A CMOS Vector Processor with a Custom Streaming Cache

Greg Faanes gjf@cray.com

Silicon Graphics

SiliconGraphics
Computer Systems

A CMOS Vector Processor With a Custom Streaming Cache

Overview:

- Goals and ground rules for the design
- CPU/Cache overview
- CPU chip
  - Scalar unit design
  - Vector unit design
- Streaming cache chip
  - Cache unit design
- Summary

SiliconGraphics
Computer Systems

Goals and Ground Rules for the Design

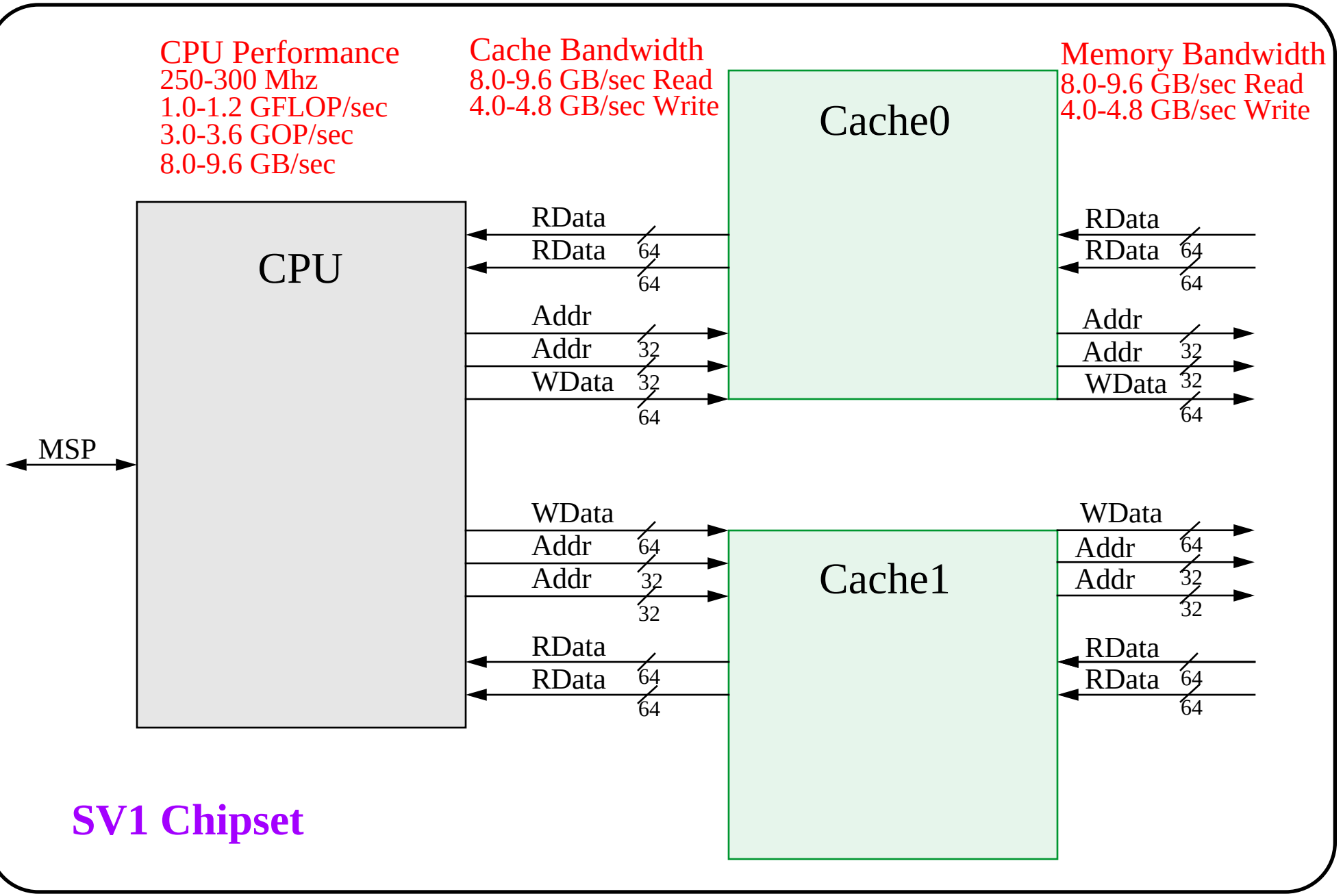
- Built to run large scale scientific applications
- Good vector performance
- High bandwidth to cache and main memory
  - unit stride
  - non-unit stride
  - gather/scatter
- YMP upward compatible
  - YMP ⇒ J90 ⇒ J90se ⇒ SV1 processor
- Few resources
  - 3 logic designers
  - “Off the shelf” technology

