Data management and visualization

Contents

What a time dependent simulation looks like

description of the problem

Data management

scripts to keep track of files, tasks, and machines

Visualization

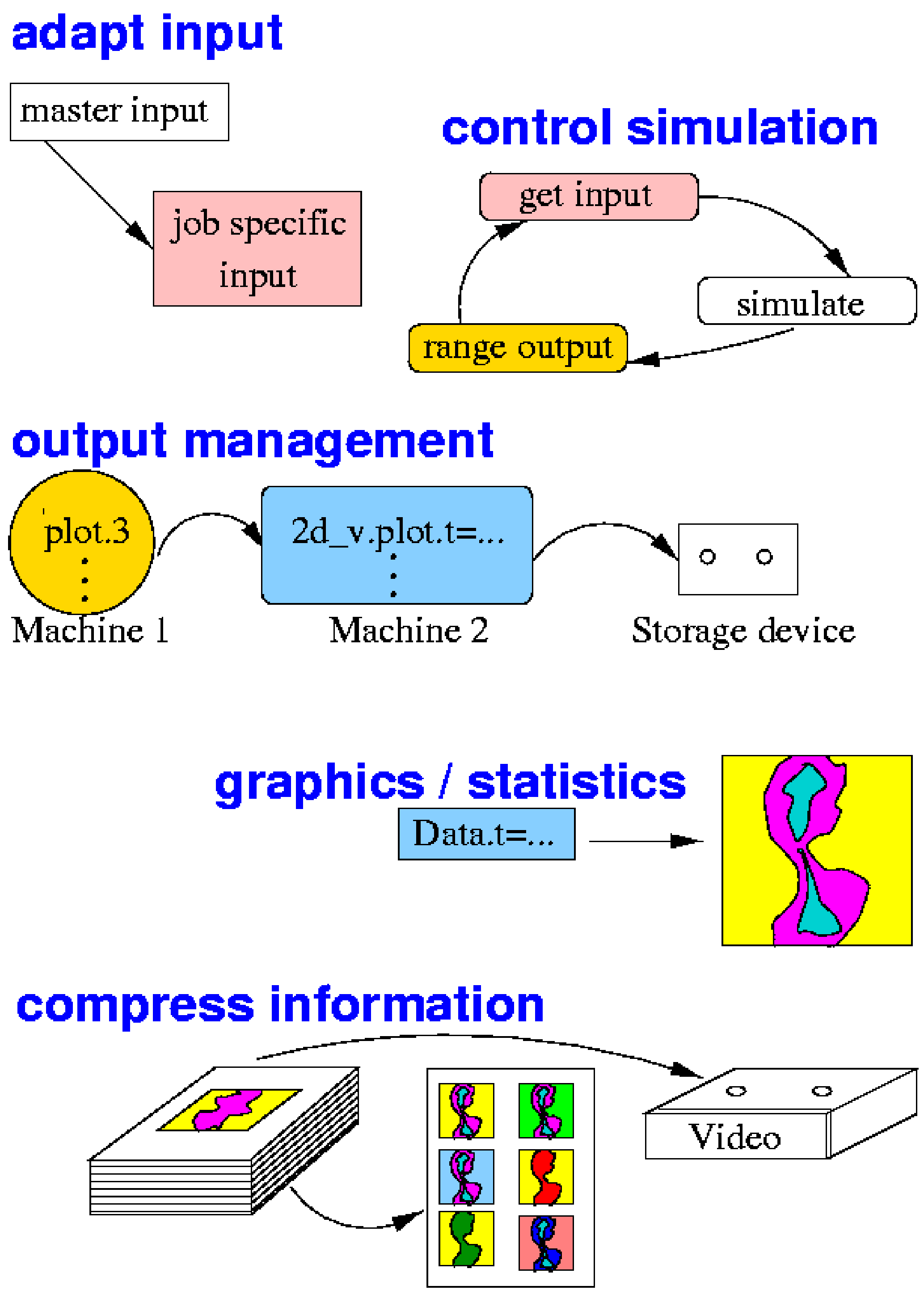
various ways from numbers to pictures to videos

What a time dependent simulation looks like

What your simulation looks like

Need:

- different machines for computing, storage, visualization
- a tape drive
- visualization tools
- scripts to organize it all automatically (by hand, you sooner or later do it wrong...)



