E3SM Energy Exascale Earth System Model

Parallel Physics-Dynamics Coupling in an Atmosphere Model

Aaron S. Donahue and Peter M. Caldwell

3rd Workshop on Physics Dynamics Coupling (PDC18)

July 12th, 2018

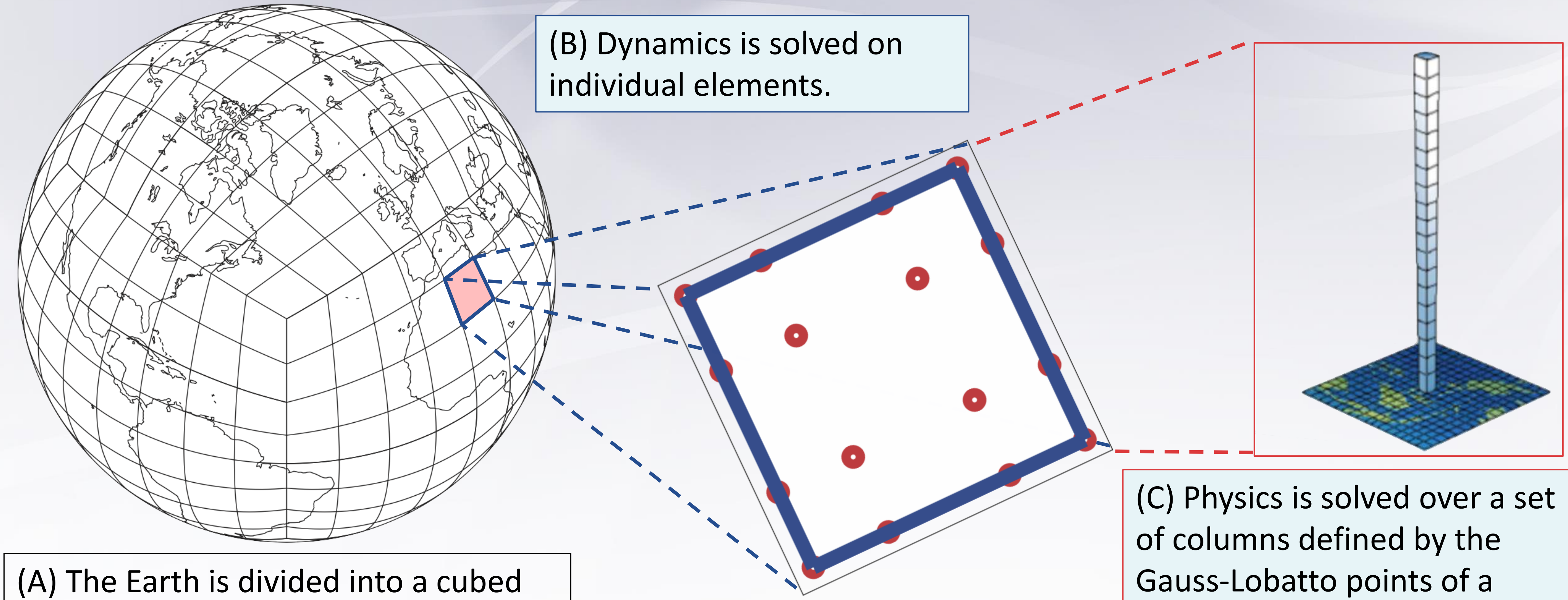
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 IM Release Number LLNL-PRES-754353

PLS AEED U.S. DEPARTMENT OF ENERGY

Outline

- Introduction
- Climate Impact
- Performance Impact
- Stability Impact
- Overstabilization
- Mass Conservation
- Conclusions

Introduction



(B) Dynamics is solved on individual elements.

(A) The Earth is divided into a cubed sphere of quadrilateral elements.

(C) Physics is solved over a set of columns defined by the Gauss-Lobatto points of a spectral element.

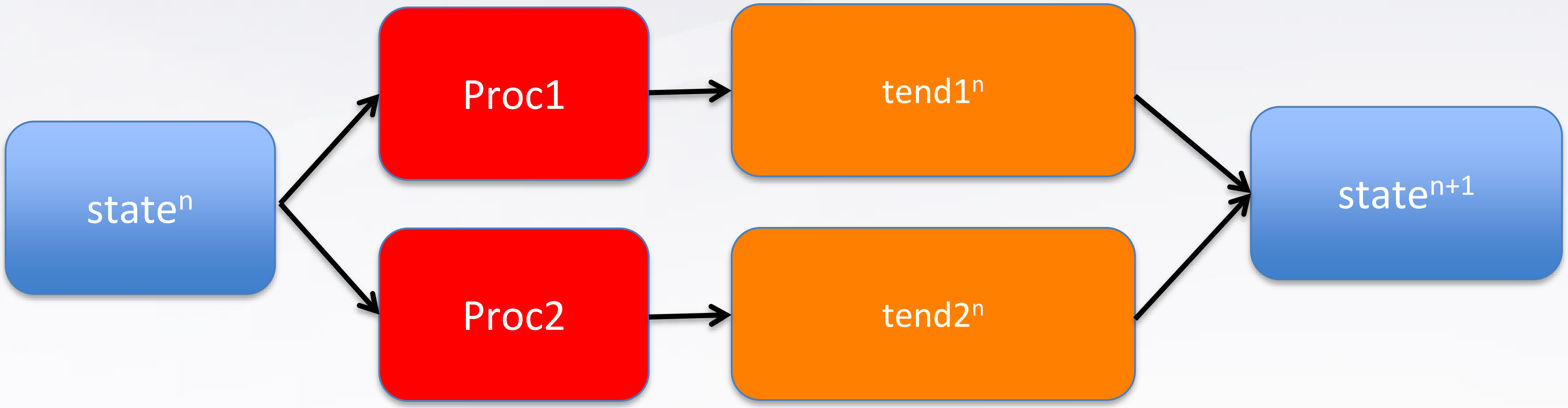
Figure: Dynamics and physics domain for the E3SM atmosphere model. (A) Cubed sphere, (B) example element, (C) example physics column. Image credit: Dennis et al. (2012) Int. J. of High Performance Computing Applications (A and B).

Coupling methods:

<u>Sequential Split (SS, aka time split/fractional steps):</u> State is updated after each process.

Used in E3SM physics and for E3SM physics/dynamics coupling.

<u>Parallel Split (PS , aka process/additive split):</u> All processes are computed from the same state.



Used in E3SM microphysics

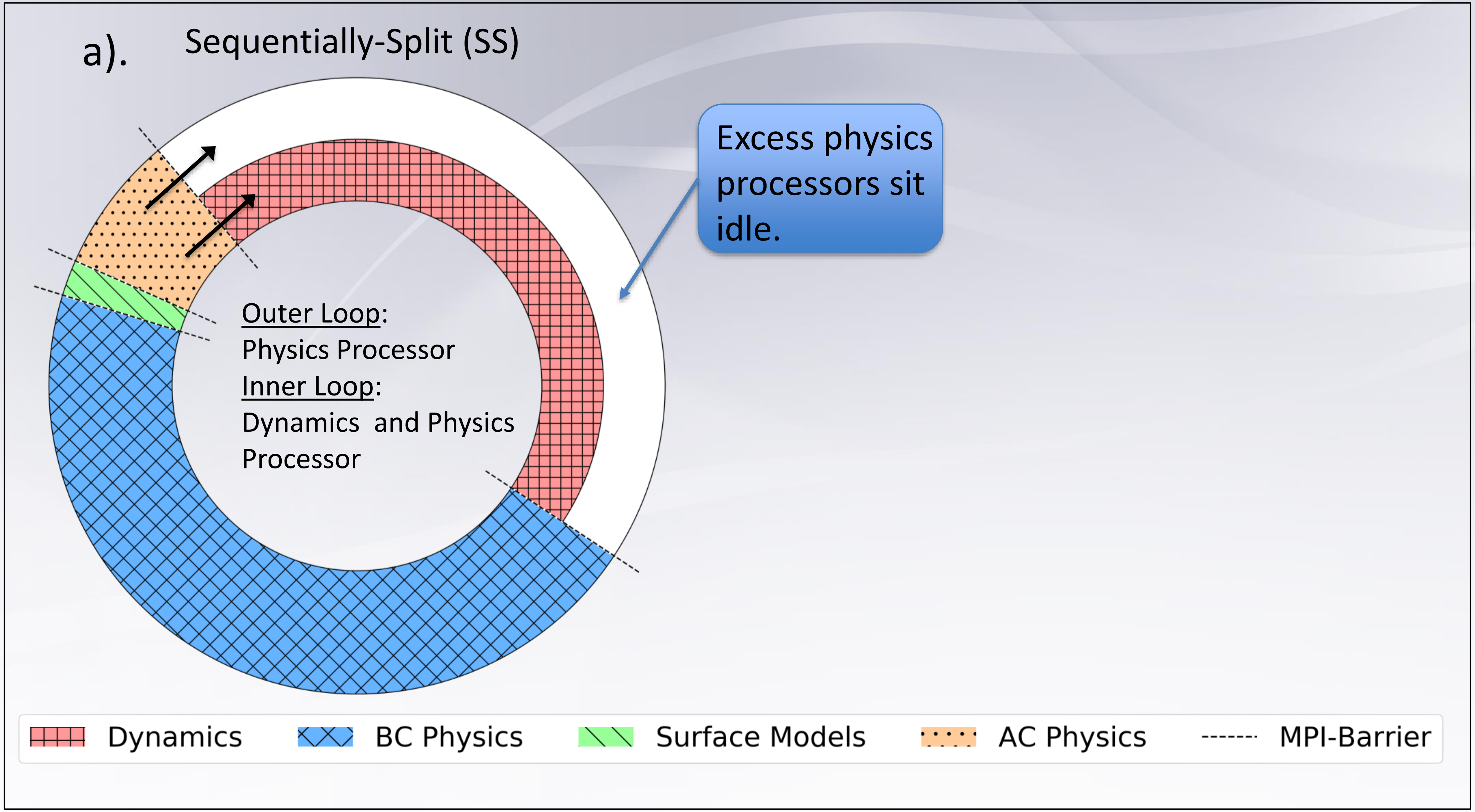


Figure: Order of operations in the E3SM atmosphere model for sequential-splitting (left) and parallel-splitting (right) for a case where more cores than elements are used. The inner loop depicts cores assigned to dynamics, the outer loop is all other cores.

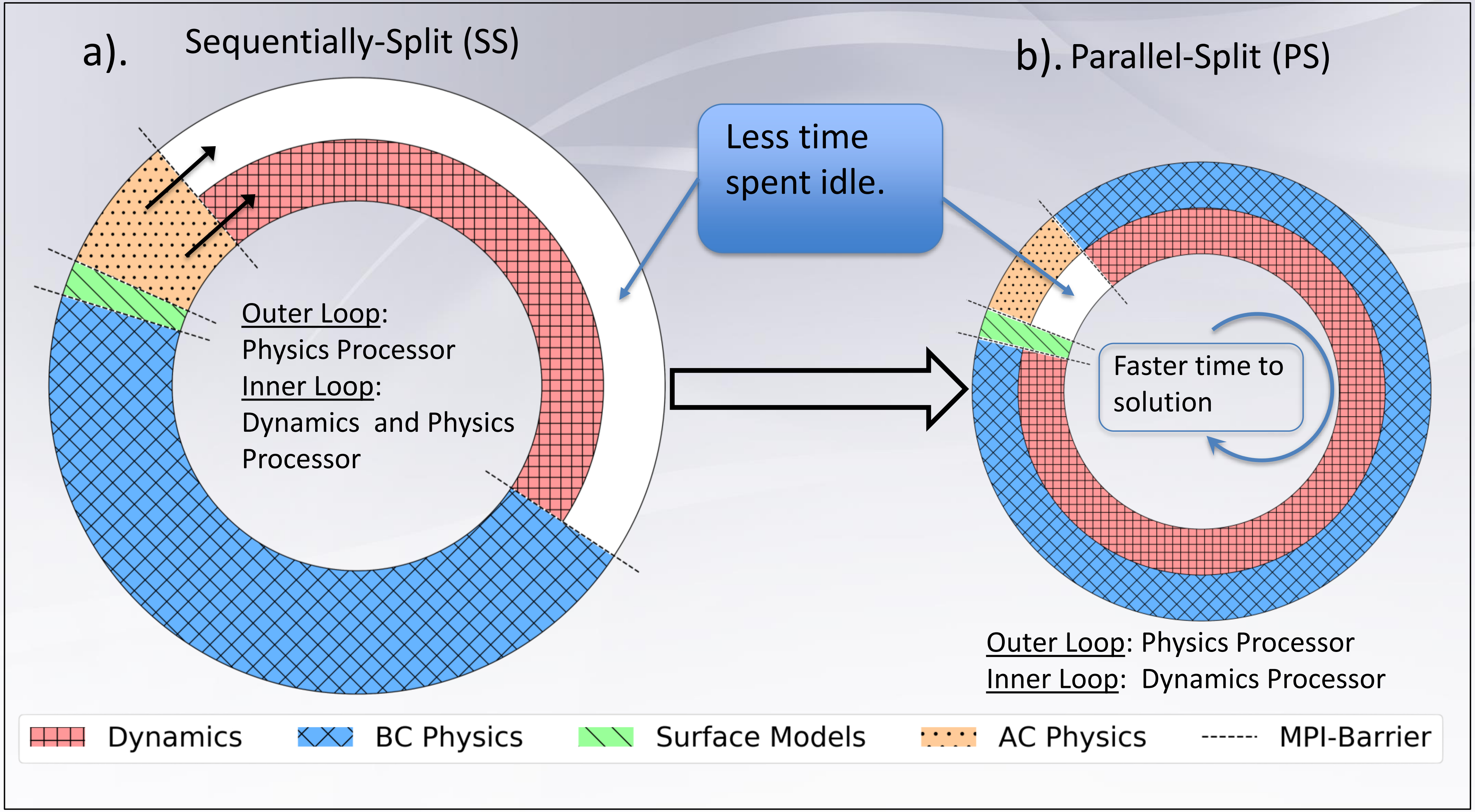


Figure: Order of operations in the E3SM atmosphere model for sequential-splitting (left) and parallel-splitting (right) for a case where more cores than elements are used. The inner loop depicts cores assigned to dynamics, the outer loop is all other cores.

Resolution	# of dynamics elements	# of physics columns
7.5°	96	866
2.7°	726	6536
1.9°	1536	13826
1.0°	5400	48602
0.25°	86400	777602

- The spectral element (SE) **dycore** is **scalable** up to the number of **elements**.
- **Physics scales** up to the total number of **columns** (which is 9x greater than the number of elements).
- Scalability of the model is **limited** to the total number of **elements**.