Not Your Ordinary Processor

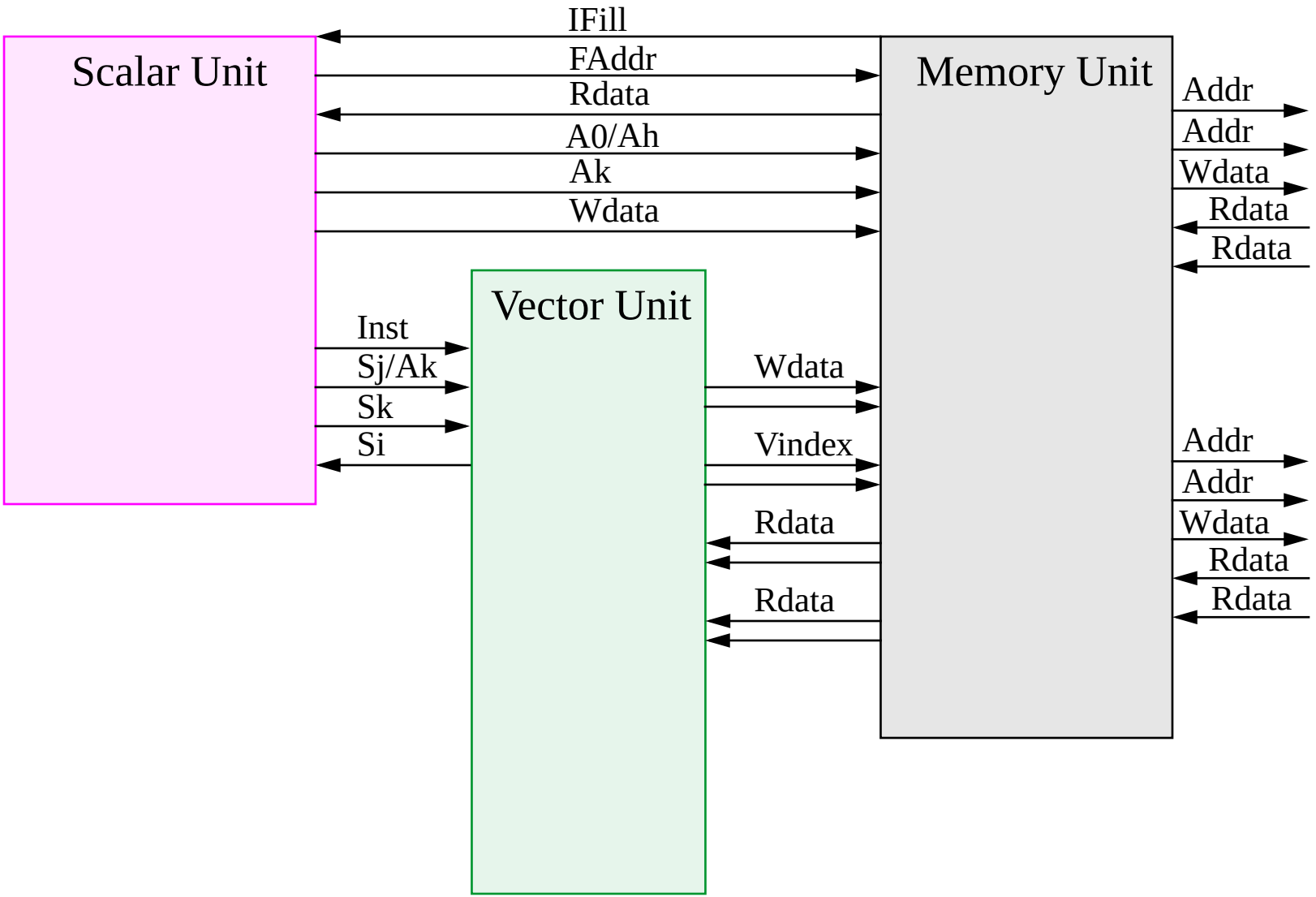
- **Not** built to run SPECINT95 fast
- Different cache and system interface
  - 1 set of pins for both cache and memory
- High performance through multiple parallel and pipelined functional units

