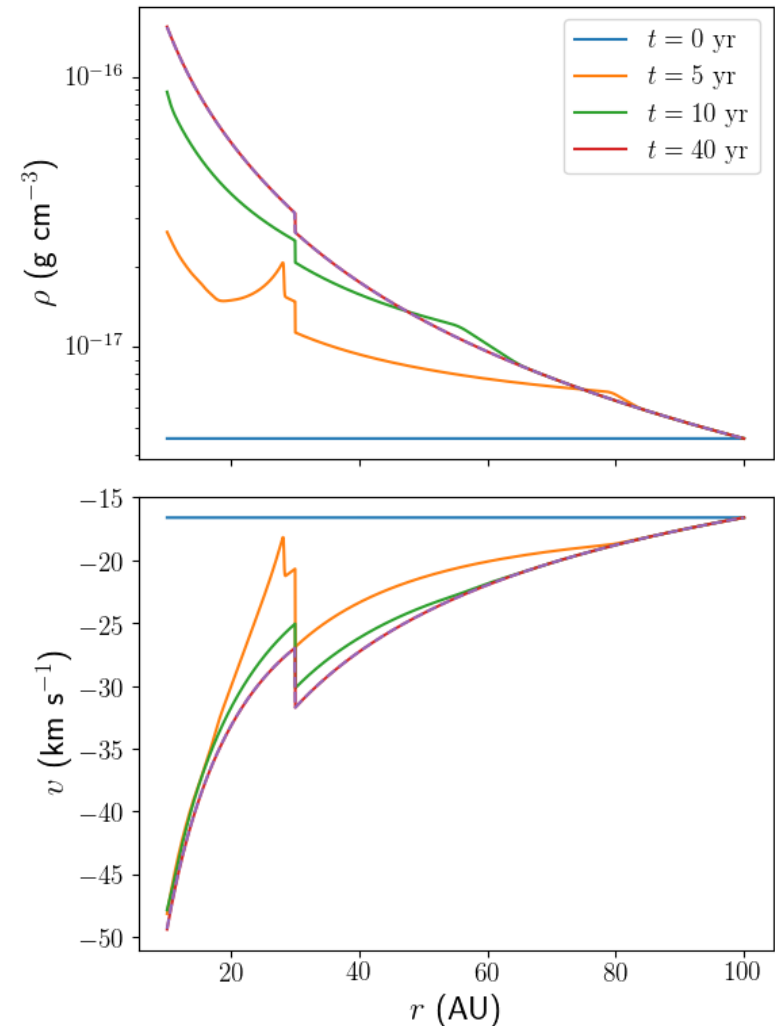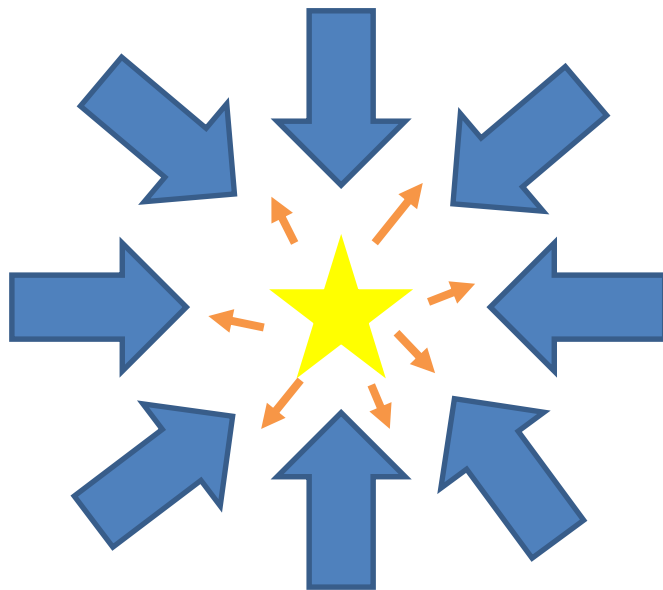# Workflow Management Systems