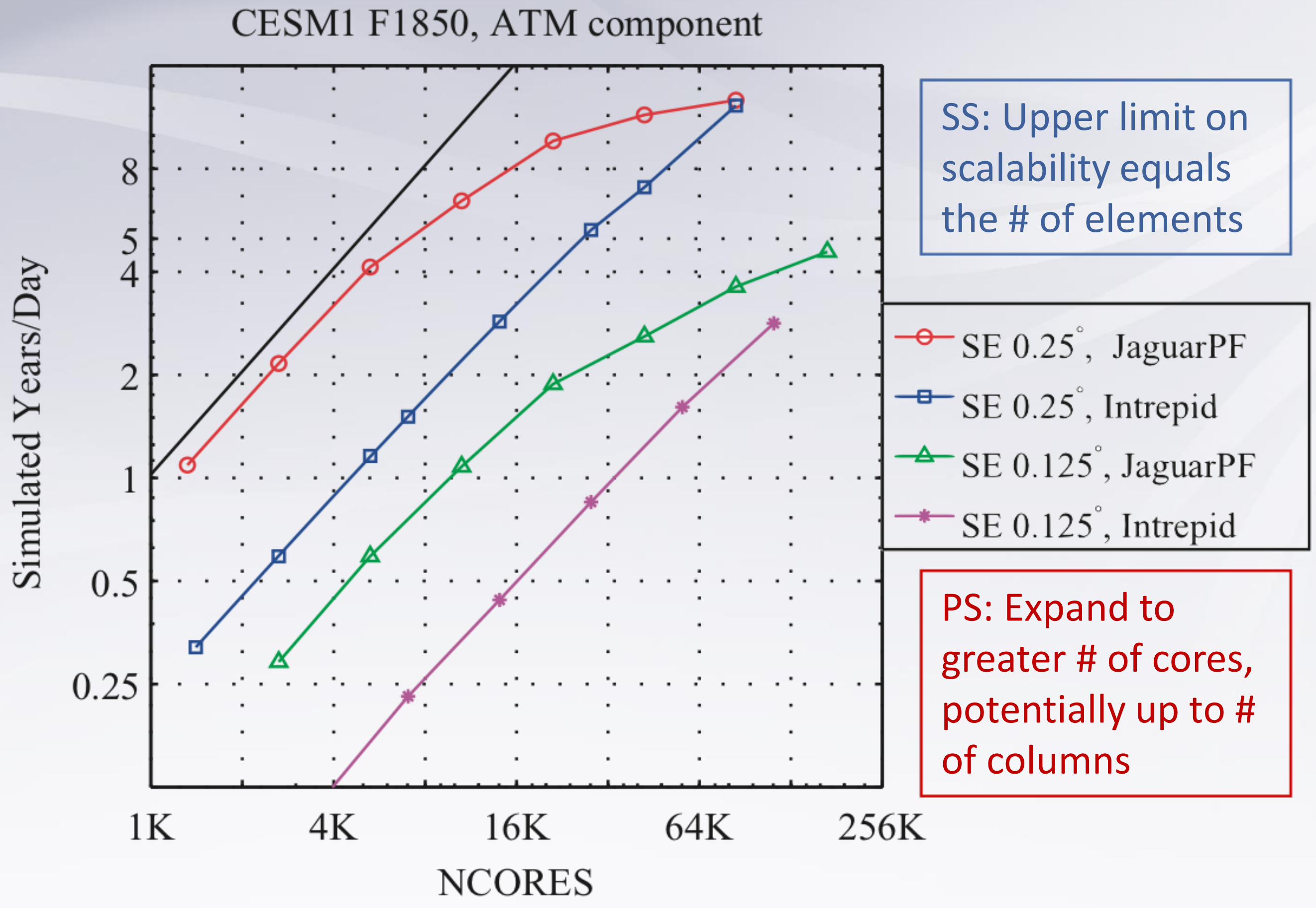
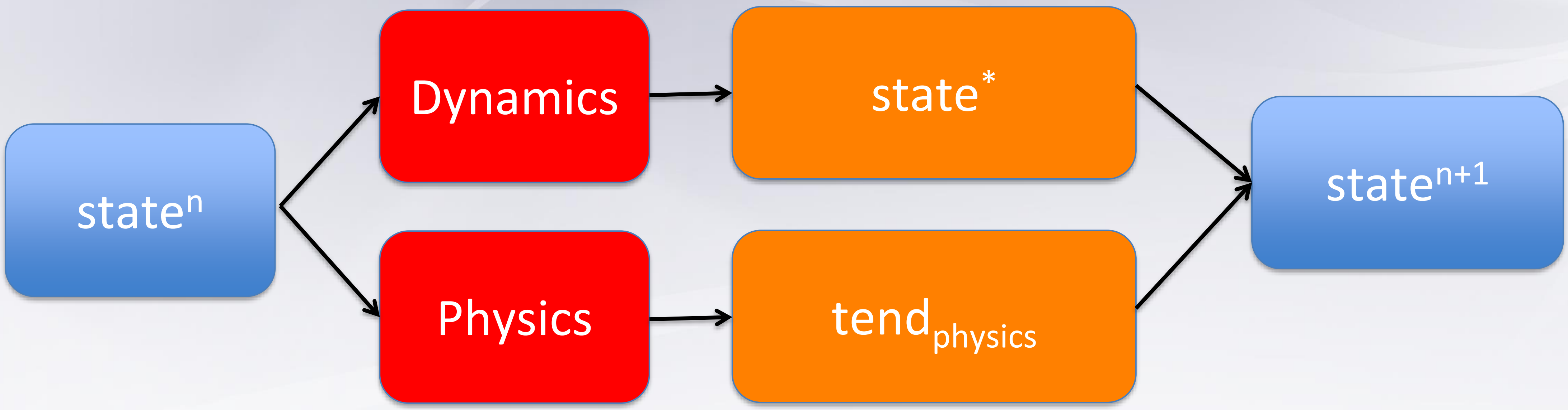


Figure: Scalability of the SE-Dycore. Image credit: Dennis et al. (2012) Int. J. of High Performance Computing Applications.

Parallel-split Implementation:

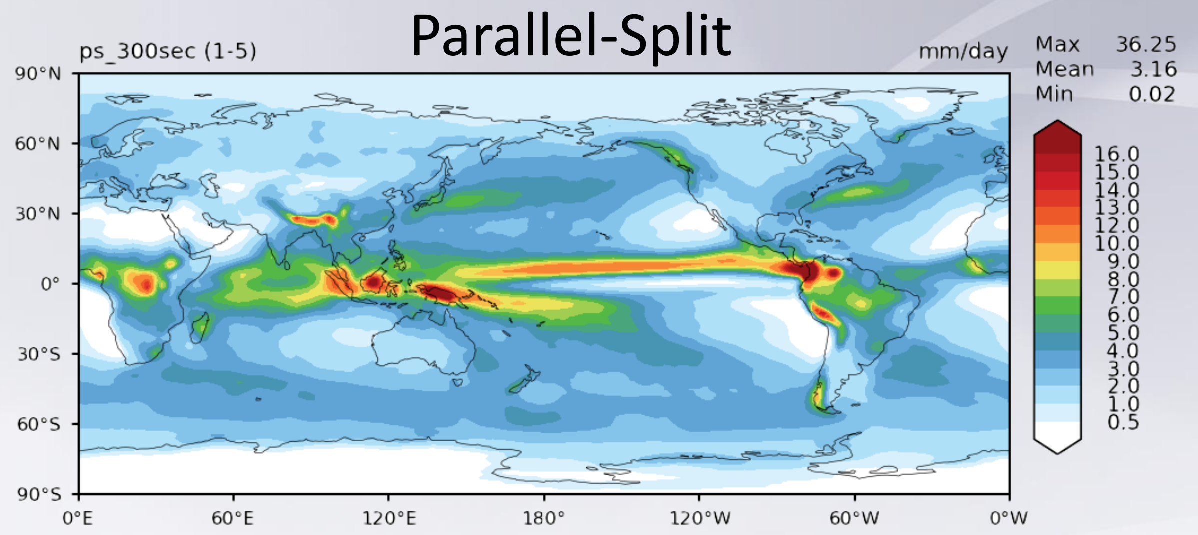


- Implementation is relatively straightforward:
  - Dynamics and Physics are passed the same state
  - Physics produces a tendency and dynamics produces an intermediate state
  - The physics tendency is then used to update the state for the next timestep

Outline

- Introduction
- Climate Impact
- Performance Impact
- Stability Impact
- Overstabilization
- Mass Conservation
- Conclusions

The climate impact of switching to PS is surprisingly small



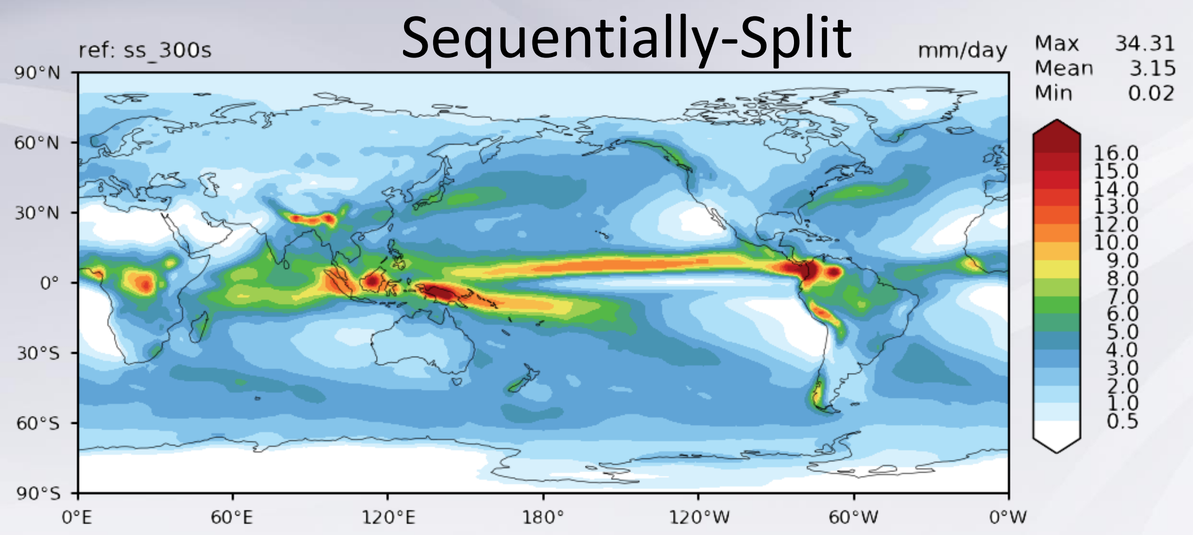
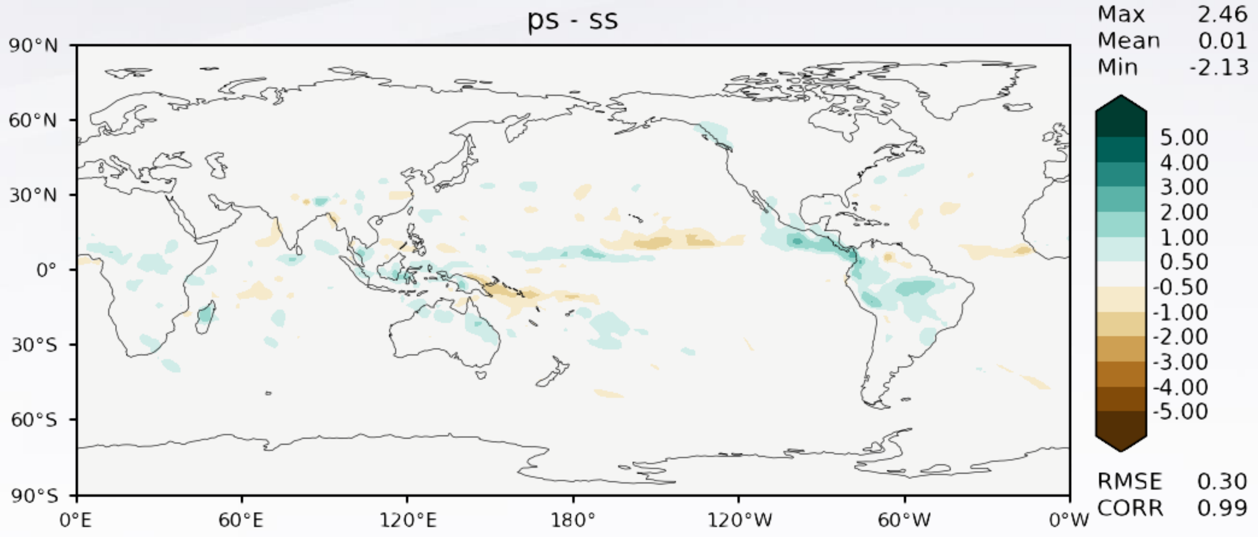
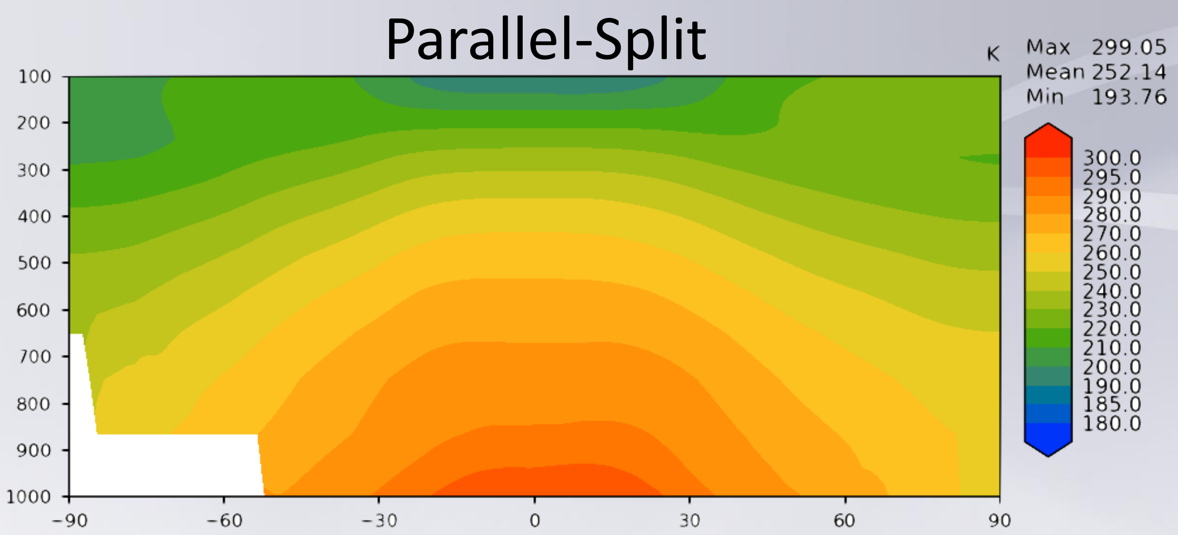
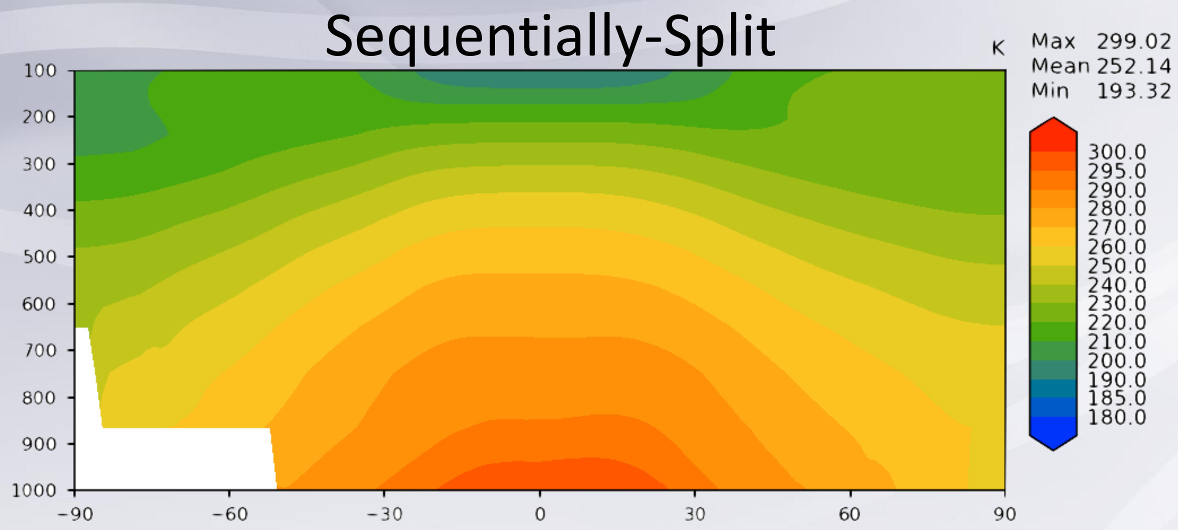
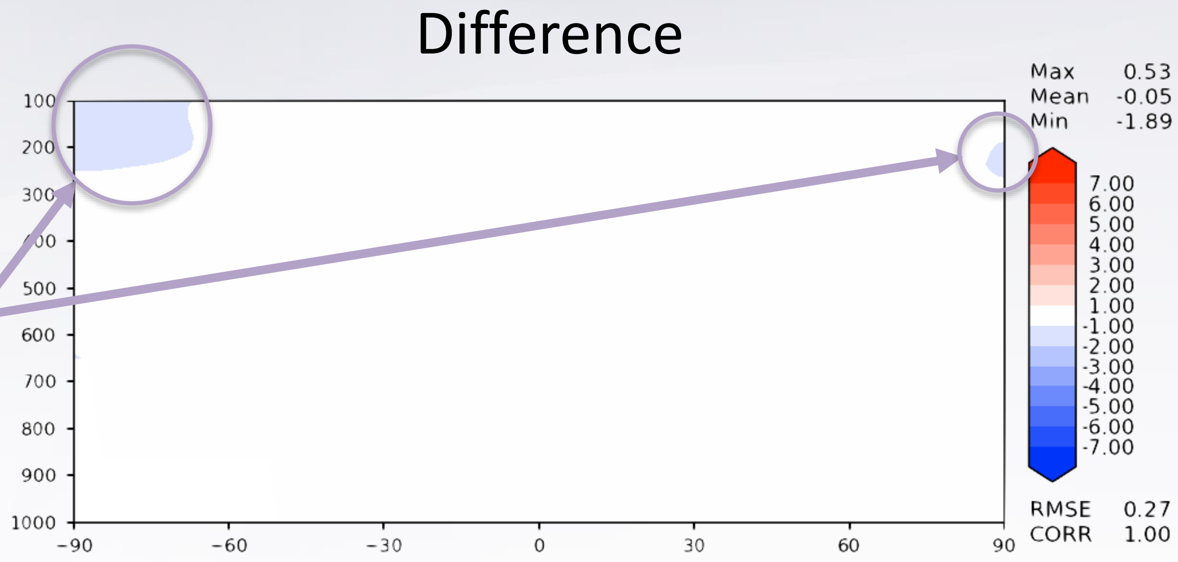


Figure: Annual total precipitation (mm/day) based on 5 year climatologies for parallel-split (top-left), sequentially-split (top-right) simulations, and their difference (bottom), $\Delta t = 300s$.