SV1 Chipset

	CPU Chip	Cache Chip
Clock Frequency	250-300 Mhz	250-300 Mhz
Performance	1.0-1.2 GFLOP/sec	128 KB and 4.0-4.8 GB/sec
Die Size	14.6 mm x 14.6 mm = 213 mm^2	13.7 mm x 13.7 mm = 188 mm^2
Technology	CMOS, 5 layer metal, 2.6 v	CMOS, 5 layer metal, 2.6 v
Transistors	9.7 million	14.9 million
Power	12 watts	8 watts
Signal Pins	635	601


CPU Performance
250-300 Mhz
1.0-1.2 GFLOP/sec
3.0-3.6 GOP/sec
8.0-9.6 GB/sec
Cache Bandwidth
8.0-9.6 GB/sec Read
4.0-4.8 GB/sec Write
Memory Bandwidth
8.0-9.6 GB/sec Read
4.0-4.8 GB/sec Write
Cache0
CPU
MSP
RData
RData
64
64
Addr
Addr
32
WData
32
64
RData
RData
64
64
Addr
Addr
32
WData
32
64
WData
Addr
64
Addr
32
32
RData
RData
64
64
WData
Addr
64
Addr
32
32
RData
RData
64
64
Cache1
SV1 Chipset

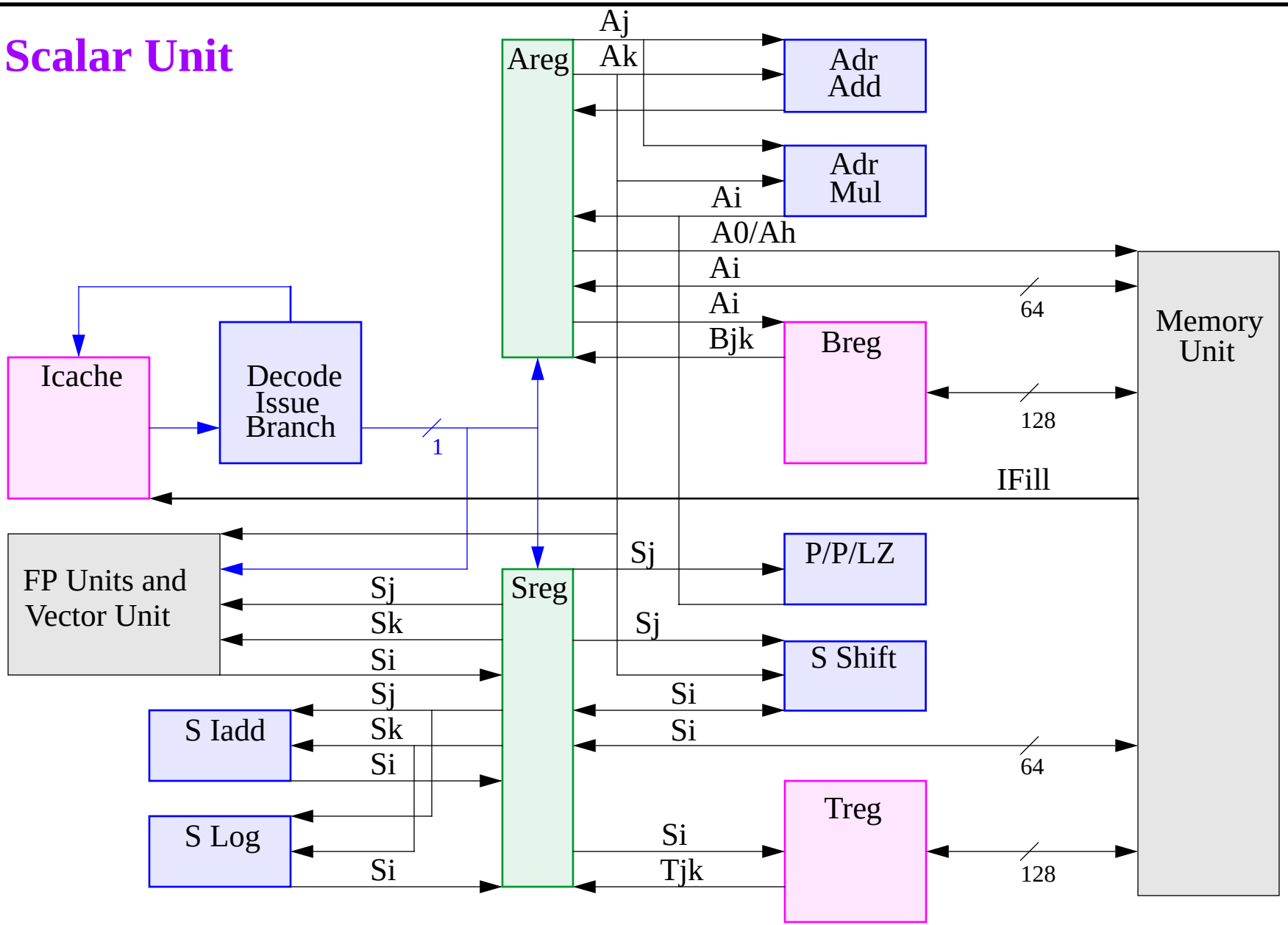
CPU Chip


Scalar Unit
Memory Unit
Vector Unit
IFill
FAddr
Rdata
A0/Ah
Ak
Wdata
Inst
Sj/Ak
Sk
Si
Wdata
Vindex
Rdata
Rdata
Addr
Addr
Wdata
Rdata
Rdata
Addr
Addr
Wdata
Rdata
Rdata

Scalar Unit

- Icache
  - 16-bit instruction parcels
  - 1/2/3 parcel instructions
  - 2 KByte size
  - 8 way associativity
  - 256 Byte lines
  - Branch prediction - always predict not taken
- Scalar
  - 8 32-bit Address Registers **Areg**
  - 64 32-bit B Registers **Breg**
  - 8 64-bit Scalar Registers **Sreg**
  - 64 64-bit T Registers **Treg**
  - Total of >1KB of scalar registers
  - 1 instruction per CP
  - No scalar cache

