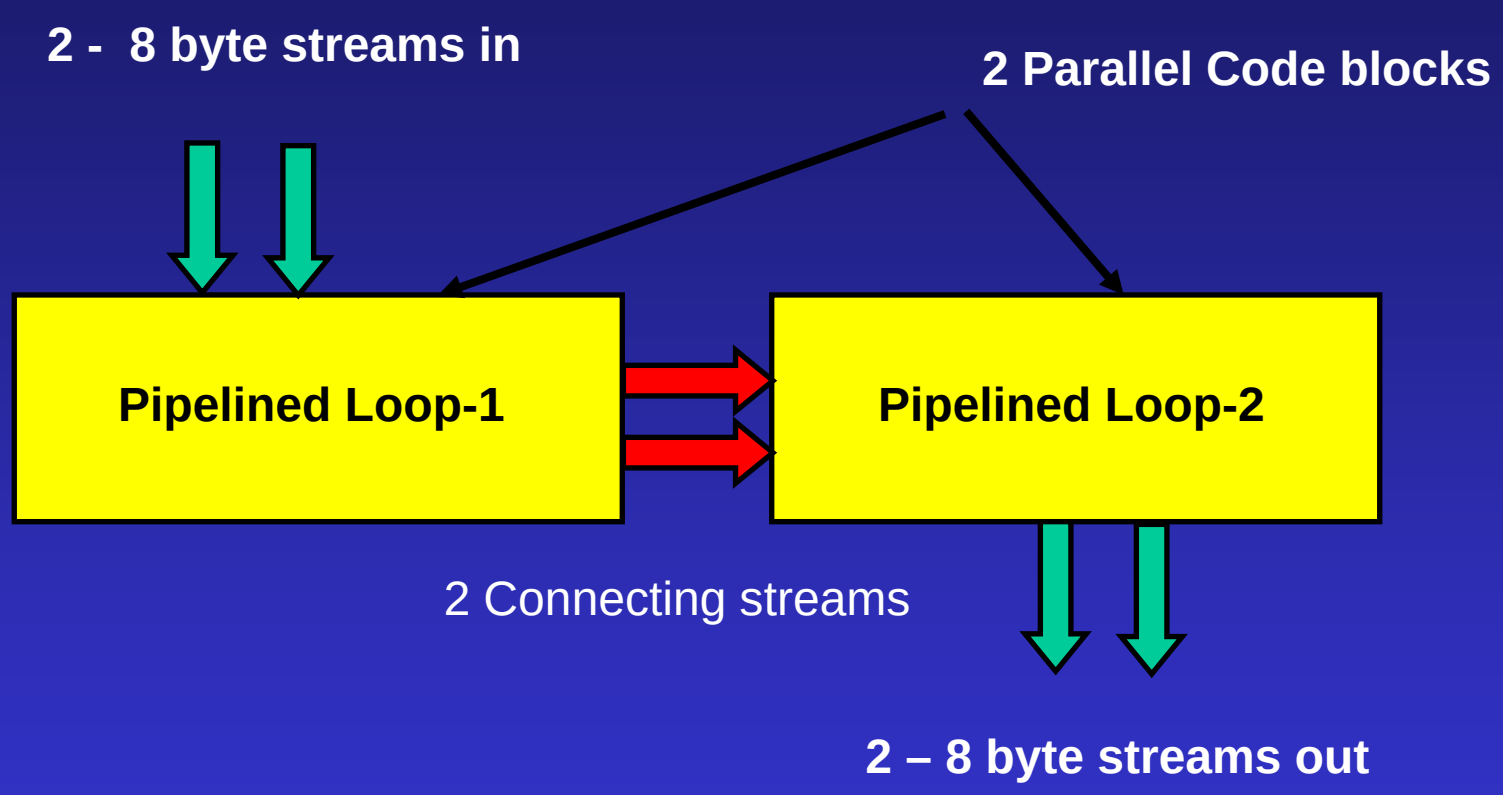
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Data Access

- On-Board Memory Access
- On-Chip Memory Access (BRAM)
- Streams (FIFOs with Flow Control)
- Caching Replacements

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

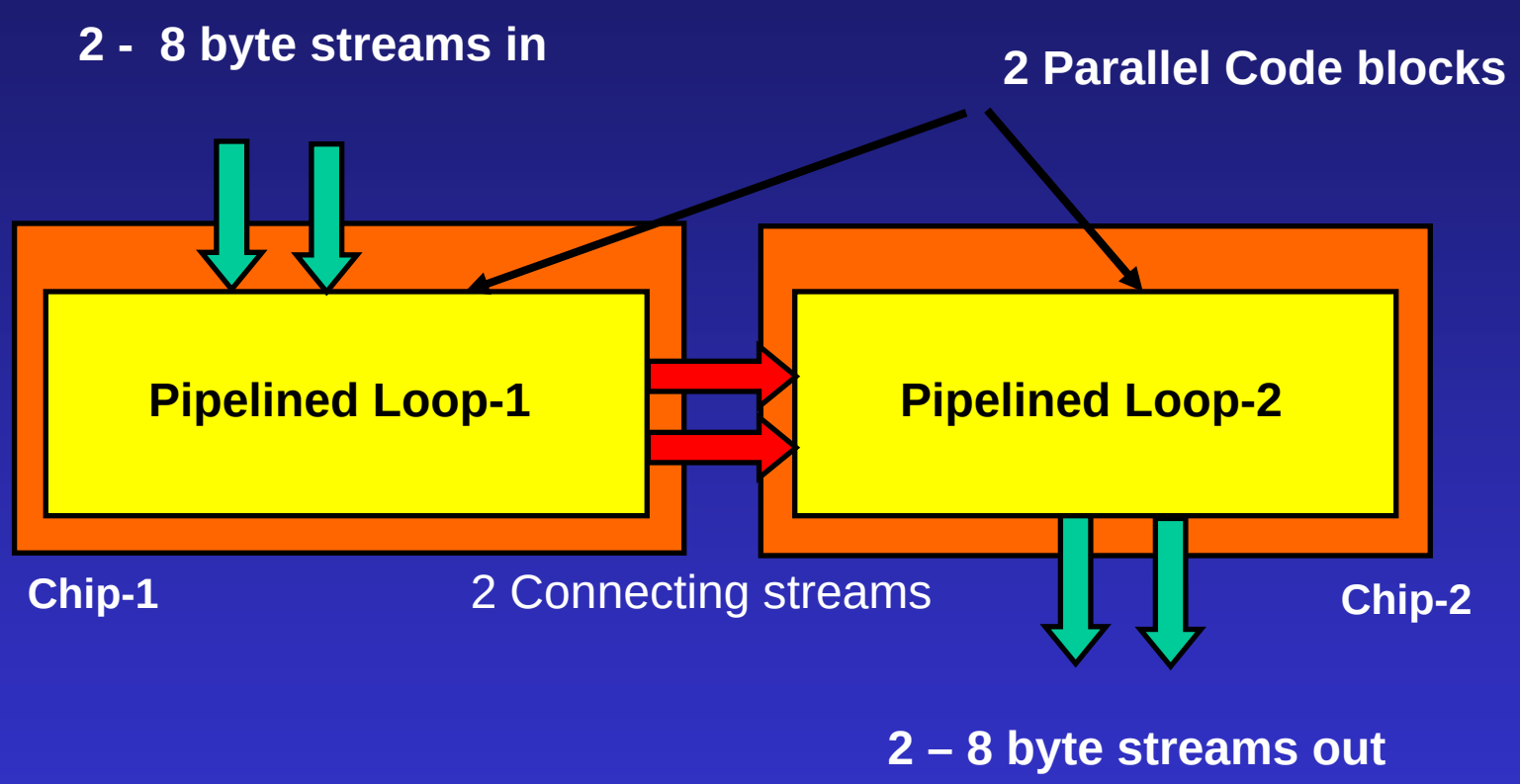
Streams – 2 Code Blocks



Stream is FIFO with flow control

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

Streams – 2 Chips

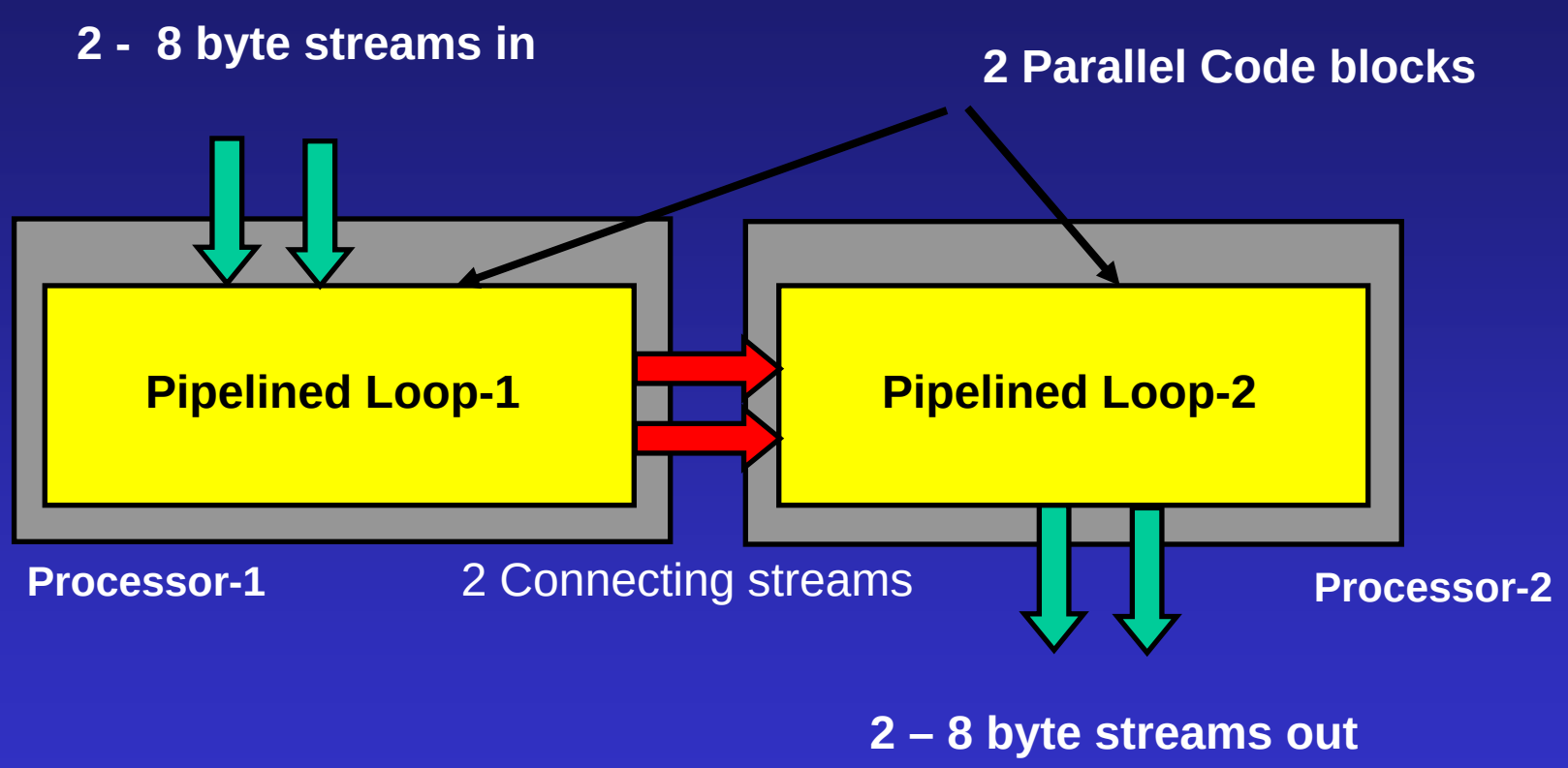


Convert to 2 chips:

1. **Define two bridge streams**
2. **Split parallel code blocks into parallel routines**

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

Streams – 2 Processors



Convert to 2 processor:

1. **Define two chain port streams**
2. **Split parallel code blocks into parallel routines**

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

Data access example

- Stencil Code
- Delay Queue Data Access
- Eliminates need for Cache
- It’s just a function call

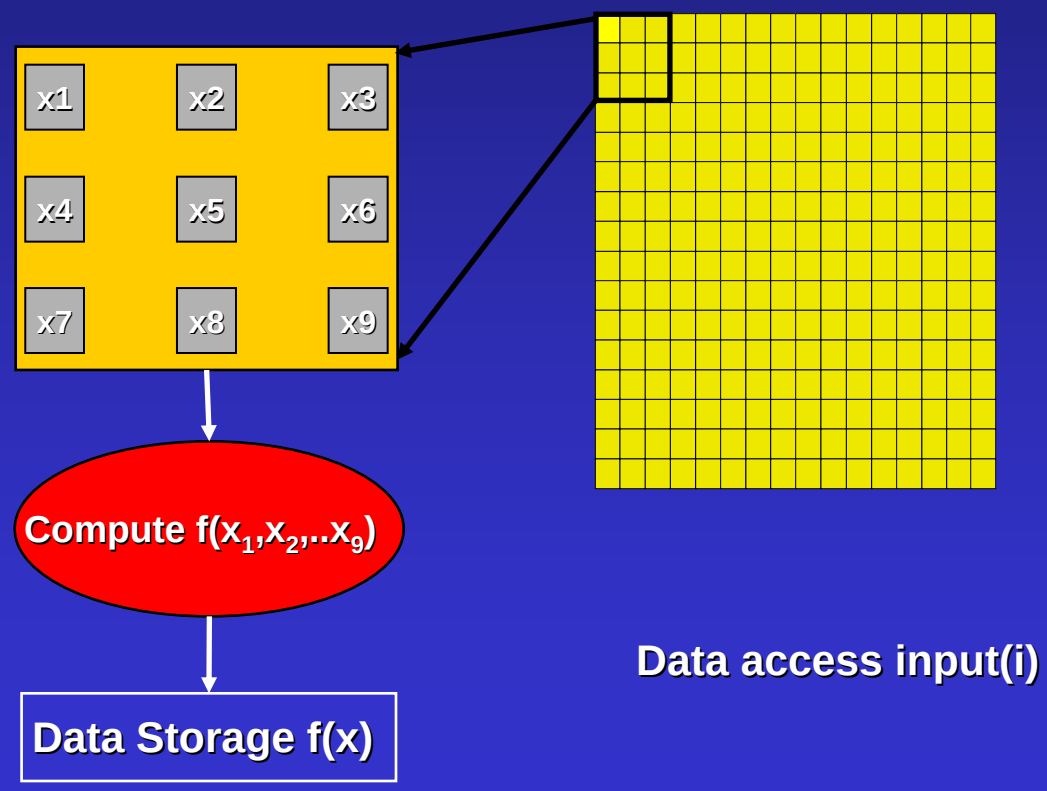
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Pipelining Stencil Code

Move a window through the image

Compute Process

9 points in stencil. 8 have been seen before. The leading point should be the only data access.



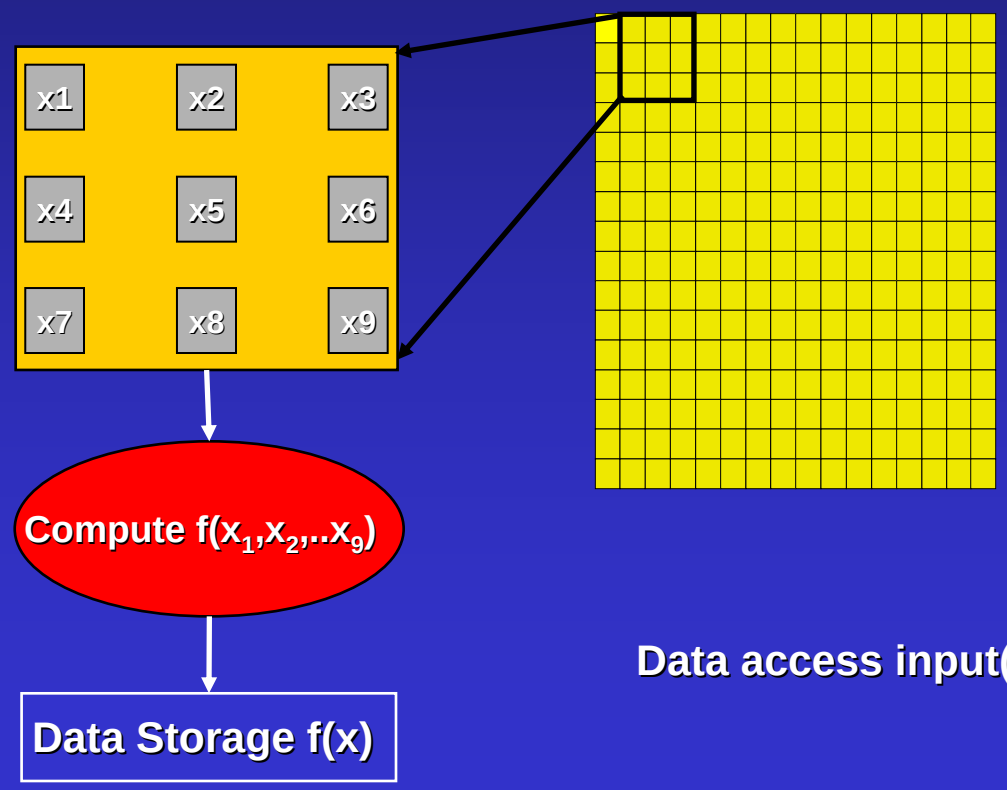
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Pipelining Stencil Code

Move a window through the image

Compute Process

9 points in stencil. 8 have been seen before. The leading point should be the only data access.



Data access input(i)

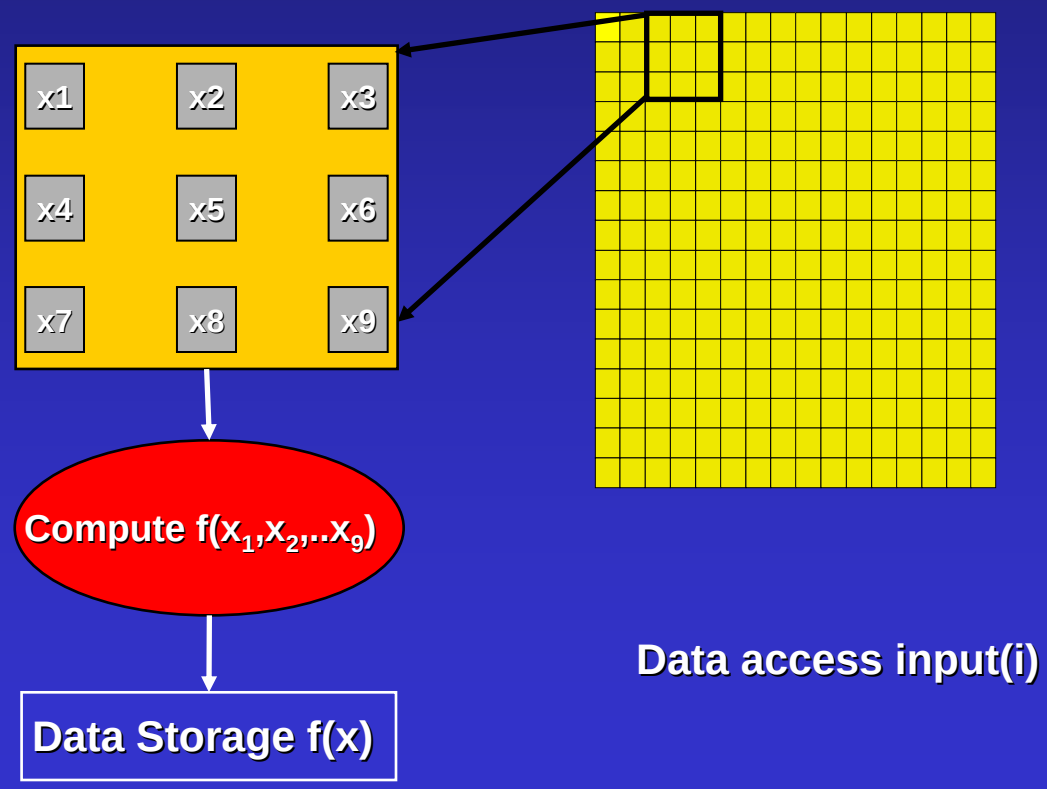
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Pipelining Stencil Code

Move a window through the image

Compute Process

9 points in stencil. 8 have been seen before. The leading point should be the only data access.