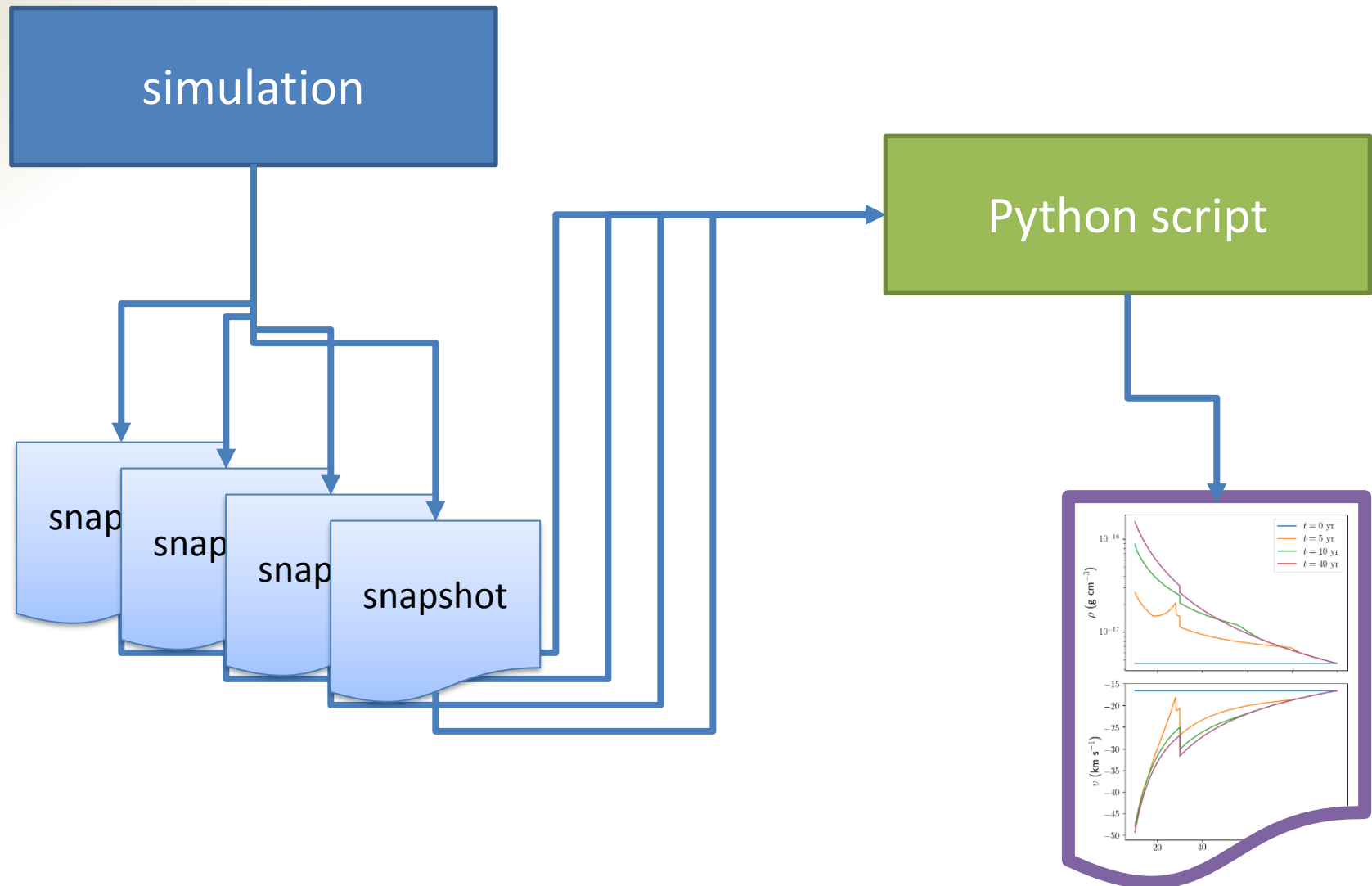## Bert Vandenbroucke

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

