



University of  
St Andrews | FOUNDED  
1413 |

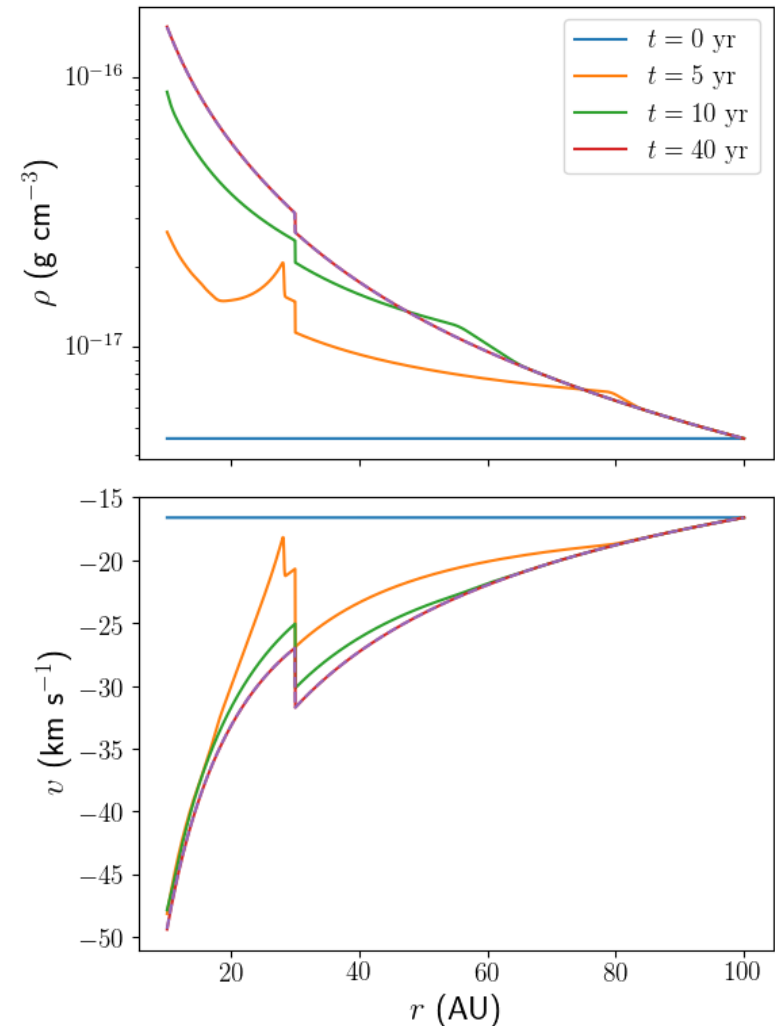
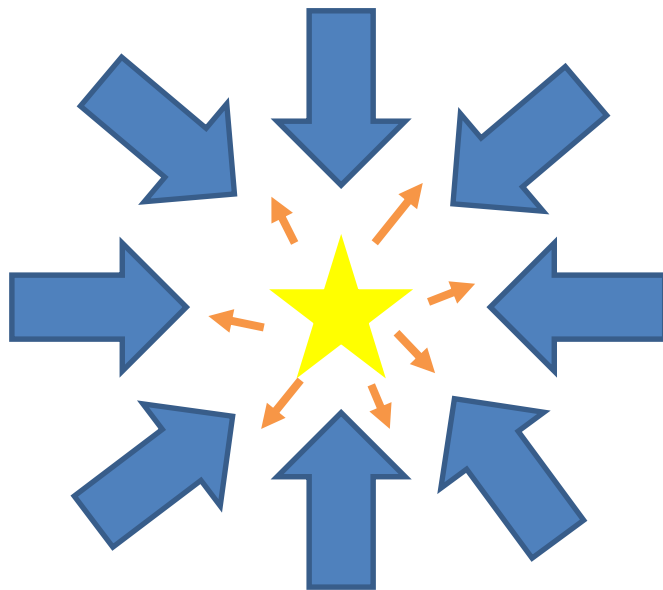
# Workflow Management Systems

