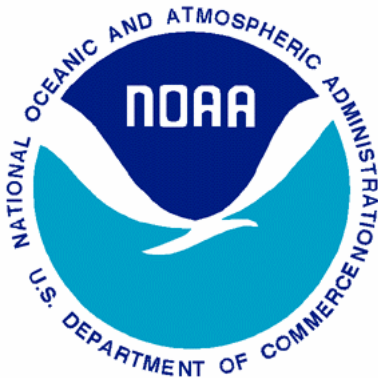# Recent Developments for Parallel CMAQ

Jeffrey Young*
Atmospheric Sciences Modeling Division
NOAA - Air Resources Laboratory
Research Triangle Park, NC
* On assignment to the National Exposure Research Laboratory, U.S. EPA.

David Wong
SAIC – NESCC/EPA

CMAQ Model Peer Review Meeting
R.T.P., NC
December 17, 2003

AQF-CMAQ

- Running in "quasi-operational" mode at NCEP

- 26+ minutes for a 48 hour forecast (on 33 processors)

Standard (May 2003 Release)

DRIVER
  read IC's into CGRID
  begin output timestep loop

    advstep (determine sync timestep)
    couple
    begin sync timestep loop
      SCIPROC
        X-Y-Z advect
        adjadv
        hdiff
        decouple
        vdiff ⟶ DRYDEP
        cloud ⟶ WETDEP
        gas chem
        (aero ⟶ VIS)
        couple
    end sync timestep loop
    decouple
    write conc and avg conc ⟶ CONC, ACONC

  end output timestep loop

DRIVER
  set WORKERs, WRITER
  if WORKER, read IC's into CGRID
  begin output timestep loop

    if WORKER
      advstep (determine sync timestep)
      couple
      begin sync timestep loop
       SCIPROC
         X-Y-Z advect
         hdiff
         decouple
         vdiff
         cloud
         gas chem
         (aero)
         couple
      end sync timestep loop
      decouple
      MPI send conc, aconc, drydep, wetdep, (vis)
    if WRITER
      completion-wait: for conc, write conc ⟶ CONC;
                  for aconc, write aconc ⟶ ACONC; etc.

  end output timestep loop

AQF CMAQ

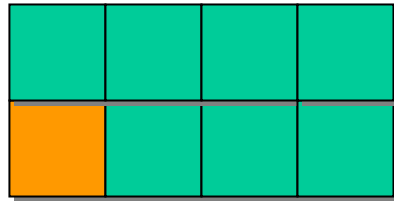Parallel I/O

2002 Release

0   1   2

2003 Release
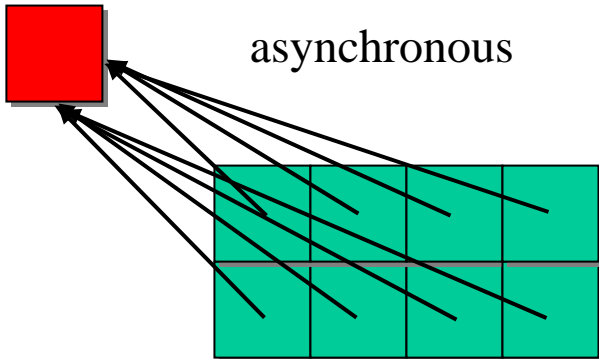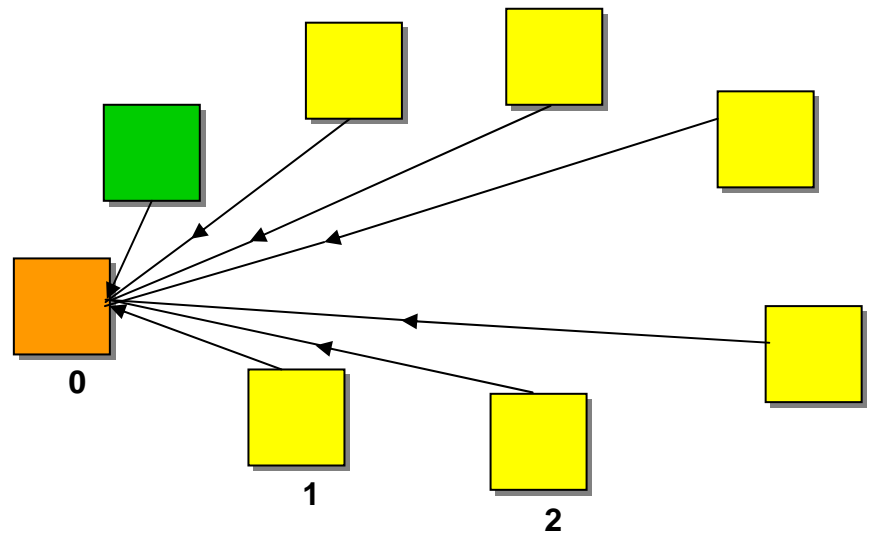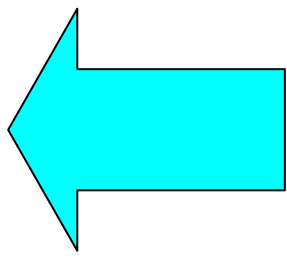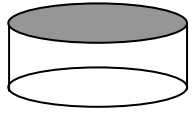
AQF

asynchronous
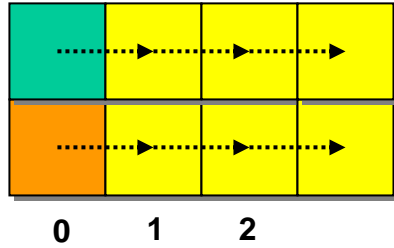
Computation only

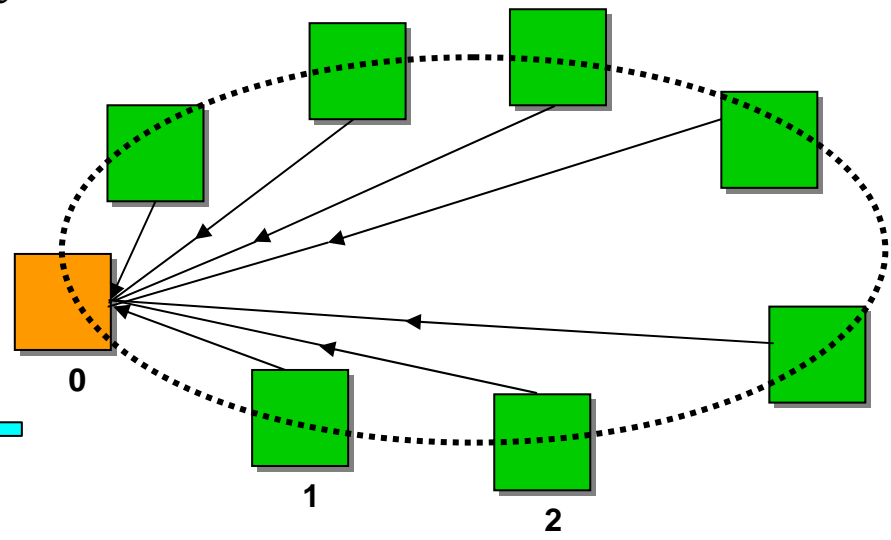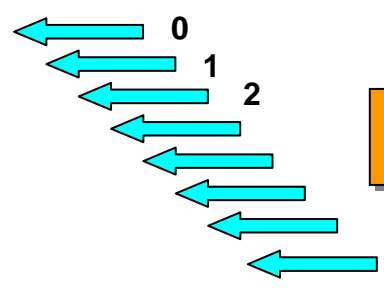Read, write and computation