The most significant differences between the solutions are observed in the temperature at high altitudes (~200mb) and concentrated in the polar regions.

Interannual variability at high latitudes is quite large, so differences seen here may not be significant.

Figure: Annual zonal temperature (K) based on 5 year climatologies for parallel-split (top-left), sequentially-split (top-right) simulations, and their difference (bottom), $\Delta t = 300s$.

Outline

- Introduction
- Climate Impact
- Performance Impact
- Stability Impact
- Overstabilization
- Mass Conservation
- Conclusions

Proper physics/dynamics load balancing extends and application of the PS method extends the scalability of the model.

At the highest core counts communication costs begin to dominate limiting the scalability of either coupling method.

Comparison of parallel- and sequentially- split performance;

- at low core counts communication costs limit the performance of PS,
- at large core counts PS is up to 20-40% faster than sequentially-split.

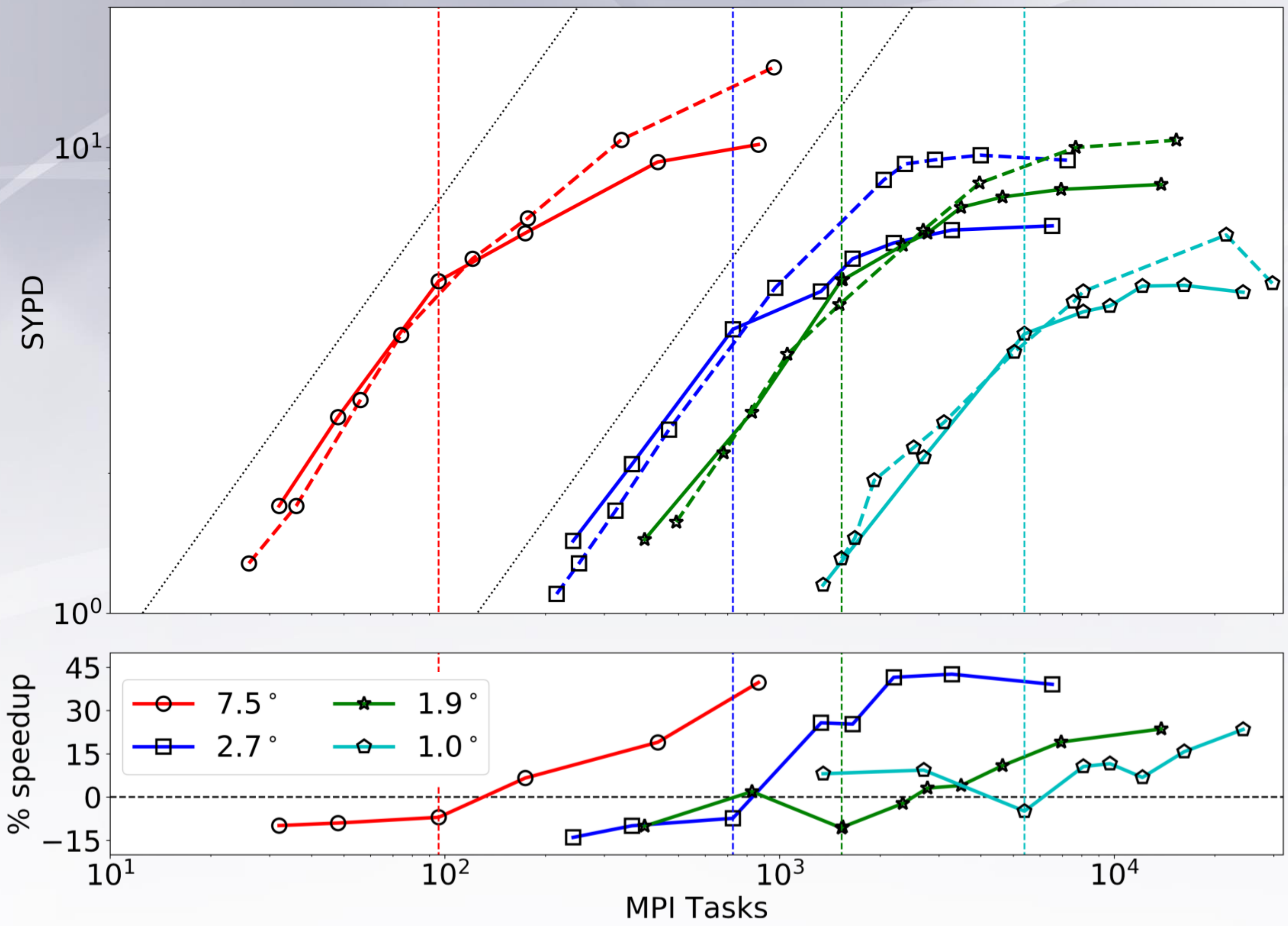


Figure: Scalability of the model and % speedup for parallel-split runs vs. sequentially-split runs. Horizontal dashed lines represent the scalability limit of the spectral element dynamics core. All runs use a timestep of $\Delta t = 300s$.

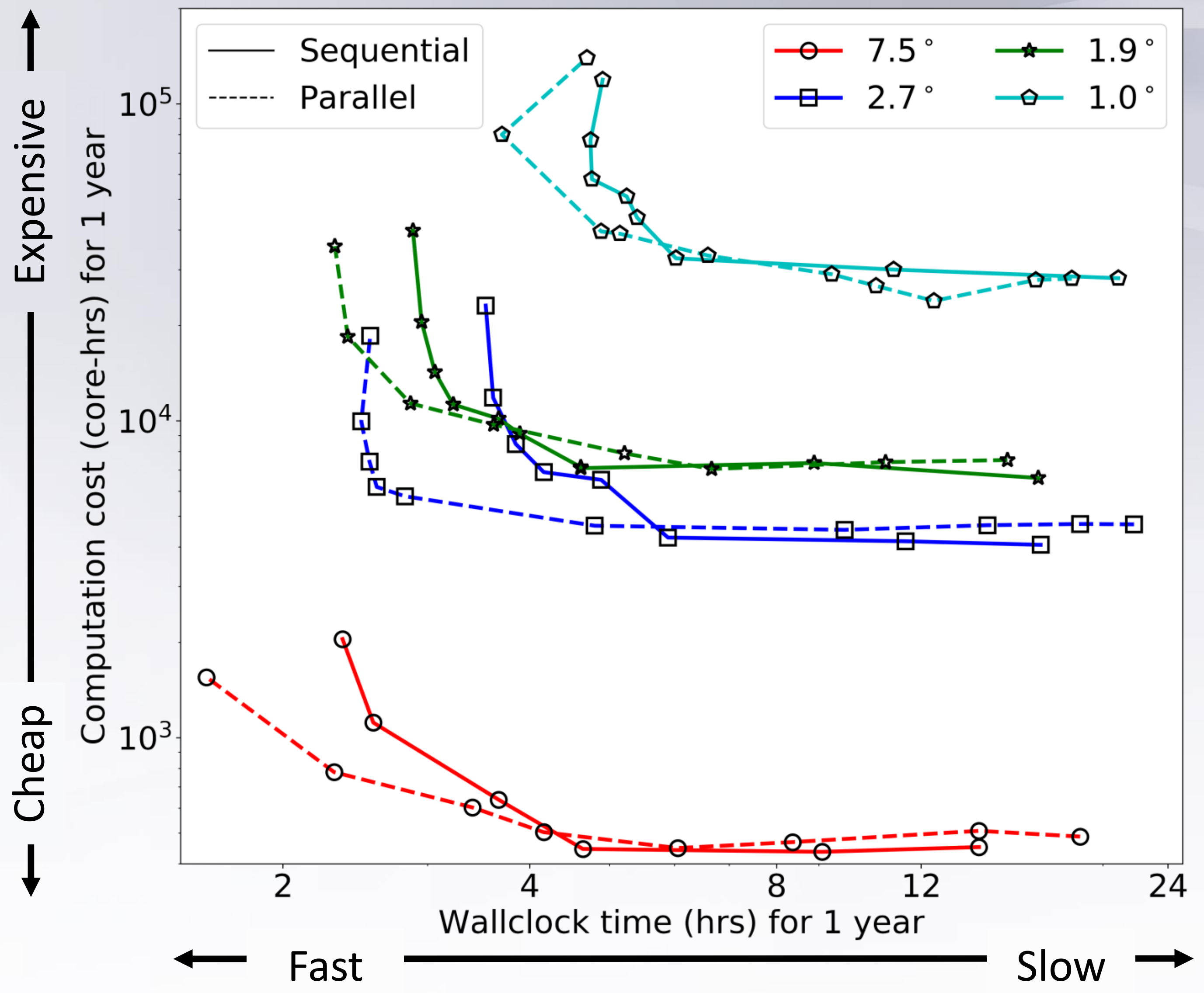


Figure: Model performance comparison for sequentially-split (solid lines) and parallel-split (dashed lines) simulations. All runs use a timestep of $\Delta t = 300s$.

Parallel-split coupling allows for faster runtimes at similar computational cost as slower sequentially-split coupling runs.

- For similar computation cost, PS offers more throughput.
- For a given throughput PS is cheaper.

Outline

- Introduction
- Climate Impact
- Performance Impact
- Stability Impact
- Overstabilization
- Mass Conservation
- Conclusions

Stability Impact

PS consistently requires a smaller timestep to remain stable. On average this value is **6 times** smaller than what is required for SS simulations.

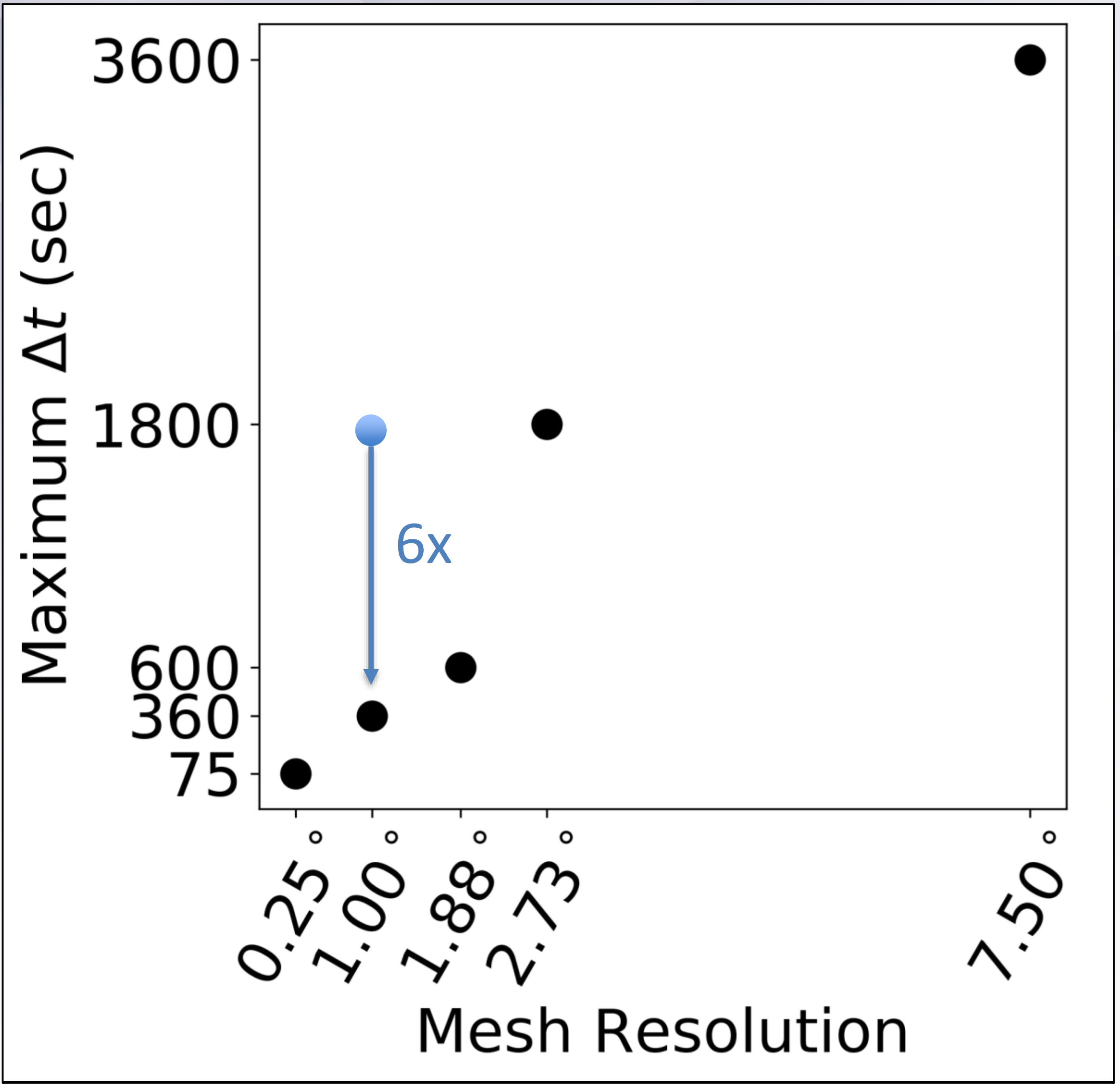


Figure: Suggested Δt for parallel-split coupling by average mesh resolution.

Stability Impact

PS consistently requires a smaller timestep to remain stable. On average this value is **6 times** smaller than what is required for SS simulations.

WHY?

- Timestep Stability Criteria?
- Mass Conservation?

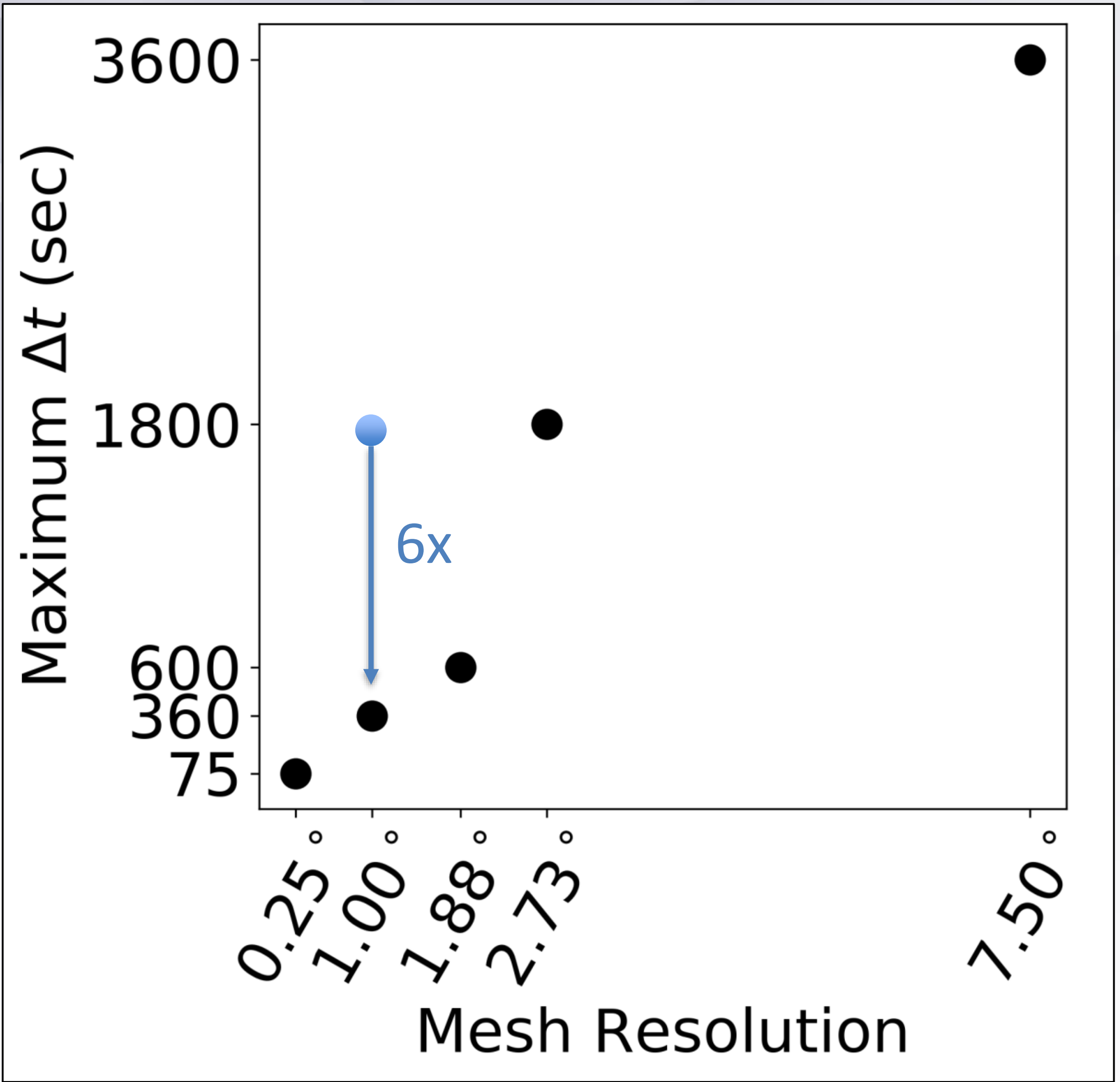


Figure: Suggested Δt for parallel-split coupling by average mesh resolution.