Bert Vandenbroucke

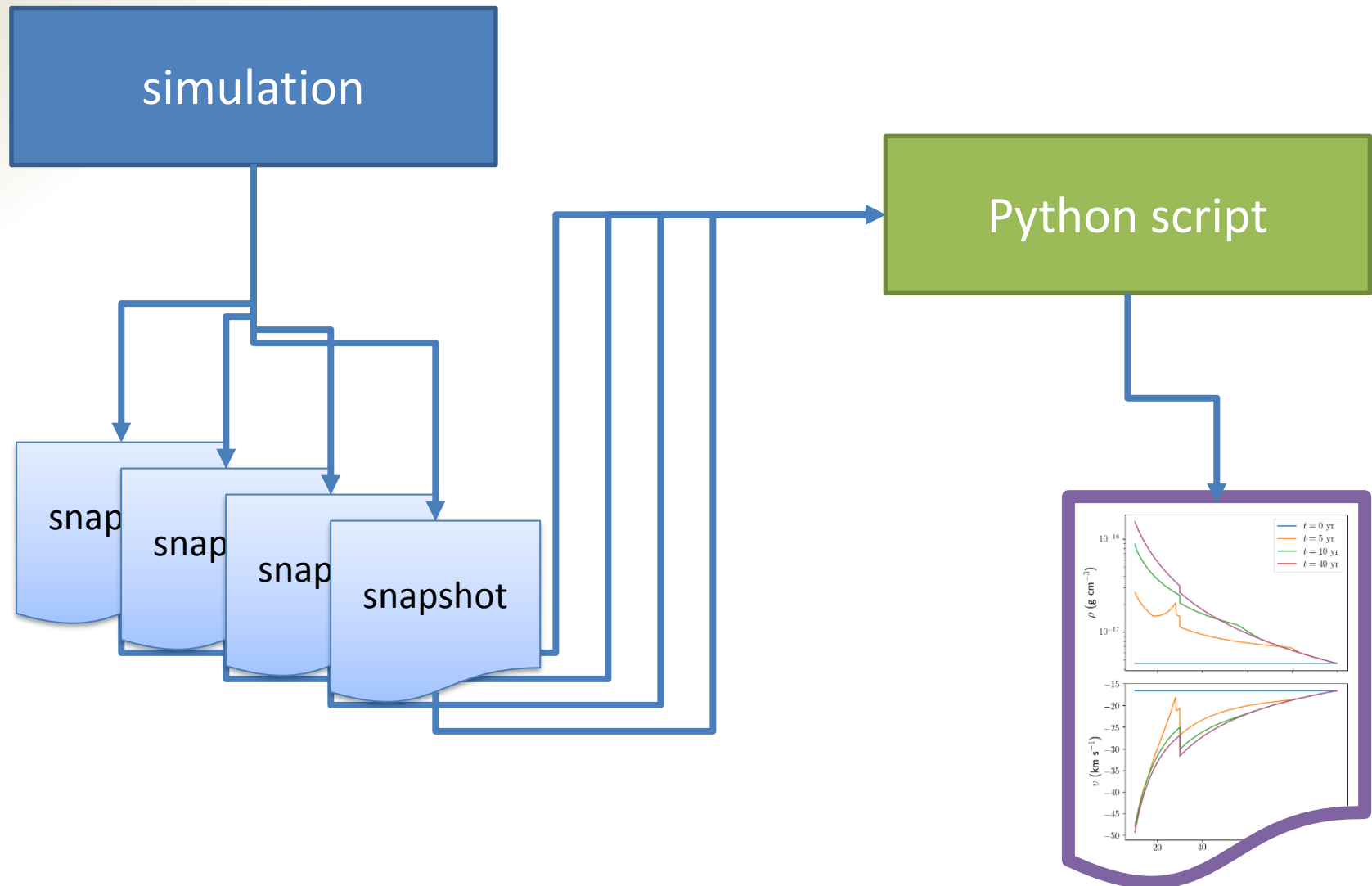
[bv7@st-andrews.ac.uk](mailto:bv7@st-andrews.ac.uk)