Read and computation

Write only

Data transfer by message passing

2002 Release

2003 Release

CMAQ-AQF

0

1

2

asynchronous
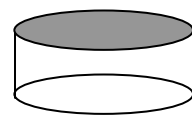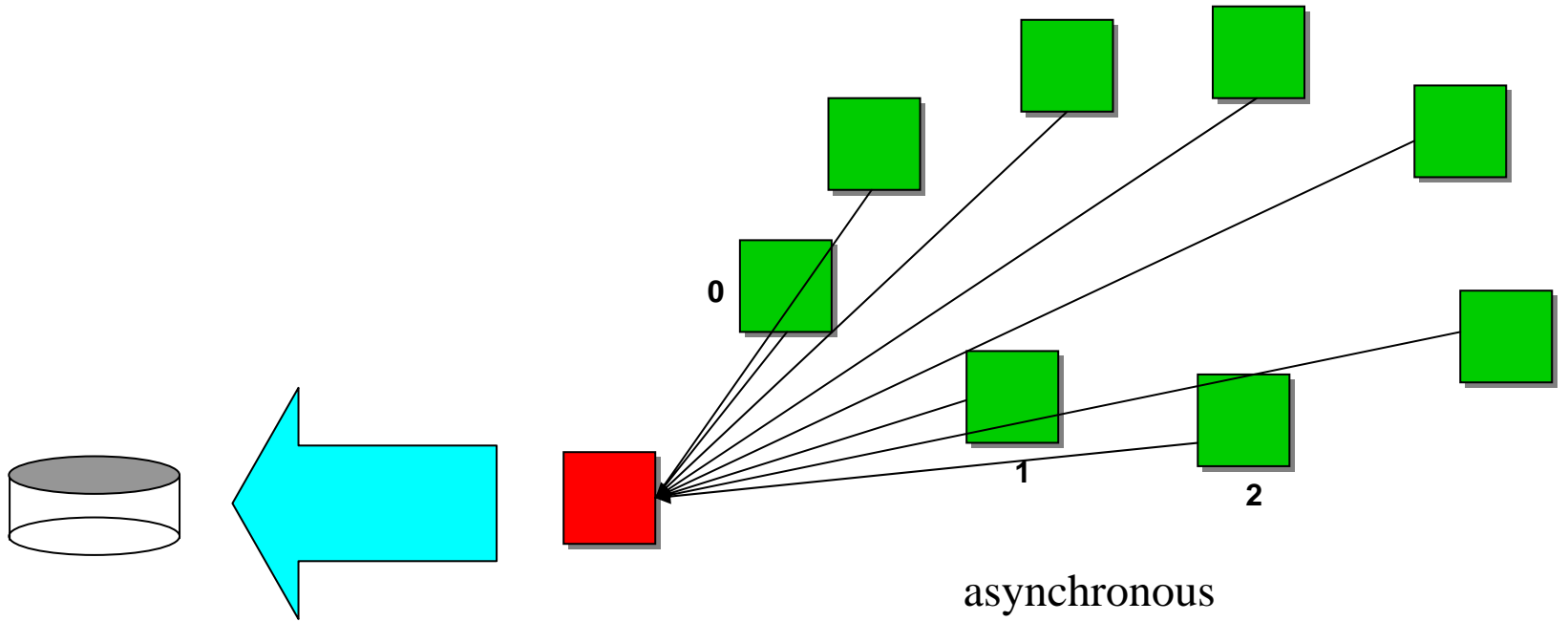
Power3 Cluster (NESCC's LPAR)

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

11	12	13	14	15

11	12	13	14	15

Memory

Switch
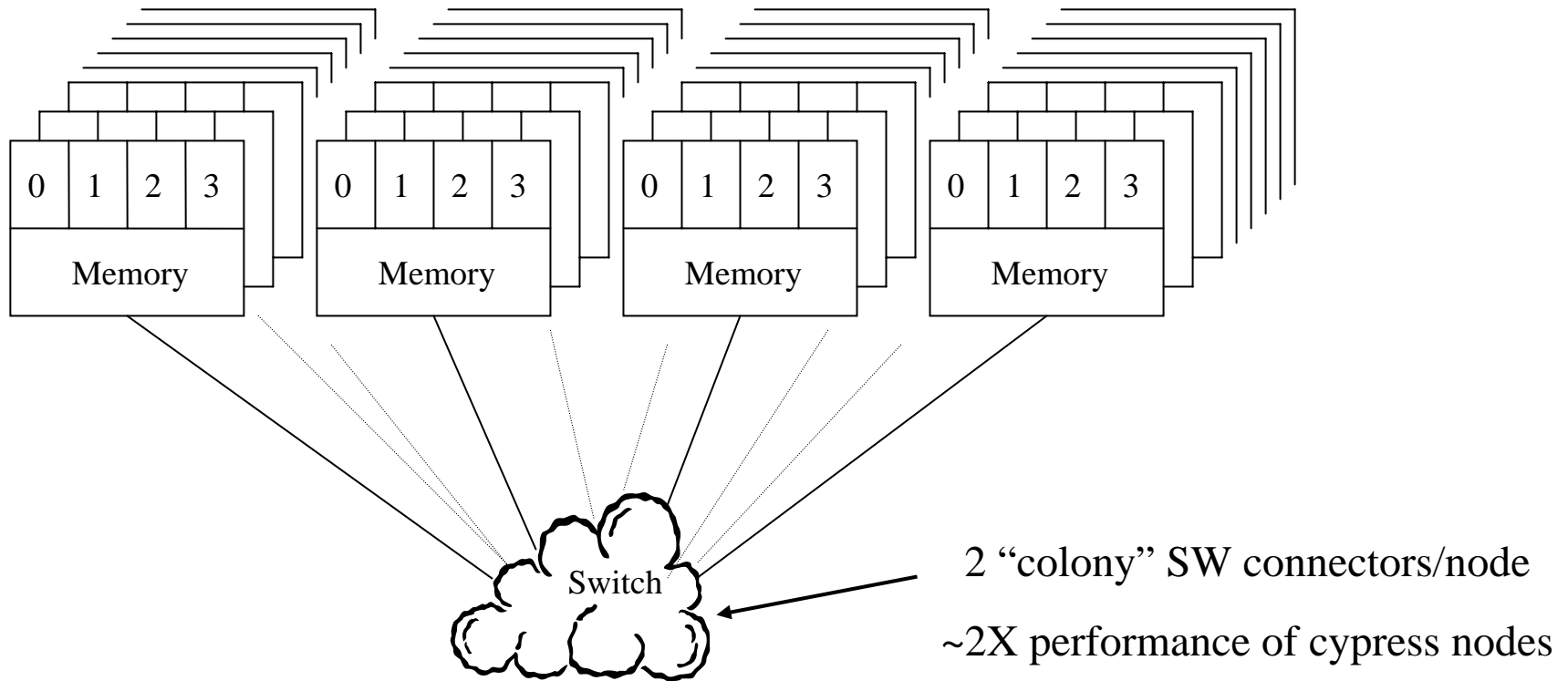
2X slower than NCEP's
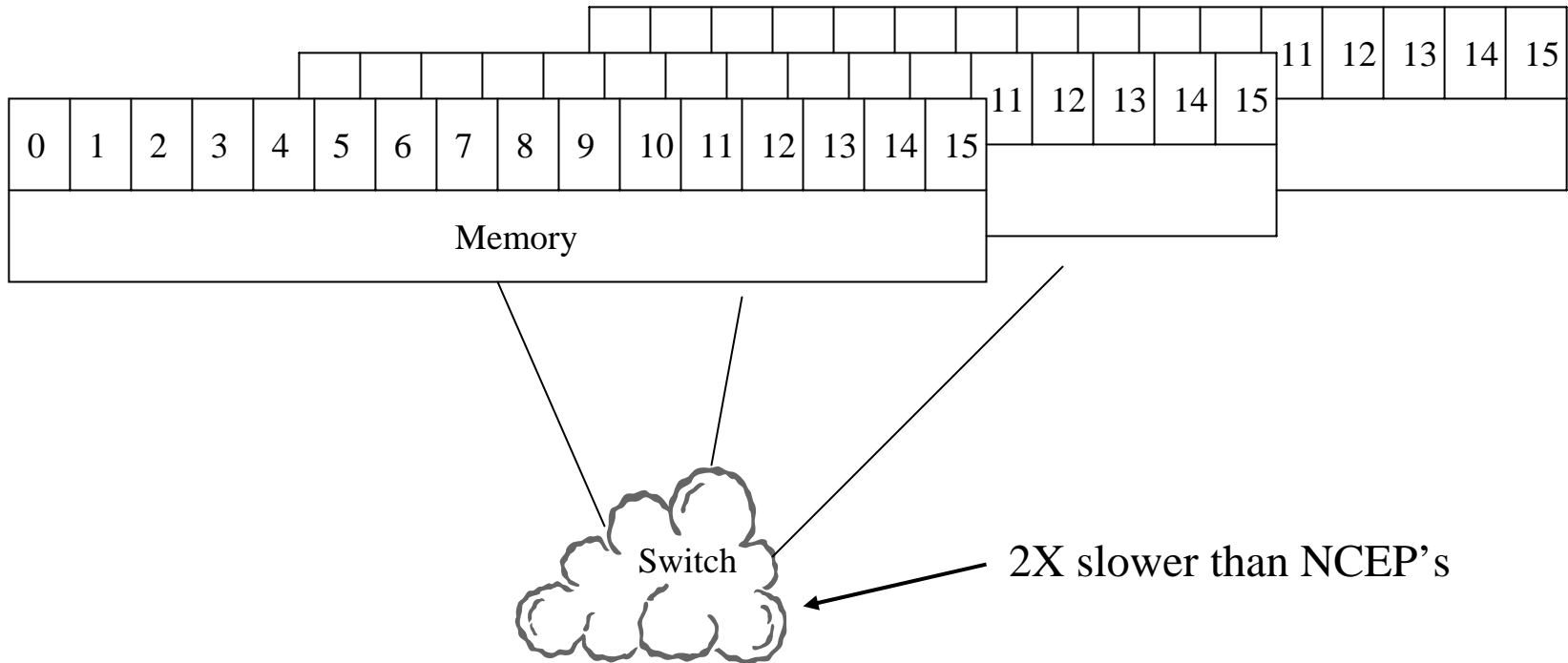
The platform (cypress00, cypress01, cypress02) consists of 3 SP-Nighthawk nodes