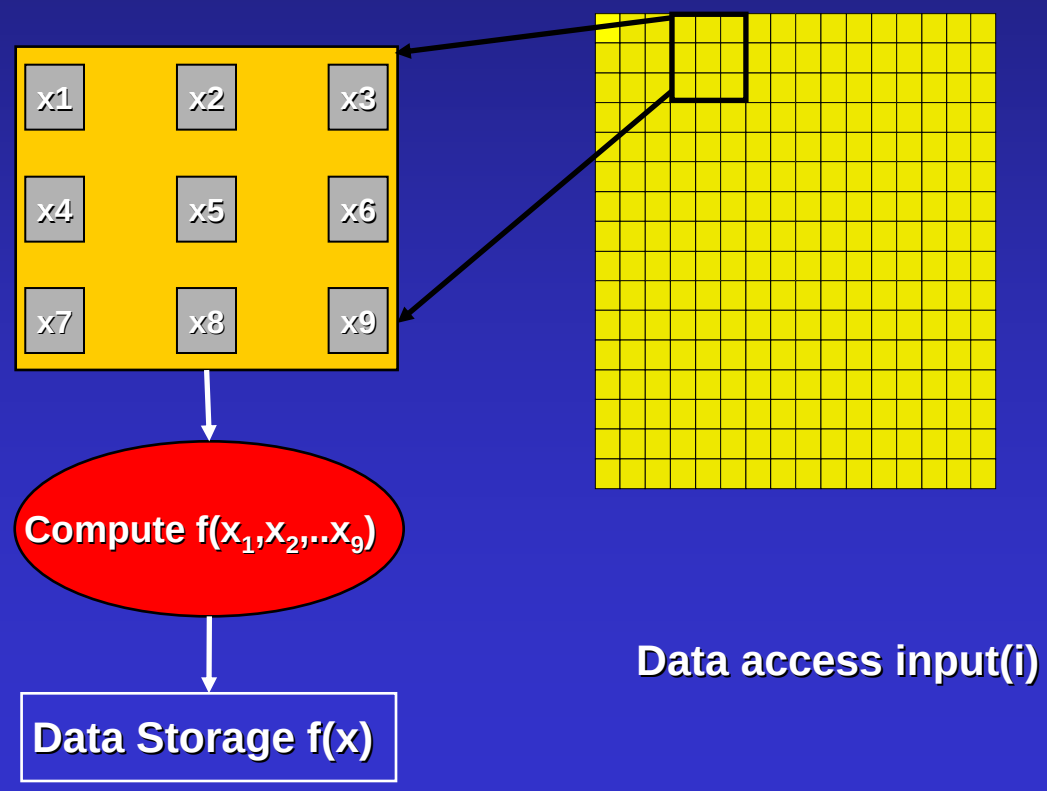
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Pipelining Stencil Code

Move a window through the image

Compute Process

9 points in stencil. 8 have been seen before. The leading point should be the only data access.



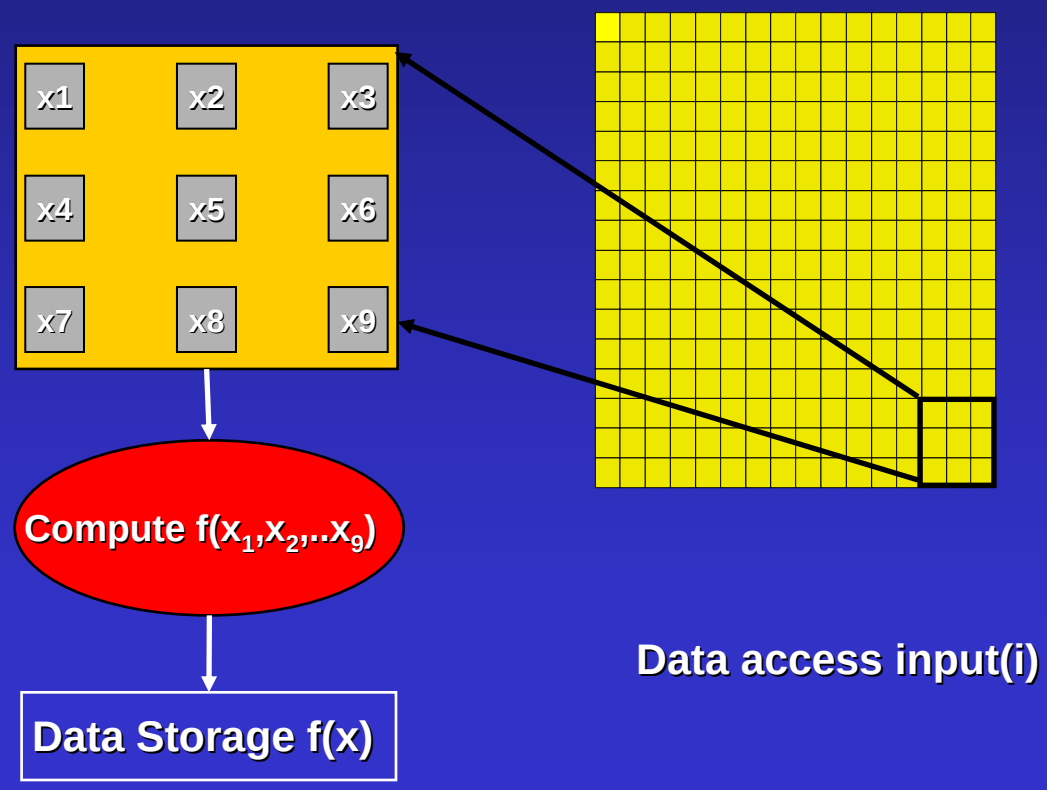
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Pipelining Stencil Code

Move a window through the image

Compute Process

9 points in stencil. 8 have been seen before. The leading point should be the only data access.



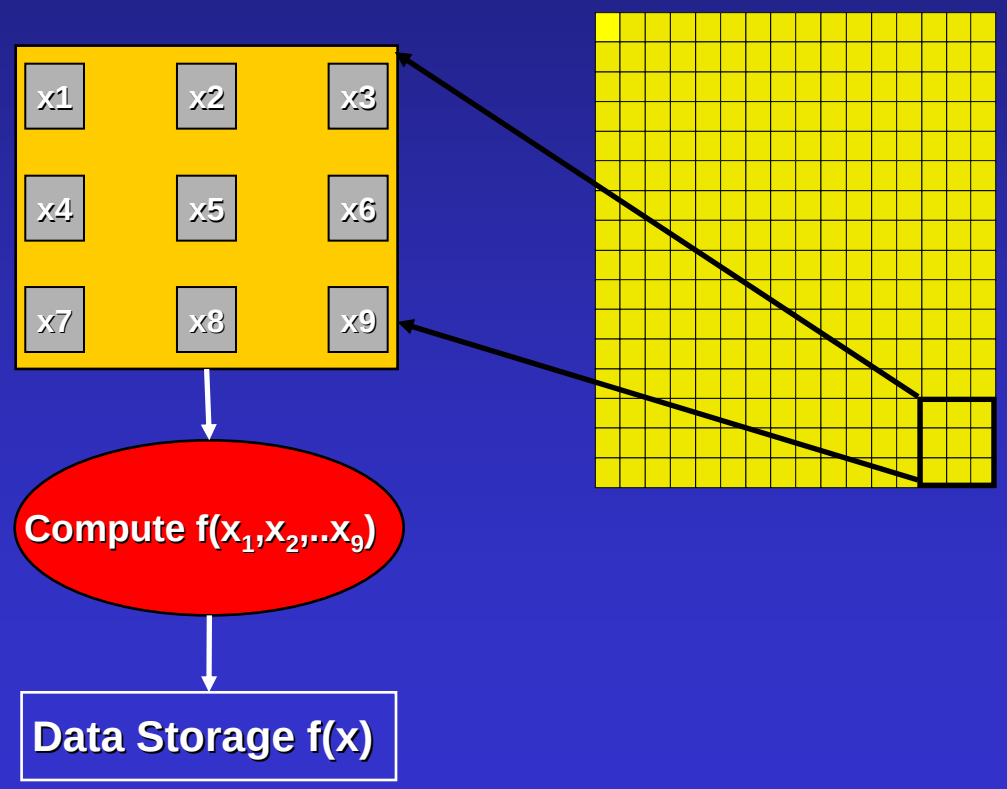
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Pipelining Stencil Code

Move a window through the image

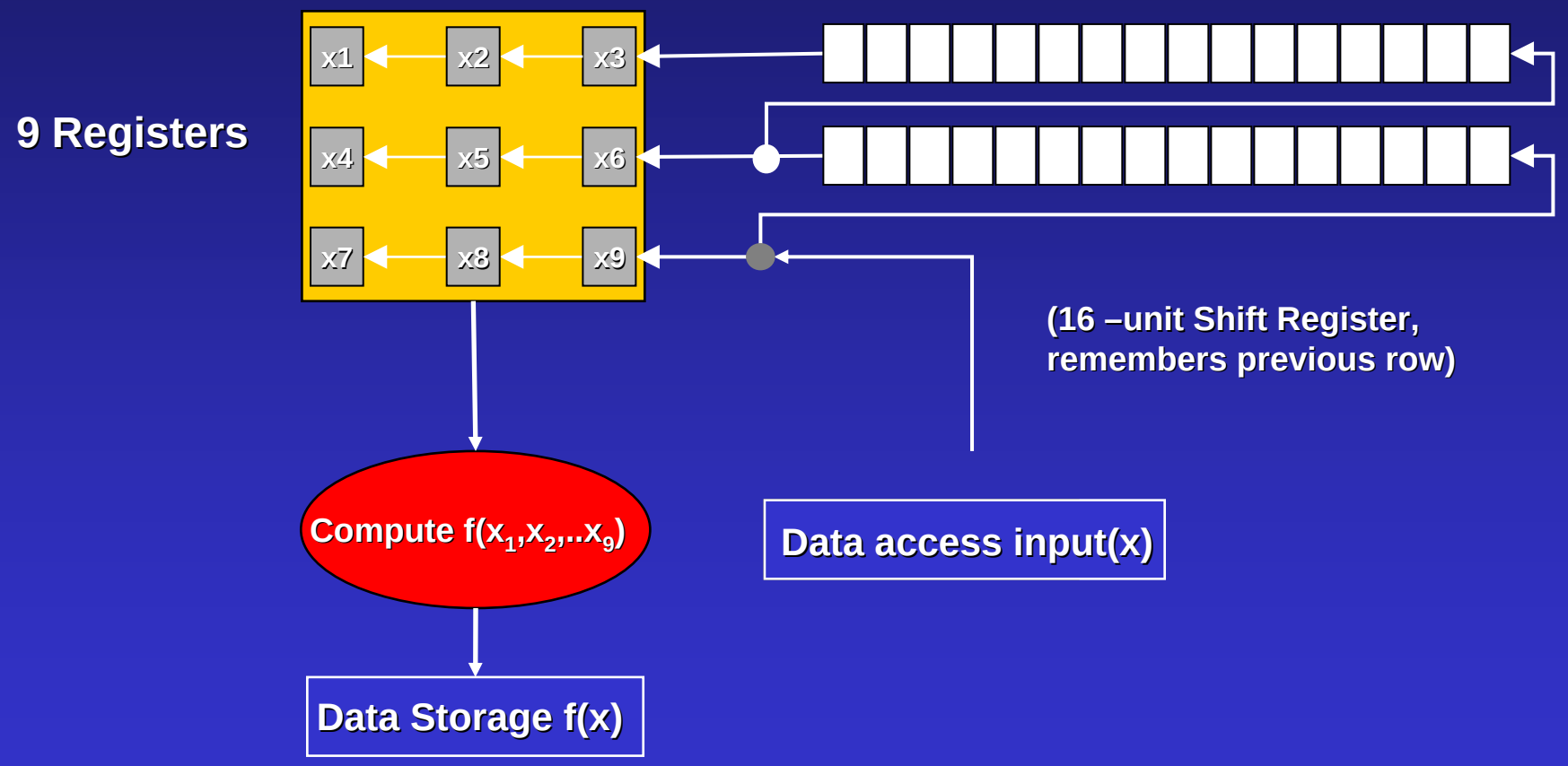
Compute Process

9 points in stencil. 8 have been seen before. The leading point should be the only data access.



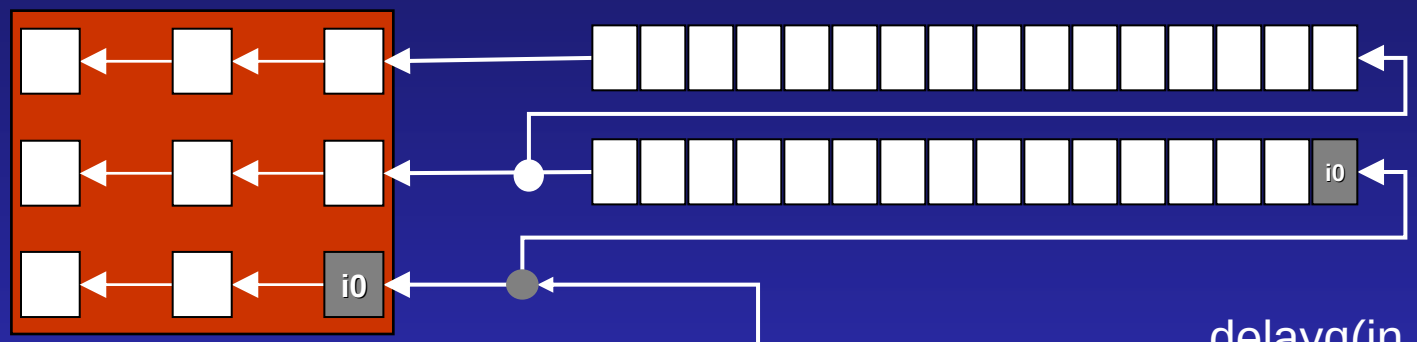
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Stencil Data Flow



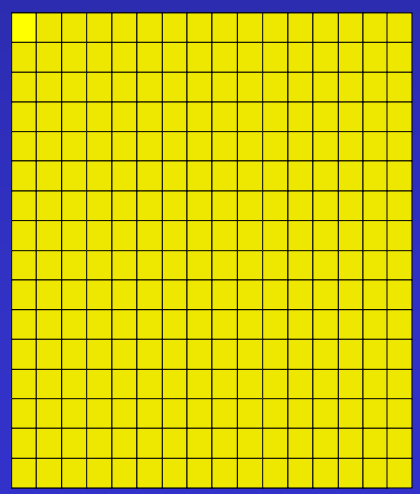
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Stencil Data Flow

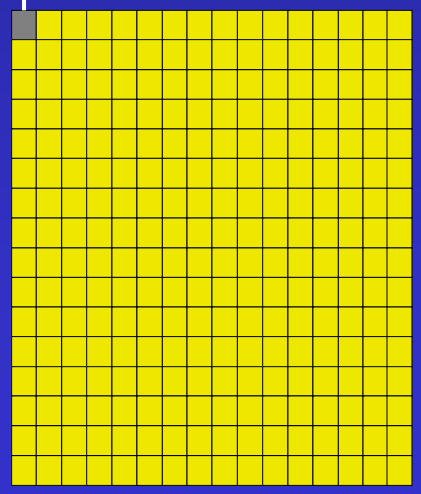


Compute $f(x_1,x_2,..x_9)$

delayq(in,&out);



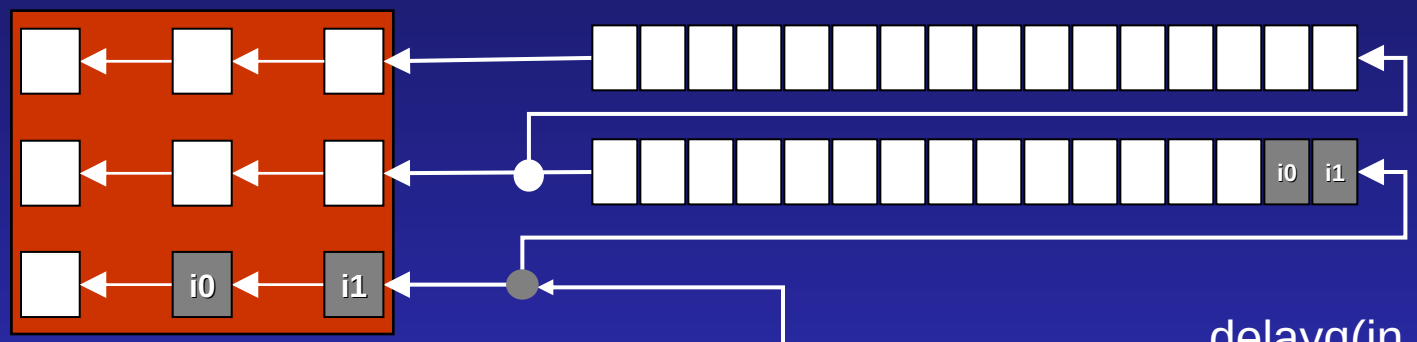
Output(i)



Data access input(i)

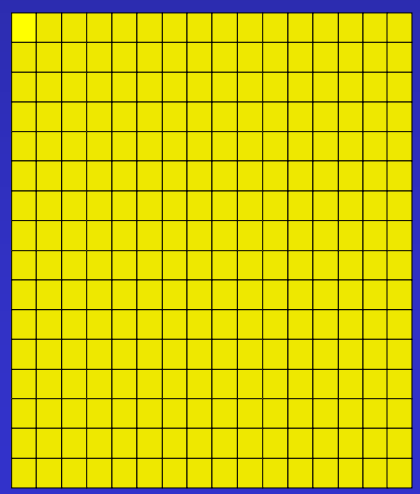
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Stencil Data Flow

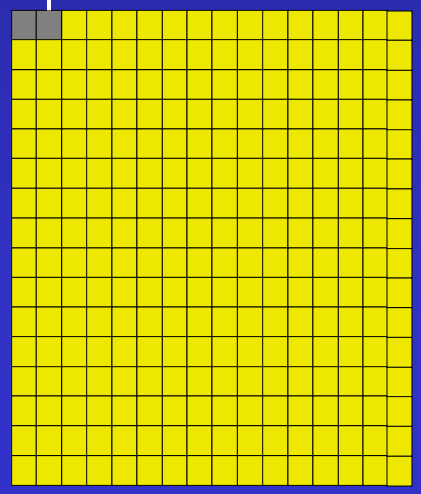


Compute $f(x_1,x_2,..x_9)$

delayq(in,&out);