# Workflows: an example

The two state Bondi problem



# A basic workflow



# What are workflows?

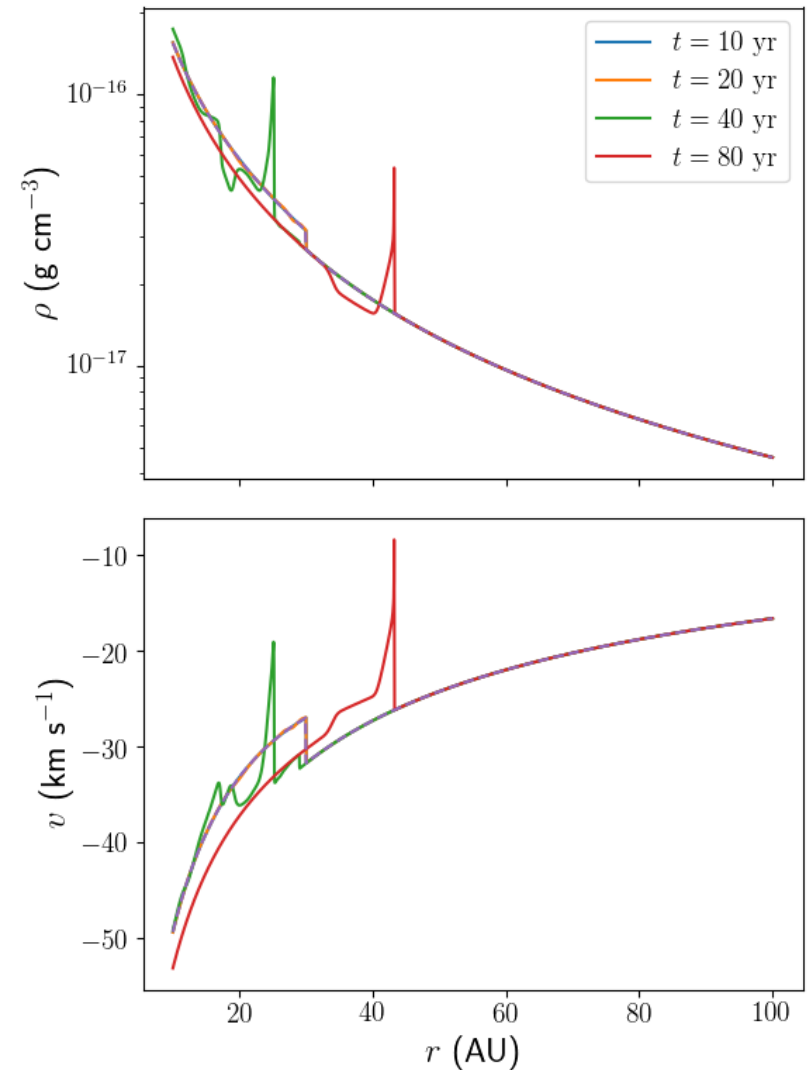
The way to go from INPUT to OUTPUT



# A more advanced example

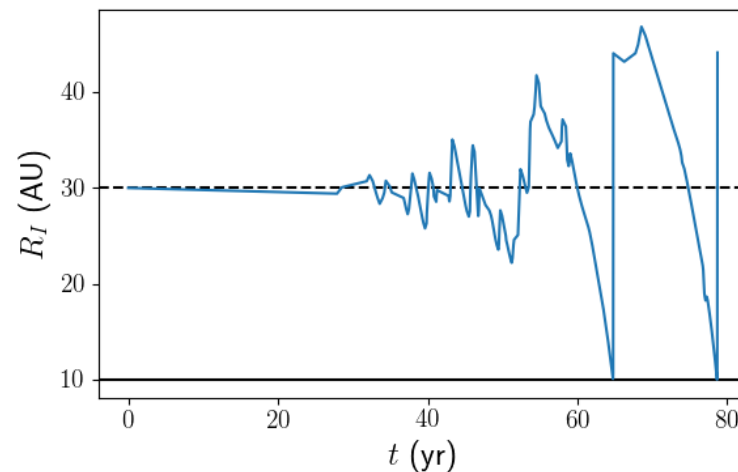
Stability test for the two state  
Bondi problem