All cpu's share user applications with file servers, interactive use, etc.

Power4 p690 Servers (NCEP's LPAR)

0	1	2	3

Memory

0	1	2	3

Memory

0	1	2	3

Memory

0	1	2	3

Memory

Switch

2 "colony" SW connectors/node

~2X performance of cypress nodes

Each platform (snow, frost) is composed of 22 p690 (Regatta) servers
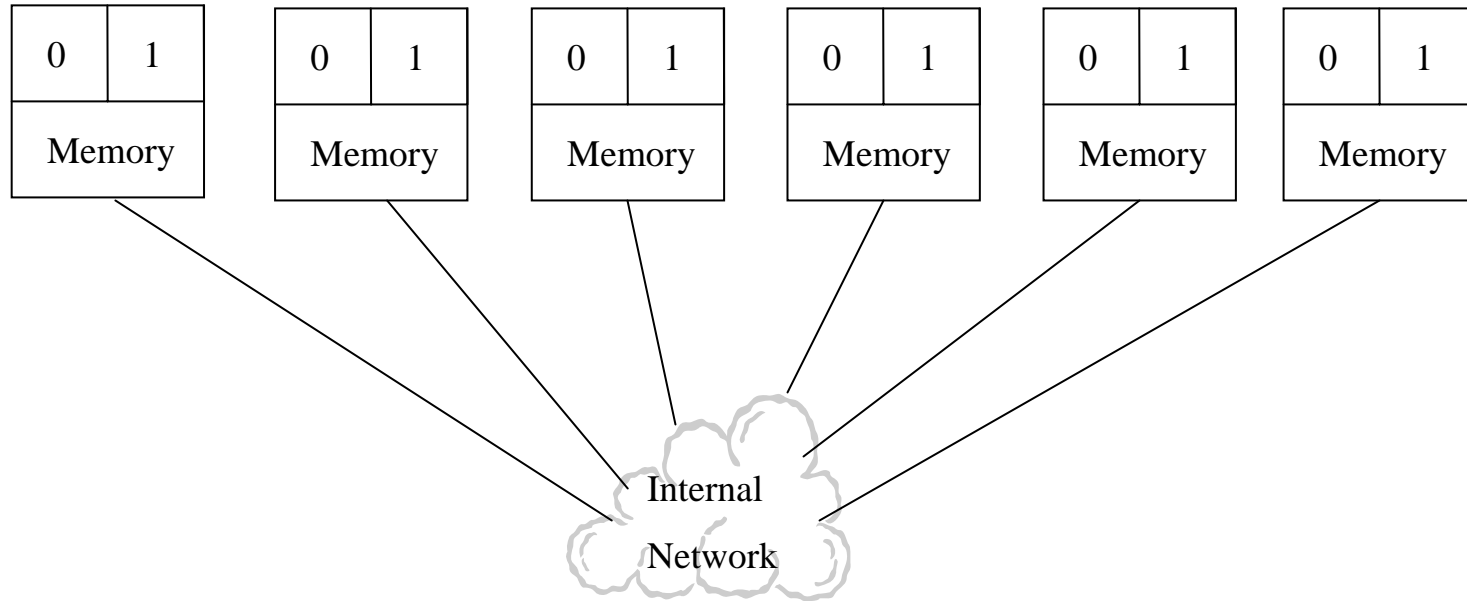
Each server has 32 cpu's LPAR-ed into 8 nodes per server (4 cpu's per node)

Some nodes are dedicated to file servers, interactive use, etc.