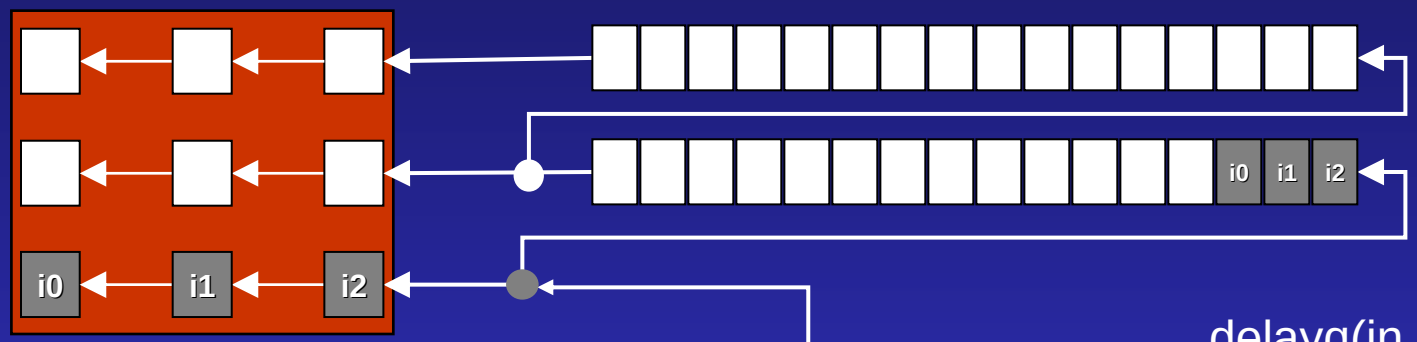
Output(i)



Data access input(i)

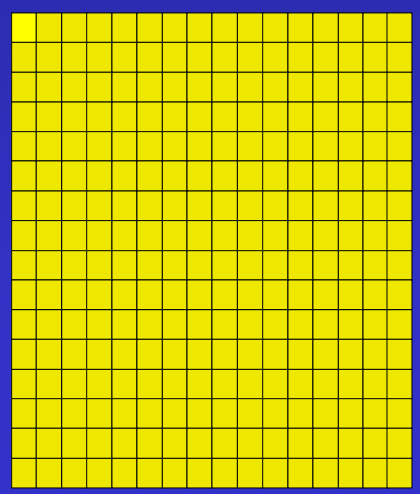
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Stencil Data Flow

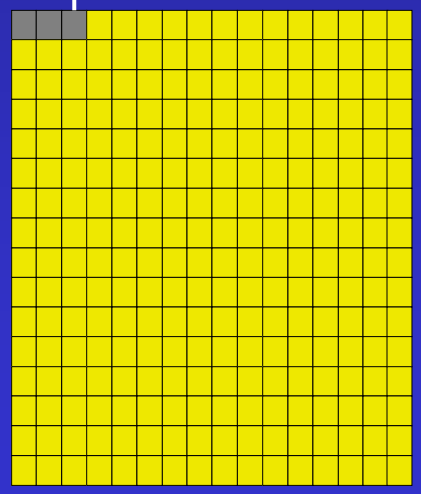


Compute $f(x_1,x_2,..x_9)$

delayq(in,&out);



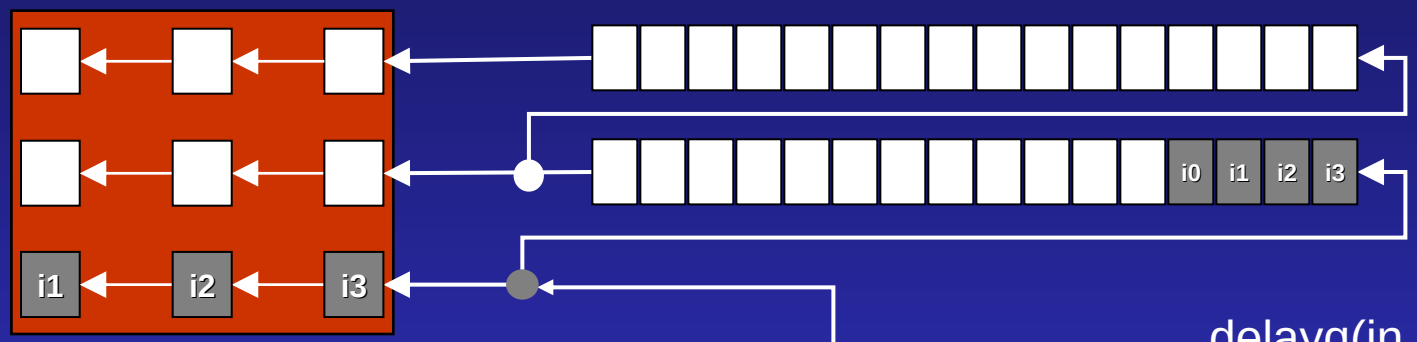
Output(i)



Data access input(i)

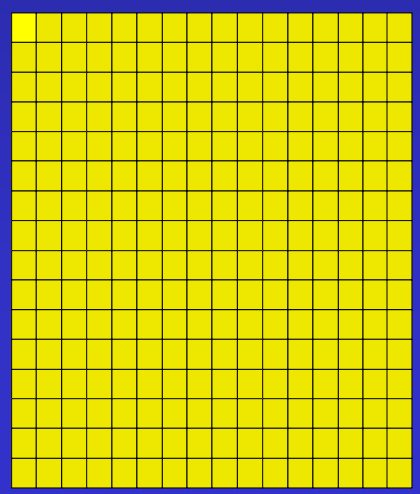
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

Stencil Data Flow

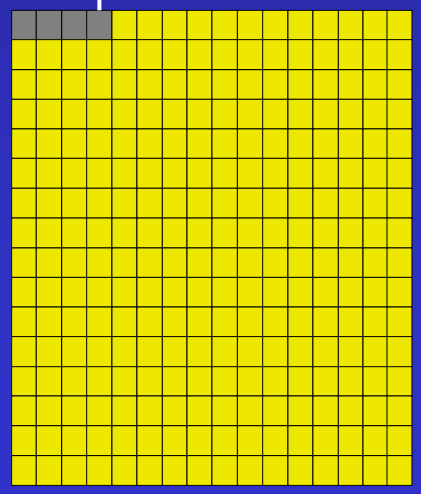


Compute $f(x_1,x_2,..x_9)$

delayq(in,&out);



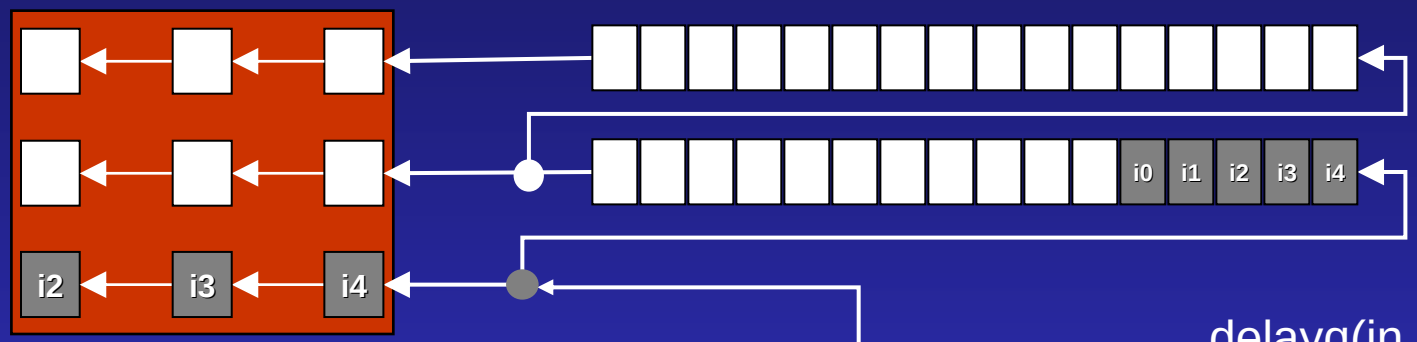
Output(i)



Data access input(i)

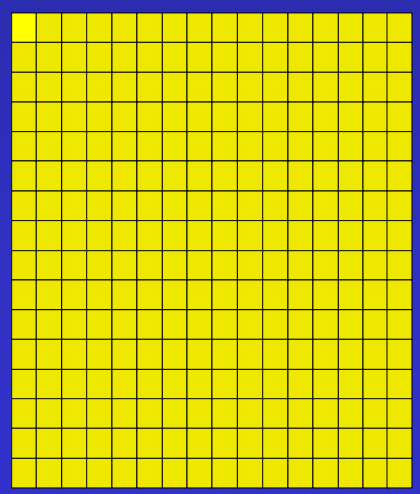
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Stencil Data Flow

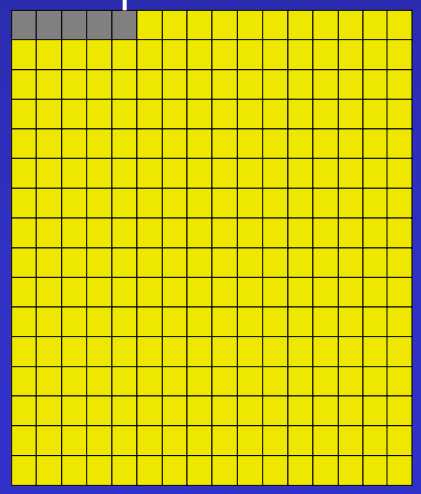


Compute $f(x_1,x_2,..x_9)$

delayq(in,&out);



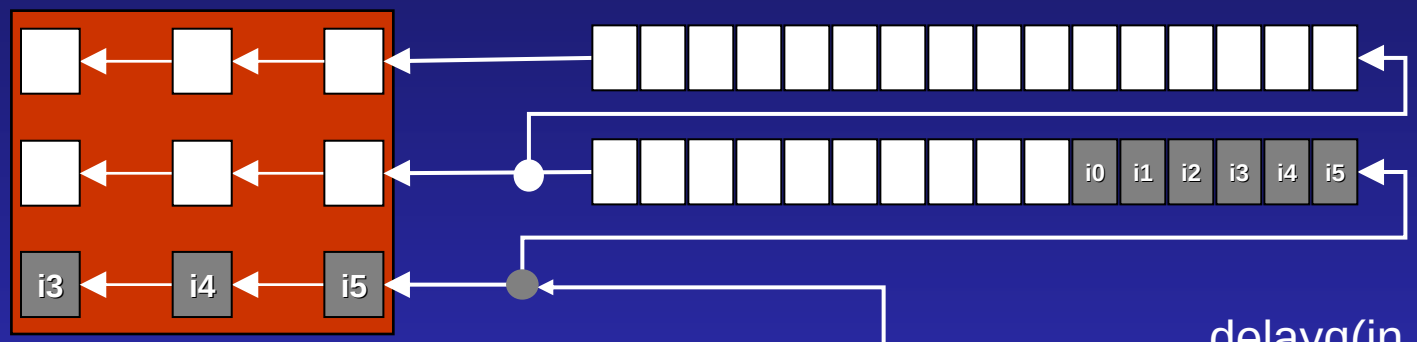
Output(i)



Data access input(i)

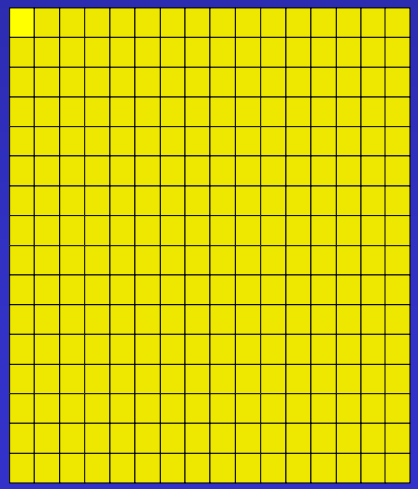
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Stencil Data Flow

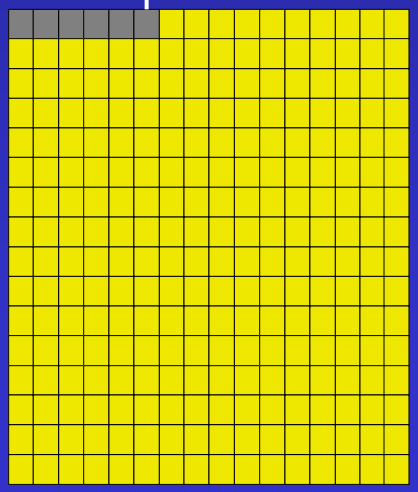


Compute $f(x_1,x_2,..x_9)$

delayq(in,&out);



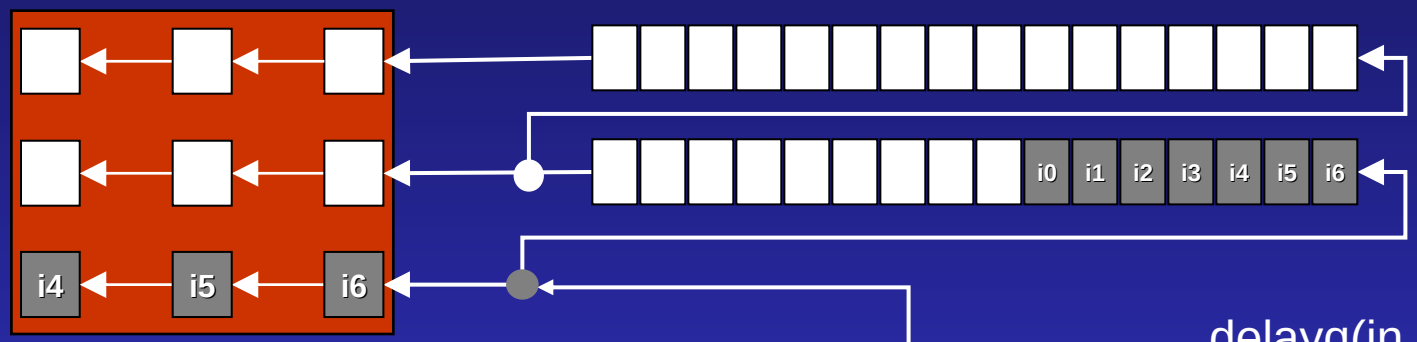
Output(i)



Data access input(i)

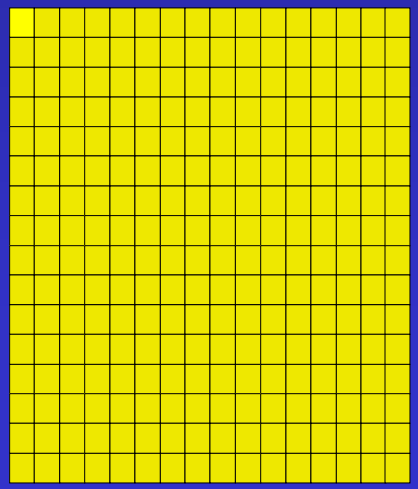
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Stencil Data Flow

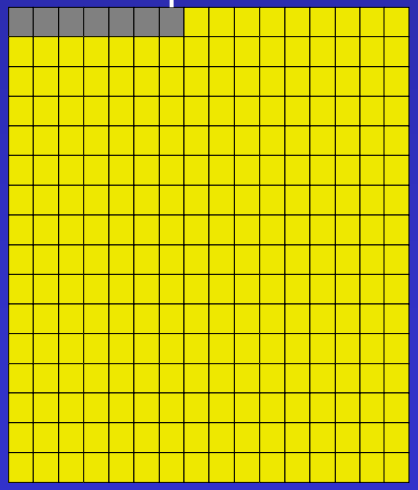


Compute $f(x_1,x_2,..x_9)$

delayq(in,&out);



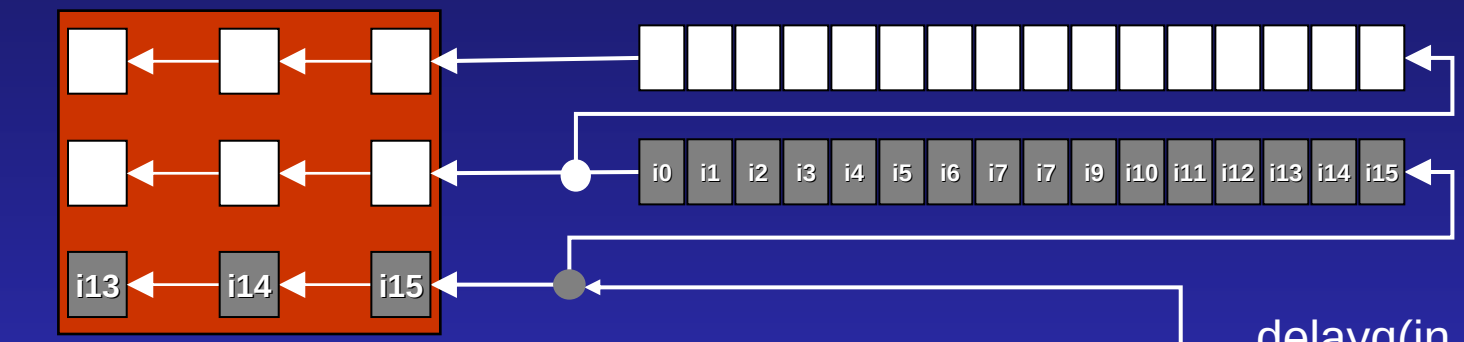
Output(i)



Data access input(i)

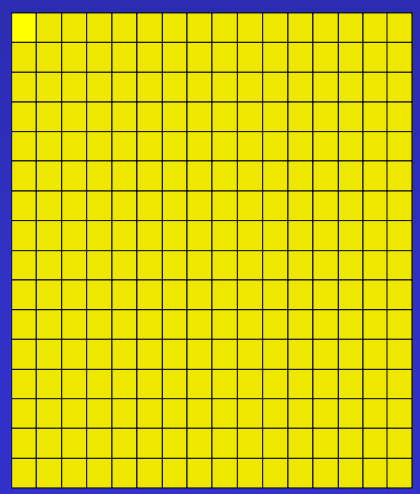
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Stencil Data Flow

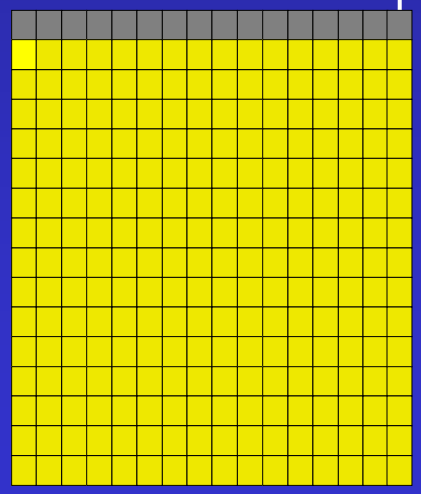


Compute $f(x_1,x_2,..x_9)$

delayq(in,&out);