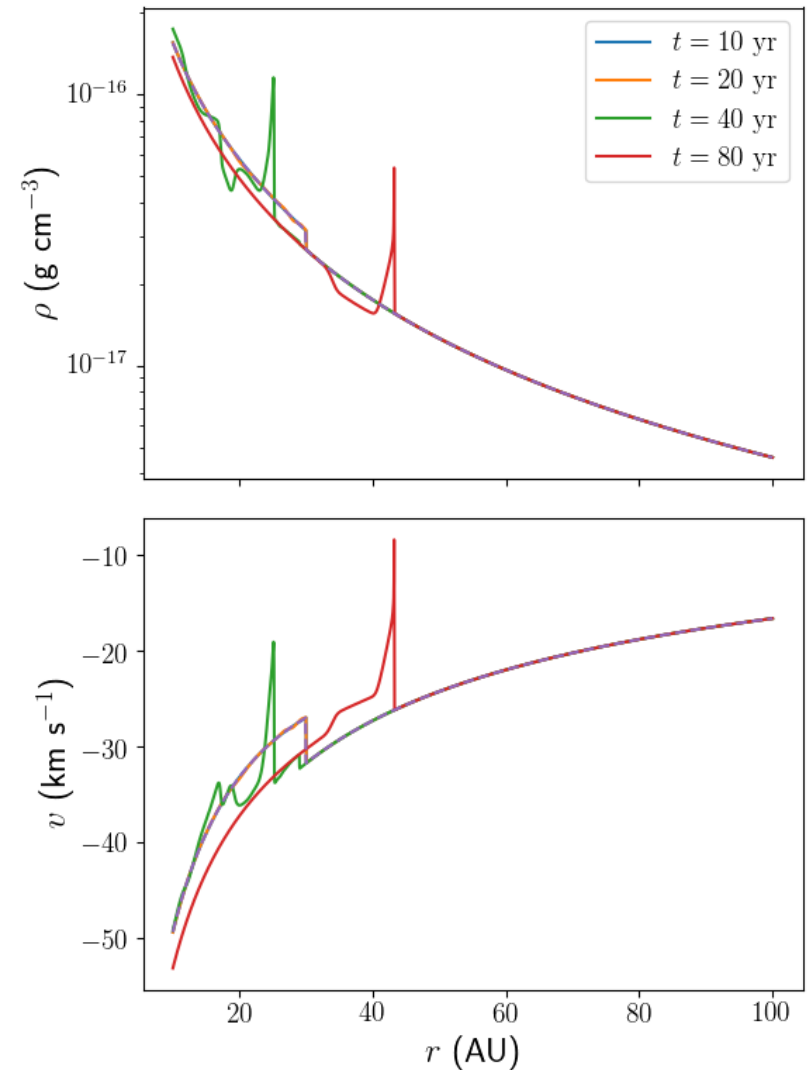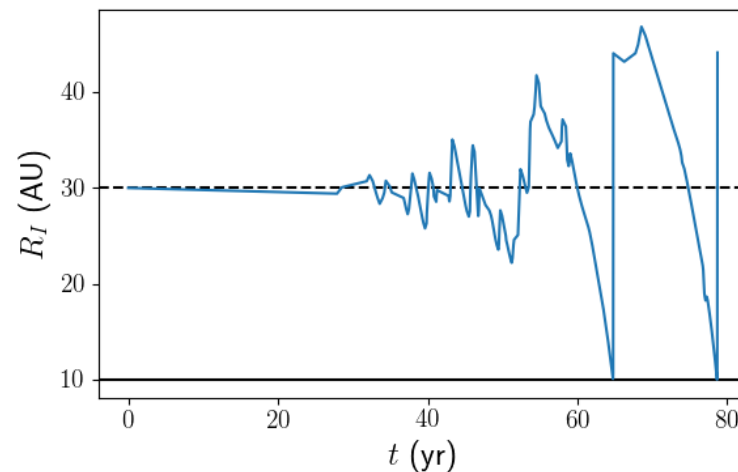INPUT — WORKFLOW → OUTPUT