There are effectively 20 servers for general use (160 nodes, 640 cpu's)

"ice" Beowulf Cluster

Pentium 3 – 1.4 GHz

0	1
Memory	

0	1
Memory	

0	1
Memory	

0	1
Memory	

0	1
Memory	

0	1
Memory	

Internal
Network

Isolated from outside network traffic

"global" MPICH Cluster

Pentium 4 XEON – 2.4 GHz

global	global1	global2	global3	global4	global5

0	1
Memory	

0	1
Memory	

0	1
Memory	

0	1
Memory	

0	1
Memory	

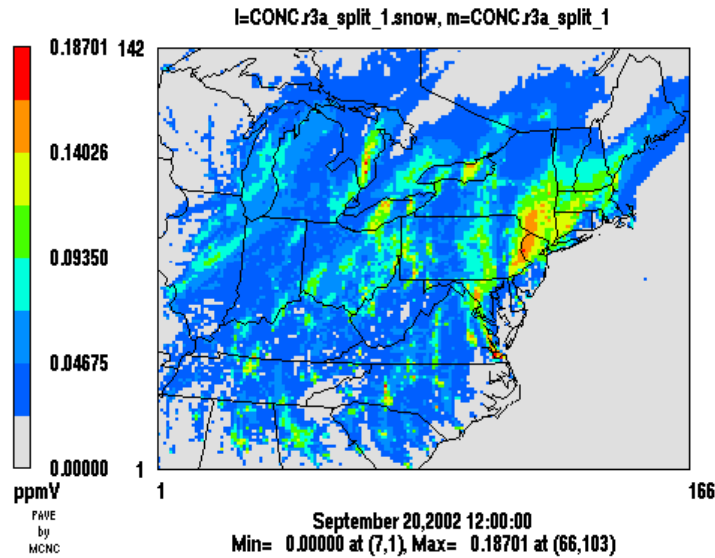| 0	1
Memory	

Network
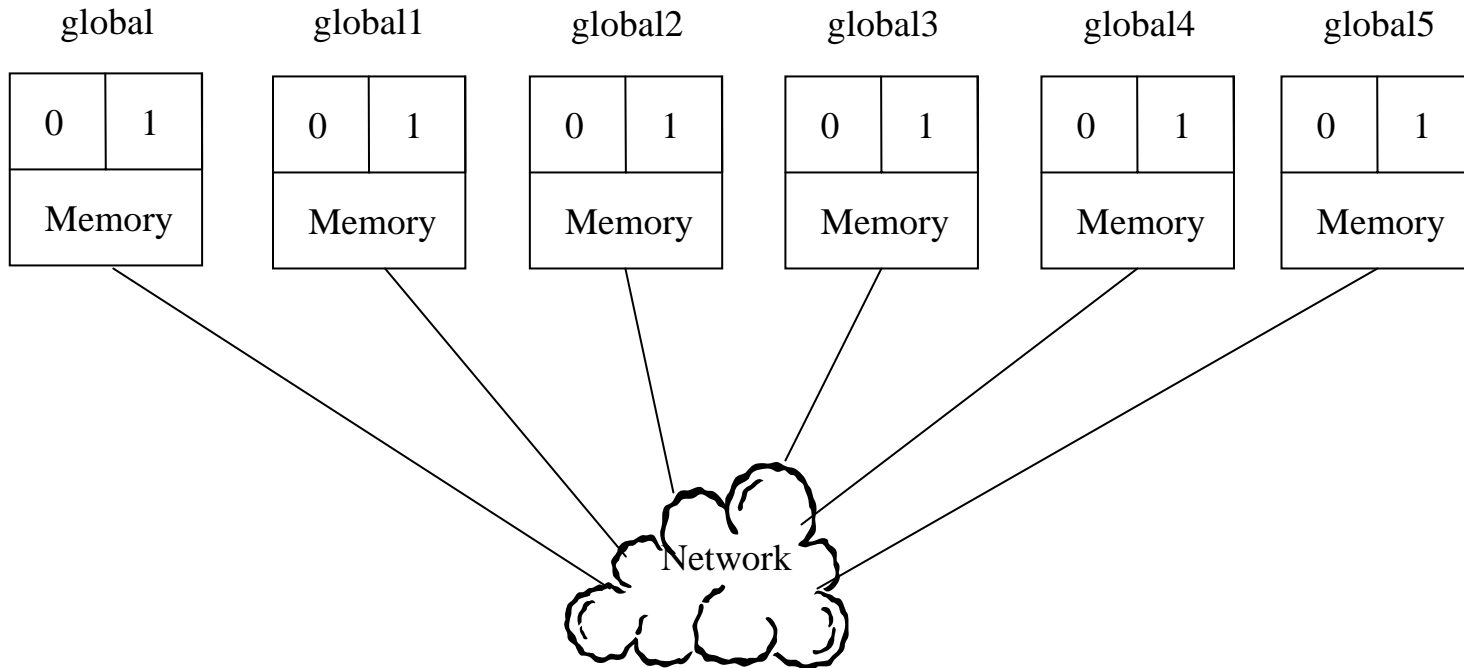
<u>RESULTS</u>

- <u>5 hour</u>, 12Z – 17Z and <u>24 hour</u>, 12Z – 12Z runs

- 20 Sept 2002 test data set used for developing the AQF-CMAQ
  Input Met from ETA, processed thru PRDGEN and PREMAQ

- CB4 mechanism, no aerosols

- Pleim's Yamartino advection for AQF-CMAQ

- PPM advection for May 2003 Release

- 166 columns X 142 rows X 22 layers at 12 km resolution
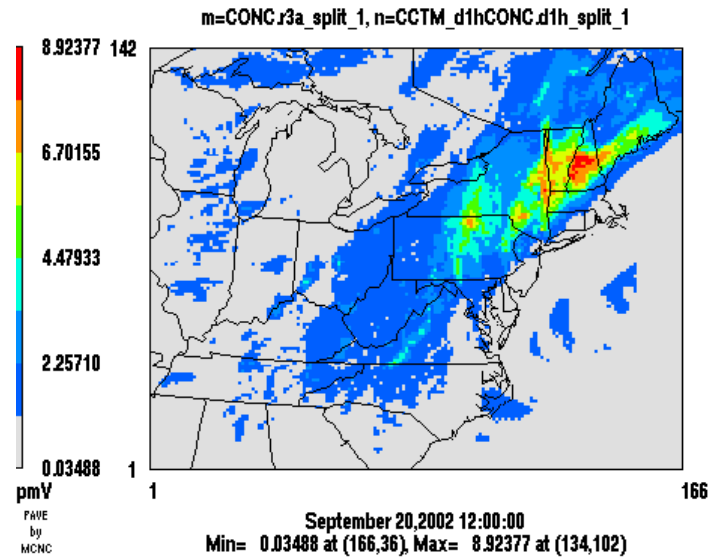
Domain seen on following slide

24–Hour AQF vs. 2003 Release

Layer 1 1000*max(O3l−O3m)

l=CONC.r3a_split_1.snow, m=CONC.r3a_split_1

September 20,2002 12:00:00
Min= 0.00000 at (7,1), Max= 0.18701 at (66,103)

Layer 1 1000*max(O3m−O3n)

m=CONC.r3a_split_1, n=CCTM_d1hCONC.d1h_split_1

September 20,2002 12:00:00
Min= 0.03488 at (166,36), Max= 8.92377 at (134,102)