Uses the output of the other  
simulation as input

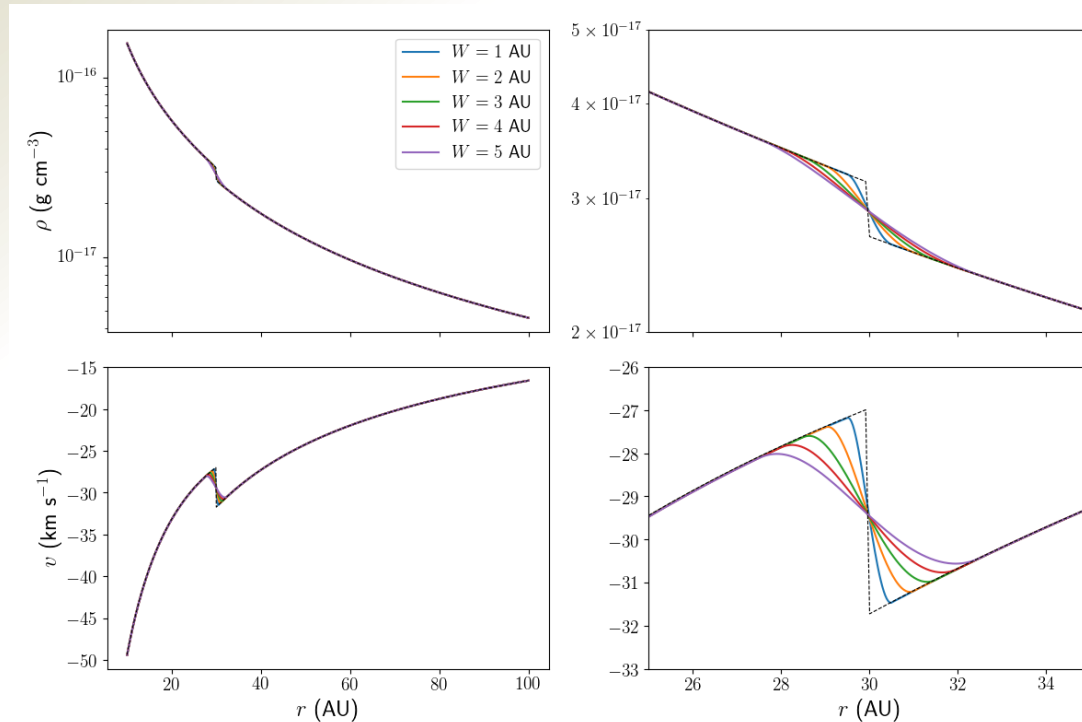




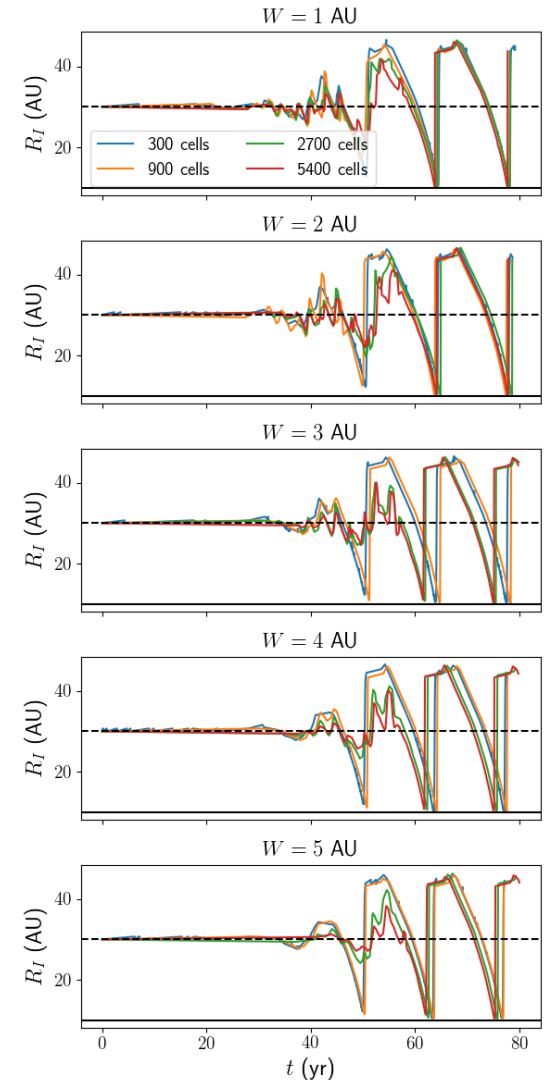
# Even more advanced: convergence



# Even more advanced: convergence

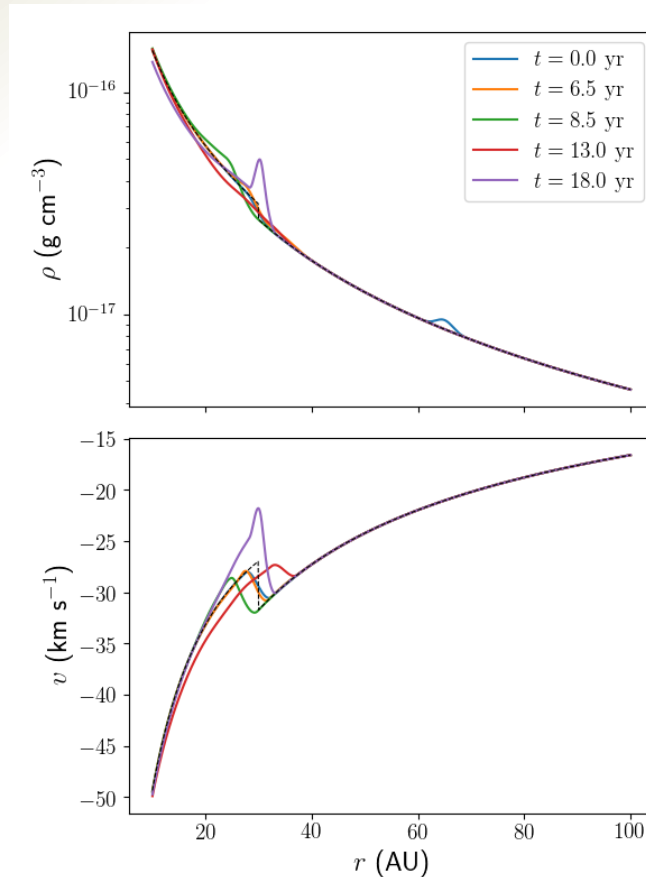


Set up smooth transitions with different widths  
This suppresses the instability for 30 years