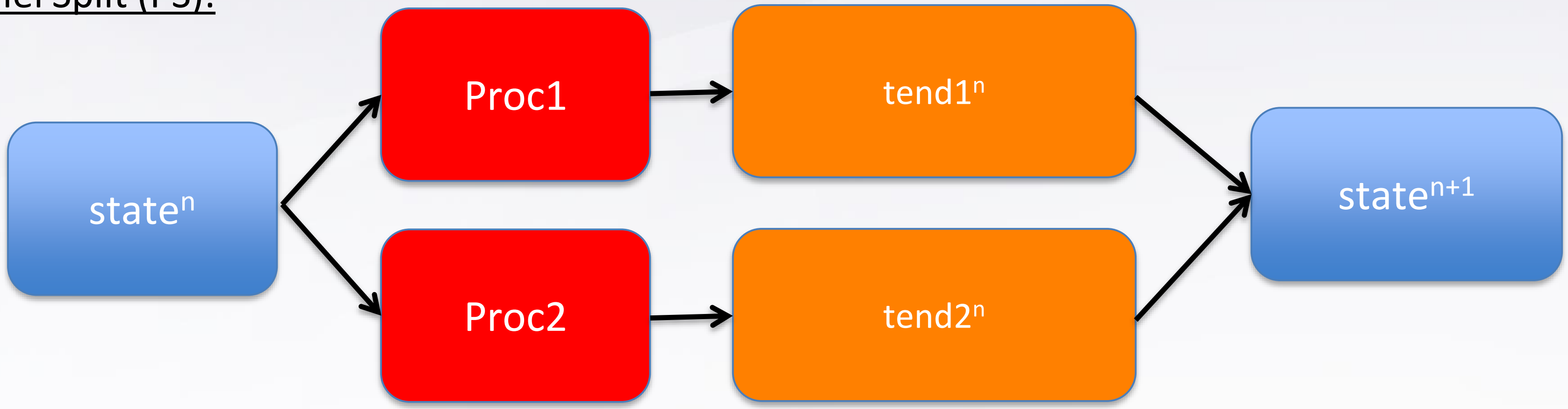
Stability Criteria:

Sequential Split (SS):

Δt constraint based on Proc1

Δt constraint based on Proc2

Parallel Split (PS):



Δt constraint based on the combined tendencies of Proc1 and Proc2

Simple Example: Advection-Diffusion

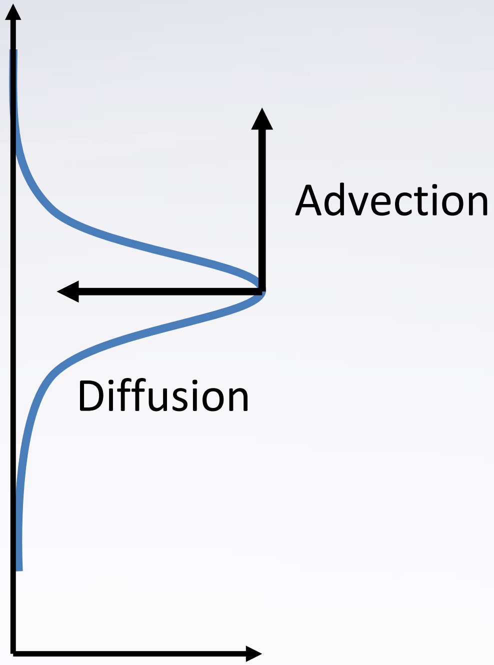
$$\frac{\partial q}{\partial t} + \alpha \frac{\partial q}{\partial z} + Dq = 0$$

<u>Stability Criteria</u>

- Advection only: $\Delta t \leq \frac{\Delta z}{\alpha}$

  provided the proper upwind scheme is applied.
- Diffusion only: $\Delta t \leq \frac{2}{D}$

  for $D > 0$

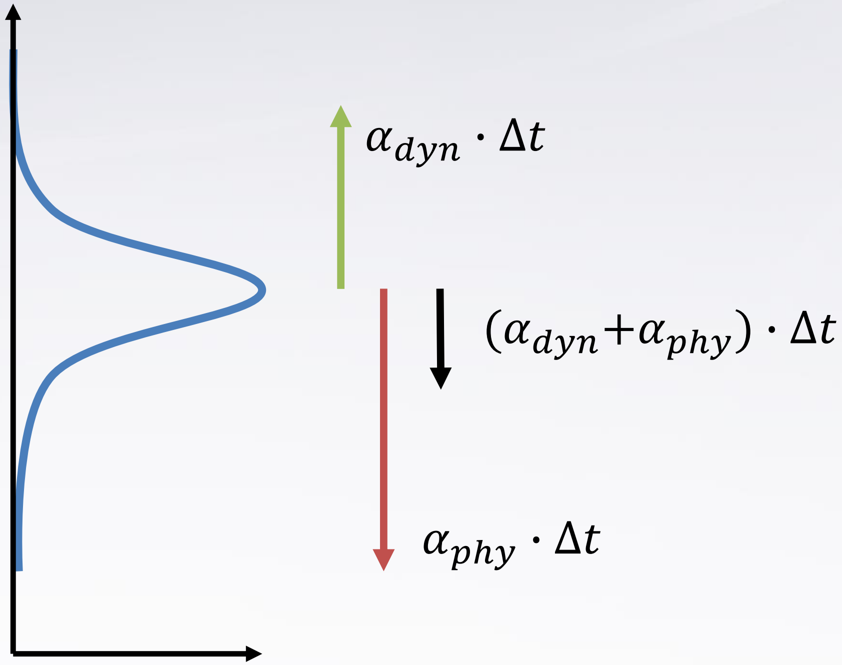


Simple Example: Advection-Diffusion

$$\frac{\partial q}{\partial t} + \alpha \frac{\partial q}{\partial z} + Dq = 0$$

Stability Criteria

- Advection only: $\Delta t \leq \frac{\Delta z}{\alpha}$
  provided the proper upwind scheme is applied.
- Diffusion only: $\Delta t \leq \frac{2}{D}$
  for $D > 0$



PS can have a smaller or larger timestep requirement, depending on the signs of advection terms.

Simple Example: Advection-Diffusion

$$\frac{\partial q}{\partial t} + \alpha \frac{\partial q}{\partial z} + Dq = 0$$

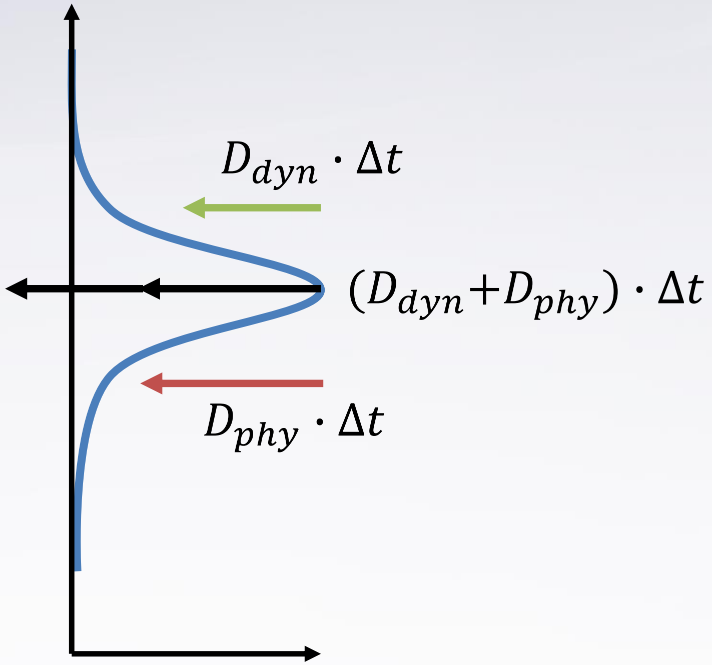
<u>Stability Criteria</u>

- Advection only: $\Delta t \leq \frac{\Delta z}{\alpha}$

  provided the proper upwind scheme is applied.
- Diffusion only: $\Delta t \leq \frac{2}{D}$

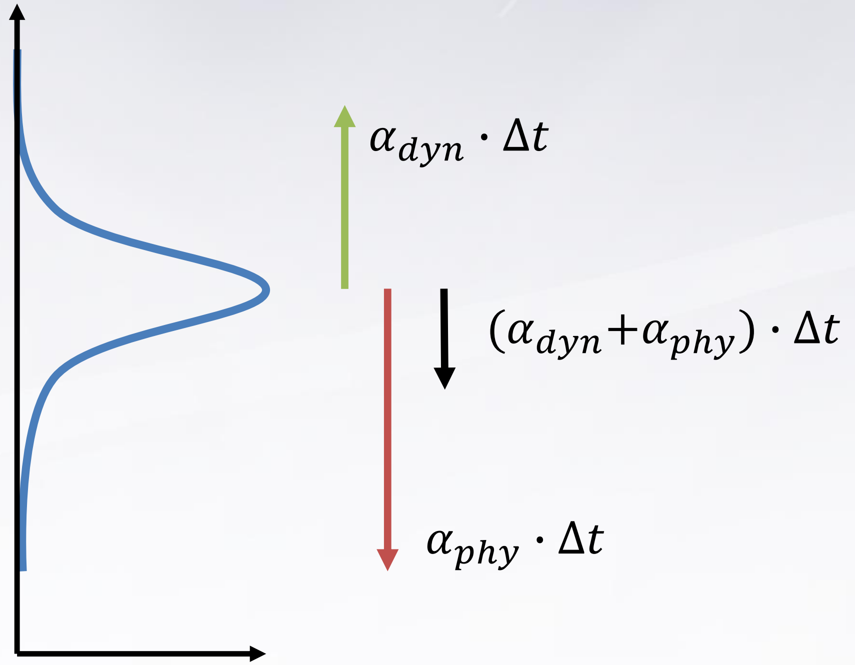
  for $D > 0$

PS has a stricter timestep requirement

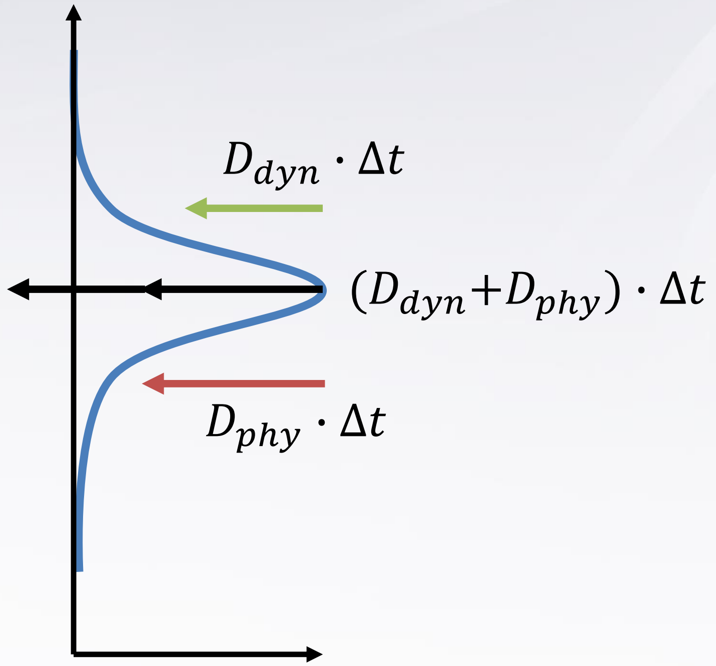


Simple Example: Advection-Diffusion

$$\frac{\partial q}{\partial t} + \alpha\frac{\partial q}{\partial z} + Dq = 0$$



PS has a different timestep requirement than SS



Outline

- Introduction
- Climate Impact
- Performance Impact
- Stability Impact
- Overstabilization
- Mass Conservation
- Conclusions

Stability Impact

At the default timestep of $\Delta t = 1800s$ the model generates $2\Delta x$ instabilities throughout the domain.

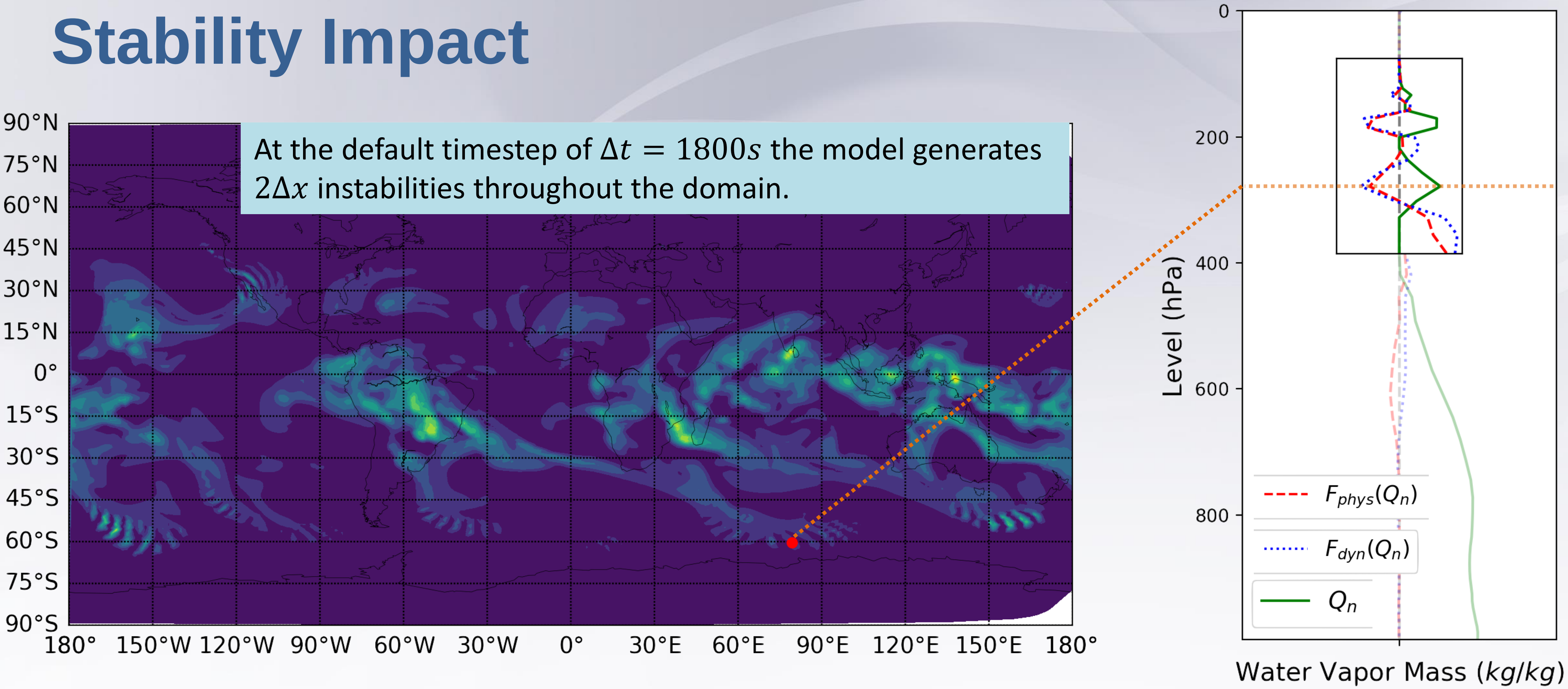


Figure: Global water vapor mass Q_n (kg/kg) at 278 hPa (left) and vertical water vapor mass with water vapor tendencies for physics, $F_{phys}(Q_n)$, and dynamics, $F_{dyn}(Q_n)$ (right). Default timestep, $\Delta t = 1800s$, mesh resolution=1°.