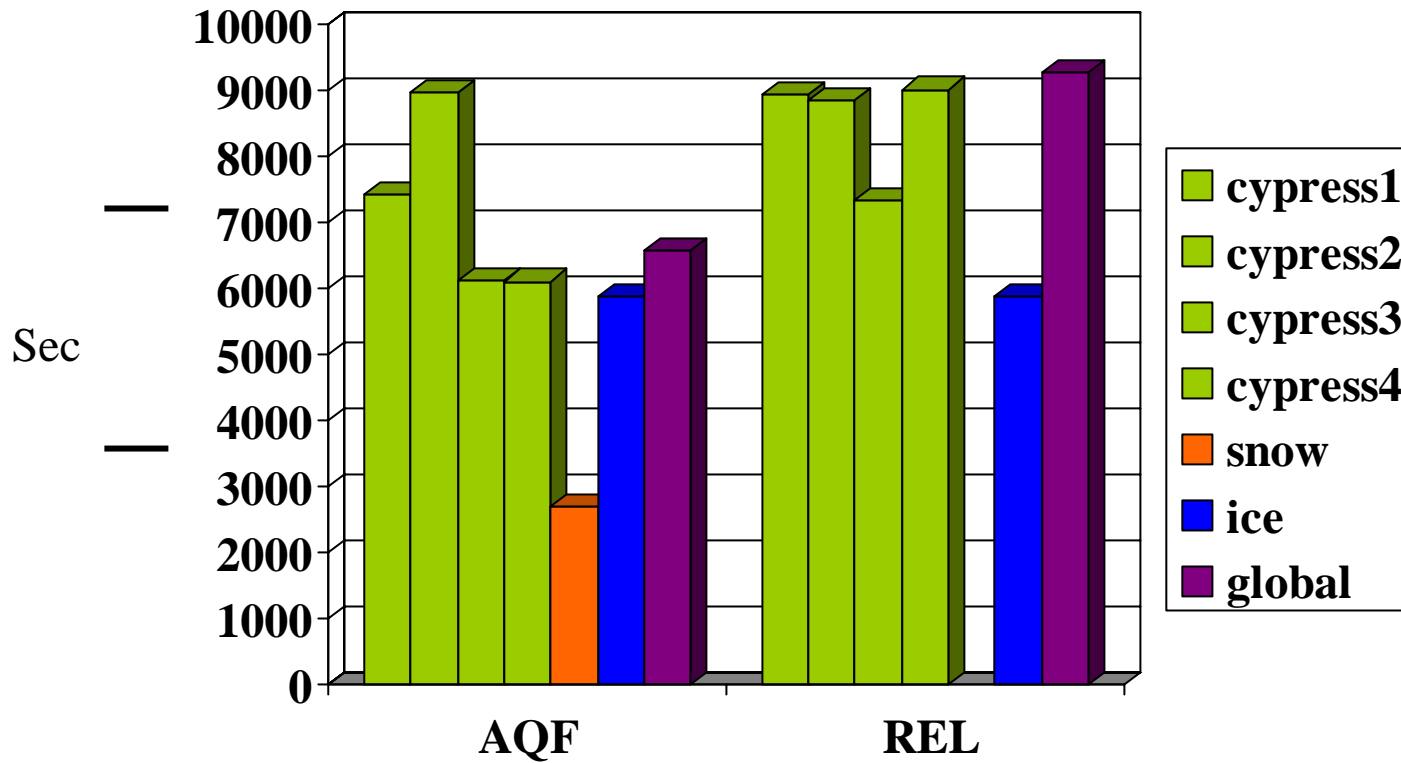Less than 0.2 ppb diff between
AQF on cypress and snow