A more advanced example

Stability test for the two state Bondi problem

Uses the output of the other simulation as input
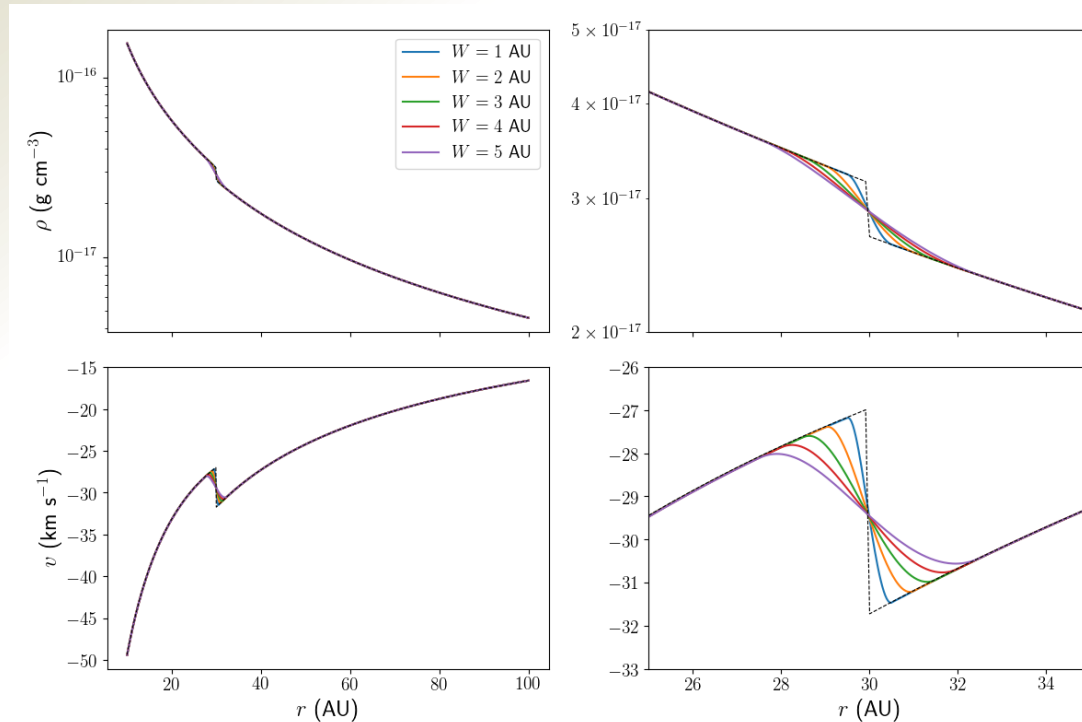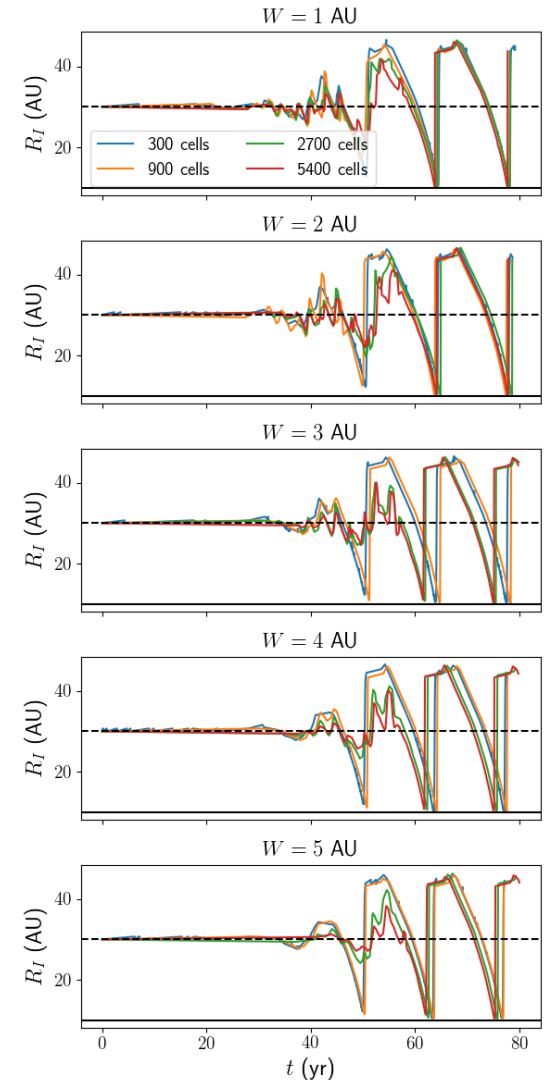
A more advanced workflow