Similarly, physics and dynamics tendencies for $2\Delta z$ instabilities ”overstabilize” the solution. Causing more instabilities to form.

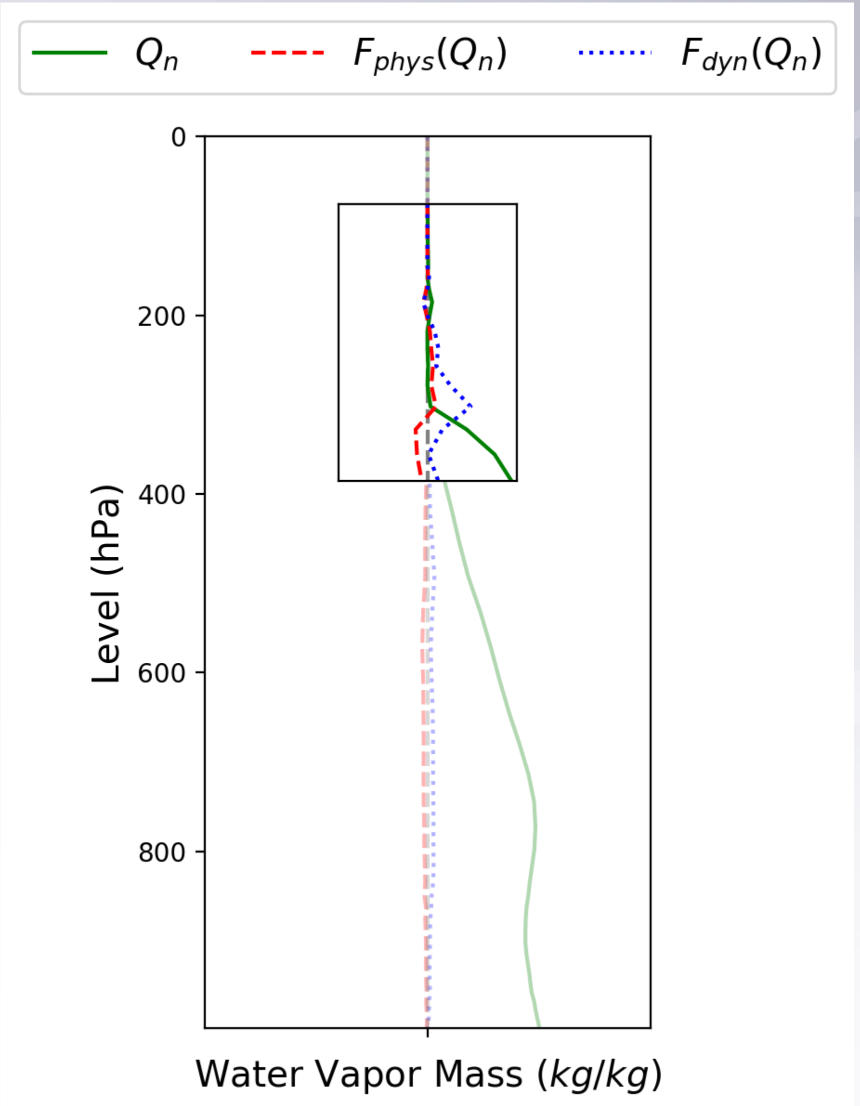


Figure: Vertical water vapor mass Q_n (kg/kg), with water vapor tendencies for physics, $F_{phys}(Q_n)$, and dynamics, $F_{dyn}(Q_n)$ at every timestep leading to instabilty. Default timestep, $\Delta t = 1800s$, mesh resolution=1°.

An instability forms in the vertical column.

Both Physics and Dynamics act to damp sharp gradients to stabilize the solution.

The combined tendencies of physics and dynamics "**overstabilize**" the solution generating a separate instability in the vertical.

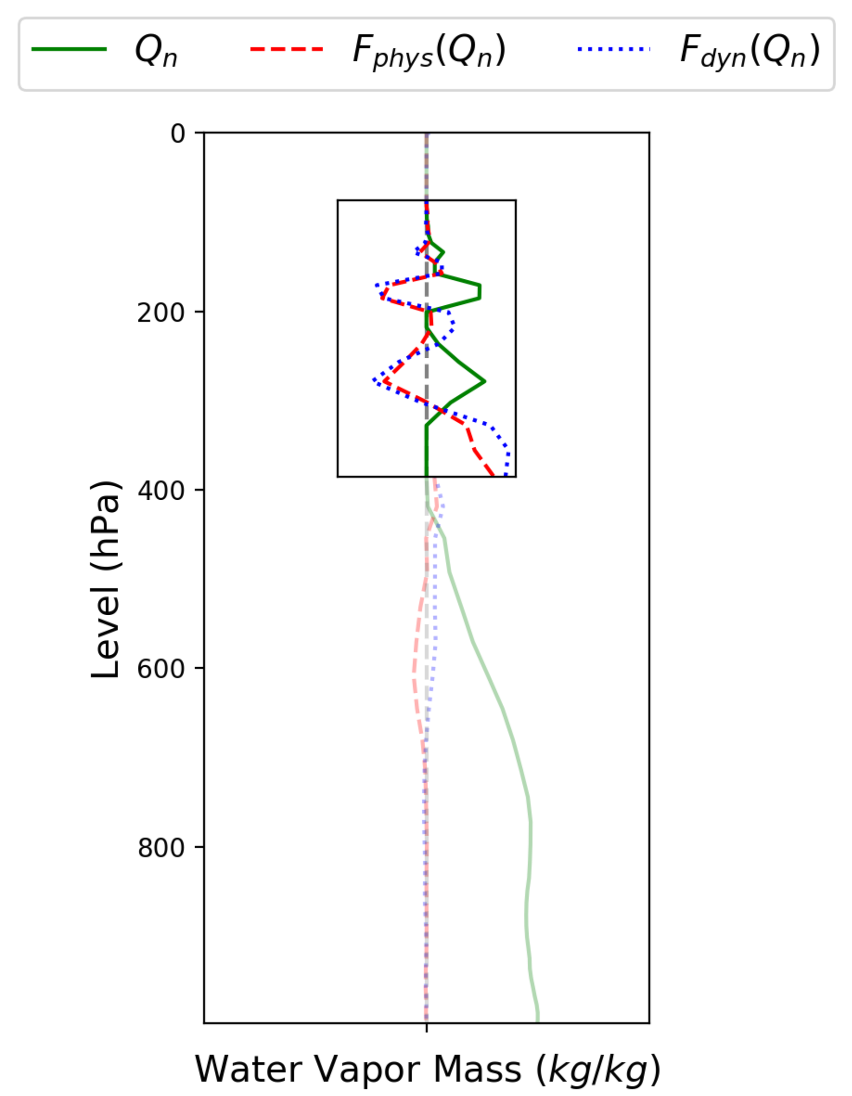


Figure: Vertical water vapor mass Q_n (kg/kg), with water vapor tendencies for physics, $F_{phys}(Q_n)$, and dynamics, $F_{dyn}(Q_n)$ at every timestep leading to instabilty. Default timestep, $\Delta t = 1800s$, mesh resolution=1°.

An instability forms in the vertical column.

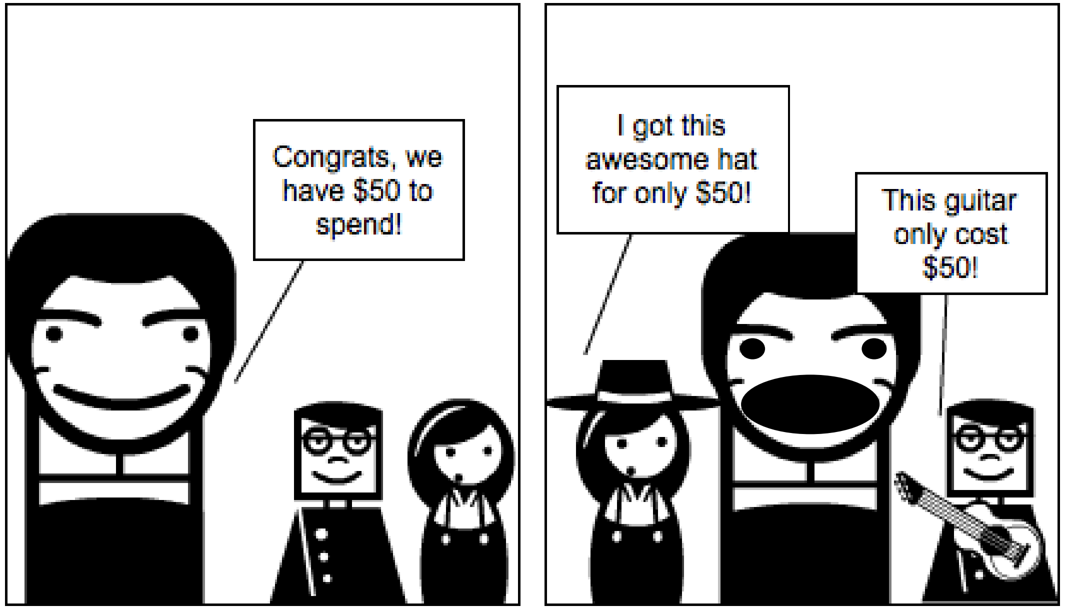
Both Physics and Dynamics act to damp sharp gradients to stabilize the solution.

The combined tendencies of physics and dynamics "**overstabilize**" the solution generating a separate instability in the vertical.

Outline

- Introduction
- Climate Impact
- Performance Impact
- Stability Impact
- Overstabilization
- Mass Conservation
- Conclusions

Issue - Overconsumption

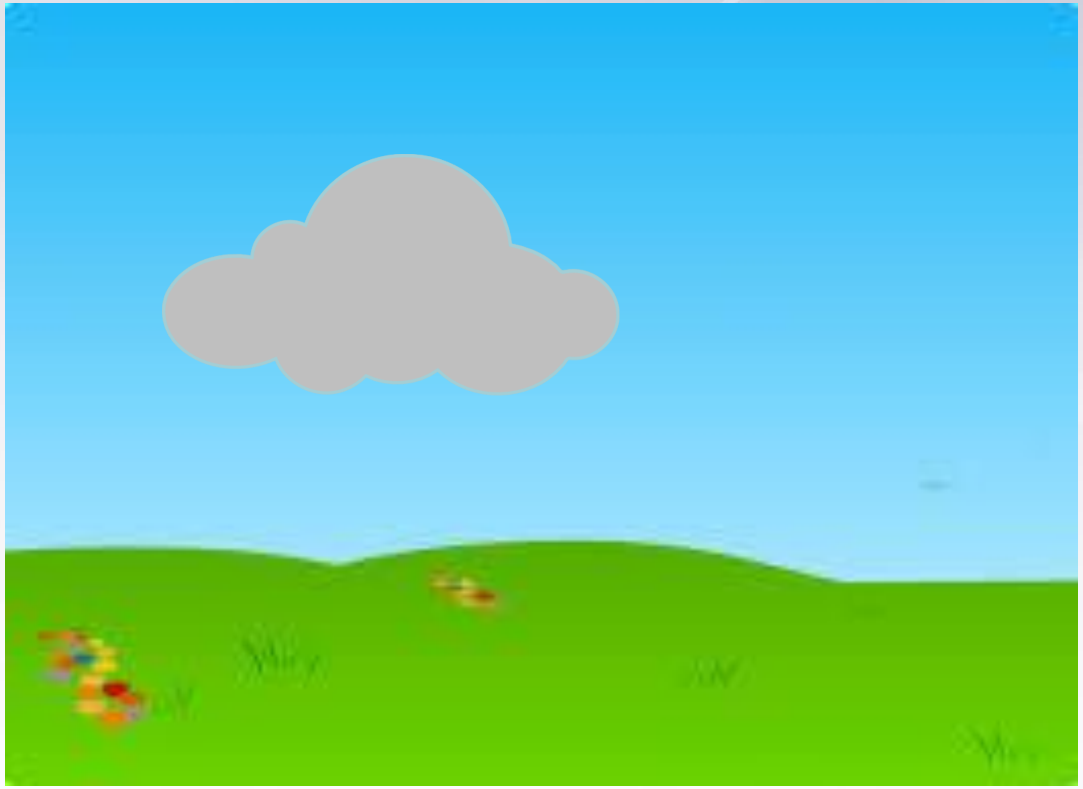


Parallel yet isolated processes can over-deplete resources when combined.

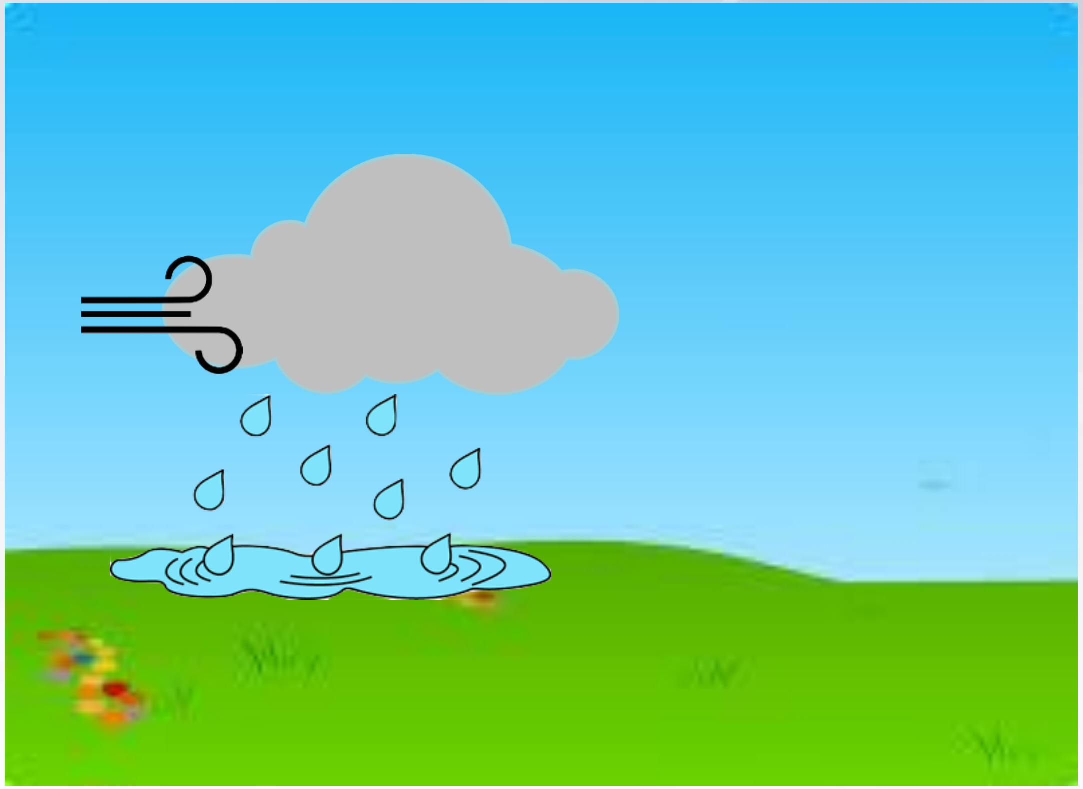
This problem is encountered any time parallel time splitting is used

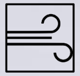
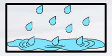
- Two processes can independently remove mass from the same point, which when compounded can lead to an overconsumption of the local resources, i.e. negative mass.
- At the phys/dyn level this occurs frequently, as much as 1/3 of the pts. for liquid cloud water.
- **We currently just set negative values to zero.**
  - Note, this actually occurs for sequential tendency splitting in SE dynamics as well
  - “Clipping” negative tendencies violates conservation.
  - Is this acceptable? Is there a better way?