Almost 9 ppb max diff between
Yamartino and PPM

AQF vs. 2003 Release
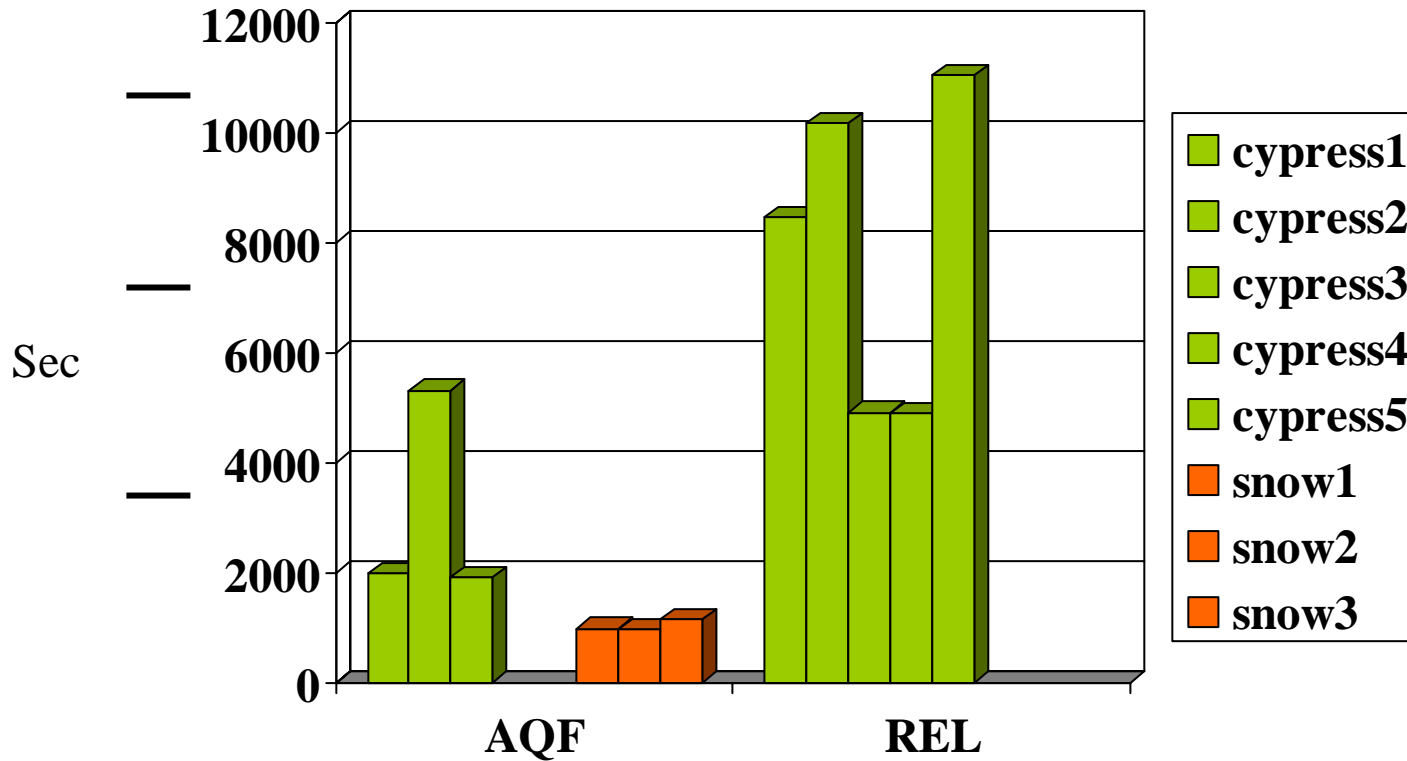
Absolute Run Times

24 Hours, 8 Worker Processors

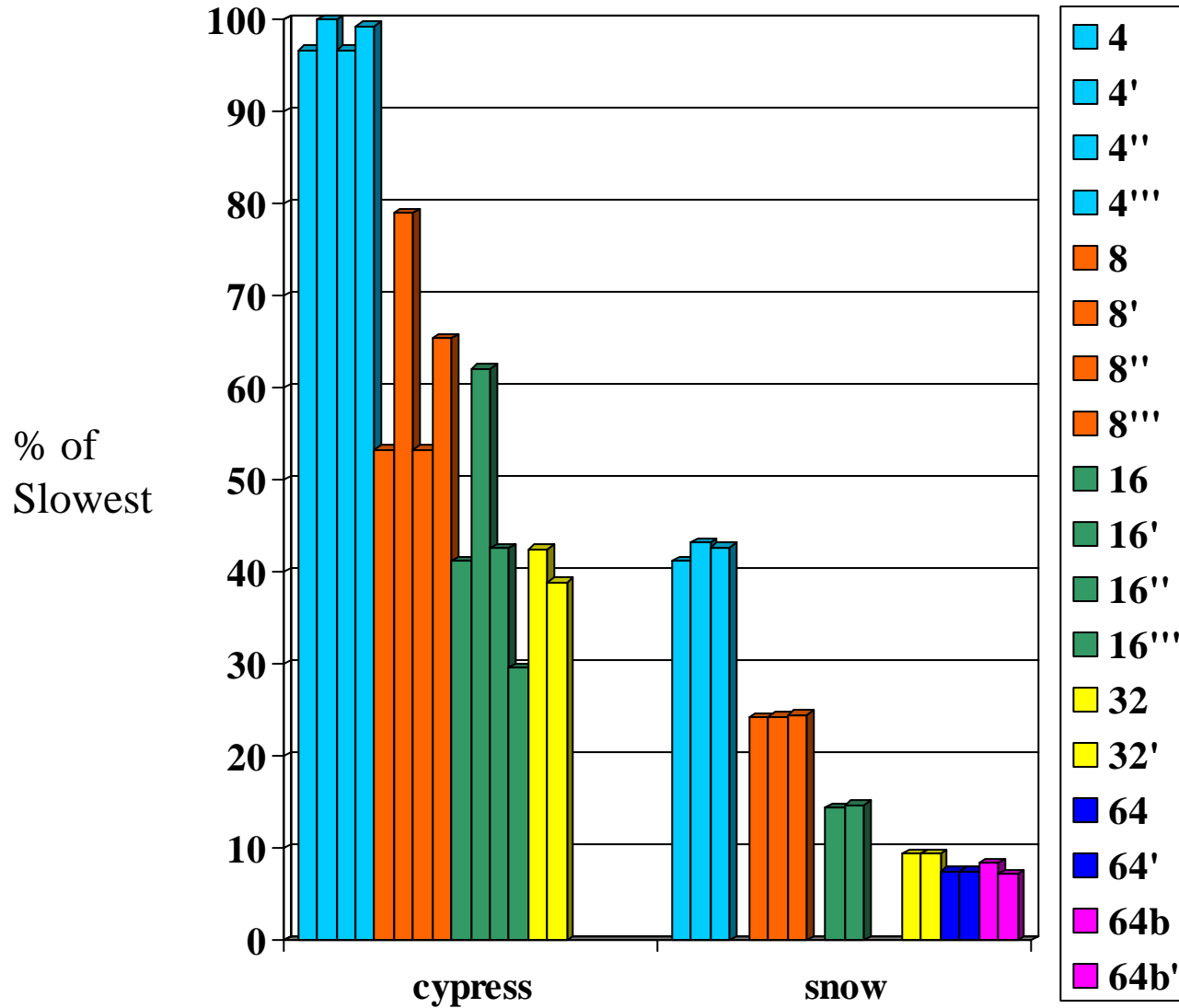# AQF vs. 2003 Release on cypress and AQF on snow

Absolute Run Times
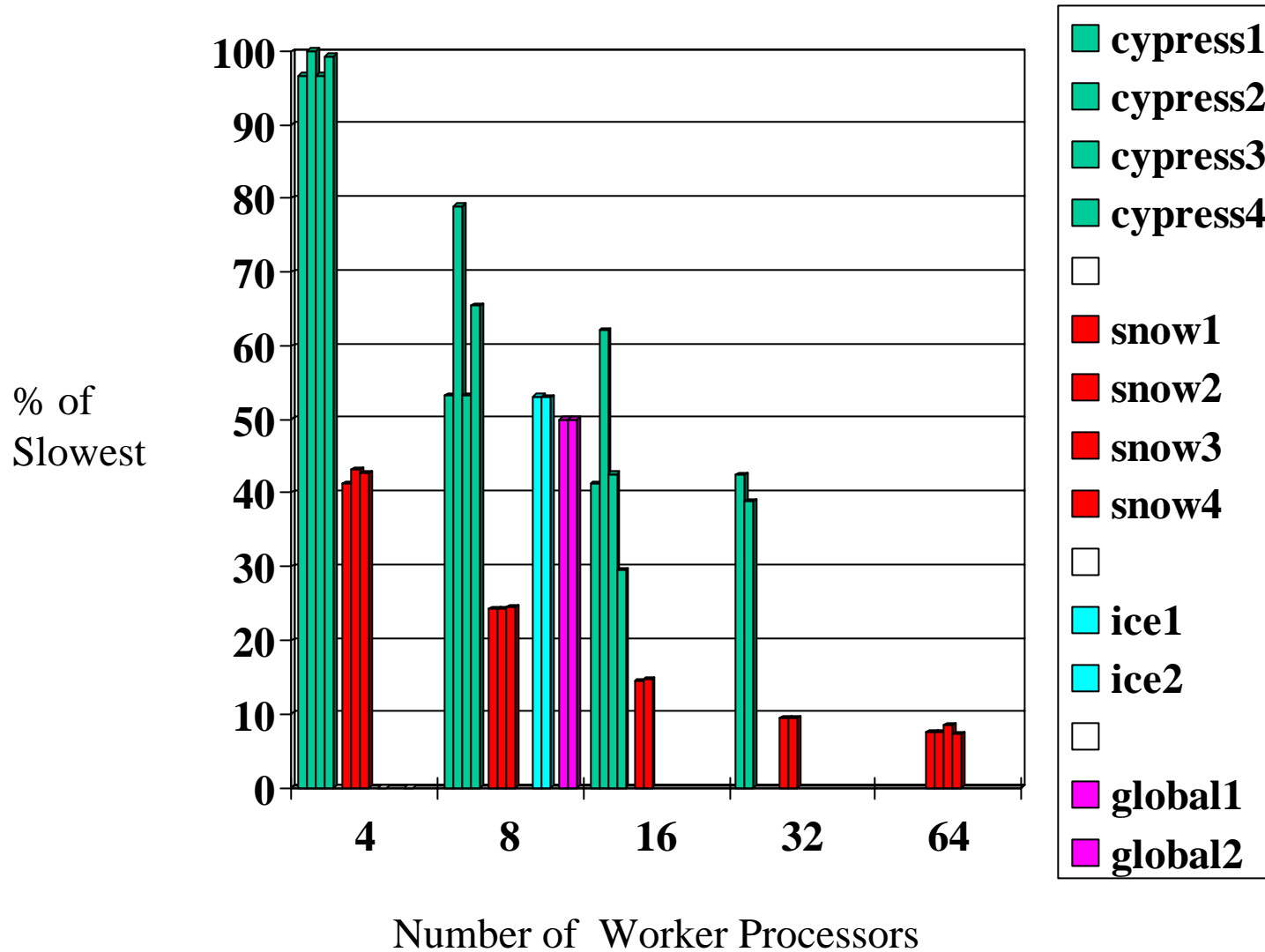
24 Hours, 32 Worker Processors