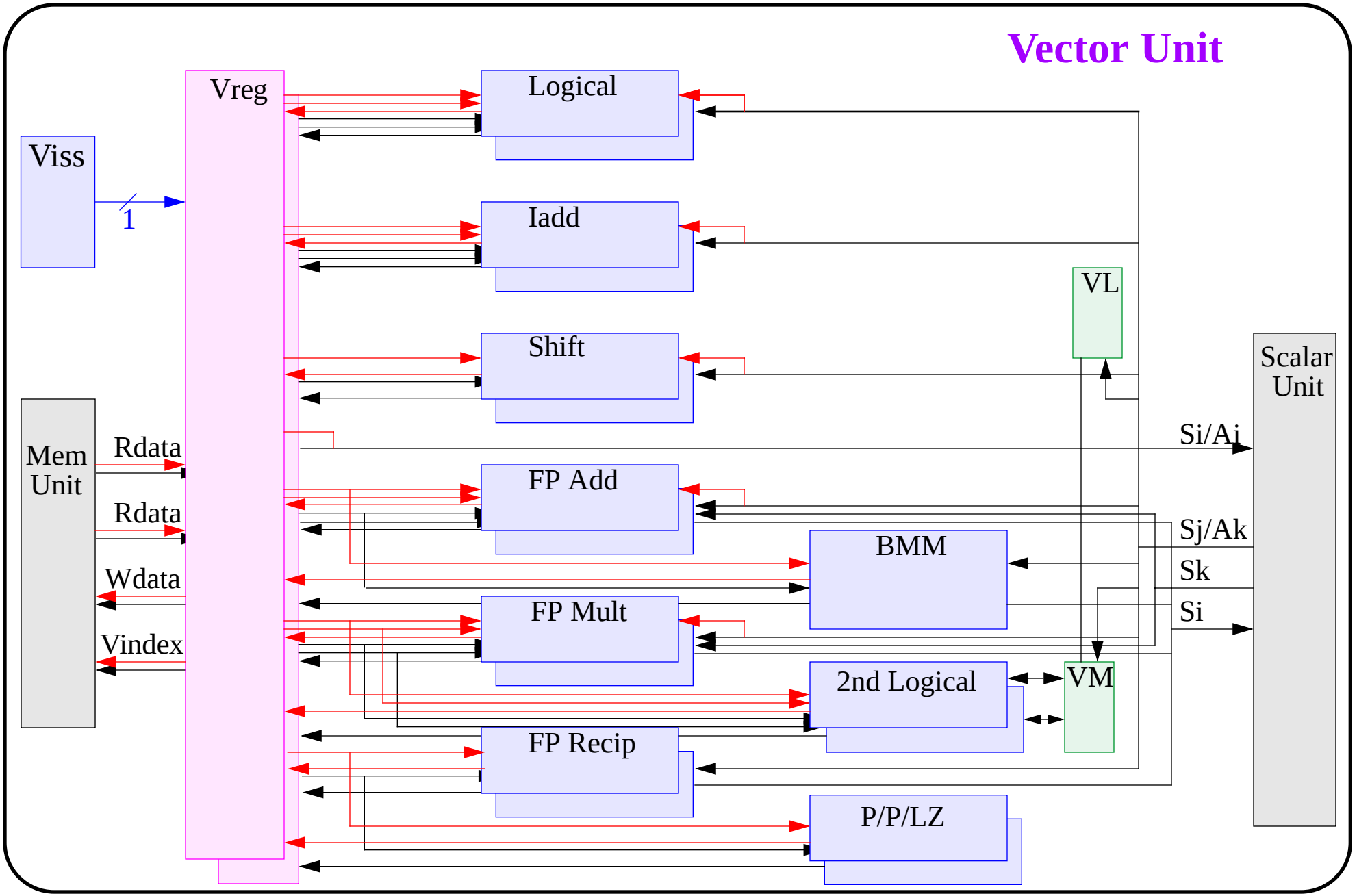
Scalar Unit


Aj
Areg
Ak
Adr
Add
Adr
Mul
Ai
A0/Ah
Ai
Ai
Bjk
64
Breg
128
Memory
Unit
Icache
Decode
Issue
Branch
1
IFill
FP Units and
Vector Unit
Sj
Sreg
P/P/LZ
Sj
Sk
Si
Sj
S Shift
Si
Sj
Si
S Iadd
Sk
Si
64
Treg
S Log
Si
Tjk
128
Si

Vector Unit

- 8 64-bit Vector Registers **Vreg**
- 64 elements per Vreg
- Total of 4KB of vector registers
- 1 64-bit Vector Mask **VM**
- 1 Vector Length Register **VL**
- Full chaining and tailgating
- 1 instruction per CP
- 2 FP add units
- 2 FP multiply units
- 2 FP reciprocal units
- 2 integer add units
- 4 logical units
- 2 shift units
- 2 pop/parity/leading zero units
- 1 BMM unit - 2 results per cp