Even more advanced: convergence

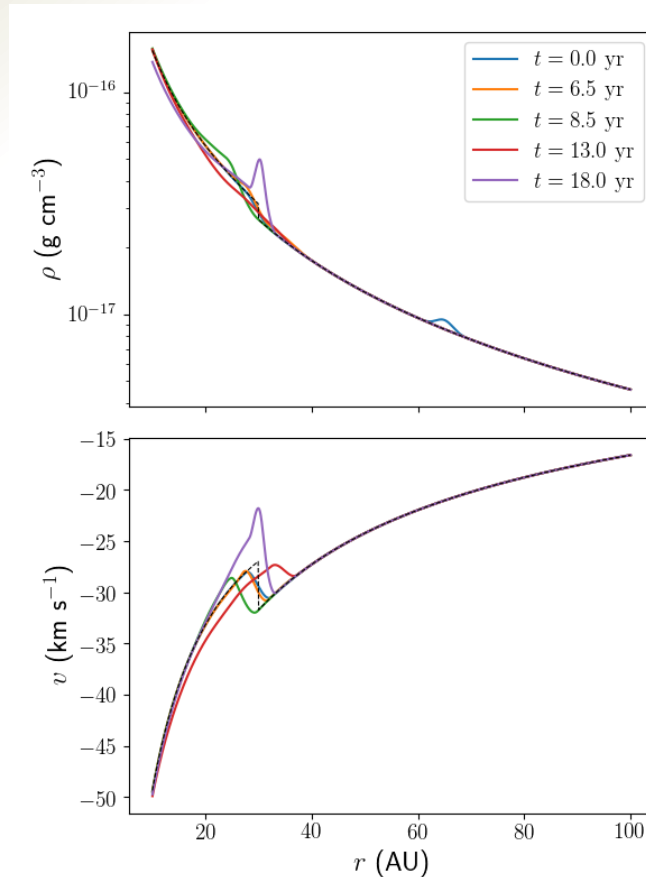# Even more advanced: convergence

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