AQF-CMAQ cypress vs. snow
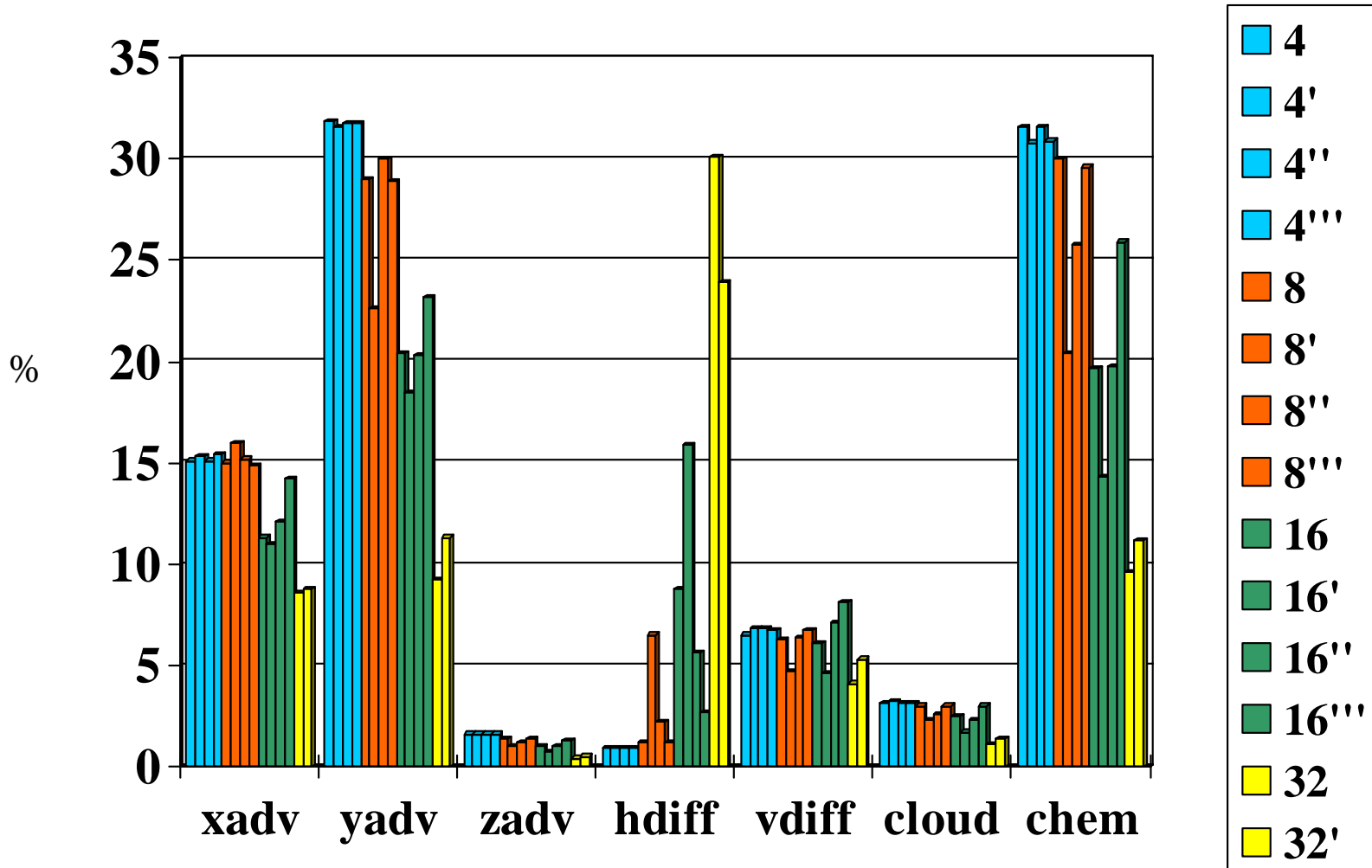
Relative Run Times   5 Hours

% of
Slowest

Legend:
- 4
- 4'
- 4''
- 4'''
- 8
- 8'
- 8''
- 8'''
- 16
- 16'
- 16''
- 16'''
- 32
- 32'
- 64
- 64'
- 64b
- 64b'

cypress   snow

AQF-CMAQ on Various Platforms

Relative Run Times   5 Hours

% of
Slowest

Number of Worker Processors

Legend:
- cypress1
- cypress2
- cypress3
- cypress4
- snow1
- snow2
- snow3
- snow4
- ice1
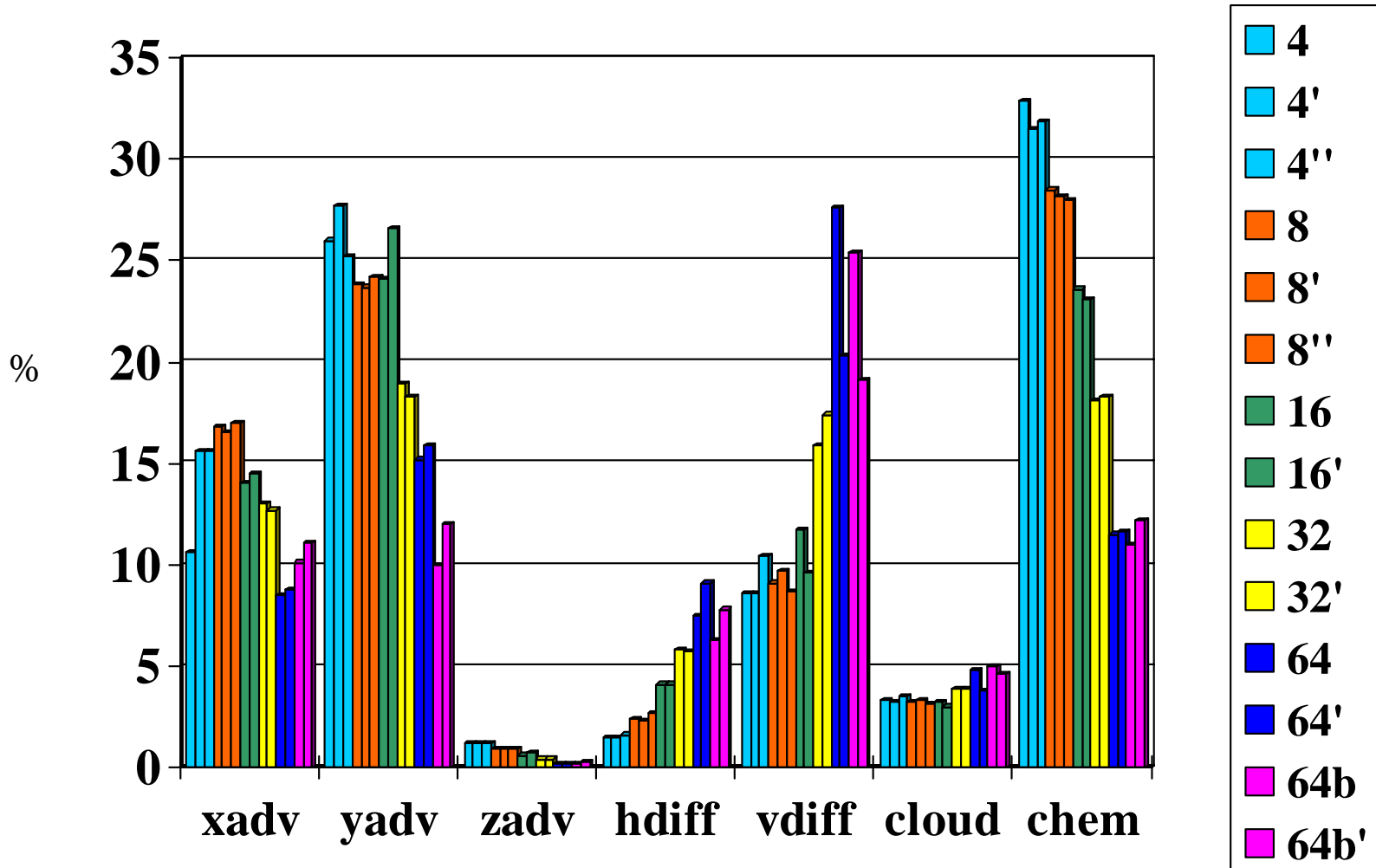- ice2
- global1
- global2