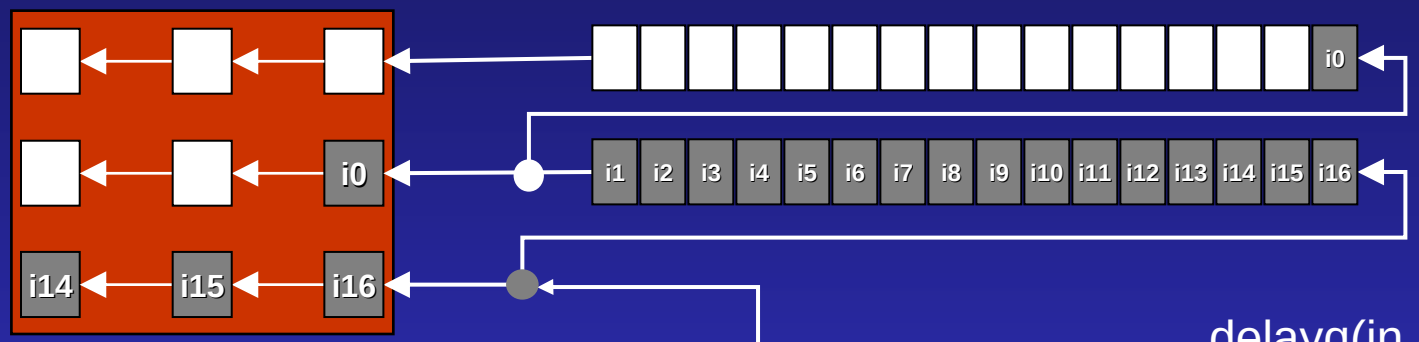
Output(i)



Data access input(i)

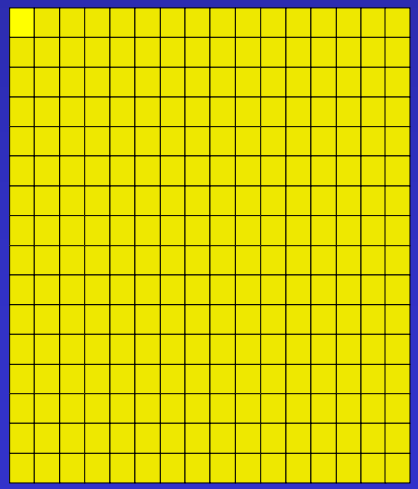
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Stencil Data Flow

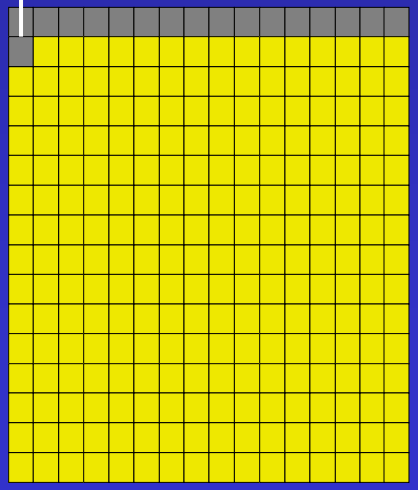


Compute $f(x_1,x_2,..x_9)$

delayq(in,&out);



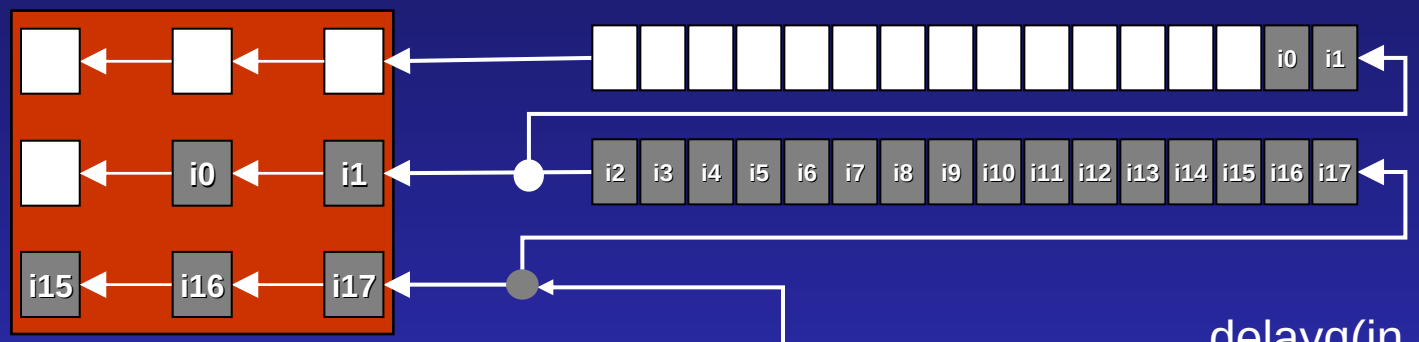
Output(i)



Data access input(i)

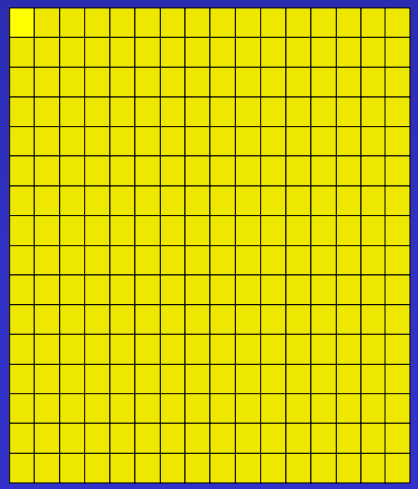
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

Stencil Data Flow

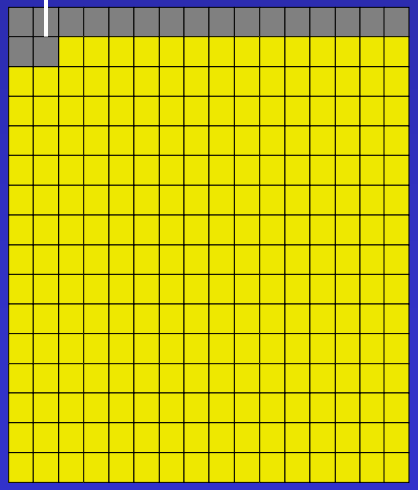


Compute $f(x_1,x_2,..x_9)$

delayq(in,&out);



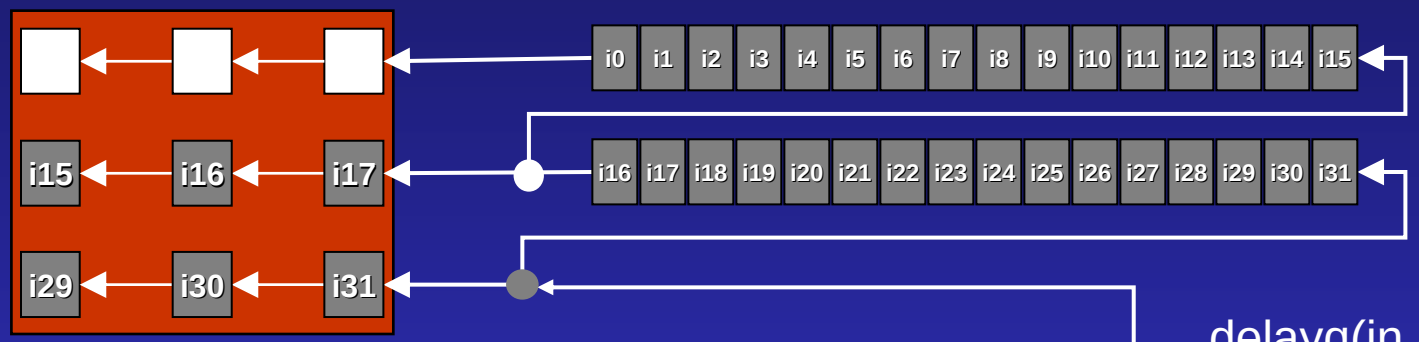
Output(i)



Data access input(i)

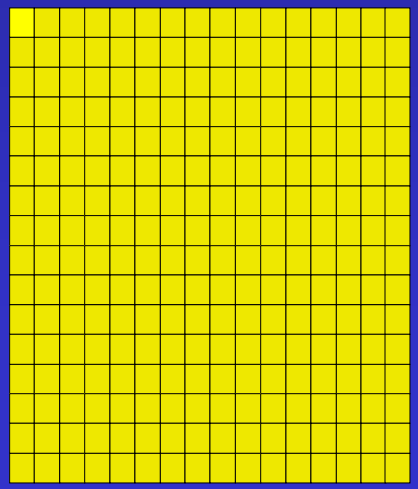
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Stencil Data Flow

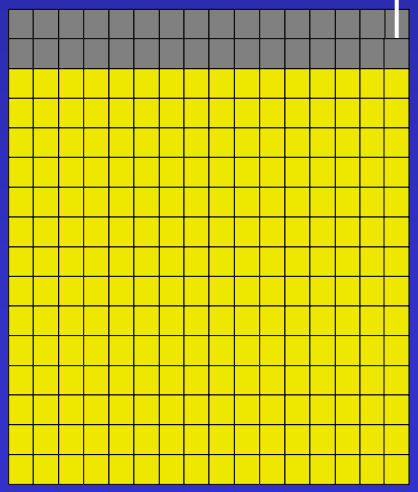


Compute $f(x_1,x_2,..x_9)$

delayq(in,&out);



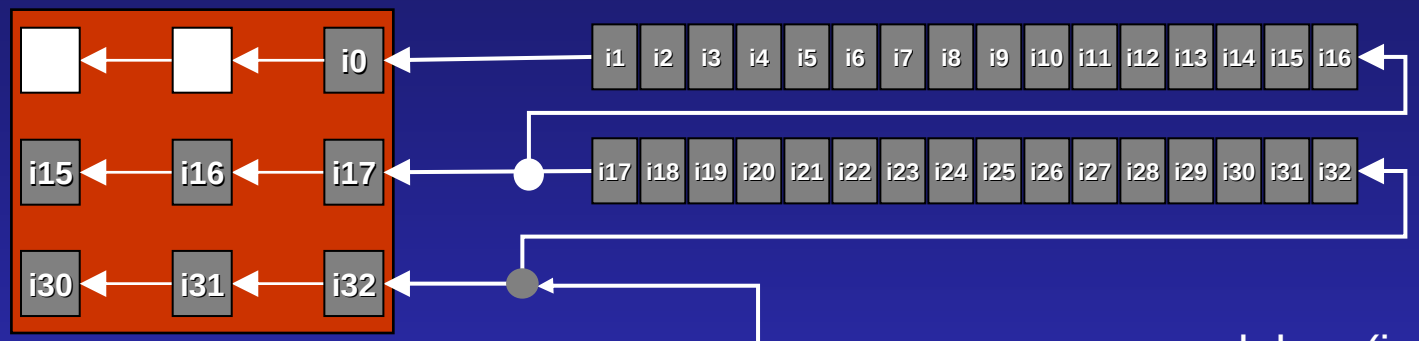
Output(i)



Data access input(i)

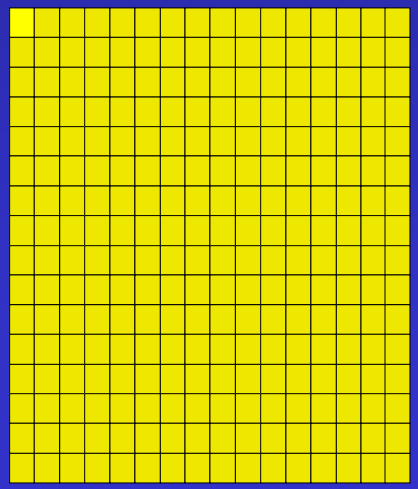
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Stencil Data Flow

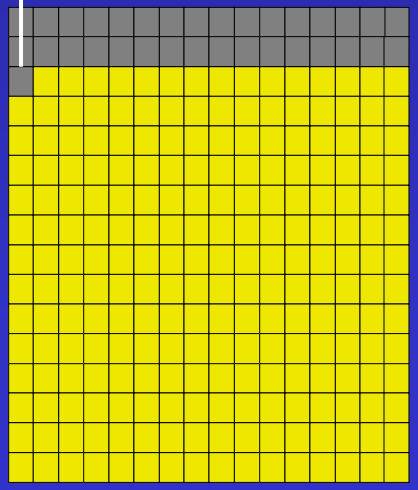


Compute $f(x_1,x_2,..x_9)$

delayq(in,&out);



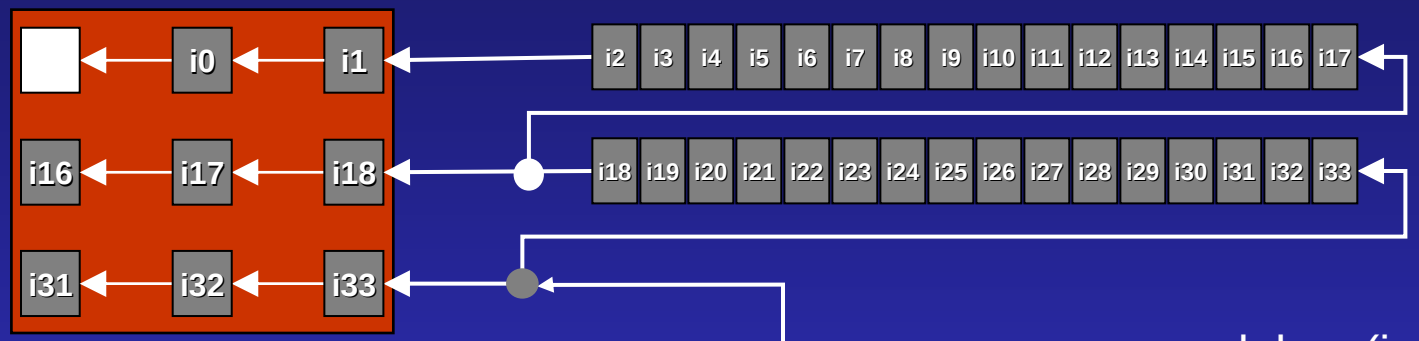
Output(i)



Data access input(i)

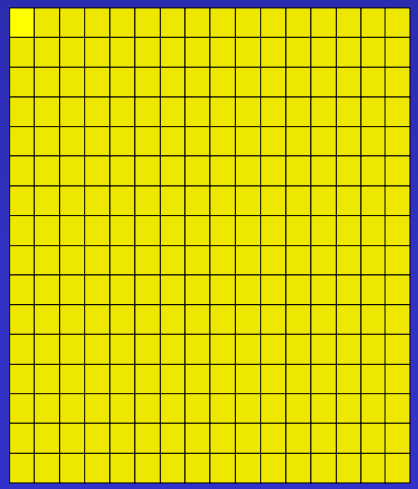
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Stencil Data Flow

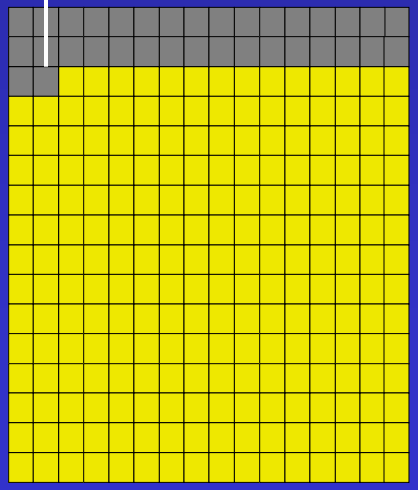


Compute $f(x_1,x_2,..x_9)$

delayq(in,&out);



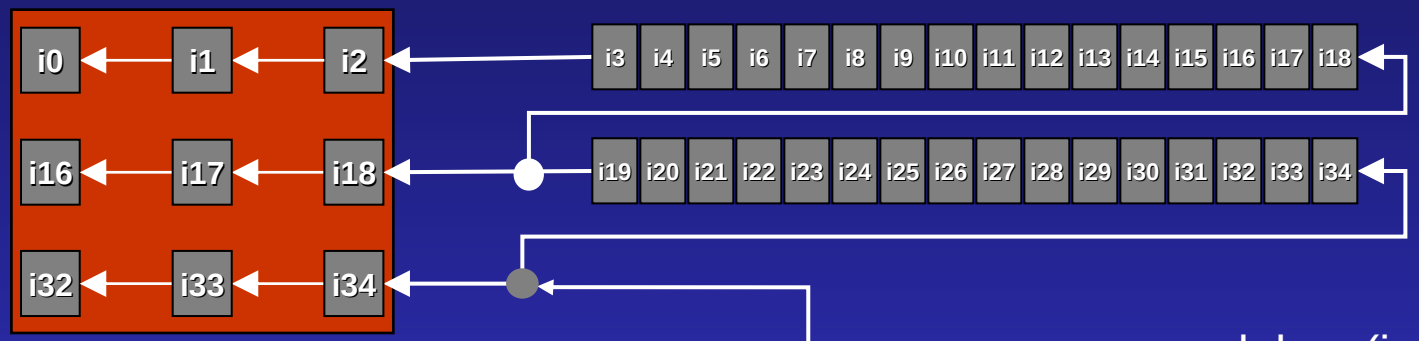
Output(i)



Data access input(i)

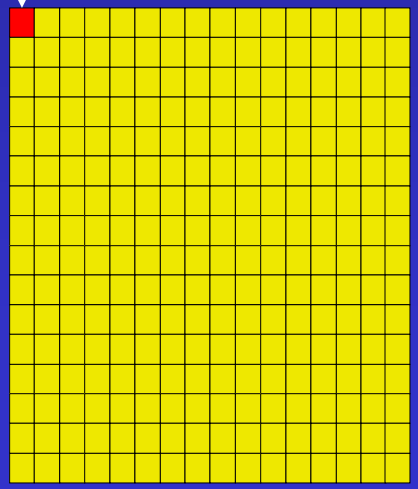
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Stencil Data Flow

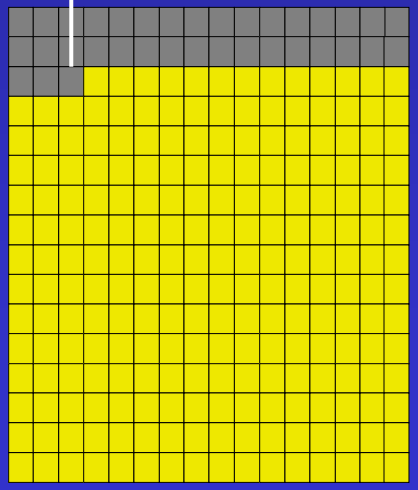


delayq(in,&out);

