Parallel IO 2 Update
Kate Thayer-Calder & Jim Edwards
NCAR CSEG
Overview

• What is Parallel IO (PIO) and why do we need it?
• What’s New in PIO2?
• Does PIO2 improve performance?
• Current status, To-Do’s, and CESM2
What is PIO and why do we need it?

- A library to organize data from its structures in memory to the desired structure on disk.

Of course, it’s not that simple…

TPW

PRECT
What is PIO and why do we need it?

- Computational data is distributed across multiple processors, and processes are load balanced to make efficient use of resources.

P1 - Columns 1,2,3,4,5…

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

P2 - Columns 1,2,3,4,5…

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

P3 - Columns 1,2,3,4,5…

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

P4 - Columns 1,2,3,4,5…

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
What is PIO and why do we need it?

- PIO acts as a translator that takes in distributed multi-processor computational data and spits out simple easy-to-write arrays for the NetCDF libraries (and vice-versa for reading in data).
What is PIO and why do we need it?

- And you can imagine how much more complicated this story is at higher and higher resolutions...
What’s new in PIO2?

1) Library rewritten in C to improve portability and flexibility with different software products.
What’s new in PIO2?

2) PIO2 supports an extra method for handling the computational data translation.

**Old (PIO1) Method:** BOX

![Diagram showing the old PIO1 method with processes P1 to P6 connected to IO Task 1 and IO Task 2, with NetCDF library]

*Higher MPI overhead, but fastest netcdf operations. Good for large data, smaller processor counts.*
What’s new in PIO2?

2) PIO2 supports an extra method for handling the computational data translation.

New (PIO2) Method: **SUBSET**

*Less MPI overhead, but more netcdf operations. Good for medium data, large processor counts.*
What’s new in PIO2?

3) PIO is now hosted on GitHub with doxygen-generated documentation and automated nightly tests on CDash.

https://github.com/NCAR/ParallelIO

http://ncar.github.io/ParallelIO/

http://my.cdash.org/index.php?project=PIO
Does PIO2 improve performance?

**IO WRITES**

**IO READS**
Does PIO2 improve performance?

Can we tune IO in CESM to get better performance from PIO2?

- Default B1850 PIO1 run for 1 year: \# simulated years / cmp-day = 6.254
- Default B1850 PIO2 run for 1 year: \# simulated years / cmp-day = 6.210
- Change CPL from default 40 IO Tasks to 5 IO Tasks: \# simulated years / cmp-day = 6.314 (1% improve)
Does PIO2 improve performance?

- Is a straight average across all decomps the correct way to look at this?

- What if most model IO time is spent on a single large or complex IO operation - it would be best to optimize only that one.

- Tune by decomp size or complexity?

- Currently adding custom IO timers to CAM to see which decomps are related to the highest IO cost.
Current Status
PIO2 and CESM2

• PIO2 is an option in the latest versions of CIME, but PIO1 is still the default.

• Fixed a few problems that were causing alpha & beta tests to fail with PIO2 but still have some outstanding issues with CLM.

• Planned to be included in CESM2.

• If you’re interested in doing some experiments in PIO2, contact me and I can help you get started: katec@ucar.edu
Thank you! Questions?
How to tune PIO2

- PIO_VERSION=2 (env_build.xml)
- {ATM|CPL|GLC|ICE|LND|OCN|ROF|WAV} = ###
- ###_PIO_NUMTASKS
- ###_PIO_REARRANGER (1=box, 2=subset)
- ###_PIO_ROOT (root PE)
- ###_PIO_STRIDE (num tasks btn IO tasks)
- ###_PIO_TYPENAME (netcdf, pnetcdf, netcdf4p, netcdf4c)