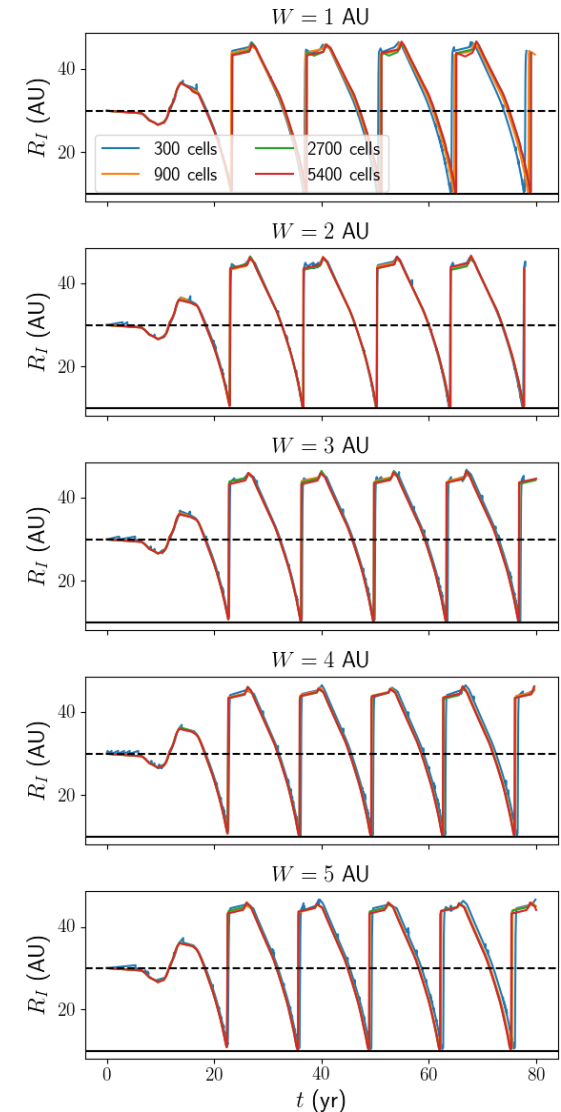




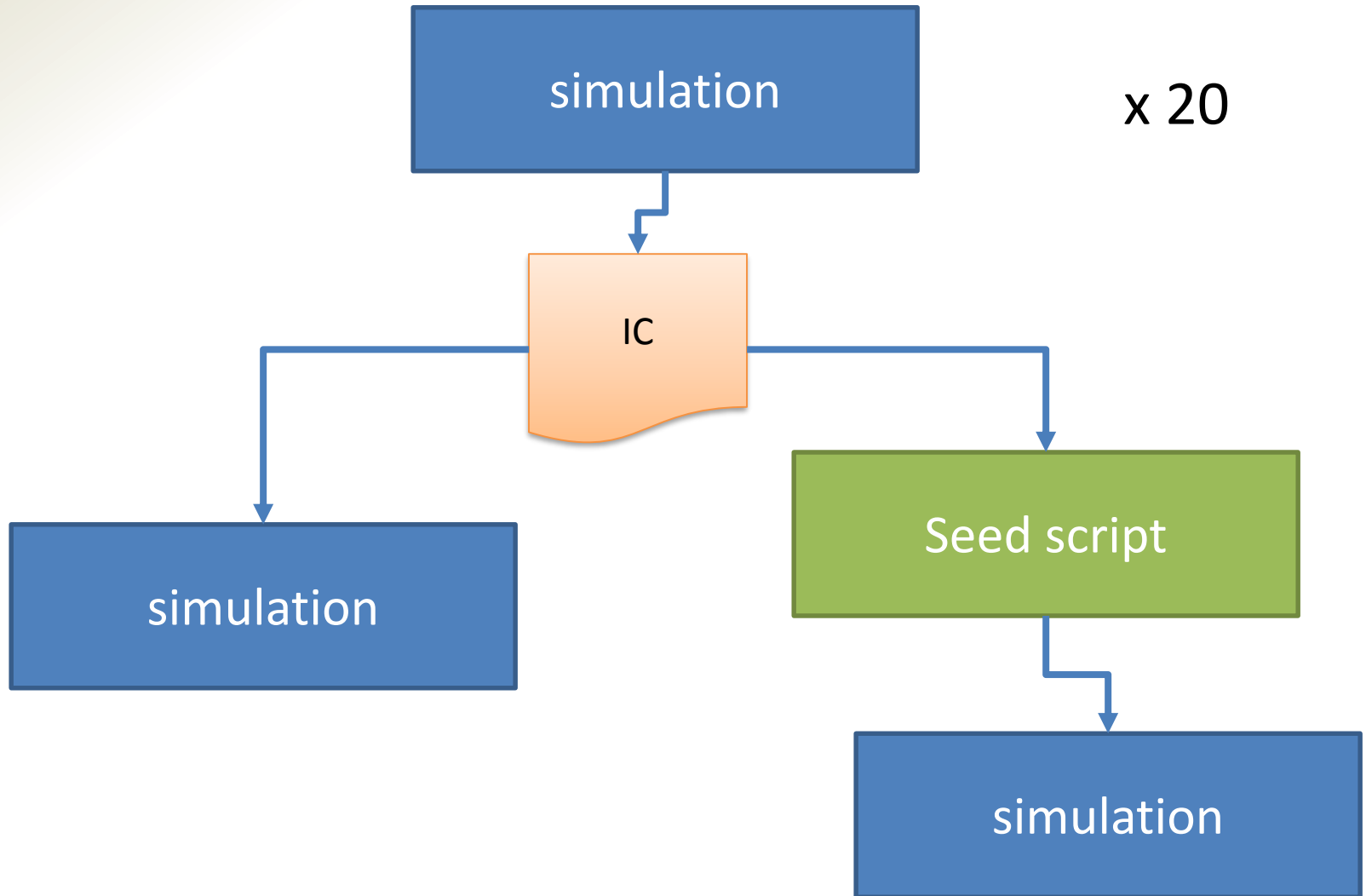
# Even more advanced: convergence



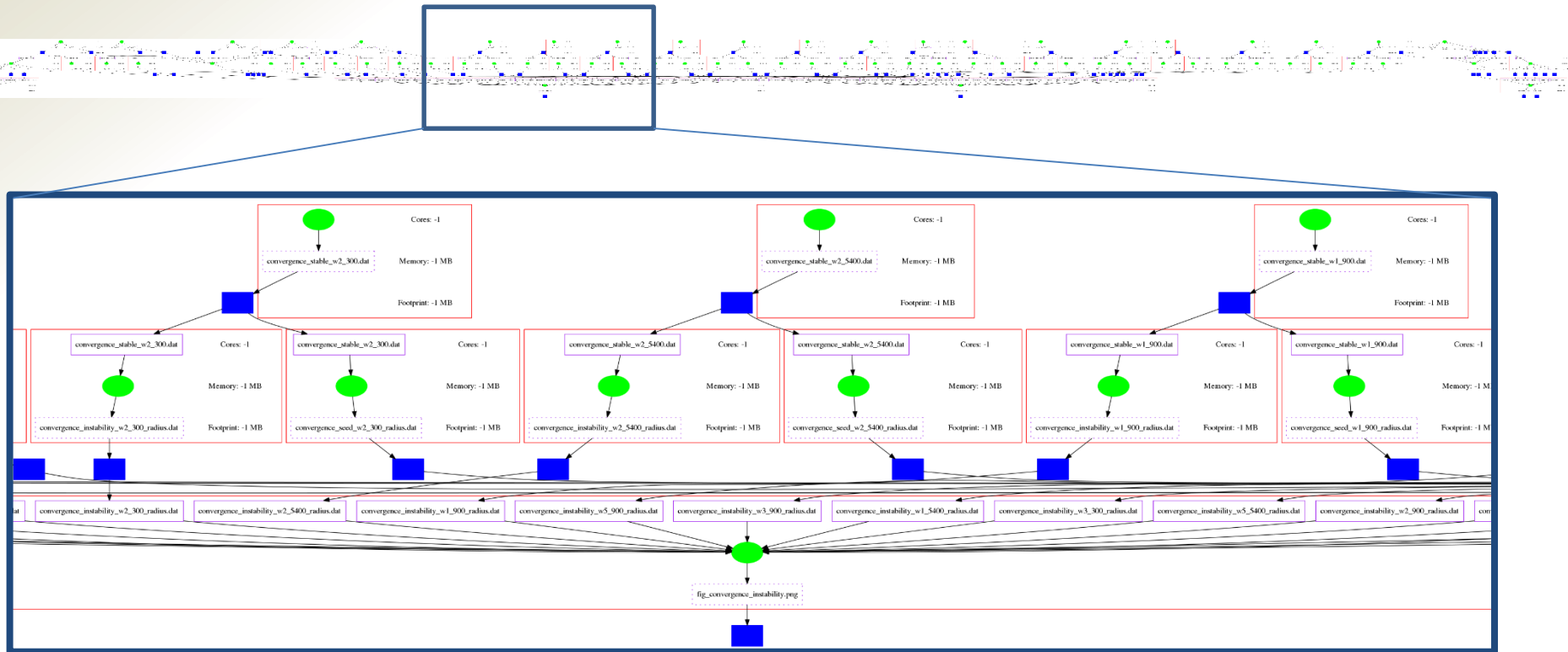
Add a density perturbation to seed a controlled instability