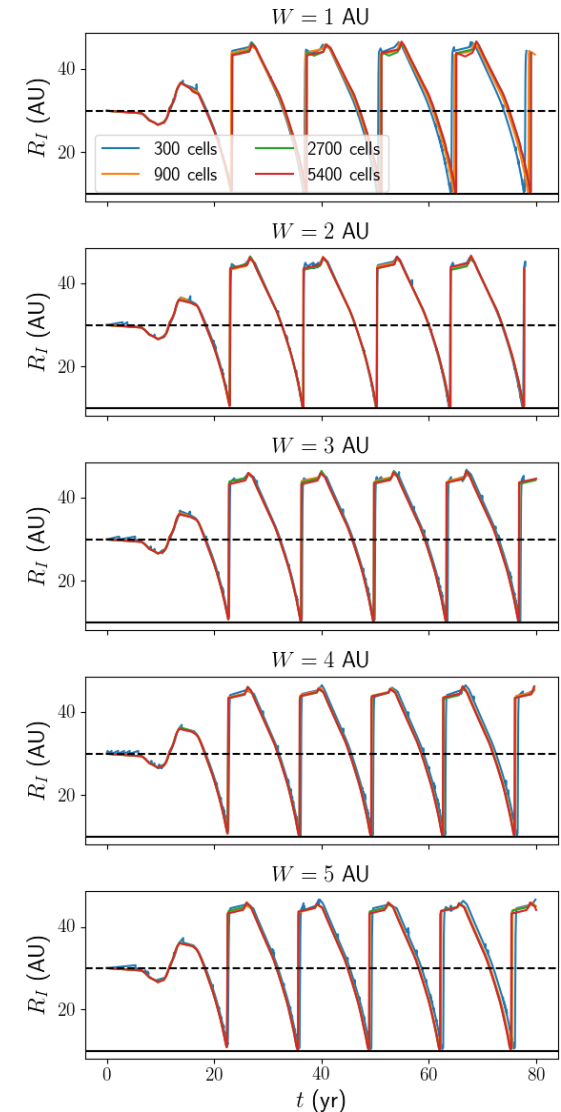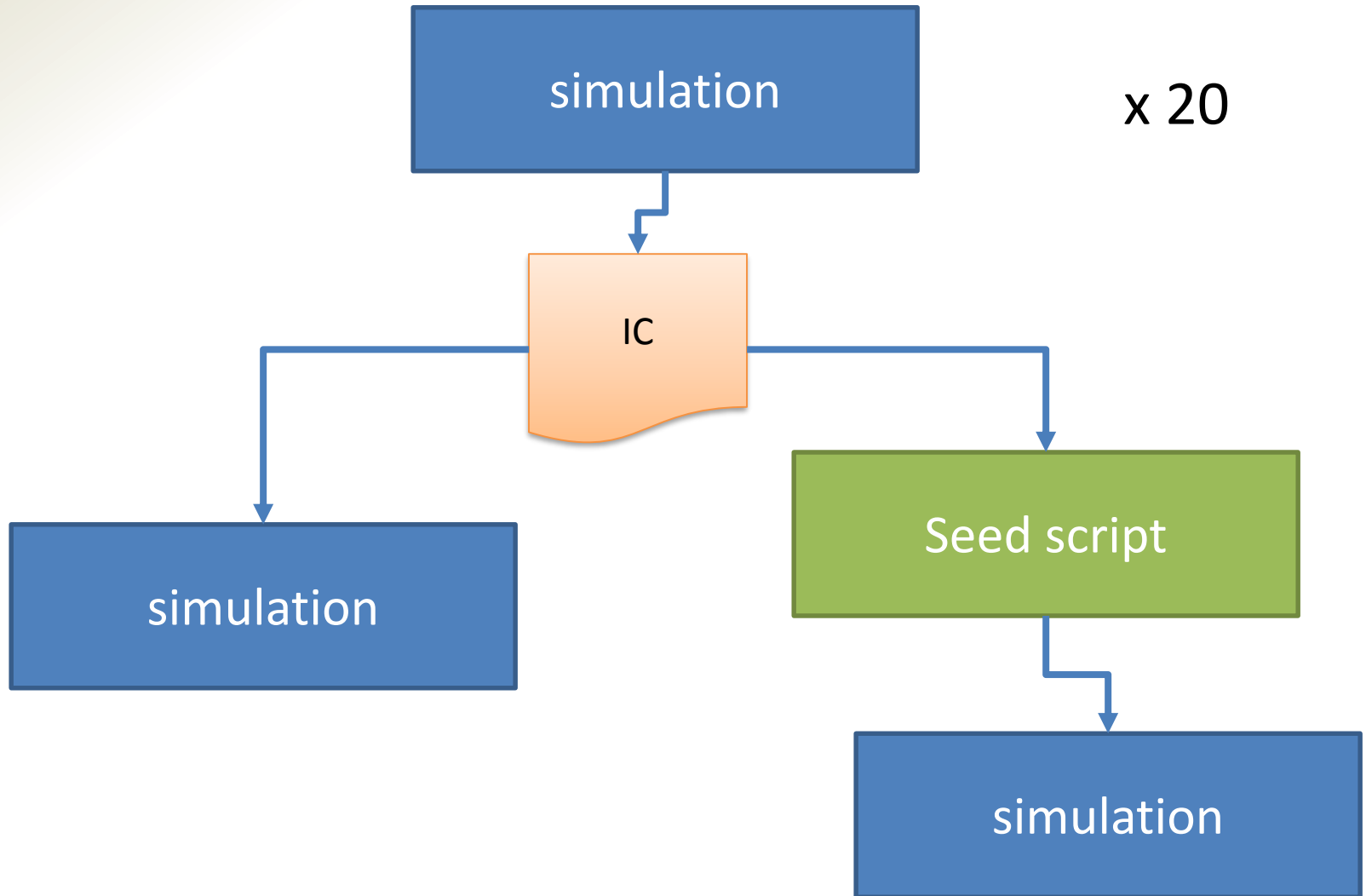This suppresses the instability for 30 years