Compute $f(x_1,x_2,..x_9)$



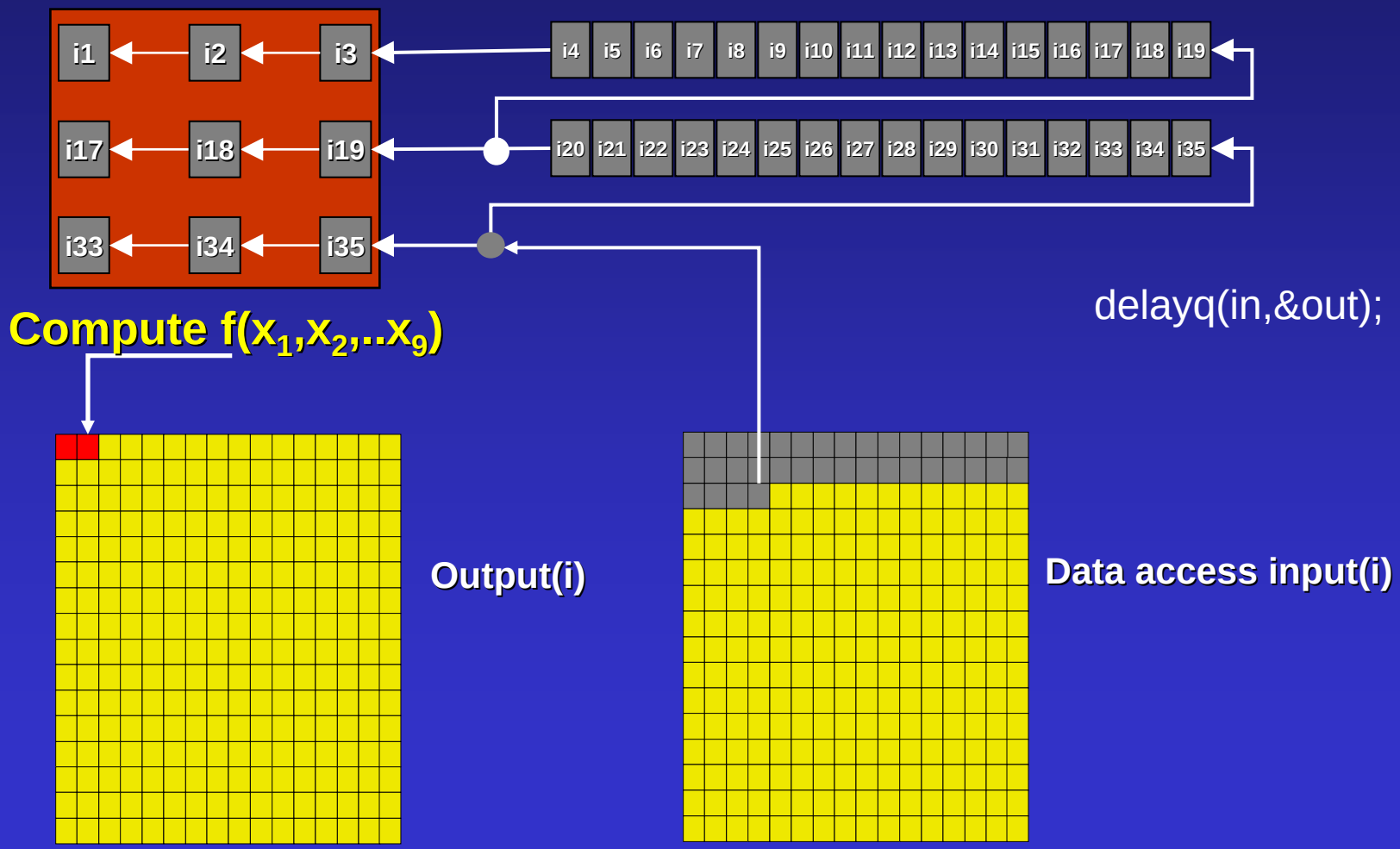
Output(i)



Data access input(i)

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

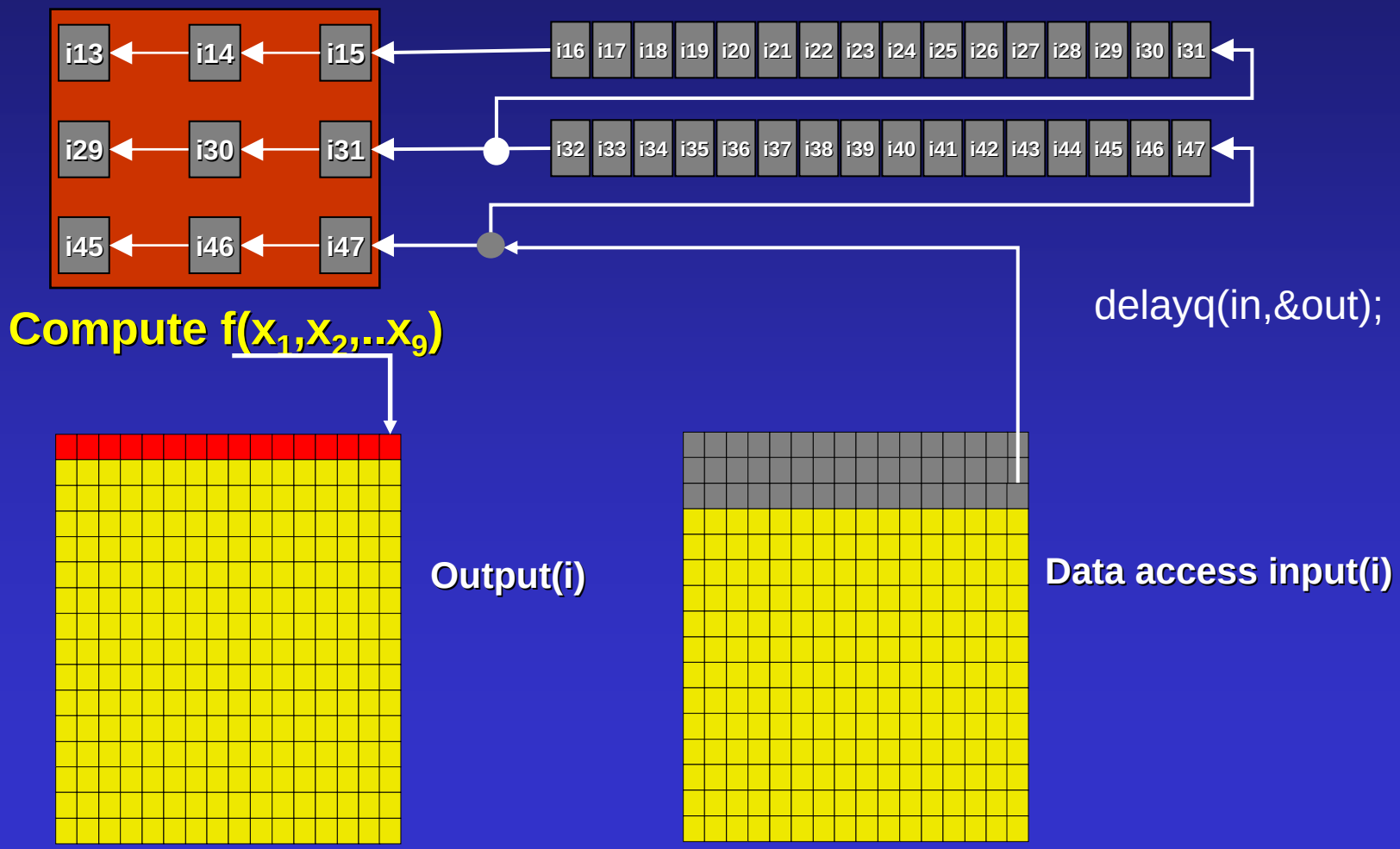
Stencil Data Flow



delayq(in,&out);

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

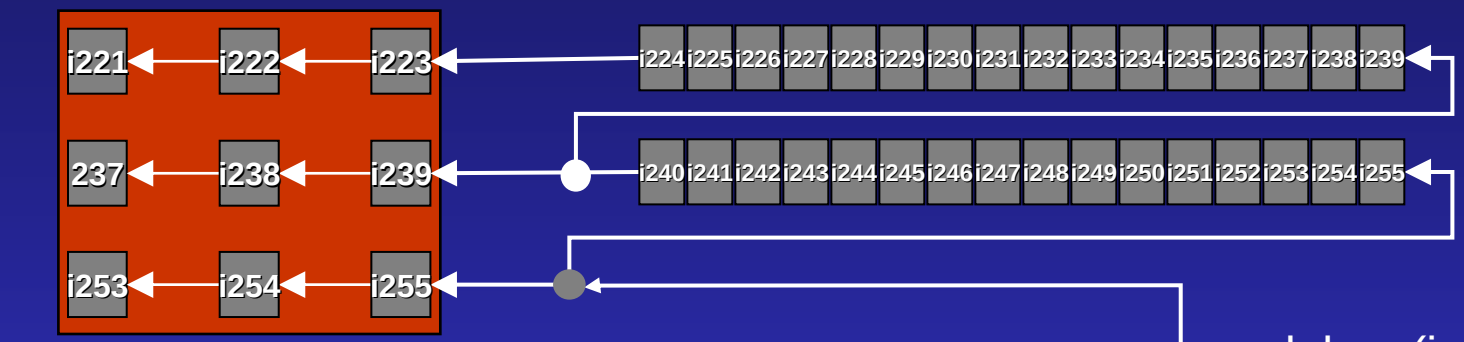
Stencil Data Flow



delayq(in,&out);

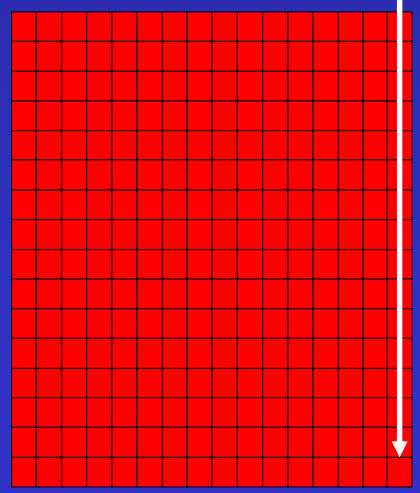
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Stencil Data Flow

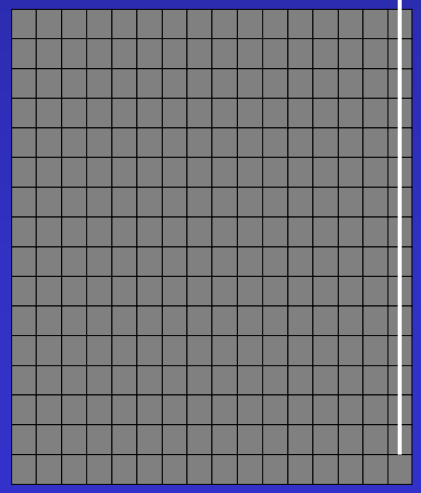


Compute $f(x_1,x_2,..x_9)$

delayq(in,&out);



Output(i)



Data access input(i)

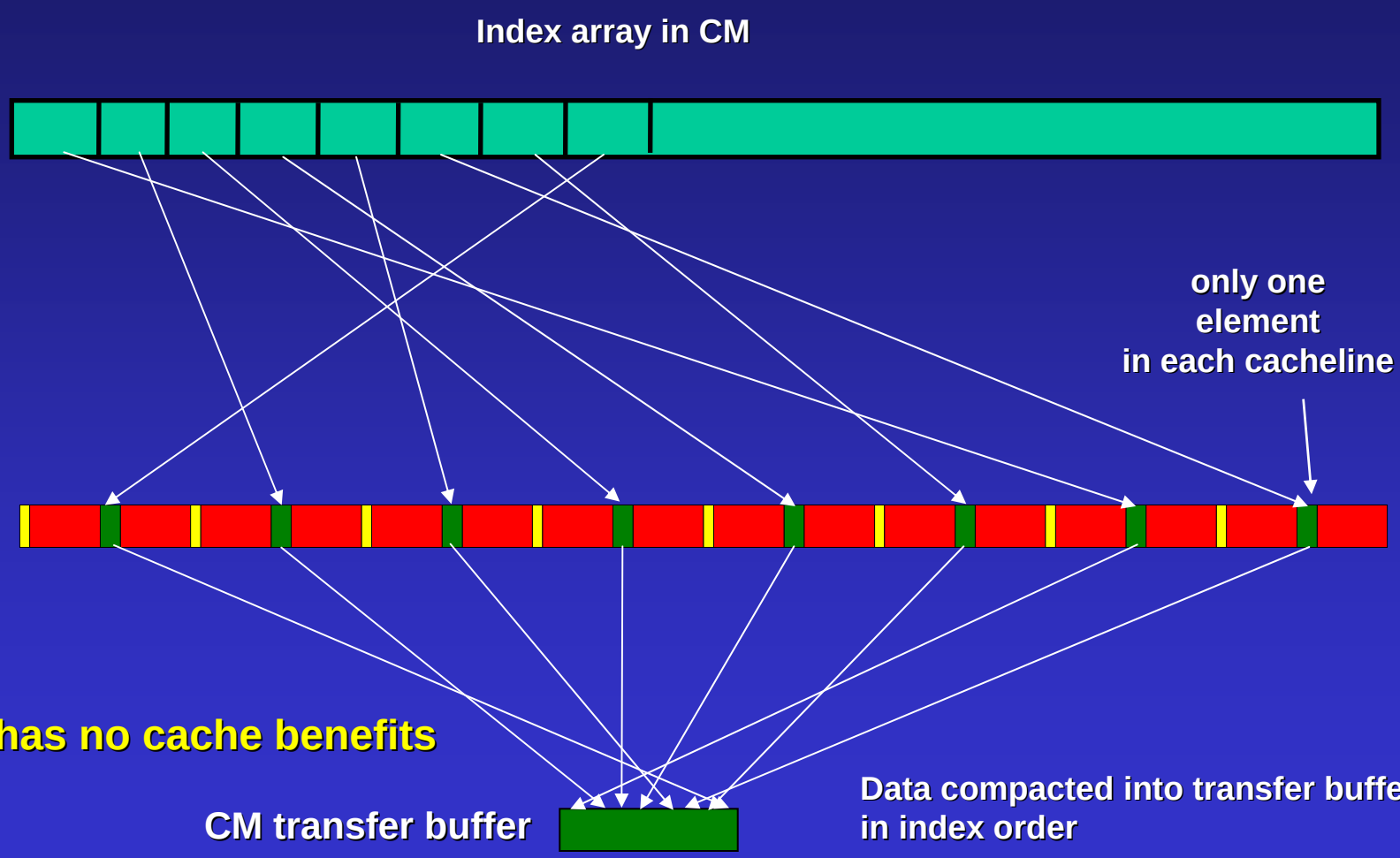
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Data Prefetch

- A functional Unit
- Maximized interconnect bandwidth
- FPGAs in memory control

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Prefetch - Gather



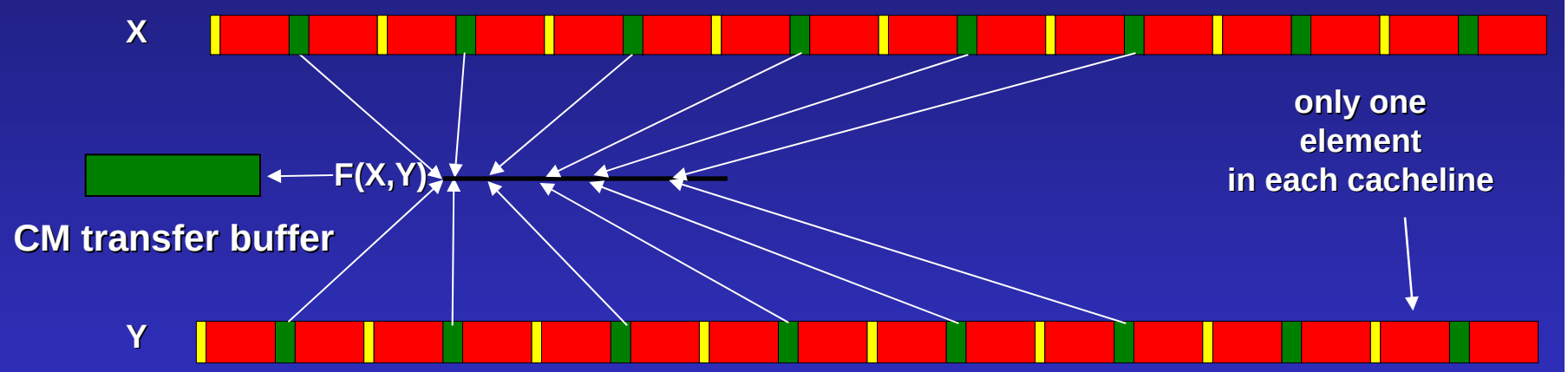
µP has no cache benefits

16x Transfer bandwidth efficiency advantage

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Prefetch - Strided with Compute

Strided data arrays in CM



32x Transfer bandwidth efficiency advantage

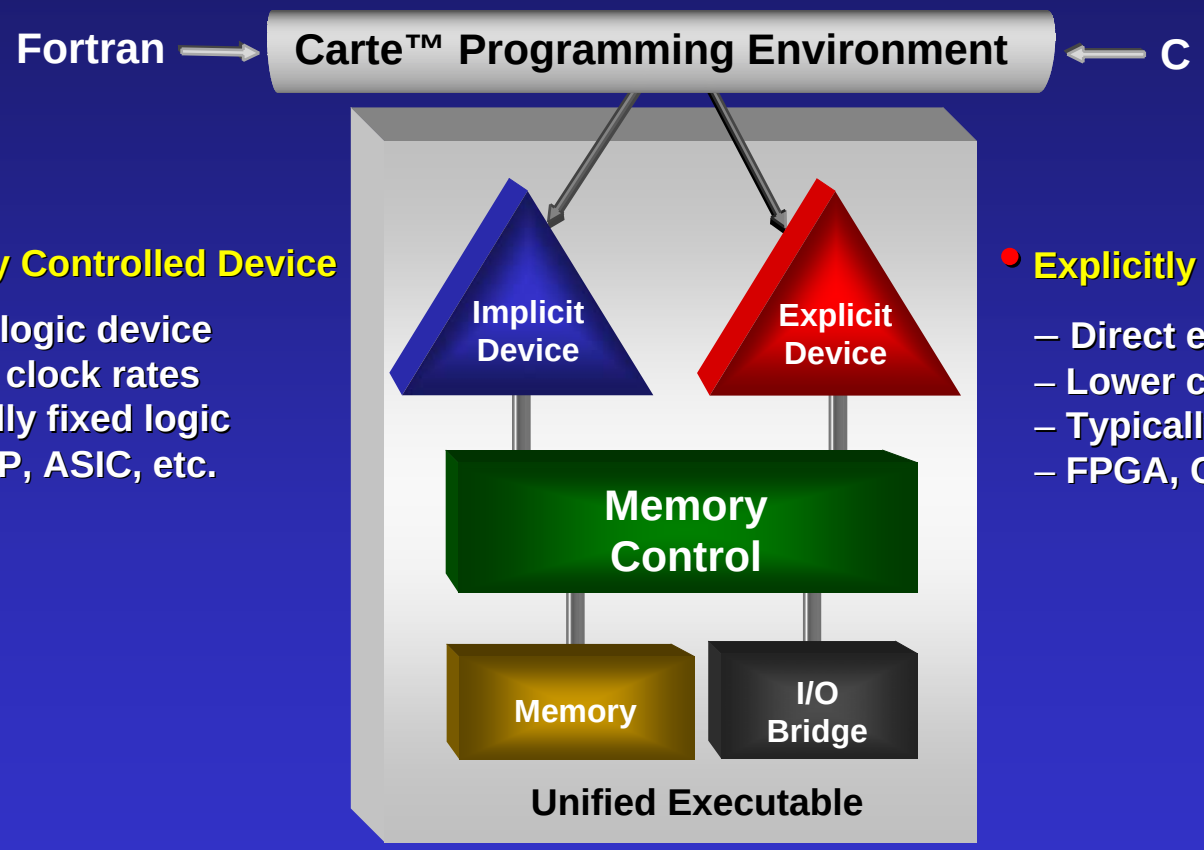
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