Issue - Overconsumption



Issue - Overconsumption



Dynamics  Physics 

Sequential Splitting:

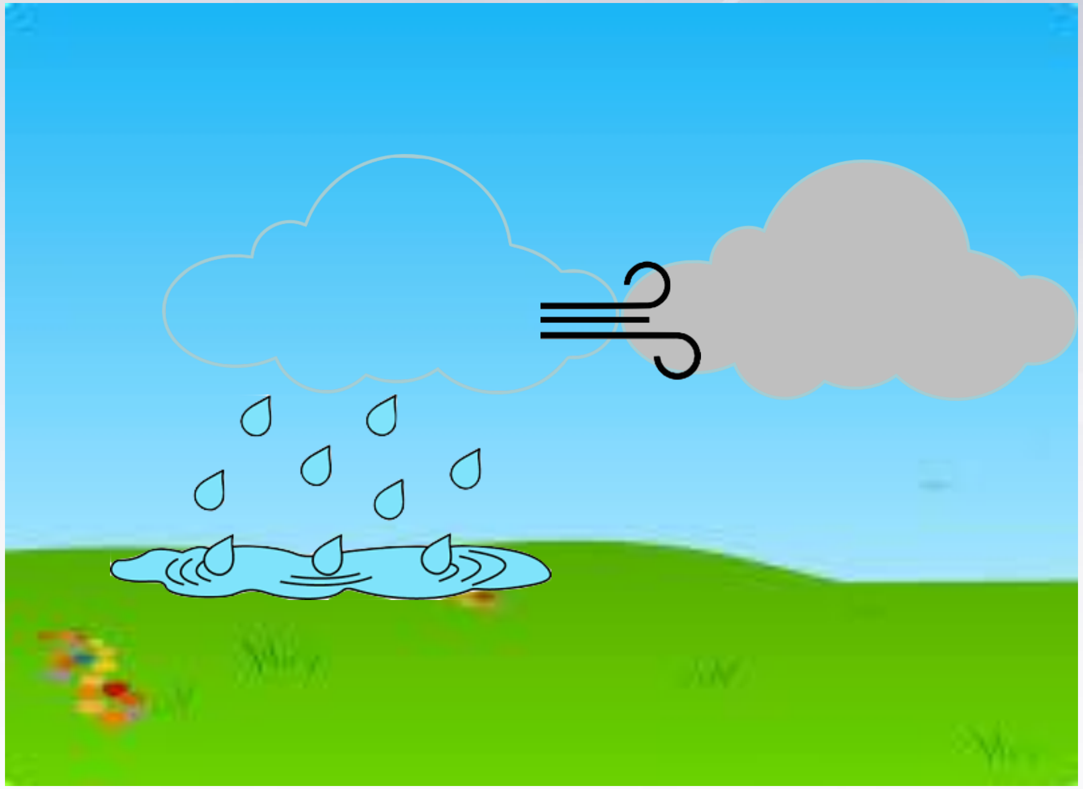
- Dynamics moves mass, then physics removes mass

Parallel Splitting:

- Dynamics prescribes a tendency to move mass.
- Physics prescribes a tendency to remove mass.
- Both are applied to the same location in space.

Potential Mass Conservation Issues

Issue - Overconsumption



A. **Clipping:** Setting all negative masses to zero.

B. **Mass Fixer:**

   i. **Weighted Horizontal Distribution:** Drawing mass from neighboring nodes horizontally.

   ii. **Weighted Vertical Distribution:** Drawing mass from neighboring levels vertically.

   iii. **Full Element Distribution:** Drawing mass from all points within an element.

Mass Conservation

Dynamics

$$\frac{\partial q}{\partial t} + U\frac{\partial q}{\partial x} = 0$$

Physics

$$\frac{\partial q}{\partial t} + Dq = 0$$

C. Consistent-Parallel Splitting:

- Dynamics prescribes a tendency to move mass.
- Physics prescribes a tendency to remove mass.
- Physics tendency is advected along with dynamics.

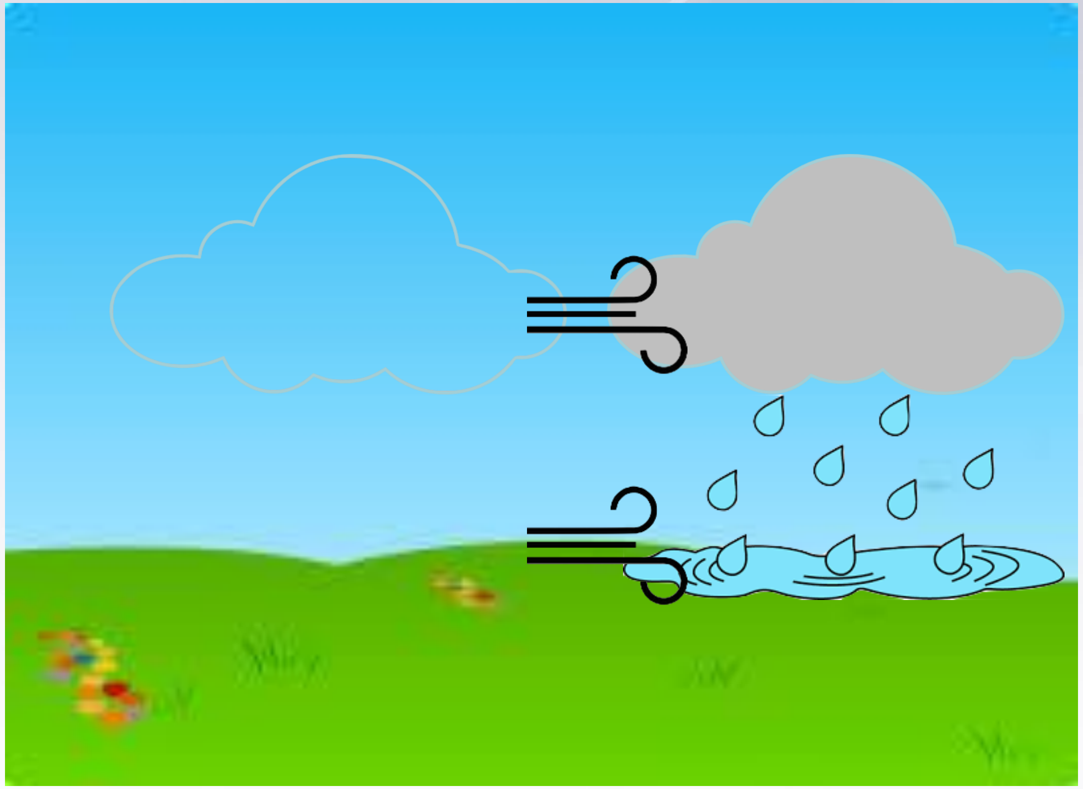
Mass Conservation

Dynamics

$$\frac{\partial q}{\partial t} + U\frac{\partial q}{\partial x} = 0$$

Physics

$$\frac{\partial q}{\partial t} + Dq = 0$$



C. Consistent-Parallel Splitting:

- Dynamics prescribes a tendency to move mass.
- Physics prescribes a tendency to remove mass.
- Physics tendency is advected along with dynamics.

Conclusions

- After implementing parallel physics/dynamics, we found:
  - Little change to model climate*.
  - For a given number of core hours, PS allows for faster time to solution than SS (at higher core counts)*.
  - Parallel splitting requires smaller Δt for stability, canceling the benefit of faster throughput.
    - Due to overcompensation between physics and dynamics.
    - PS leads to mass conservation errors.

* At the timestep needed for stability in parallel-splitting

Conclusions

- After implementing parallel physics/dynamics, we found:
  - Little change to model climate*.
  - For a given number of core hours, PS allows for faster time to solution than SS (at higher core counts)*.
  - Parallel splitting requires smaller Δt for stability, canceling the benefit of faster throughput. A smaller timestep isn't all bad -> more accurate.
    - Due to overcompensation between physics and dynamics.
    - Apply stabilization mechanisms (i.e. hyperviscosity) after physics/dynamics coupling.
    - PS leads to mass conservation errors.
    - Better handling of mass conservation (i.e. tendency advection)

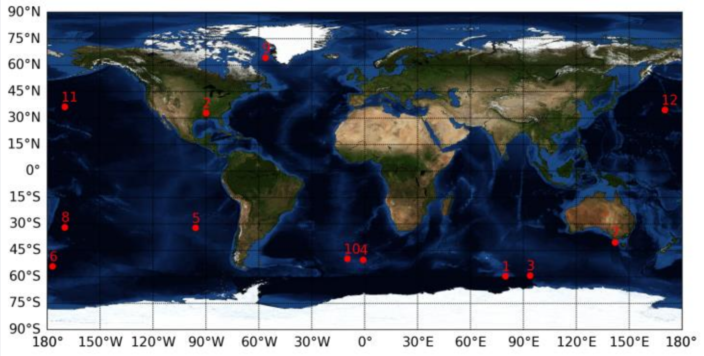
* At the timestep needed for stability in parallel-splitting

Acknowledgements:

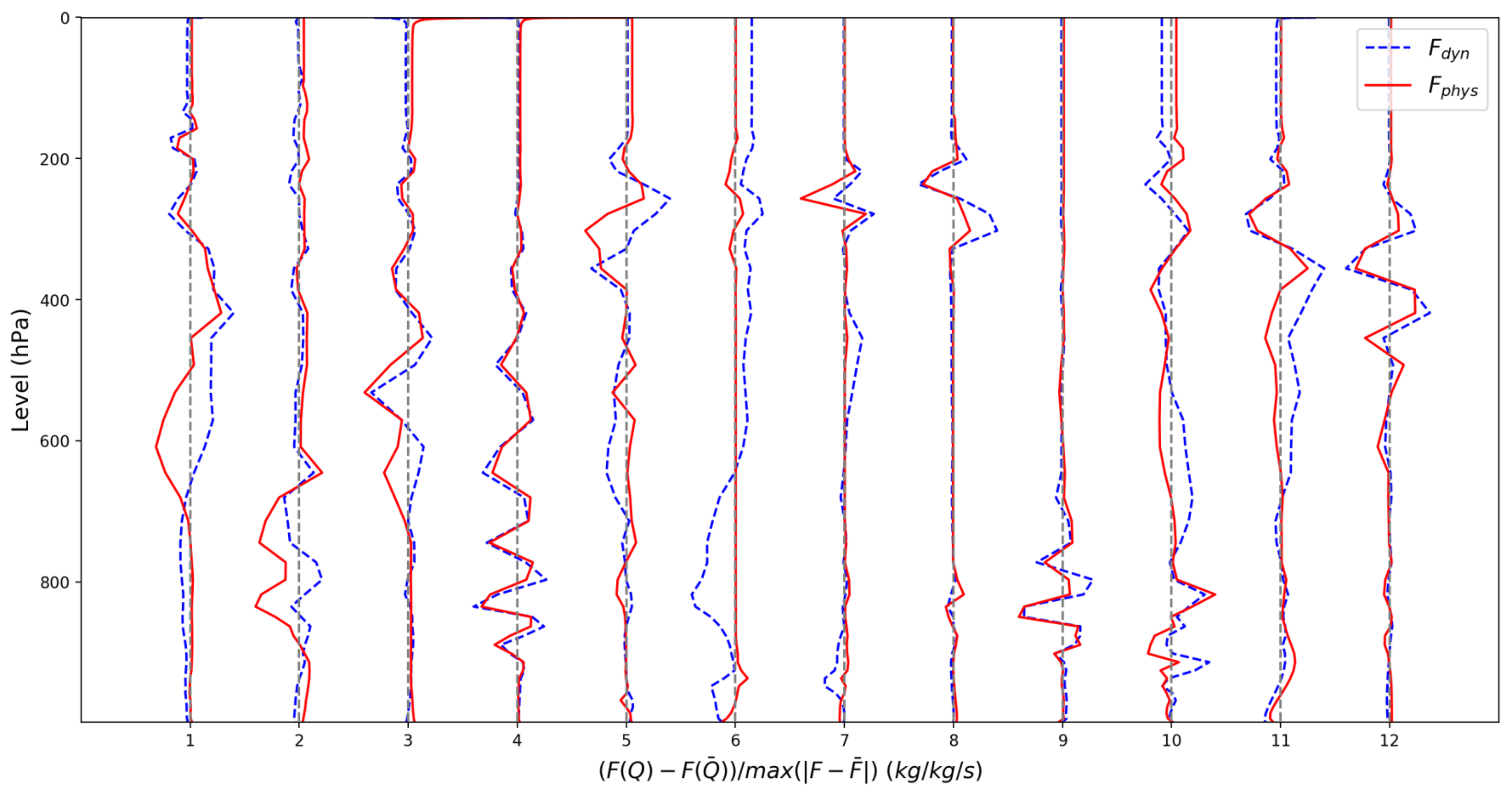
This work was supported by the DOE Code Modernization and Validation-Software Engineering (CMDV-SE) project.

Questions?

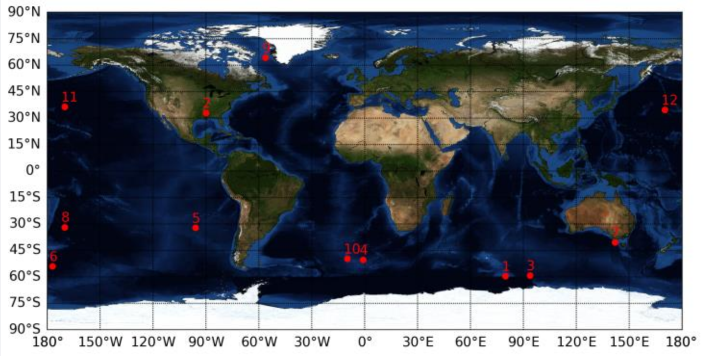
Extra Slides



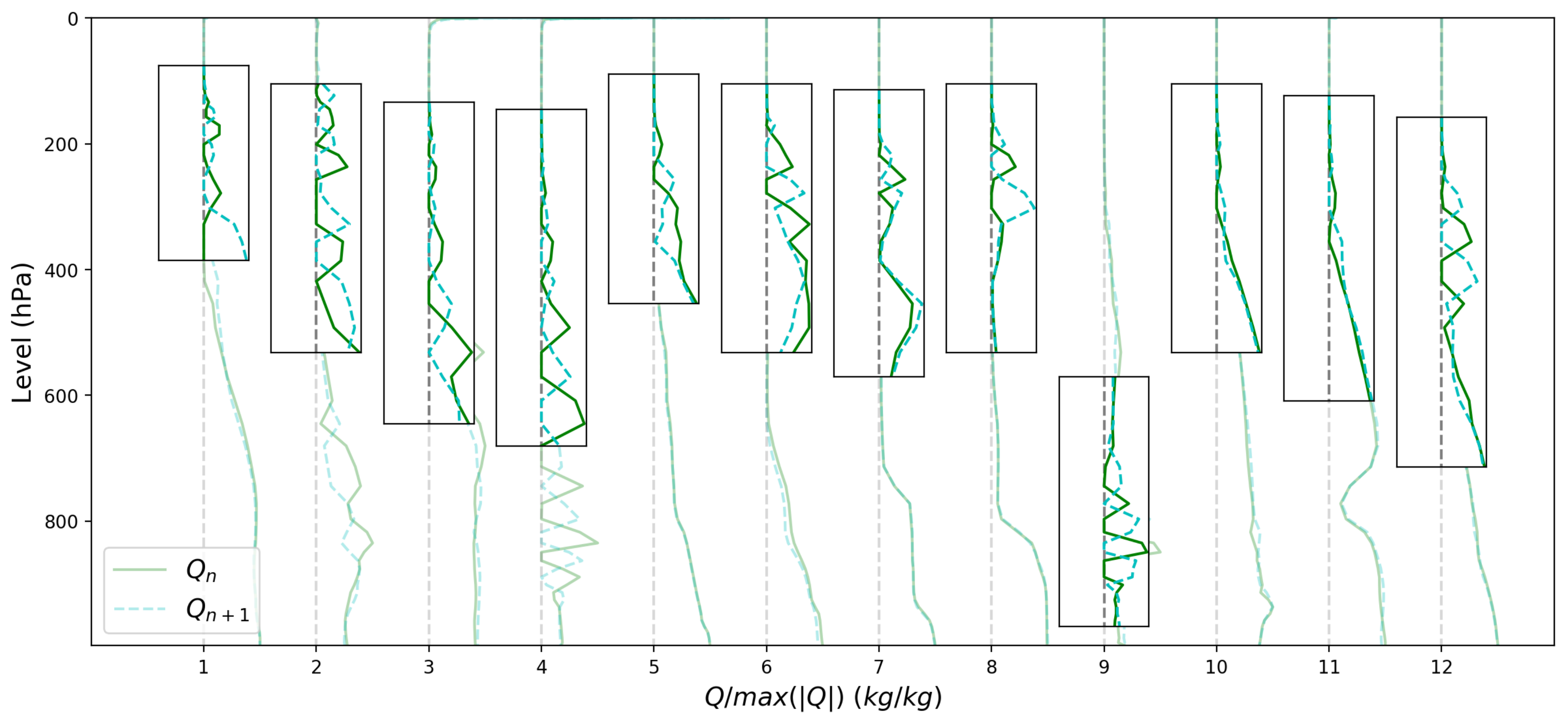
90°N
75°N
60°N
45°N
30°N
15°N
0°
15°S
30°S
45°S
60°S
75°S
90°S
180°
150°W
120°W
90°W
60°W
30°W
0°
30°E
60°E
90°E
120°E
150°E
180°
1
2
3
4
5
6
7
8
9
10
11
12