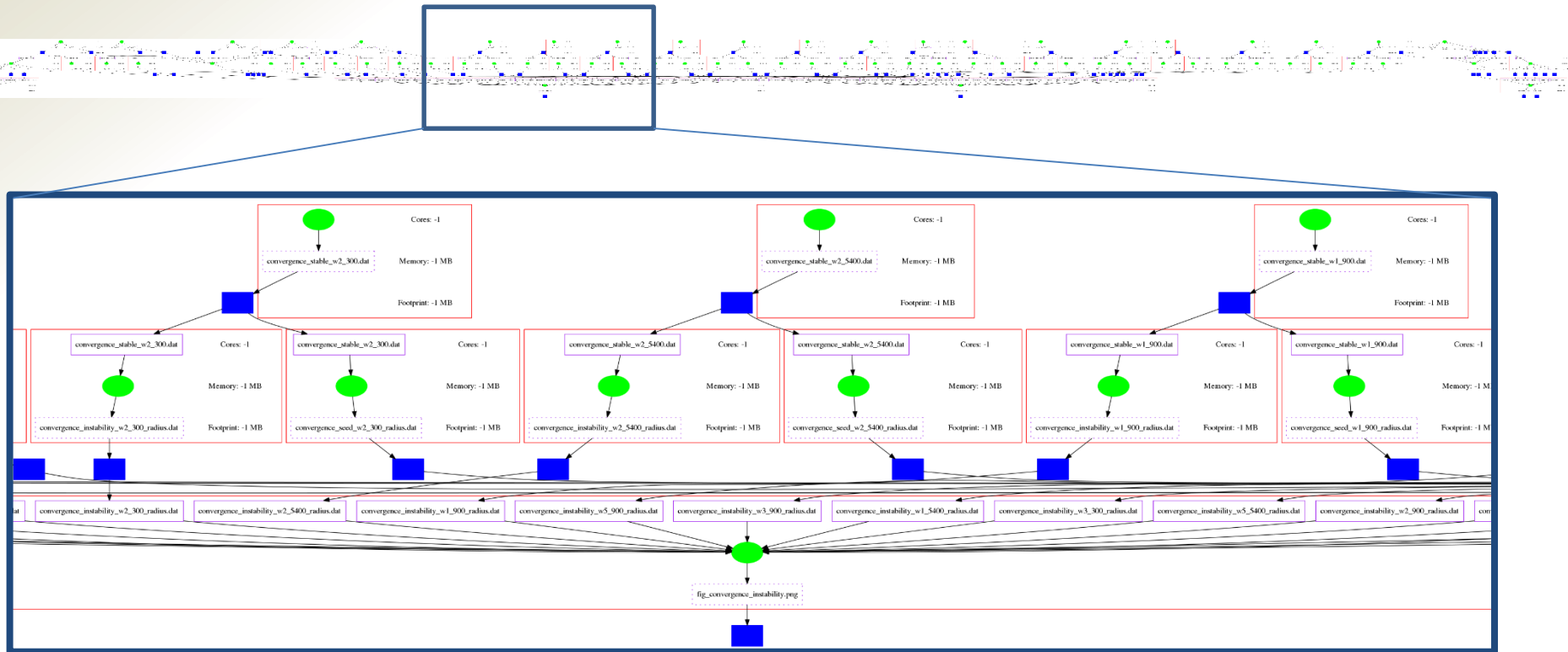# Even more advanced: convergence

Add a density perturbation to seed a controlled instability

Advanced workflow

The workflow

68 different tasks with various complexities
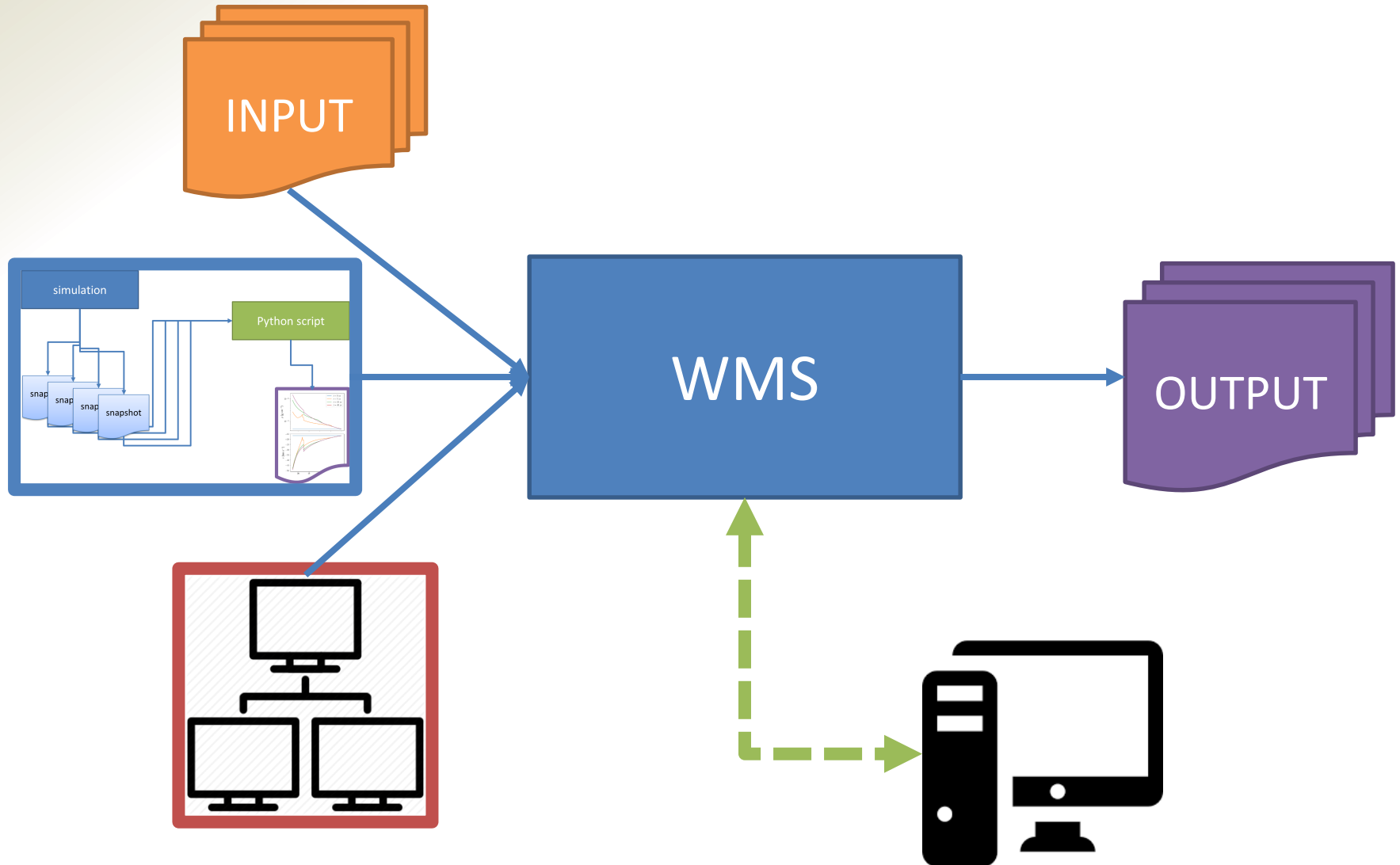
6 input files   80 intermediate output files

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

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/bwvdnbro/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