The Explicit Reconfigurable Processor

We call it MAP®

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

IMPLICIT+EXPLICIT™ Architecture



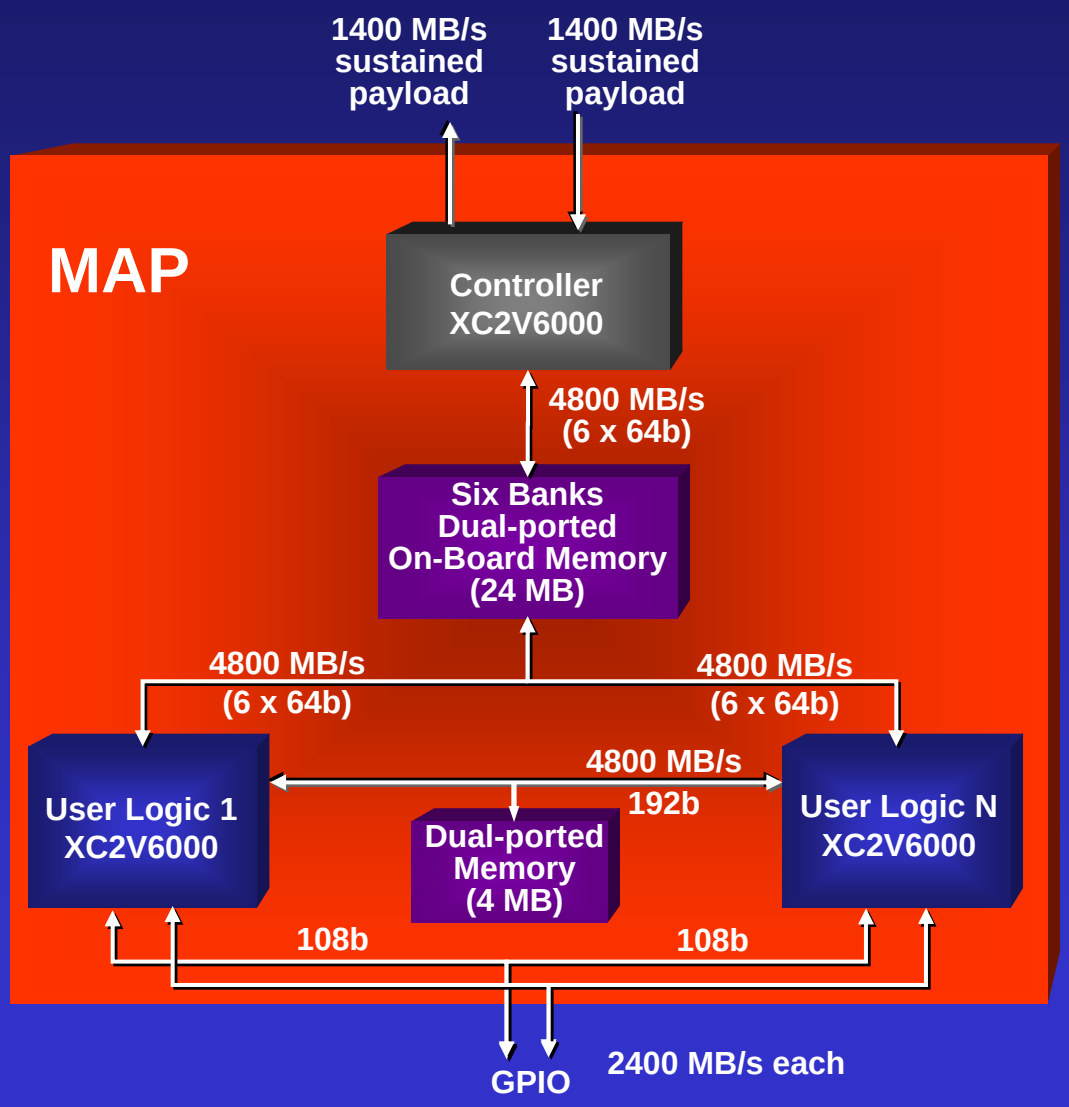
- **Implicitly Controlled Device**
  - Dense logic device
  - Higher clock rates
  - Typically fixed logic
  - µP, DSP, ASIC, etc.

- **Explicitly Controlled Device**
  - Direct execution logic
  - Lower clock rates
  - Typically reconfigurable
  - FPGA, CPLD, OPLD, etc.

SRC's explicitly controlled processor is called MAP®

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

MAP® Implementation



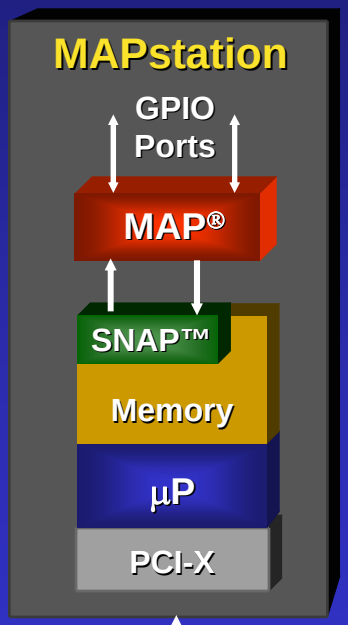
- Direct Execution Logic (DEL) made up of one or more User Logic devices
- Control circuits allow explicit control of memory prefetch and data access
- Multiple banks of On-Board Memory maximizes local memory bandwidth
- GPIO ports allow direct MAP to MAP chain connections or direct data input
- Multiple DMA engines support
  - Distributed SRAM in User Logic
    - 264 KB @ 844 GB/s
  - Block SRAM in User Logic
    - 648 KB @ 260 GB/s
  - On-Board SRAM
    - 28 MB @ 9.6 GB/s
  - Microprocessor Memory
    - 8 GB @ 1400 MB/s

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

SRC MAPstation™

SRC-6 uses standard external network connections

Single MAP Workstation



MAPstation Configurations

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

Programming Model

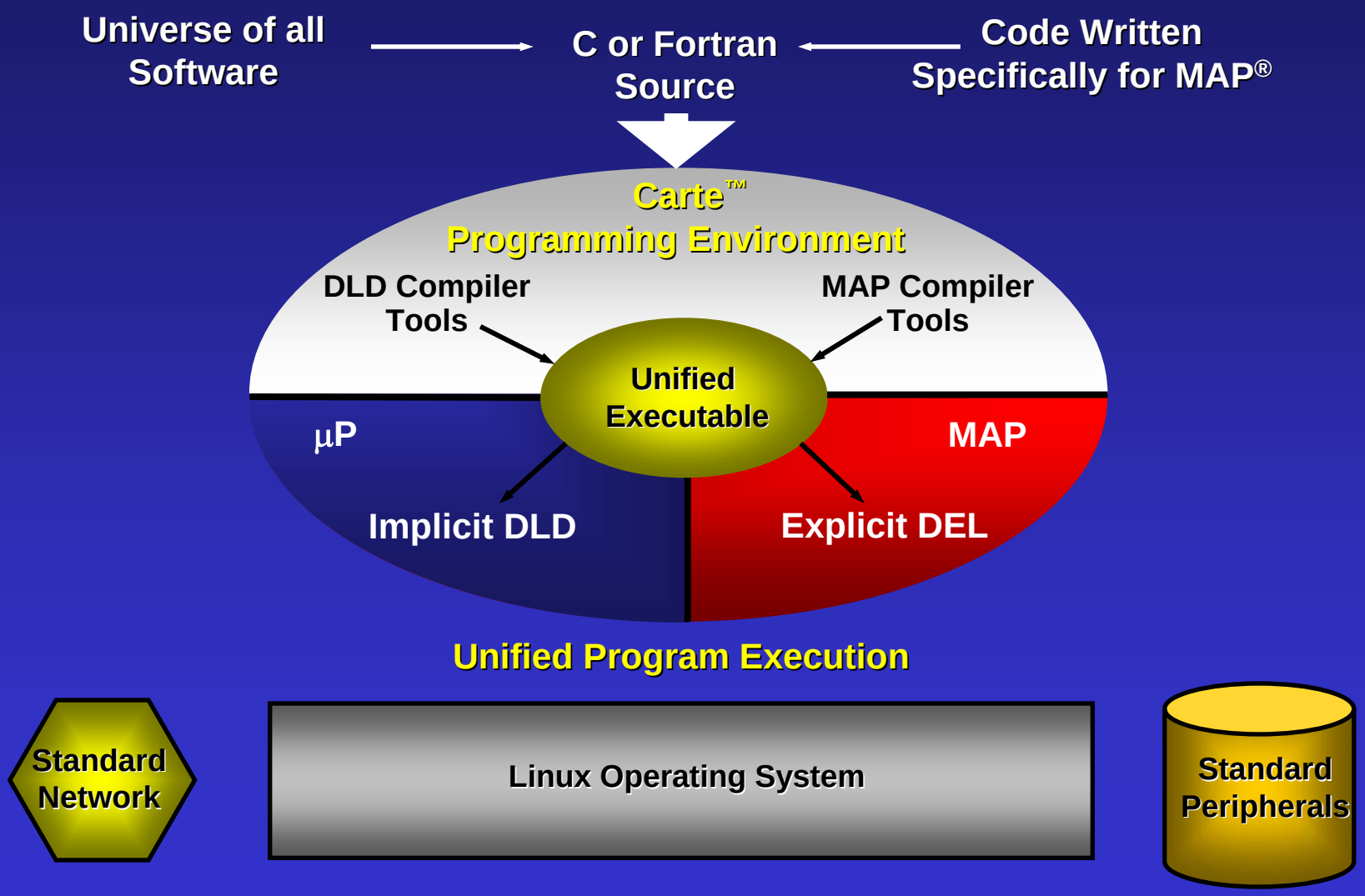
For Programmers not Logic Designers

But

Logic Designers are our Friends

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Unified Applications Environment



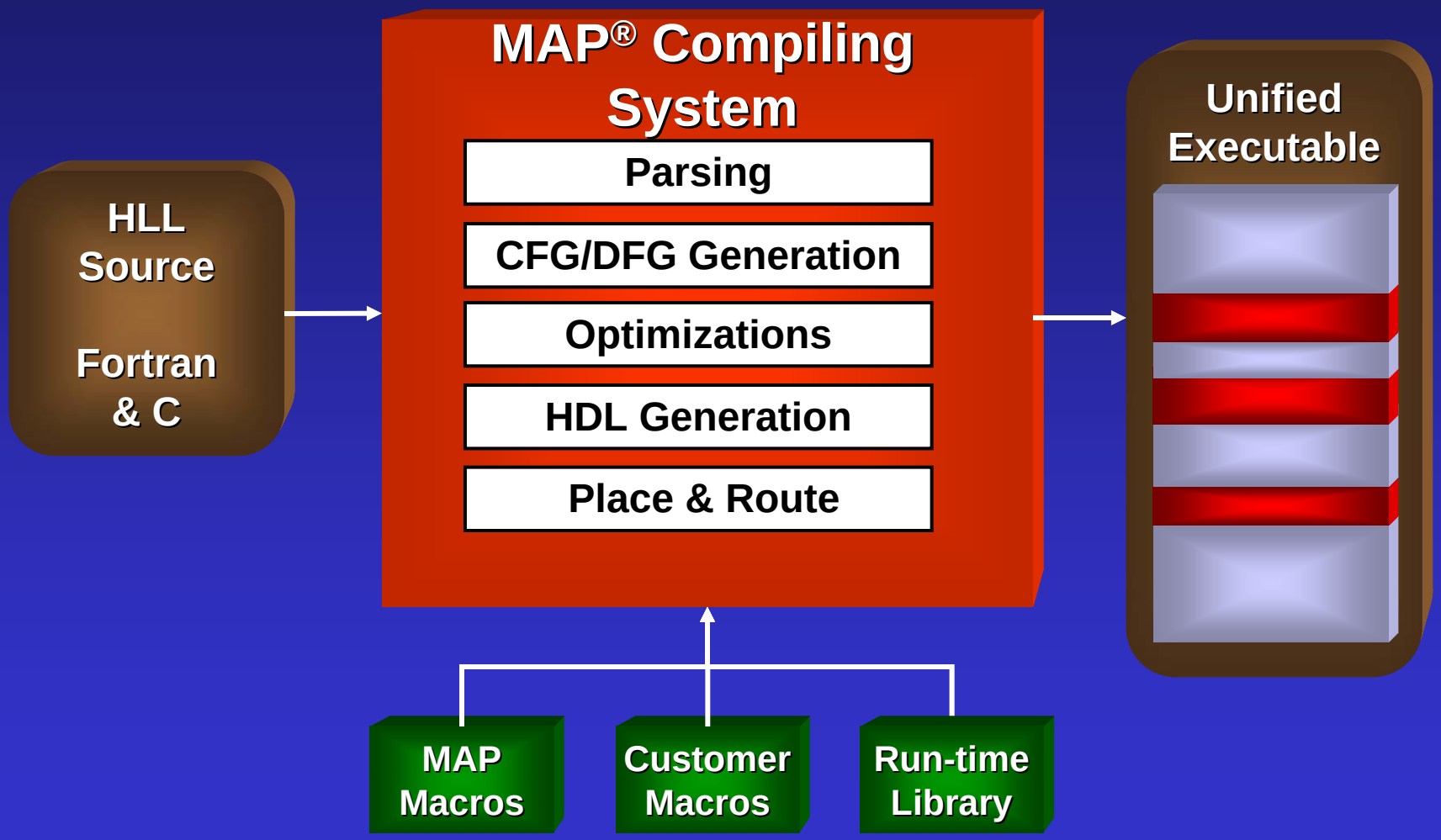
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

Carte: Programming not Logic Design

- Familiar High Level Programming Languages
- Familiar development and debugging tools
- Automatically generate MAP logic
- Automatic generation of MAP-μP integration code

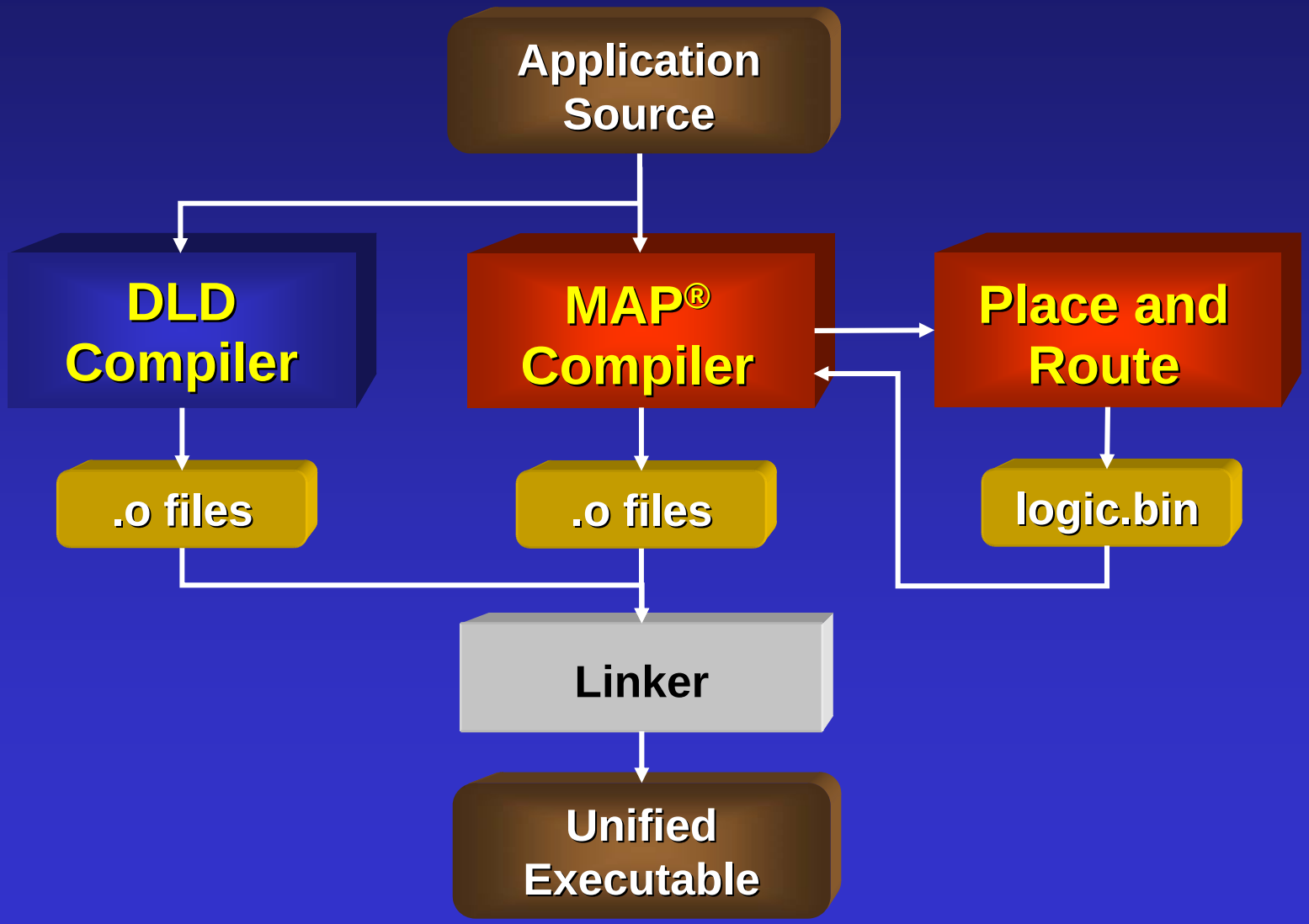
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Carte™ Programming Environment



Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

Compilation and Linking



Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Support for Development Process

- **Application Debug Environment**
  - Fast compile times
  - Familiar debugger environment
  - Debugging includes data movement to/from MAP®
  - Does not require MAP hardware to run
  - Fast enough to test a full application
- **Execution on MAP Hardware**
  - Compilation for MAP on the order 10s of minutes
  - Results match debug and simulation
- **Easy integration of Existing Direct Execution Logic**
  - Custom functional unit development supported
  - Functional unit simulation environment supported