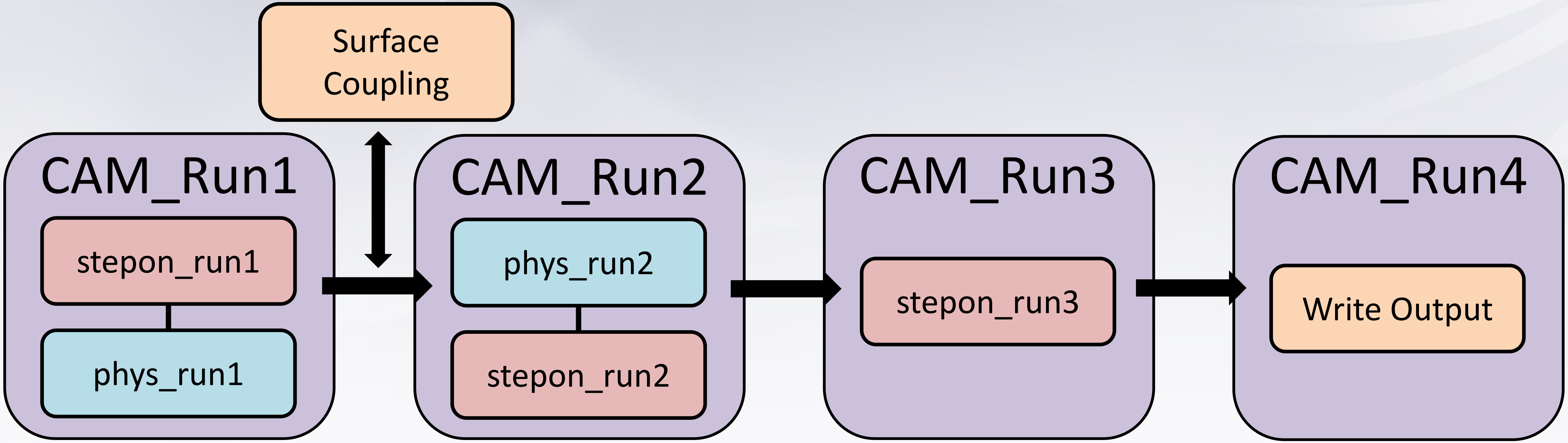
Level (hPa)
0
200
400
600
800
1
2
3
4
5
6
7
8
9
10
11
12
$(F(Q) - F(\bar{Q}))/max(|F - \bar{F}|)$ (kg/kg/s)
F_{dyn}
F_{phys}

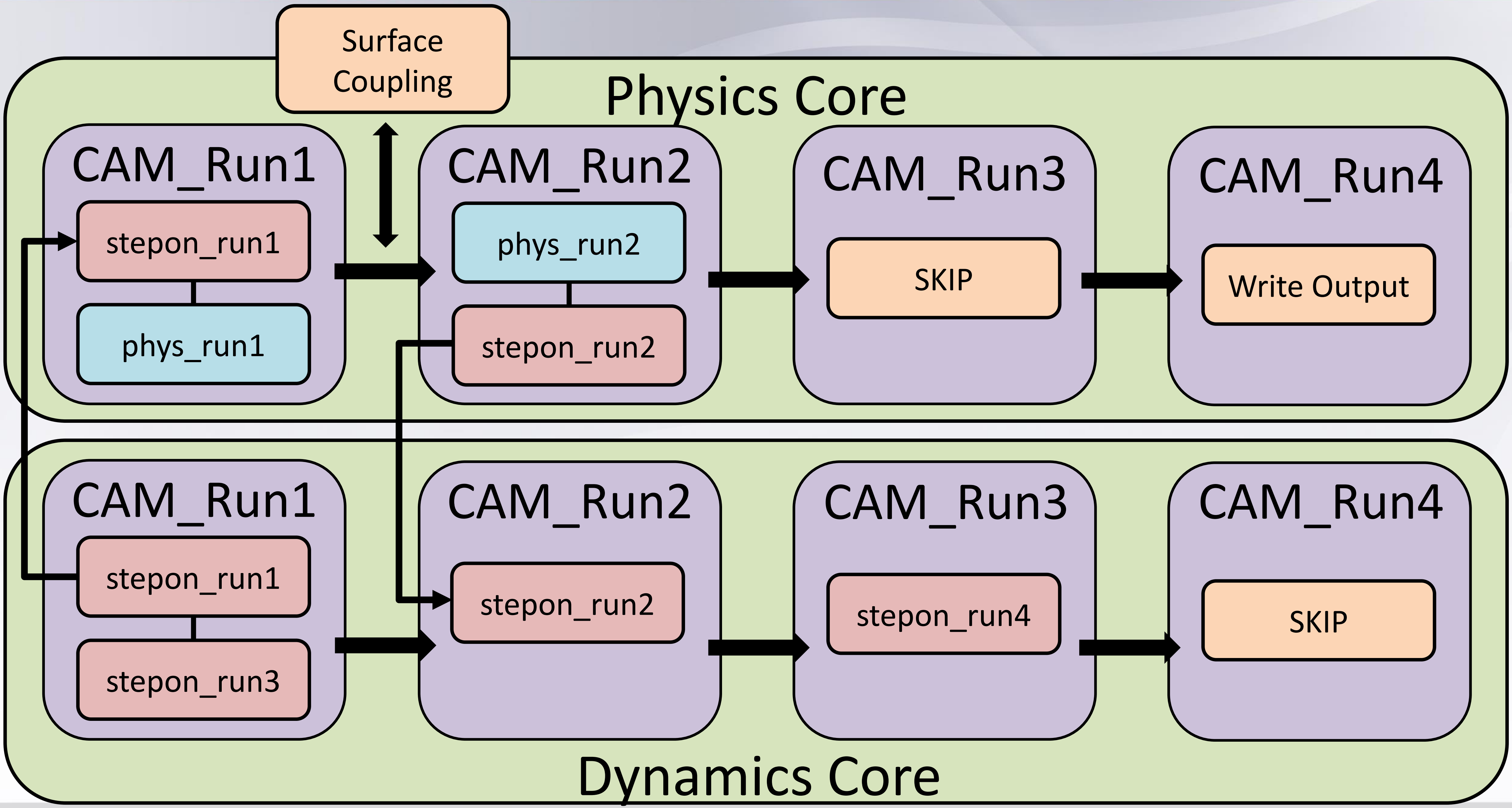


90°N
75°N
60°N
45°N
30°N
15°N
0°
15°S
30°S
45°S
60°S
75°S
90°S
180°
150°W
120°W
90°W
60°W
30°W
0°
30°E
60°E
90°E
120°E
150°E
180°
1
2
3
4
5
6
7
8
9
10
11
12



Level (hPa)
0
200
400
600
800
Q_n
Q_{n+1}
1
2
3
4
5
6
7
8
9
10
11
12
$Q/max(|Q|)$ (kg/kg)


Surface
Coupling
CAM_Run1
stepon_run1
phys_run1
CAM_Run2
phys_run2
stepon_run2
CAM_Run3
stepon_run3
CAM_Run4
Write Output


Surface Coupling
Physics Core
CAM_Run1
stepon_run1
phys_run1
CAM_Run2
phys_run2
stepon_run2
CAM_Run3
SKIP
CAM_Run4
Write Output
CAM_Run1
stepon_run1
stepon_run3
CAM_Run2
stepon_run2
CAM_Run3
stepon_run4
CAM_Run4
SKIP
Dynamics Core

Coupling Strategies:

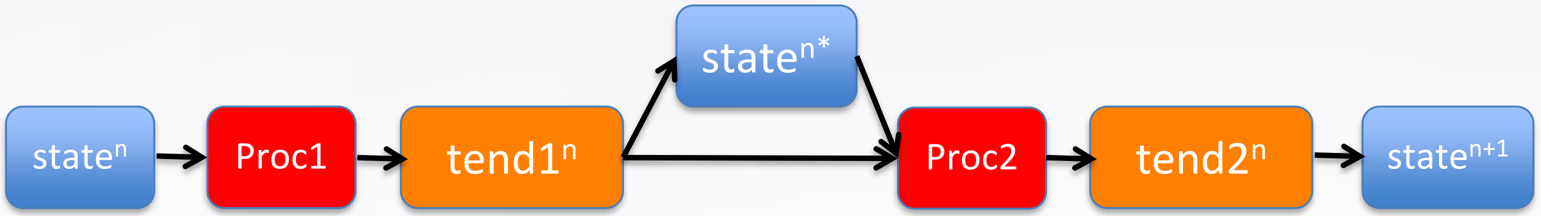
<u>A. Sequential-Update Split (SUS, aka time split/fractional steps):</u> State is updated after each process

used in E3SM physics

<u>B. Sequential-Tendency Split (STS, aka no step splitting):</u> The *tendency* from Proc1 is used by Proc2

option to couple E3SM physics and SE dynamics

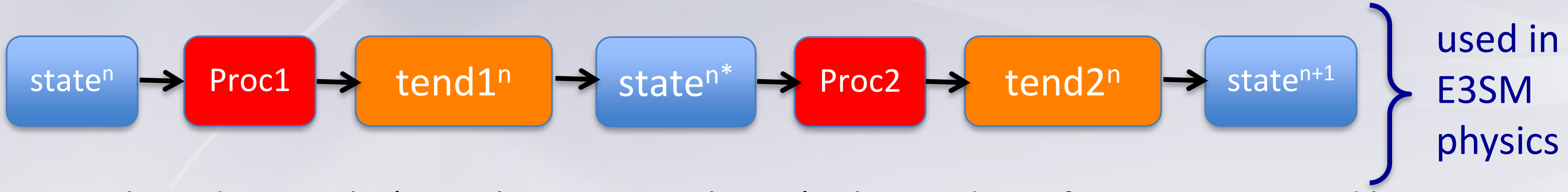
<u>C. Hybrid-Update-Tendency Split (HS, some processes SUS, some STS):</u> Mass conserved variables are updated after each process (SUS), all others use the tendency from Proc1 in Proc2 (STS).



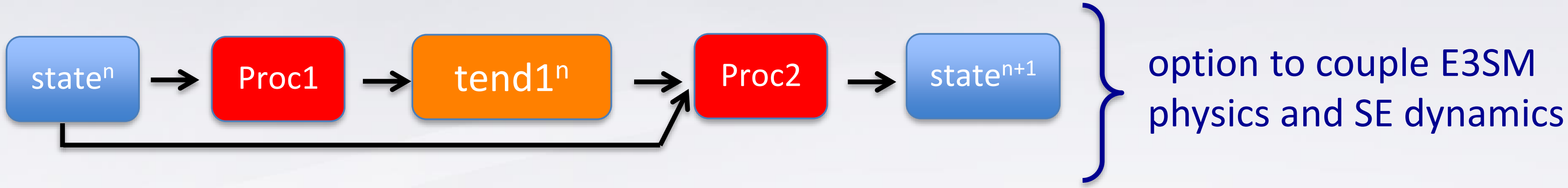
E3SM default

Coupling Strategies:

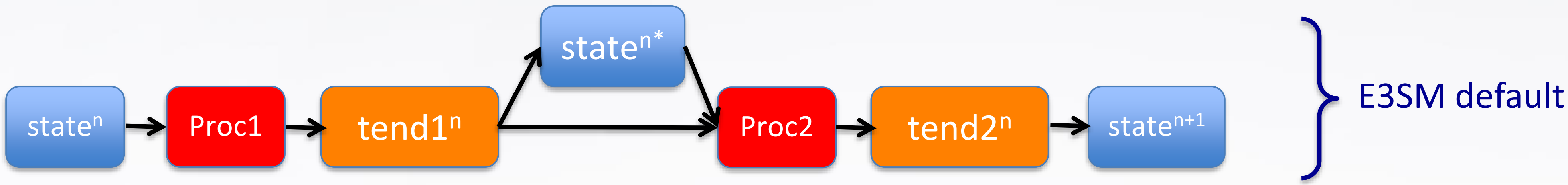
<u>A. Sequential-Update Split (SUS, aka time split/fractional steps):</u> State is updated after each process



<u>B. Sequential-Tendency Split (STS, aka no step splitting):</u> The *tendency* from Proc1 is used by Proc2



<u>C. Hybrid-Update-Tendency Split (HS, some processes SUS, some STS):</u> Mass conserved variables are updated after each process (SUS), all others use the tendency from Proc1 in Proc2 (STS).



$$f^n \Rightarrow f^* = f^n + \Delta t_D D f^n \Rightarrow f^{n+1} = f^* + \Delta t_P P f^* \Rightarrow f^{n+1}$$

$$Dynamics = F_D$$

$$Physics = F_P$$

$$f^n \Rightarrow f^{n+1} = f^n + \Delta t(F_D + F_P) \Rightarrow f^{n+1}$$

Coupling Strategies:

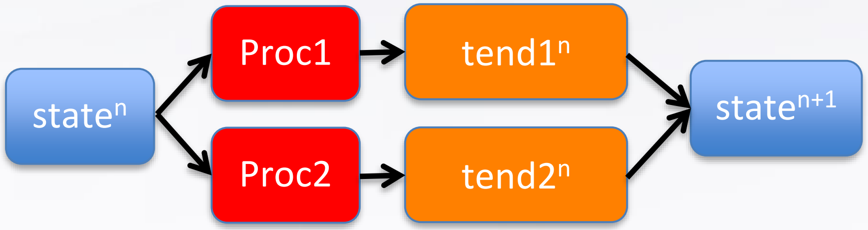
A. Sequential-Update Split (SUS, aka time split/fractional steps): State is updated after each process

used in E3SM physics

B. Sequential-Tendency Split (STS, aka no step splitting): The *tendency* from Proc1 is used by Proc2

option to couple E3SM physics and SE dynamics

C. Parallel Split (PS, aka process/additive split): All processes are computed from the same state



used in E3SM microphysics

What is the advantage of switching to Parallel Split over Sequential Splitting?

