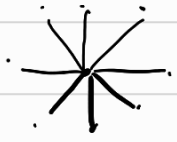
> W5 - SIMD, network topology

1) 128 x 128 data
64 cu
5 ns

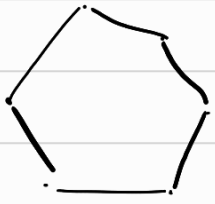
$\frac{128 \times 128}{64} \times 5 = 1280$ ns (C)

2) Vector processing removes overheads of n-1 opcode fetches, decodes and operand loads

Topology

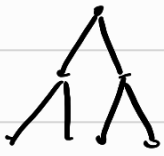


Star
2 hops



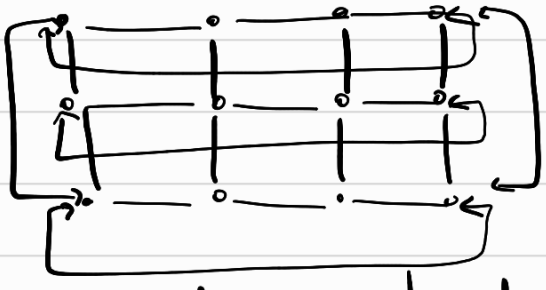
ring
$\frac{n}{2}$ hops

hypercube
$\log_2(n)$ hops



tree
$2\log_2(\frac{n+1}{2})$ hops
= 2 x height

mesh
$2(\sqrt{n}-1)$ hops



torus / wrap around mesh
$\sqrt{n}$ hops

3) - 500 nodes, line topology
- 1000 Mbps links
- 5MB data block in 500 byte packets ↳ 10000 packets
- 0.1 s transmission delay

Best: 1 node: $\frac{5MB\ (bits)}{1000\ Mbps} = 0.04s + 0.1 = 0.14s$

Worst: 499 nodes $= \left(\frac{5 \times 10^6 \times 8}{1000 \times 10^6} + 0.1\right) \times 499 = 69.86s$

4) - Binary tree topology: 31 nodes
- 100 Mbps links
- 0.1 delay
- 5 MB message
- Time for all leaf nodes to send 5MB message?

16 leaf nodes 2^5 height = 5-1, needs 4 links

1 leaf node → root: $\left(\frac{5 \times 10^6 \times 8}{100 \times 10^6} + 0.1\right) \times 4 = 2s$

$16 \times 2 = 32s$

> W6 Cluster Performance

$P = NCFR$

P: performance (M/G) FLOPS
N: no. nodes
C: no. CPUs
F: operations / clock period
R: clock rate

$S = \frac{T(1)}{T(N)} = \frac{1 \text{ processor time}}{n \text{ processor time}}$

scalability

$S \simeq N$ = Good scalability

$E = \frac{P(N)}{N}$ $\frac{\text{Performance}}{\text{processors}}$

$E \simeq 100\%$ good

1) $P = 640$ GFLOPS

$N = 10$

$C = 4$

$F = 8$

$R = ?$

$P = NCFR$

$R = \frac{P}{NCF}$

$= \frac{640}{10 \times 4 \times 8} = 2$ GHz

2)

P_1	P_2
$N = 25$	$N = 25$
$C = 96$	$C = 12$
$F = 16$	$F = 8$
$R = 2.4$ GHz	$R = 2$ GHz

$P = P_1 + P_2$

$P = 96960$ GFLOPS

3)
- Ring w/ 6 nodes
- 1 Gbps links
- 2 MB message / 10 seconds
- 0.004s delay

delay per node for msg $= \frac{2 \times 10^6 \times 8}{1 \times 10^9} + 0.004$

$= 0.02$s

1
2
3
4
5

4) 6 nodes

1 Gbps link

1 MB needs under 0.03s

no delay

$\frac{1 \times 10^6 \times 8}{1 \times 10^9} =$ node delay $= 0.008$s

1: link disconnect = not work ?

2
3
4
5
6

> W7 - Amdahl's law, parallel performance models

$$S = \frac{1}{s + \frac{p}{n}}$$

S = speedup
s = serial component
p = parallel component
n = processors

$s + p = 1$

$s_{comp} + s_{fabric} + p = 1$

- s_{comp}: serial computation
- s_{fabric}: time for computation to move through fabric

if $s_{comp} \ll s_{fabric}$, s is not constant

Delays:
Processing: Middleware + protocols @ transmitting / receiving

Gustafson's Law

$$S = s + np$$

$$X(P) = \frac{P}{R(P) + Z}$$

$$\text{throughput} = \frac{\text{Processors}}{\text{latency} + \text{execution time}}$$

$$\text{speedup} = \frac{\text{old time}}{\text{new time}}$$

4s serial

$4 + 0.319 + \frac{2}{32}$

$= 4.38$ s

$S = \frac{10}{4.38} = 2.28$

4s 95%: $\frac{1}{0.05 + \frac{0.95}{32}} = 12.55$

$\frac{4}{t} = 12.55$ $\quad t = 0.31875$ s