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

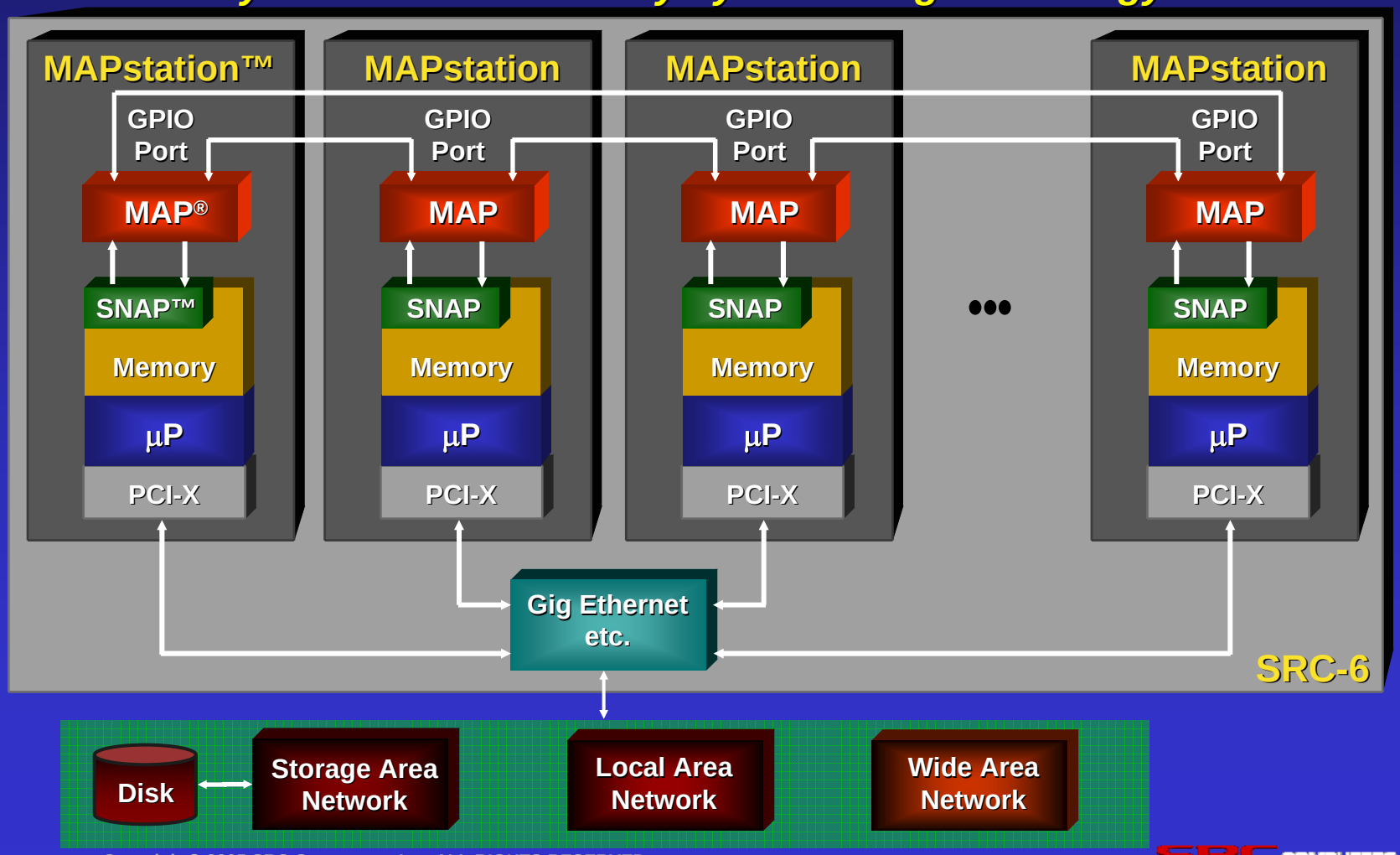
SRC System Architecture

Integration and Balance

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

SRC Cluster Based Systems

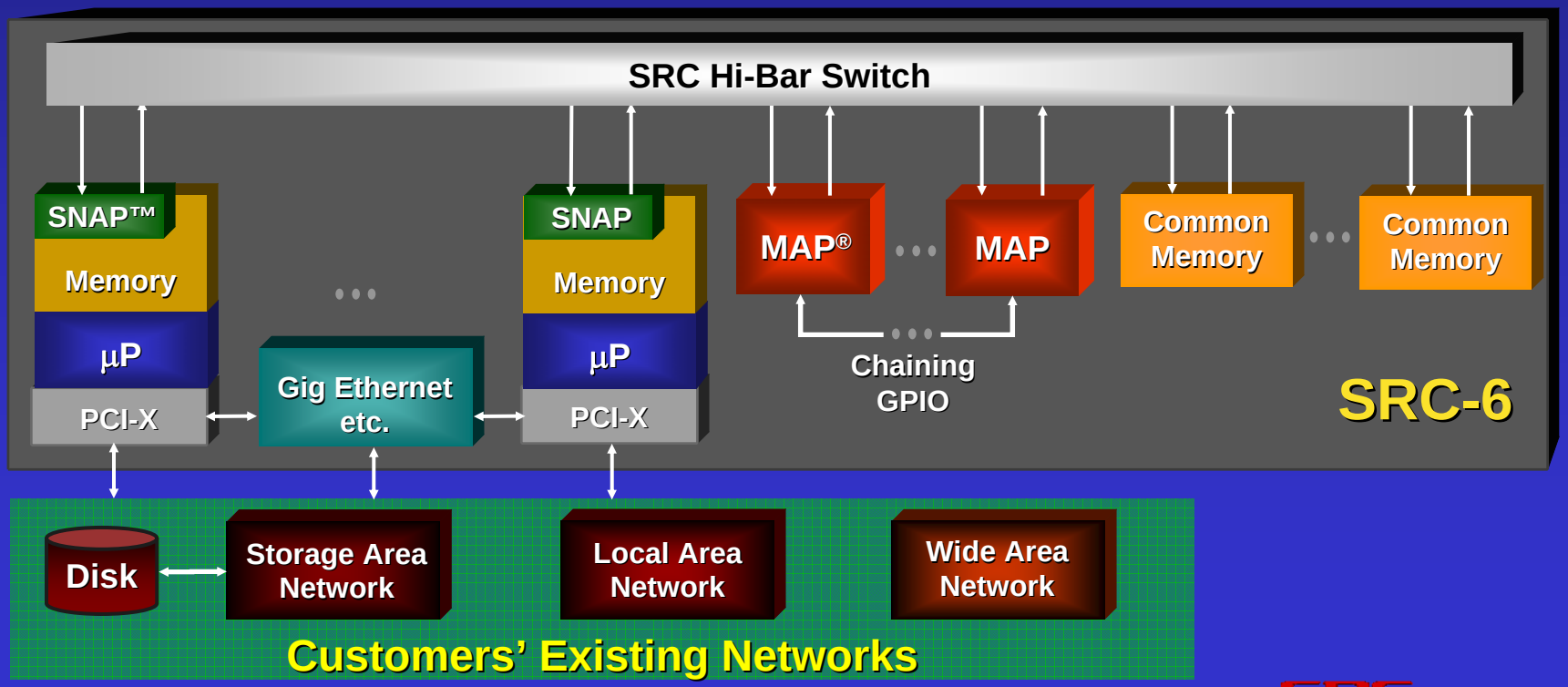
Utilizes standard clustering technology – system size limited only by clustering technology



Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

SRC Hi-Bar™ Based Systems

- Hi-Bar sustains 1.4 GB/s per port with 180 ns latency per tier
- Up to 256 input and 256 output ports with two tiers of switch
- Common Memory (CM) has controller with DMA capability
- Controller can perform other functions such as scatter/gather
- Up to 8 GB DDR SDRAM supported per CM node

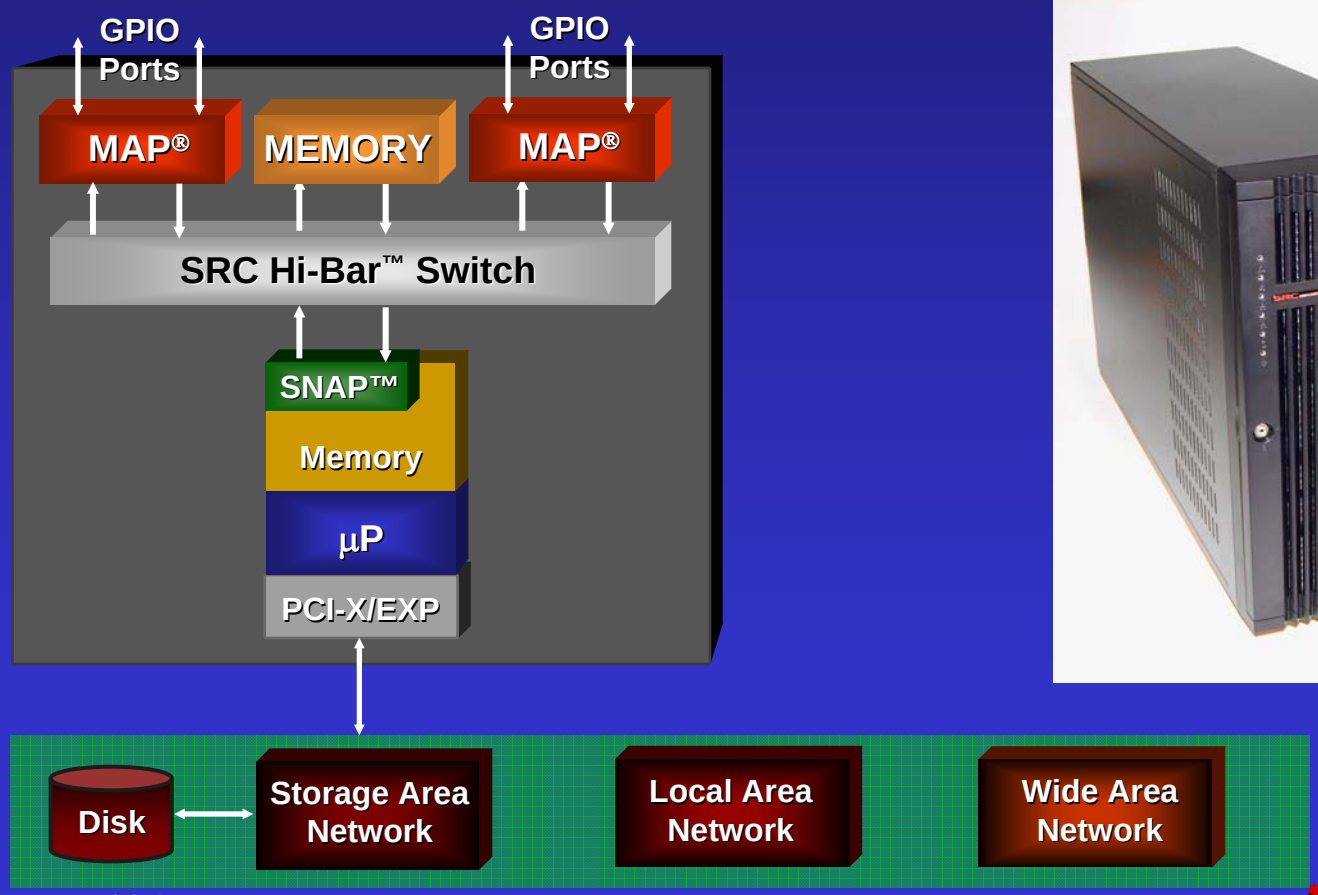


Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

SRC MAPstation™ with Hi-Bar™

MAPstation towers hold up to 3 MAP or memory nodes

MAPstation with 2 MAPs and Common Memory

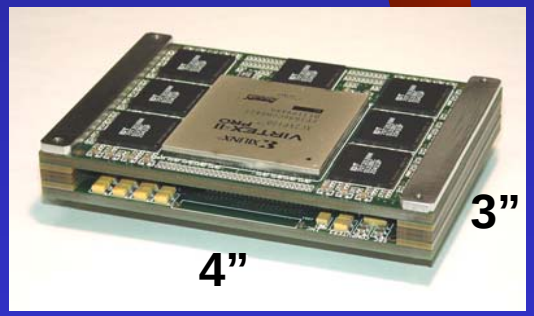


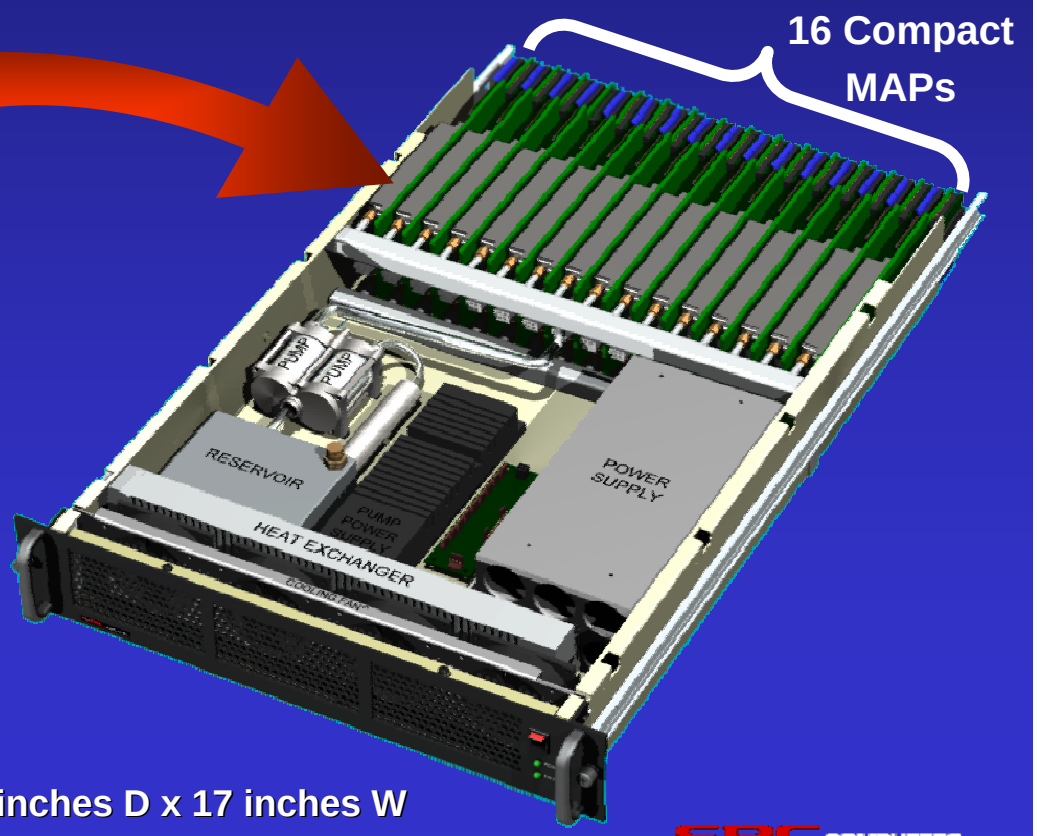
MAPstation Tower

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

2U Compact MAP™ Chassis

- Up to 16 Compact MAPs can be housed in one 2U chassis
- Increases system packaging density by 4X
- Over 200 MAPS in a single rack
- 800 Gflops per chassis
- 10 Tflops per rack





Overall dimensions are 3.50 inches H x 26 inches D x 17 inches W

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

Performance Trends

- Where do we stand today?
- What does future look like?

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

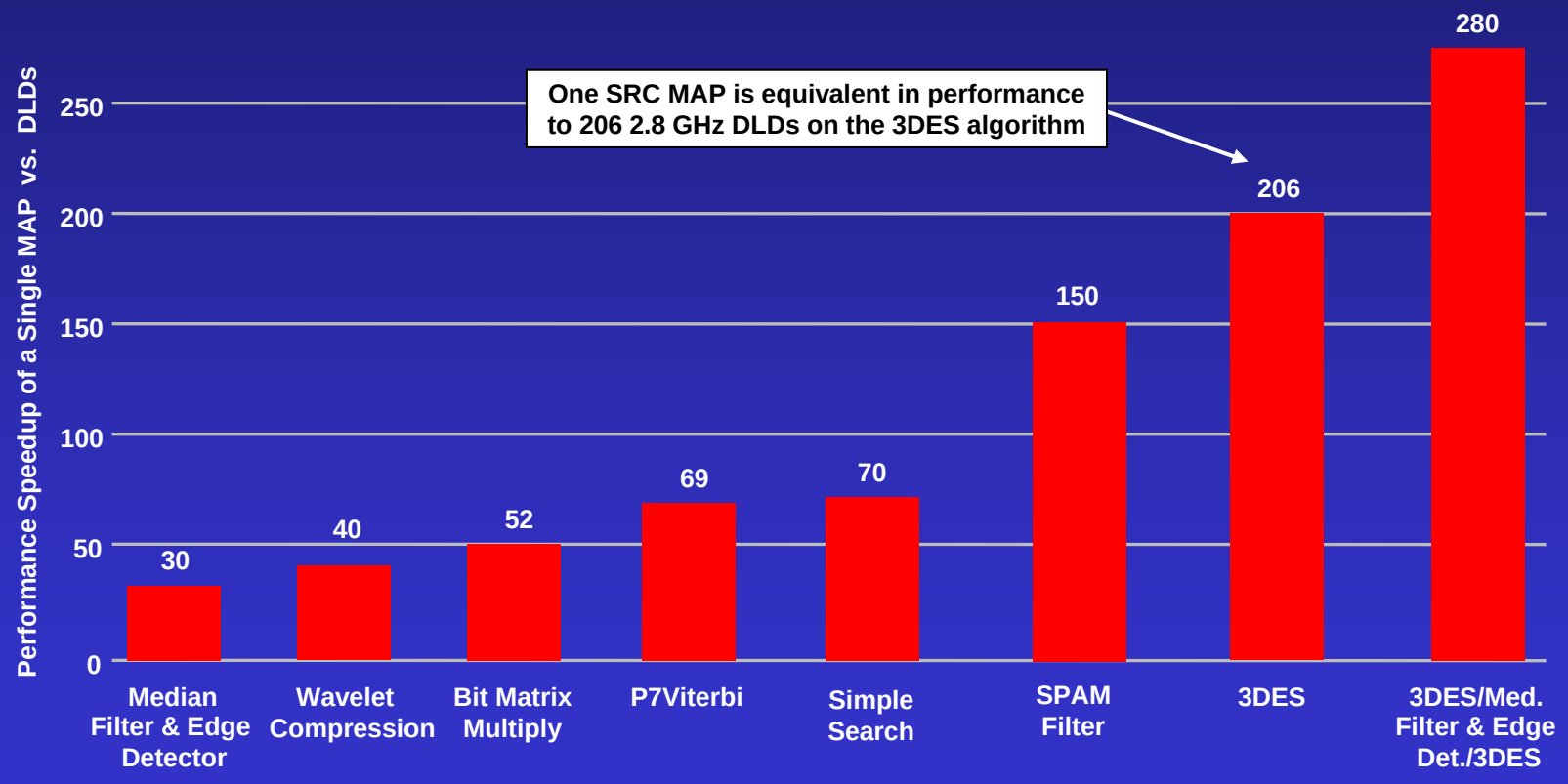
Some Comparisons

MAP vs Microprocessor

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

MAP® Logic Performance

Number of 2.8 GHz DLDs equivalent to one SRC MAP on Select Algorithms

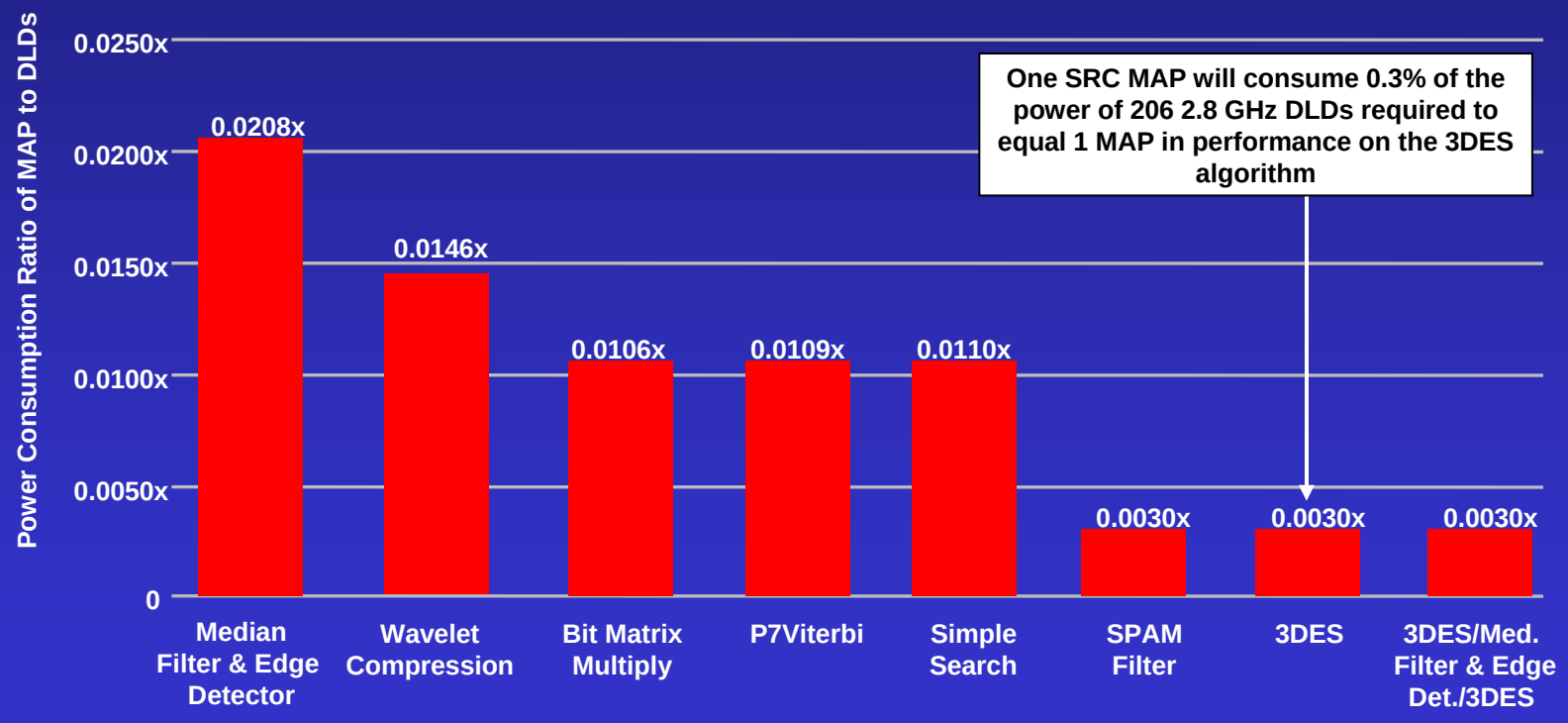


Source: SRC. - Comparisons are based on measured single microprocessor and single MAP processor performance

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

MAP® Power Savings

Power Consumption Ratio of 2.8 GHz DLDs equivalent to one SRC MAP on Select Algorithms



Based on the average commercial price of power Los Angeles, New York City, San Francisco and Washington DC of $0.12 per Kilowatt-Hour.
Comparison based on total power consumed by MAP control chip, On-Board Memory and user chips as compared to microprocessor chip and VRM only.