What you get: a huge amount of data and files

Example:
1D MHD, time evolution of 1D self- gravitating slab + Alfvén- wave:

5000 cells, 7 variables per cell
10^7 - 10^8 time steps (≈ 4 days of computation on Dec alpha)
of these data 10^4 time steps to disk → 2.5 GB of data
these time series data are organized in a few 100 files

Amount of data: intermediate, you may leave it on disk
Number of files: intermediate, but too large to administer by hand
(you make too many errors)

Example:
3D hydro, time evolution of WR+O binary star system:

$4 * 10^6$ cells, 5 variables per cell

10^3 coarse grid time steps (about 10^6 times more on finest grid,
runs for a few weeks on Dec alpha)

each coarse grid time step to disk $\rightarrow$ 160 GB of data

these time series data are organized in a few 100 files

Amount of data: large, you need a tape device for data storage
Number of files: intermediate, but too large to administer by hand
(you make too many errors)

DATA MANAGEMENT

A- MAZE provides a bunch of scripts for...

Automatically change / generate input files
(e.g. tape5.1,... ,tape5.200)

Administering the running of a time- series of jobs
(updating input files, running the job, moving output to another machine, annotating it, store it on tape)

Administering post- processing of data
(obtain statistics, transform into data formats readable by graphics package, run graphics jobs, administer input / output to graphics)

Depending on task, scripts are written in sh, csh, or perl.
These scripts alone are a few thousand lines of code.

How to keep track of files, tasks, machines?

A- MAZE provides a bunch of scripts for...

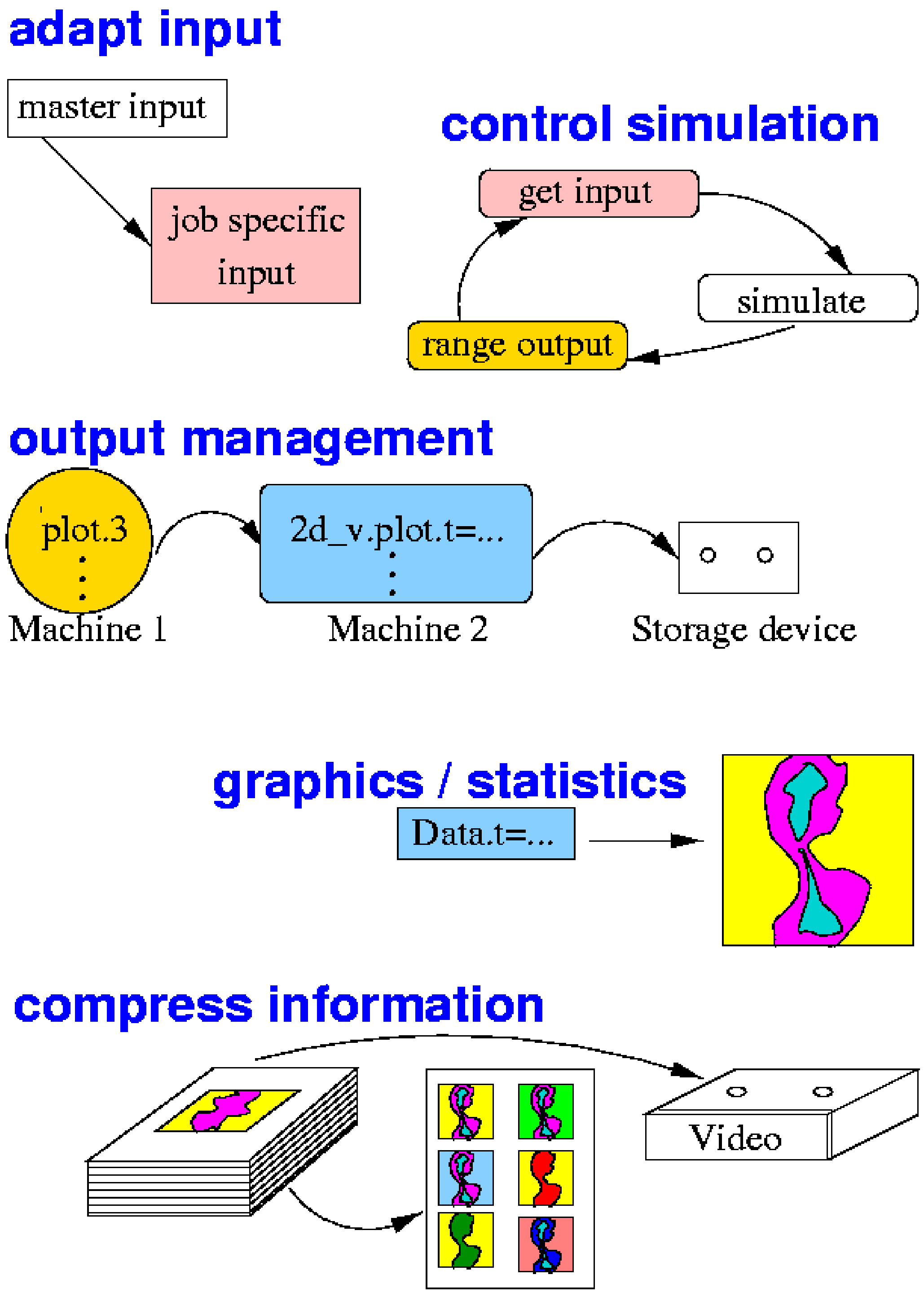
... Automatically changing / generating input files