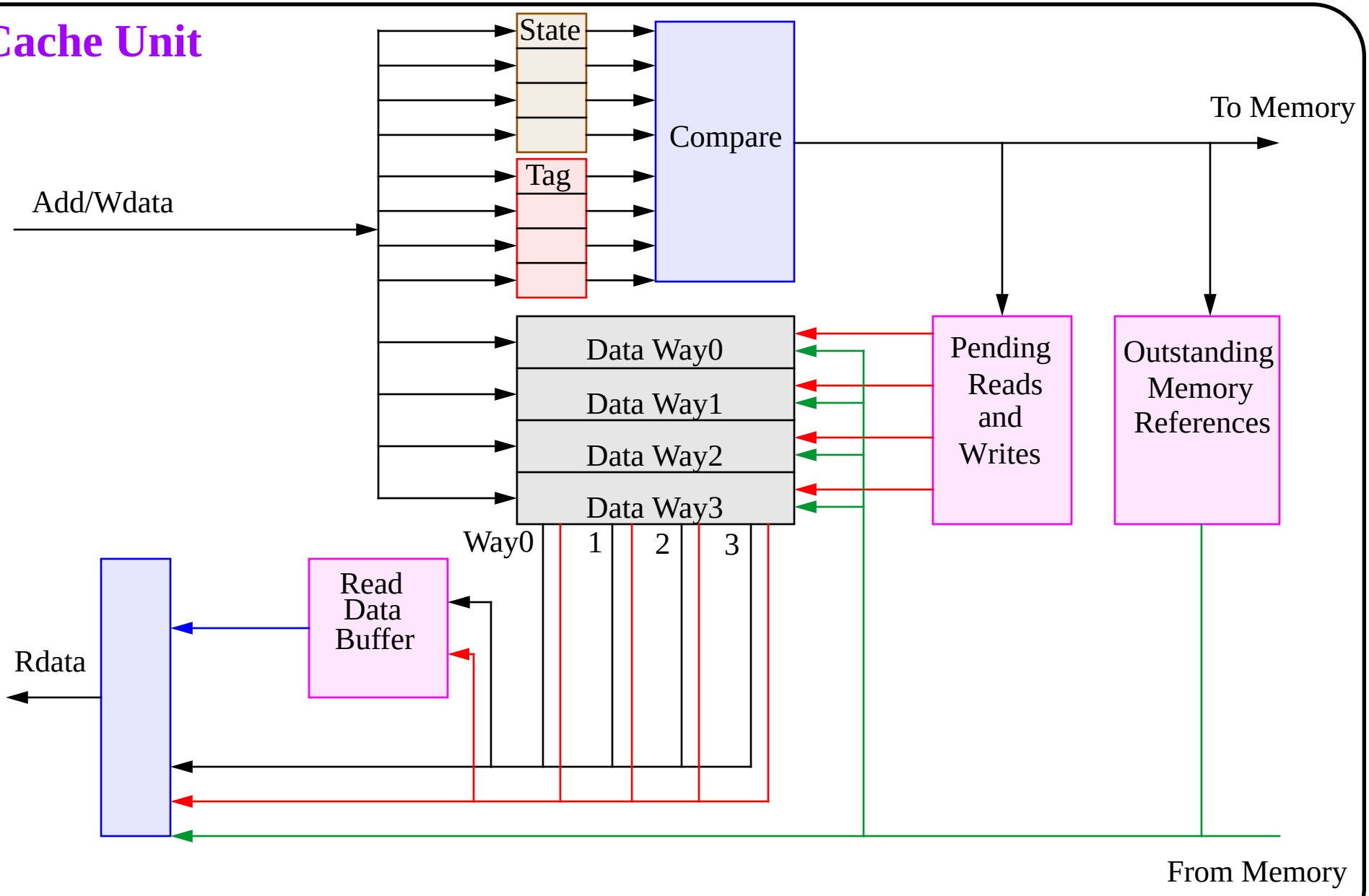
Vector Unit


Vreg
Logical
Viss
1
Iadd
VL
Shift
Scalar Unit
Mem Unit
Rdata
Si/Aj
Rdata
FP Add
Sj/Ak
BMM
Wdata
Sk
FP Mult
Si
Vindex
2nd Logical
VM
FP Recip
P/P/LZ

Streaming Cache Chip

- 128 KB
- Caches instructions, scalar data, and vector data
- 8 bytes per line
- 4-way associative
- 4.0-4.8 GB/sec bandwidth
- Write allocate/write through
- LRU replacement
- Scalar reference prefetch
- Software cache coherence
- Fast cache invalidation
- Parity protection for data and tags
- 192 outstanding references to memory

Cache Unit


State
Tag
Compare
To Memory
Add/Wdata
Data Way0
Data Way1
Data Way2
Data Way3
Pending Reads and Writes
Outstanding Memory References
Way0
1
2
3
Read Data Buffer
Rdata
From Memory

Summary

- Built to run large scale scientific applications
- YMP compatible
- Good vector performance
  - 1.0-1.2 GFLOP/sec and 3.0-3.6 GOP/sec
- High and flexible memory bandwith
  - 8.0-9.6 GB/sec stride independent bandwidth from cache
  - 8.0-9.6 GB/sec stride independent bandwith from memory
- “Off the shelf” technology
- 250-300 Mhz

SiliconGraphics
Computer Systems