Overconsumption - Example

$$\frac{\partial q}{\partial t} + \delta_D q + \delta_P q = 0$$

Dynamics:

$$\frac{\partial q}{\partial t} + \delta_D q = 0$$

Physics:

$$\frac{\partial q}{\partial t} + \delta_P q = 0$$

Overconsumption - Example

Dynamics: $\frac{\partial q}{\partial t} + \delta_D q = 0$

Physics: $\frac{\partial q}{\partial t} + \delta_P q = 0$

Sequentially-Split

Dynamics: $q^* = (1 - \delta_D \Delta t) q^n$

Physics: $q^{n+1} = (1 - \delta_P \Delta t) q^*$

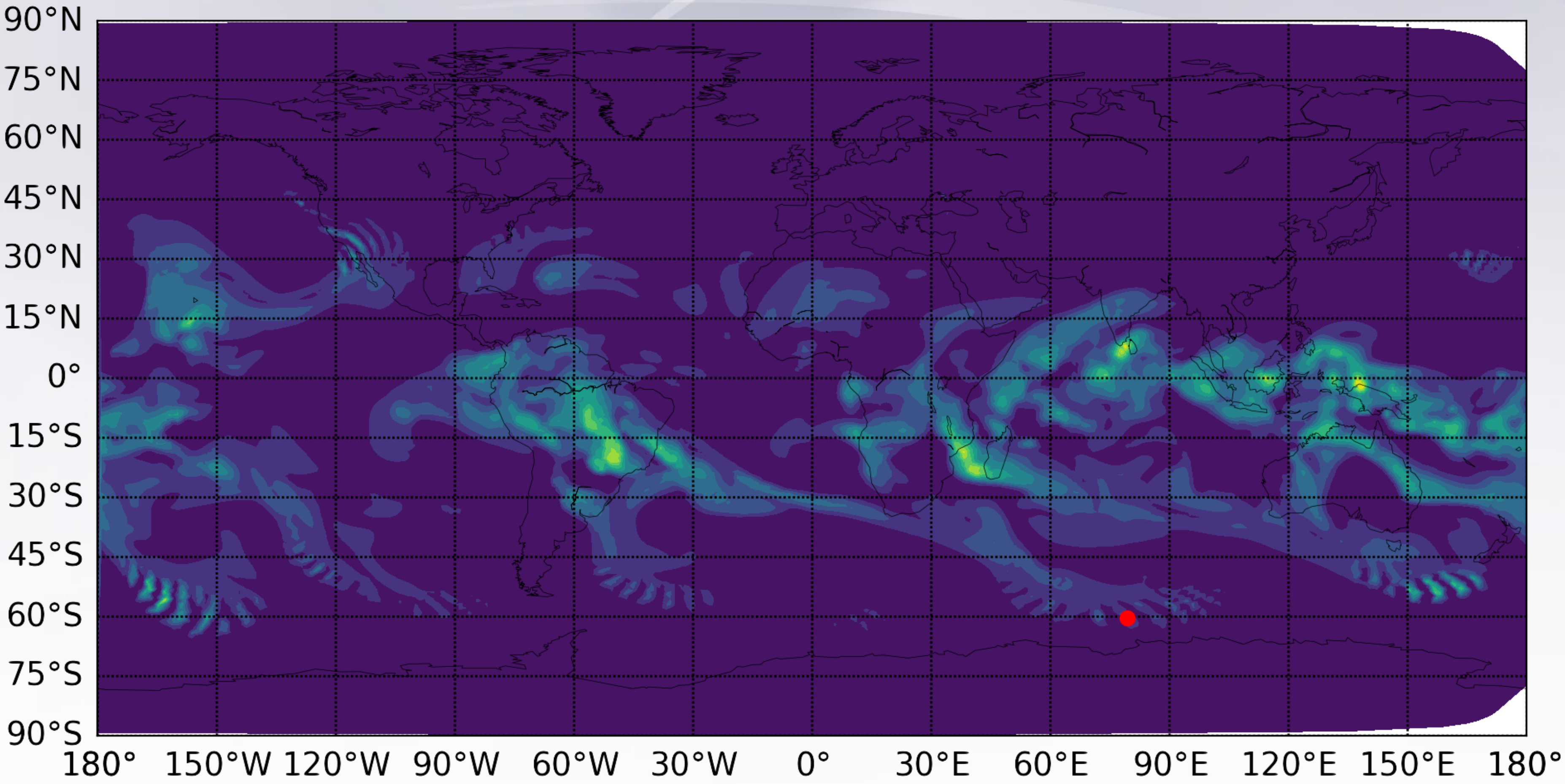
Parallel-Split

Dynamics: $F_D = -\delta_D \Delta t * q^n$

Physics: $F_P = -\delta_P \Delta t * q^n$

$$q^{n+1} = (1 - \delta_P \Delta t - \delta_D \Delta t) q^n$$

Mass Conservation

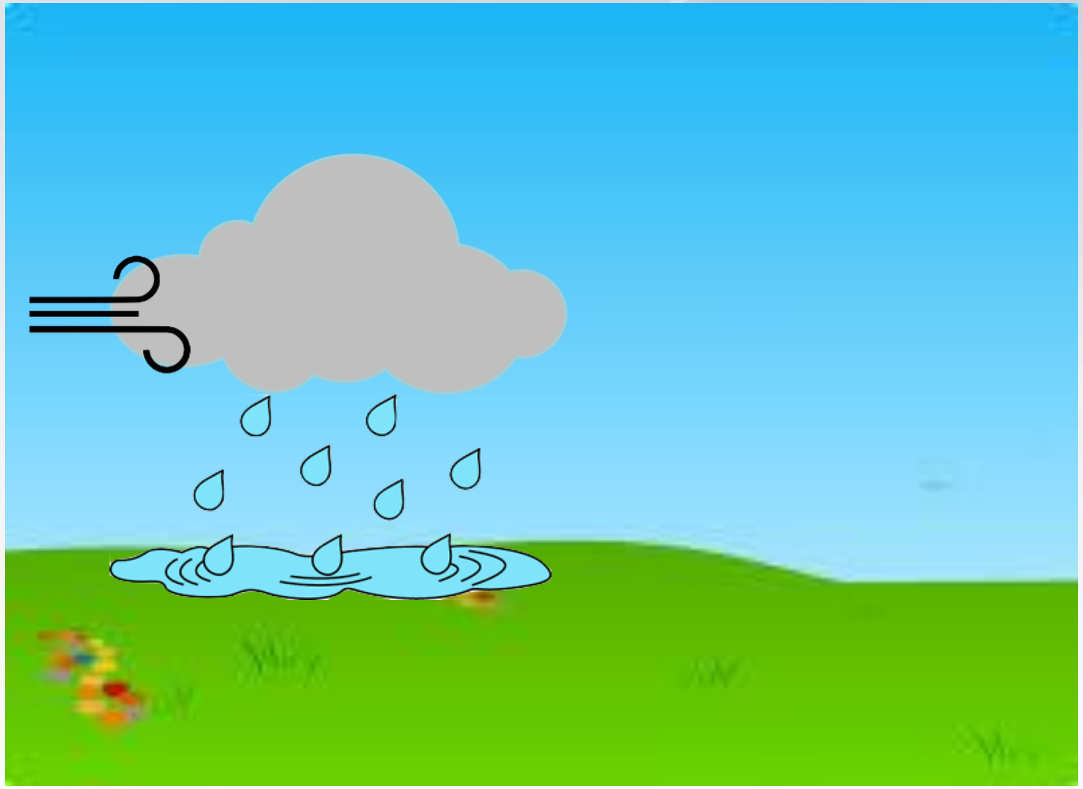


There is a clear formation of $2\Delta x$ waves forming at the mid-latitudes.

This is likely due to mass “clipping” which occurs when the conservation of mass has been violated.

Mass conservation violations increase with larger time step.

Mass Conservation



Dynamics

$$\frac{\partial q}{\partial t} + U\frac{\partial q}{\partial x} = 0$$

Physics

$$\frac{\partial q}{\partial t} + Dq = 0$$

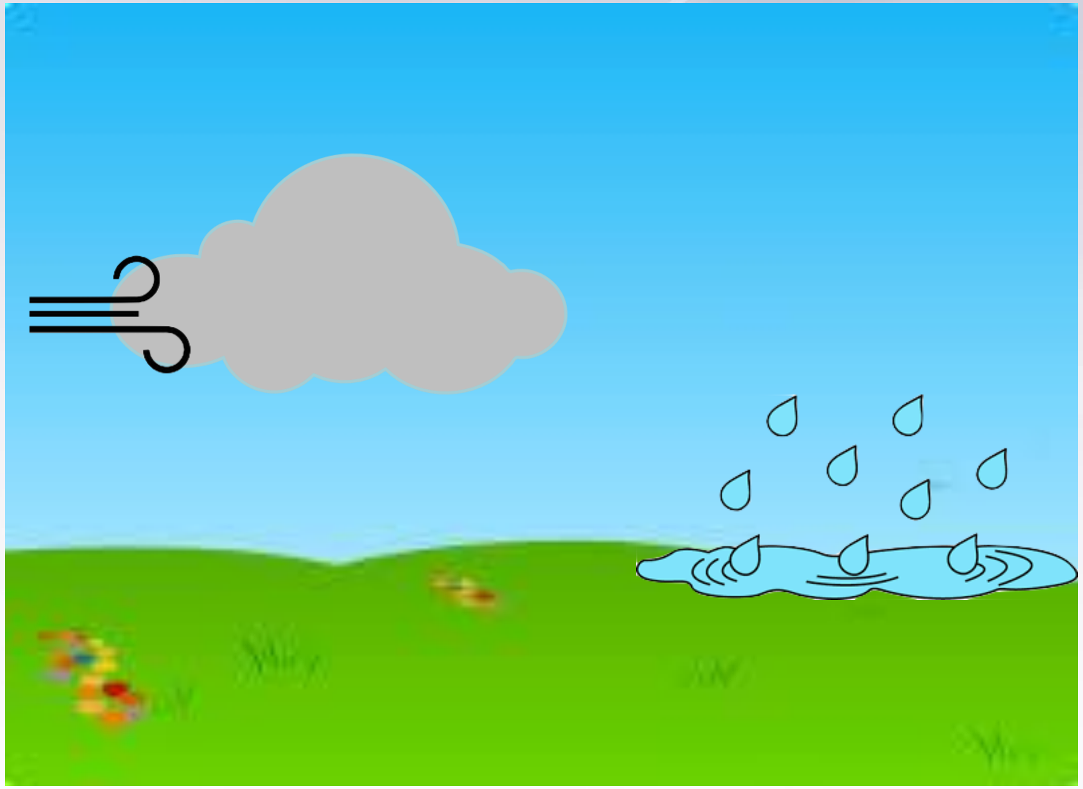
Mass Conservation

Dynamics

$$\frac{\partial q}{\partial t} + U\frac{\partial q}{\partial x} = 0$$

Physics

$$\frac{\partial q}{\partial t} + Dq = 0$$



Sequential Splitting:

- Dynamics moves mass, then physics removes mass

Mass Conservation

Dynamics

$$\frac{\partial q}{\partial t} + U\frac{\partial q}{\partial x} = 0$$

Physics

$$\frac{\partial q}{\partial t} + Dq = 0$$

Sequential Splitting:

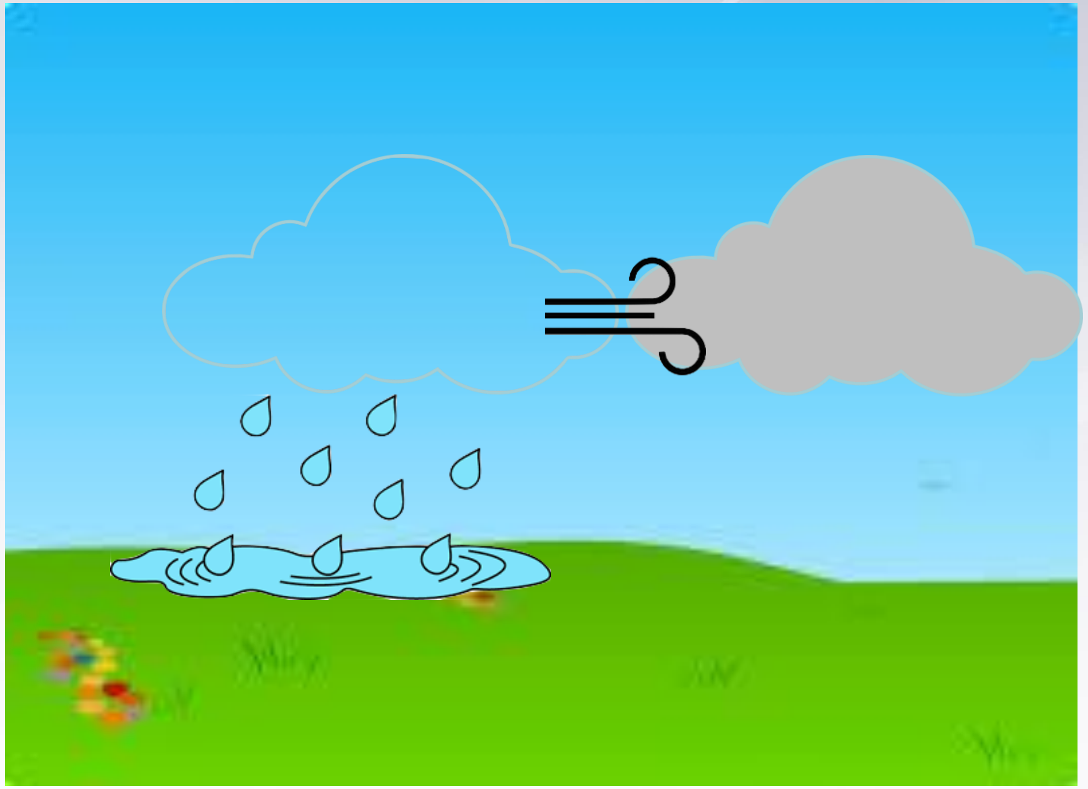
- Dynamics moves mass, then physics removes mass

Parallel Splitting:

- Dynamics prescribes a tendency to move mass.
- Physics prescribes a tendency to remove mass.
- Both are applied to the same location in space.

Potential Mass Conservation Issues

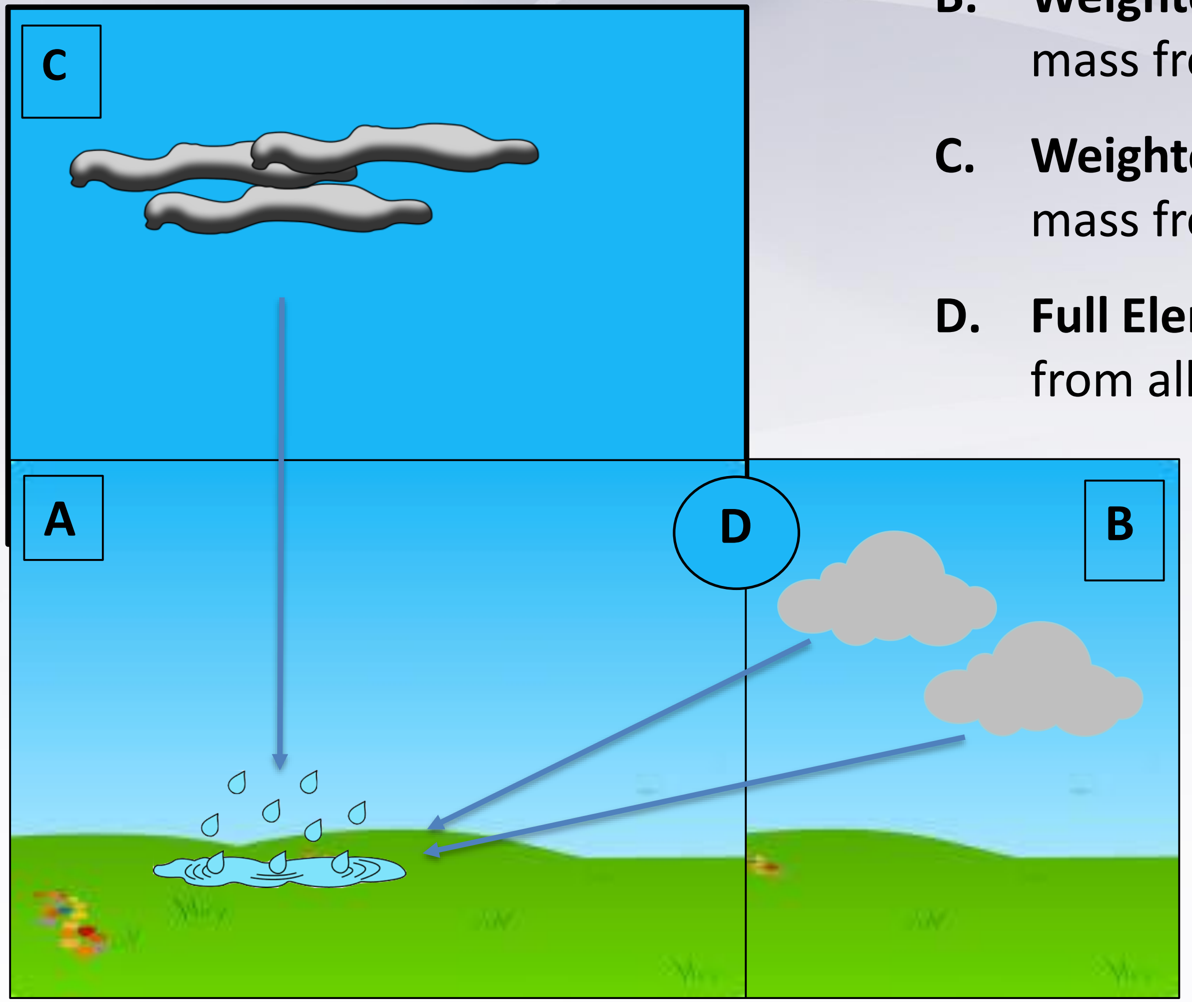
Mass Conservation



This problem is encountered any time parallel time splitting is used

- Two processes can independently remove mass from the same point, which when compounded can lead to an overconsumption of the local resources, i.e. negative mass.
- At the phys/dyn level this occurs frequently, as much as 1/3 of the pts. for liquid cloud water.
- **We currently just set negative values to zero.**
  - Note, this actually occurs in sequential tendency splitting in SE dynamics as well
  - "Clipping" negative tendencies violates conservation.
  - Is this acceptable? Is there a better way?

Mass Conservation



A. **Clipping:** Setting all negative masses to zero.

B. **Weighted Horizontal Distribution:** Drawing mass from neighboring nodes horizontally.

C. **Weighted Vertical Distribution:** Drawing mass from neighboring levels vertically.

D. **Full Element Distribution:** Drawing mass from all points within an element.

Using a weighted distribution approach will preserve global mass more effectively!

E. **Tendency Advection:** Apply dynamics to the physics tendencies.