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

A Floating Pt. Update

- **Back Projection SAR**
  - *2D Data – 5040 scans of 512 rows*
  - *16K Complex Float FFT*
  - **MATLAB implemented: 4920 sec (1x)**
  - **C - Intel FFT: 1162 sec (4.2x)**
  - **MATLAB – MAP: 34 sec (144x)**
  - **C - MAP: 33 sec (144x)**

- **More Floating Pt. to come from:**
  - **ORNL, LANL, GWU, U So Carolina, USC**
  - **NASA, Air Force**

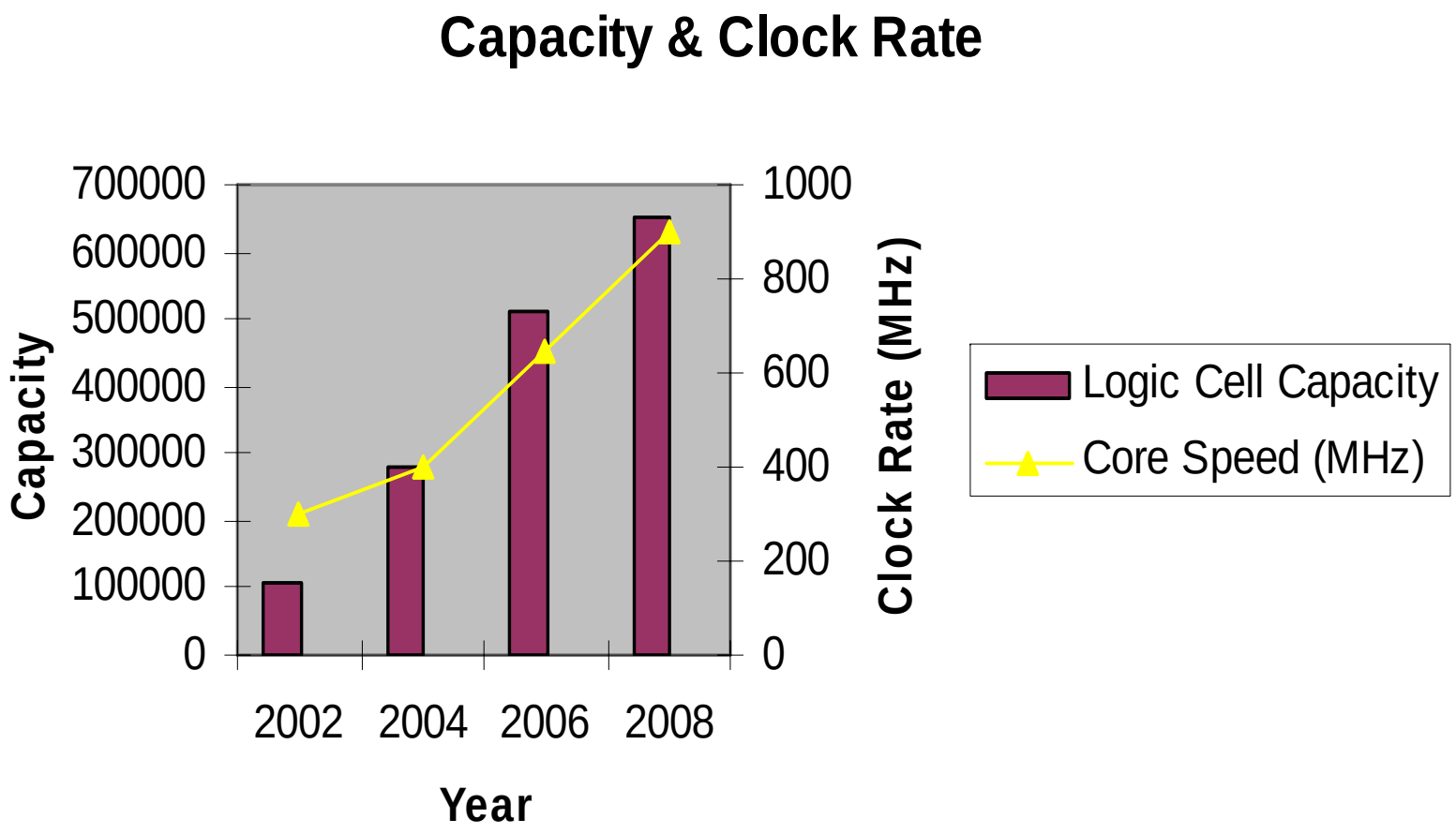
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

The FPGA Trends

What's Happening to FPGAs?

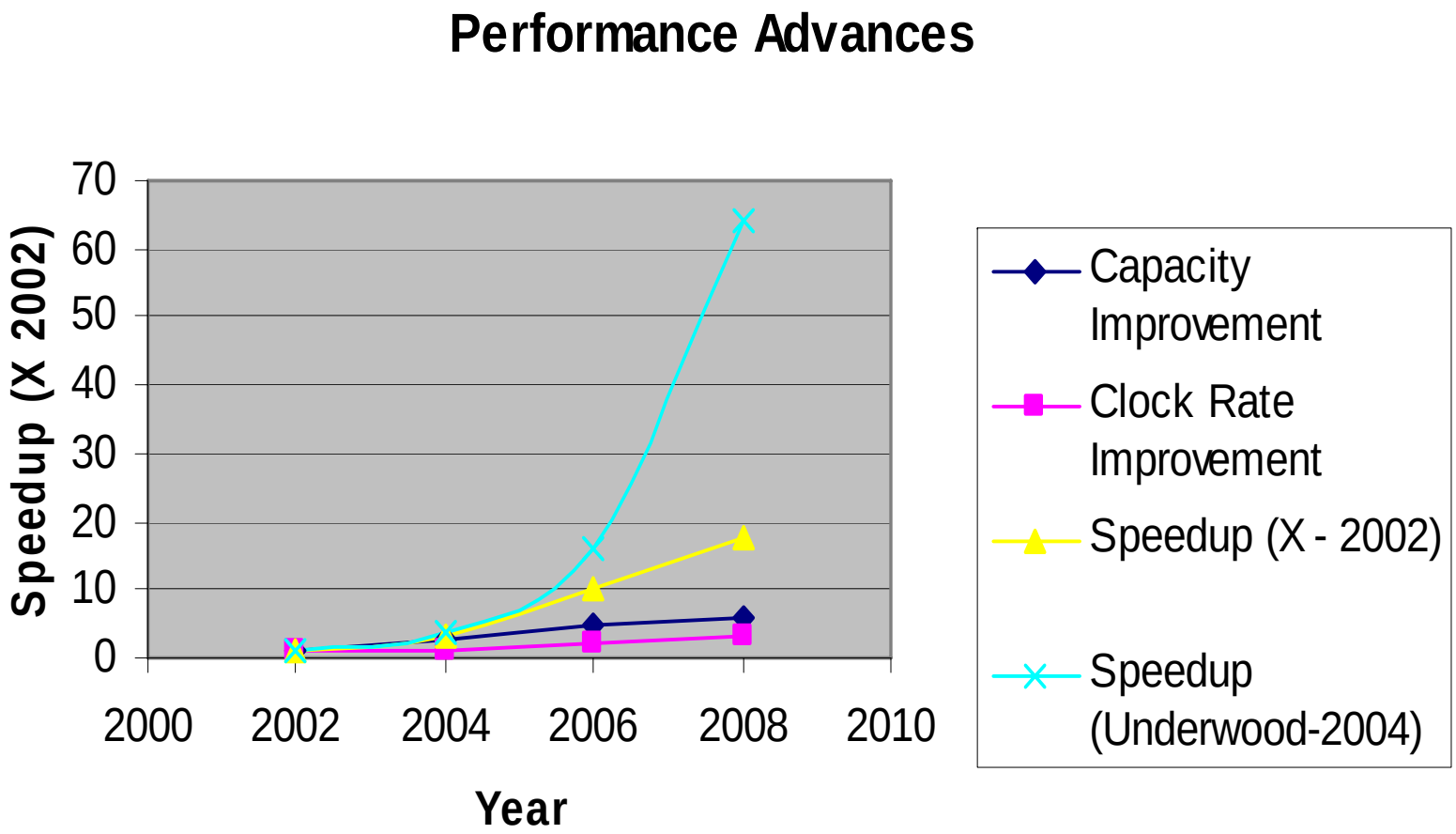
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

FPGA Industry Stats



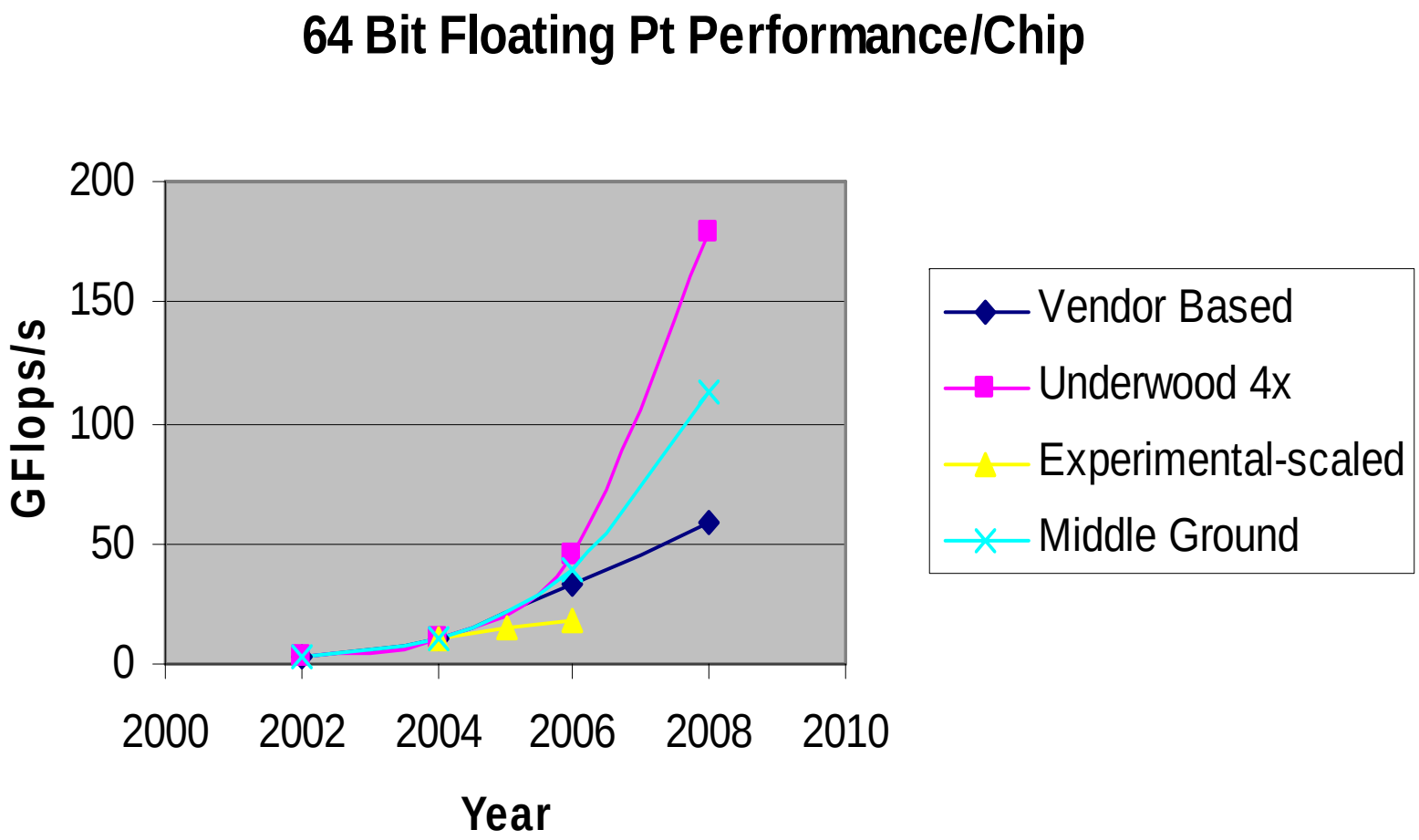
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Speedup Projections



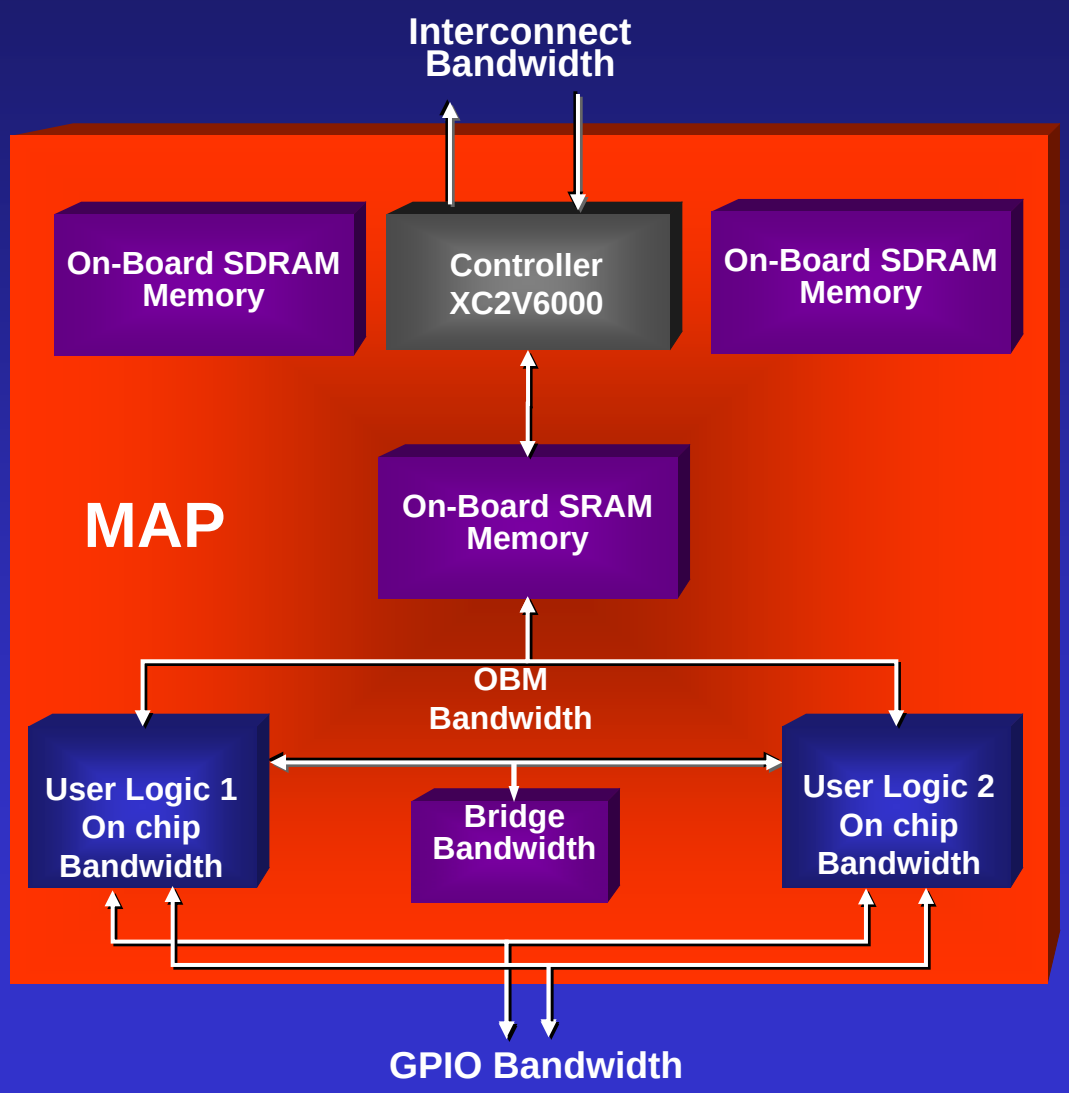
Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Projected Floating Pt Performance



Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

MAP® Processor Performance



External Processor Bandwidth:

- Interconnect
- GPIO

Internal Processor Bandwidth:

- On Board Memory
- Bridge
- BRAM Banks
- SRAM registers

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

Explicit RC Processor

Yr	Middle Ground GF/chip	Clk (MHz)	B/Flop	Tot GB/s	Ext GB/s	OBM GB/s	Int. GB/s
2004	11.2	100	2.2	49.60	8.00	4.80	36.80
2005	18.6	166	2.4	87.94	8.00	7.97	71.97
2006	39.0	166	1.3	98.99	16.40	18.59	64.00
2008	112.3	333	0.8	168.9	19.20	21.31	128.4

Rule of thumb 1 Byte BW per Flop – Is this still valid?

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

Petascale System

Yr	Middle Ground GF/chip	GF/MAP	MAPs Per Tfop	MAPs Per Pflop	Racks Per Pflop
2004	11.20	22.40	45	44643	930
2005	18.59	37.18	27	26893	560
2006	38.98	77.95	13	12829	267
2008	112.34	224.68	4	4451	93

48 MAP + Switches per rack

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

Path to Petaflops

2005 100 Teraflop system

- 2600 MAPs in 56 Racks
- 2600 CM Banks in 27 Racks
- μP's Depending on application mix

2006 1-Petaflop system

- 12829 MAPs in 267 Racks
- 12829 CM Banks in 140 Racks
- μP's Depending on application mix

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

Path to Petaflops with Help

2008-9 1-Petaflop system

- 500-1000 MAPs
- 500-1000 CM Banks
- μP's Depending on application mix

Price Tag:

R&D:	$10M
Optimization:	$15M
Production:	$20M
Recurring:	<< $20M

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED

Concluding Comments

Explicit Reconfigurable Computing:

- Makes efficient use of the chip
- Can be implemented with commodity parts
- Can yield significant performance
- Allows application specific processors
- Puts creativity in hands of users
- Can be programmed
- Is a cheaper path to Petascale

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com

SRC Computers, Inc.

Copyright© 2005 SRC Computers, Inc. ALL RIGHTS RESERVED www.srccomputers.com