# Advanced workflow



# The workflow



68 different tasks with various complexities

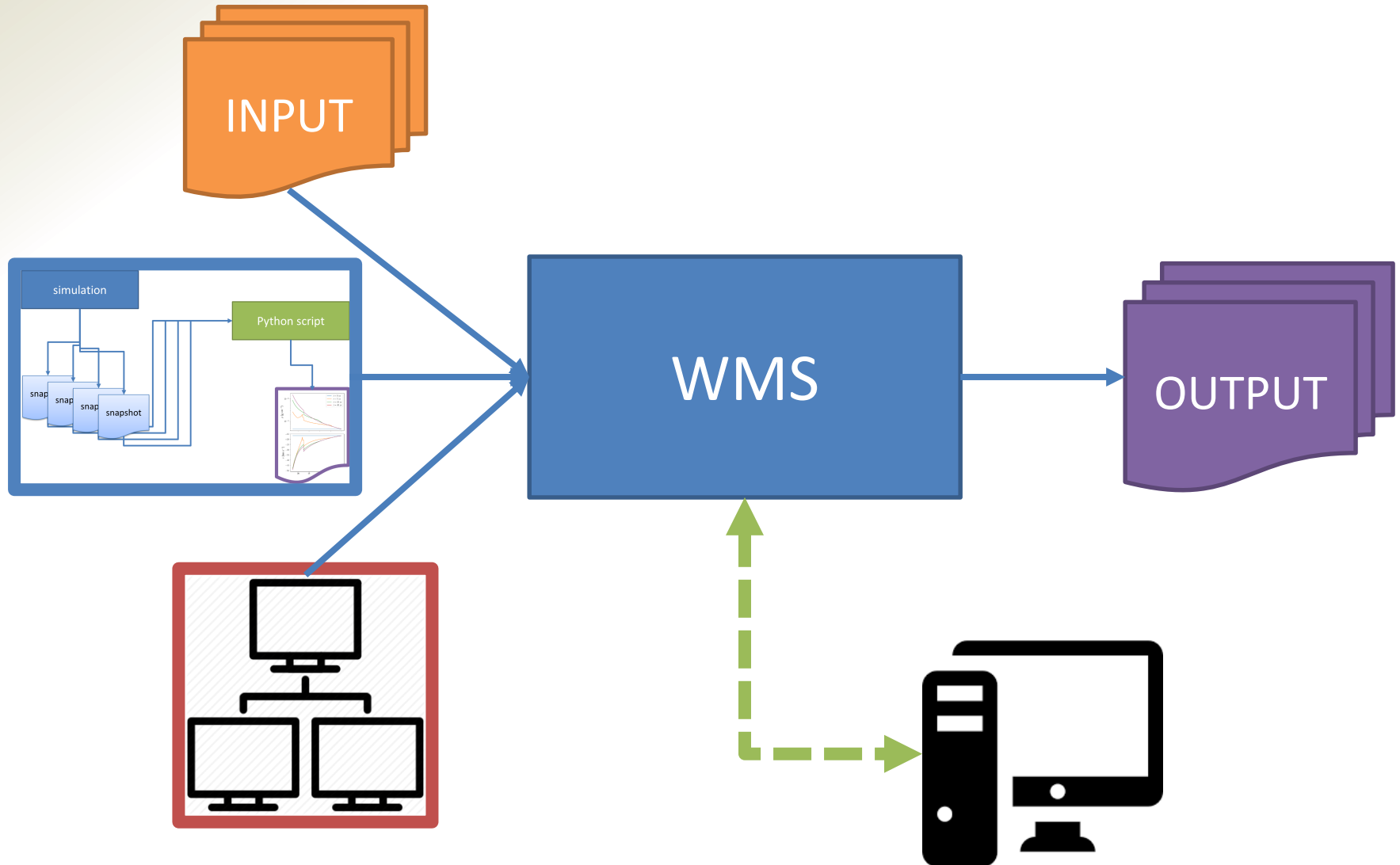
6 input files

80 intermediate output files

total computing time: 6 hours

end result: 8 figures

# Workflow management systems



# The advantages

- Efficiency: no manual steps
- Reproducibility
- Robustness:
  - automatic retry on hardware crash
  - checkpointing of current state
- Reliable dependency tracking: WMS will only rerun dependent tasks if some inputs change
- Parallelization on available resources

# The disadvantages

- Relatively poor support
- No satisfying handling of code inputs (yet)  
BUT support for containers (Docker...)
- No general standard: different systems use different language/interface, have different complexities and features

# Example WMS: Makeflow



## MAKEFLOW

Reads .makeflow workflow files  
Submits jobs to a queue system

## WORK\_QUEUE

Queueing system  
Connects to makeflow process and runs  
jobs on available hardware

# Example WMS: Makeflow

```
example.makeflow
```

```
output1: input1  
<TAB>command1
```

```
output2: output1  
<TAB>command2
```

Same syntax as GNU make (subset!)