AQF-CMAQ on cypress

Relative Wall Times   5 Hours
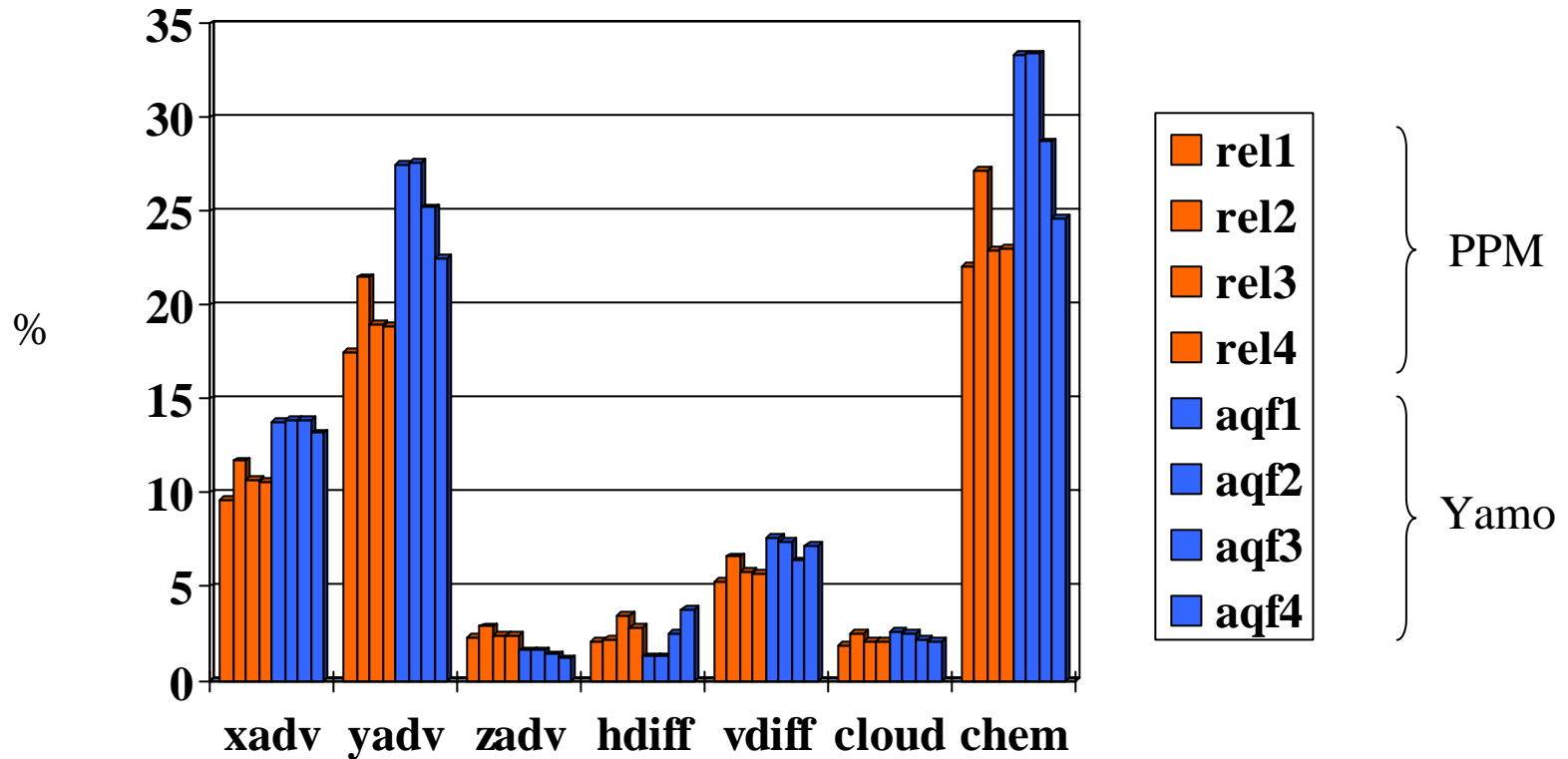
AQF-CMAQ on snow

Relative Wall Times   5 Hours

AQF vs. 2003 Release on cypress

Relative Wall Times

24 hr, 8 Worker Processors

Ongoing Work (CMAQ and AQF-CMAQ

- SNL/DOE has made substantial speedups of the current public release version of CMAQ on their Linux clusters
  - ⇒ I/O improvements
  - ⇒ Improve horizontal advection/diffusion scalability

- Layer variable horizontal advection time steps for Pleim-Yamartino

- TKE vdiff

- Native binary file access layer for I/O-API

- Add aerosols back into AQF_CMAQ

CMAQ Parallel Performance

- 118 X 118 xy grid (2 km), 21 z-layers, 84 species

- 24-hour run, output every 2 hrs

- HP Linux cluster, 3.0 GHz Xeon (dual-proc), 4 Gb/node, Myrinet
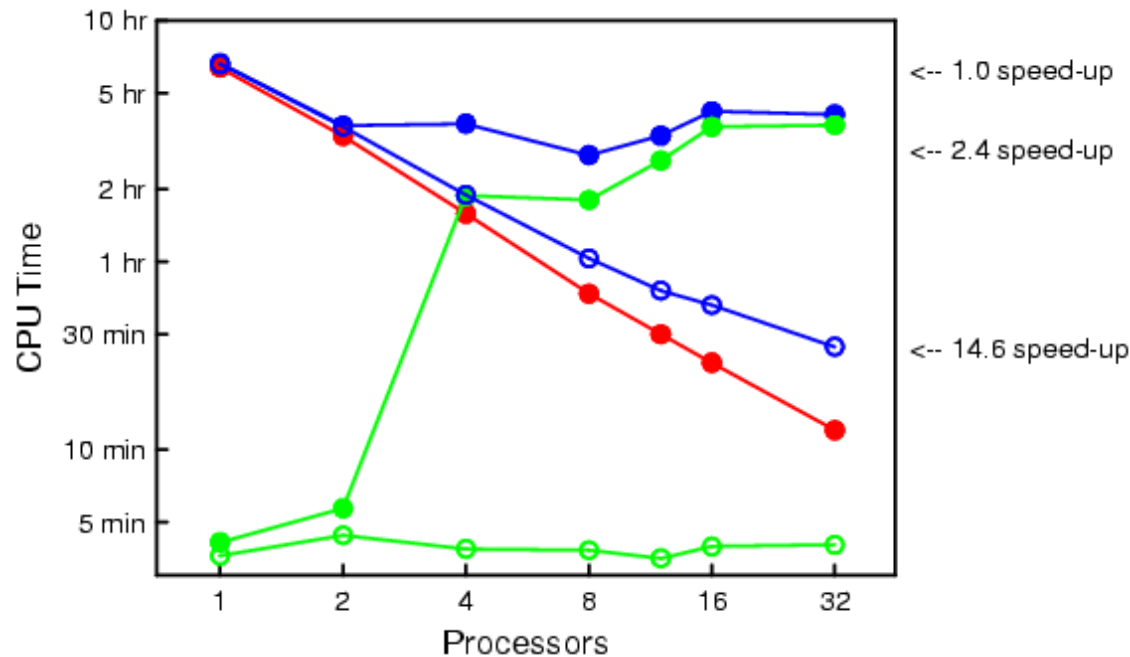
- New version on 1 proc: 1 day in 6.7 hrs   1 year in 102 days
- New version on 32 procs: 1 day in 26 min   1 year in 6.8 days