Jobs are rules

Dependencies are set by output and input specifiers

Some basic support for resource management



# Example Makeflow file

```
CATEGORY="plot"
CORES=1
DISK=1000
MEMORY=1000

figure_stable_solution.png figure_stable_solution_reldiff.png:
    stable_solution_t00.txt stable_solution_t05.txt stable_solution_t10.txt
    stable_solution_t40.txt fig_stable_solution.py
    LOCAL python fig_stable_solution.py

[...]

CATEGORY="simulation"
CORES=16
DISK=1000
MEMORY=1000

stable_solution.dat->lastsnap.dat stable_solution_t00.txt->snapshot_0000.txt
stable_solution_t05.txt->snapshot_0250.txt stable_solution_t10.txt->
snapshot_0500.txt stable_solution_t40.txt->snapshot_2000.txt:
    [...]
```

# Example Makeflow run

MAKEFLOW

WORK\_QUEUE

```

Terminal
bv7@apsient:/data/bv7/HydroCodeSpherical1D/paper_workflows$ makeflow -T wq creat
e_paper_figures.makeflow -p 9000

Terminal
parsing create_paper_figures.makeflow...
checking create_paper_figures.makeflow for consistency...
create_paper_figures.makeflow has 68 rules.
starting workflow...
listening for workers on port 9000.
submitting job: git clone https://github.com/bwvndbro/HydroCodeSpherical1D.git s
ource; cmake -DCMAKE_BUILD_TYPE=Release -Dionisation_transition_width_in_au=5.0
-Drmax_in_au=100.0 -Dncell=5400 -Dbondi_pressure_contrast=32.0 -Dinitial_ionisat
ion_radius_in_au=30.0 -Drmin_in_au=10.0 -Dic=IC_BONDI -Driemannsolver_type=RIEMA
NNSOLVER_TYPE_HLLC -Dnumber_of_snaps=2000 -Dunit_length_in_si=1.2e+13 -Dboundari
es=BOUNDARIES_BONDI -Dcourant_factor=0.05 -Dpotential=POTENTIAL_POINT_MASS -Dic_
file_name=ic.dat -Dbondi_density_in_si=1e-16 -Dionisation_transition=IONISATION_
TRANSITION_SMOOTH -Dg_internal=1.0 -Dunit_mass_in_si=2.479e+31 -Dionisation_mode
=IONISATION_MODE_CONSTANT -Dpos_FOS_BONDI -Disothermal_temperature_in_k=500.0 -D
submitting job: python fig_convergence_instability.py
submitted job 9602
Plotting convergence_instability_w1_300_radius.dat ...
Plotting convergence_instability_w1_900_radius.dat ...
Plotting convergence_instability_w1_2700_radius.dat ...
Plotting convergence_instability_w1_5400_radius.dat ...
Plotting convergence_instability_w2_300_radius.dat ...
Plotting convergence_instability_w2_900_radius.dat ...
Plotting convergence_instability_w2_2700_radius.dat ...
Plotting convergence_instability_w2_5400_radius.dat ...
Plotting convergence_instability_w3_300_radius.dat ...
Plotting convergence_instability_w3_900_radius.dat ...
Plotting convergence_instability_w3_2700_radius.dat ...
Plotting convergence_instability_w3_5400_radius.dat ...
Plotting convergence_instability_w4_300_radius.dat ...
Plotting convergence_instability_w4_900_radius.dat ...
Plotting convergence_instability_w4_2700_radius.dat ...
Plotting convergence_instability_w4_5400_radius.dat ...
Plotting convergence_instability_w5_300_radius.dat ...
Plotting convergence_instability_w5_900_radius.dat ...
Plotting convergence_instability_w5_2700_radius.dat ...
Plotting convergence_instability_w5_5400_radius.dat ...
nothing left to do.
(END)

```

```

Terminal
bv7@apsient:/data/bv7/HydroCodeSpherical1D/paper_workflows$ work_queue_worker --
cores 16 --memory 10000 --disk 100000 localhost 9000
work_queue_worker: creating workspace /tmp/worker-110090-7433
couldn't connect to 127.0.0.1:9000: Connection refused
couldn't connect to 127.0.0.1:9000: Connection refused

```

# Makeflow extras

- Visualization of .makeflow file with `makeflow_viz`
- Analysis of workflow with `makeflow_analyse`
- Monitoring of execution with `makeflow_monitor`, `work_queue_status` and `work_queue_graph_log`

# Other workflow systems

- Pegasus WMS: <https://pegasus.isi.edu/>  
used for e.g. LIGO analysis
- Copernicus: <http://copernicus-computing.org/>
- Kepler: <https://kepler-project.org/>
- ...

# Conclusion

- Workflows are what you do every day
- Workflow management systems help you do them
- Work well for small problems with lots of tasks on own hardware if you get it to work
- Look very promising for future of HPC

# Extra resources

- Bondi workflow:  
[https://github.com/bwvdbro/HydroCodeSpherical1D/tree/master/paper\\_workflows](https://github.com/bwvdbro/HydroCodeSpherical1D/tree/master/paper_workflows)  
(uses Makeflow WMS)
- Blue Waters WMS seminar series:  
<https://bluewaters.ncsa.illinois.edu/webinars/workflows>
- bv7@st-andrews.ac.uk