... Administering the running of a time- series of jobs

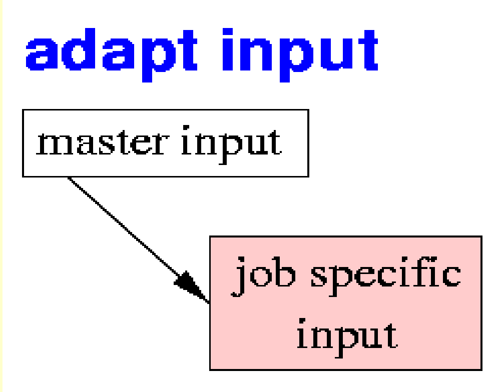
... Administering post-processing of data

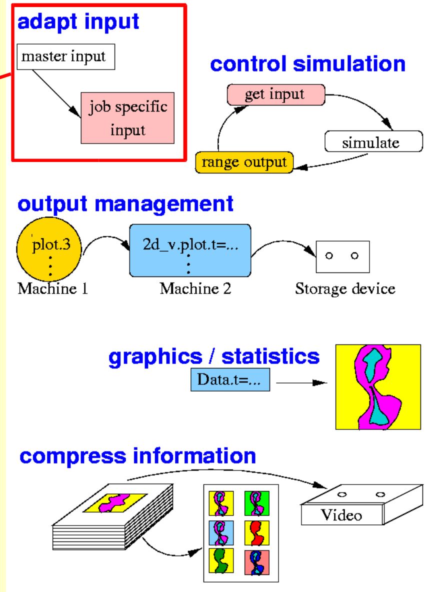
Depending on task, scripts are written in sh, csh, or perl.

These scripts alone are a few thousand lines of code.



Data- management 1



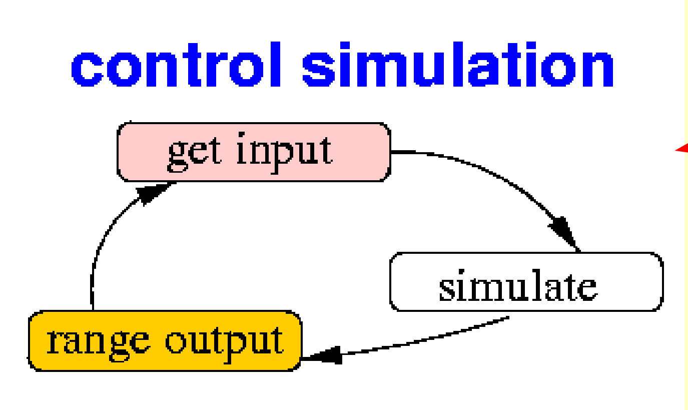


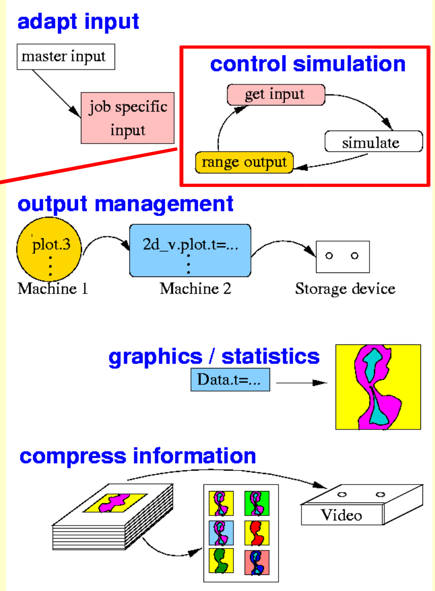
time- dependent simulations:

- must split your simulation into digestible pieces
  (some hours - 1 day, ~ # 100Mbytes of data)

- must generate necessary input- files (up to some 1000 perhaps)
  do this automatically by means of scripts (e.g. perl)

Data- management 2

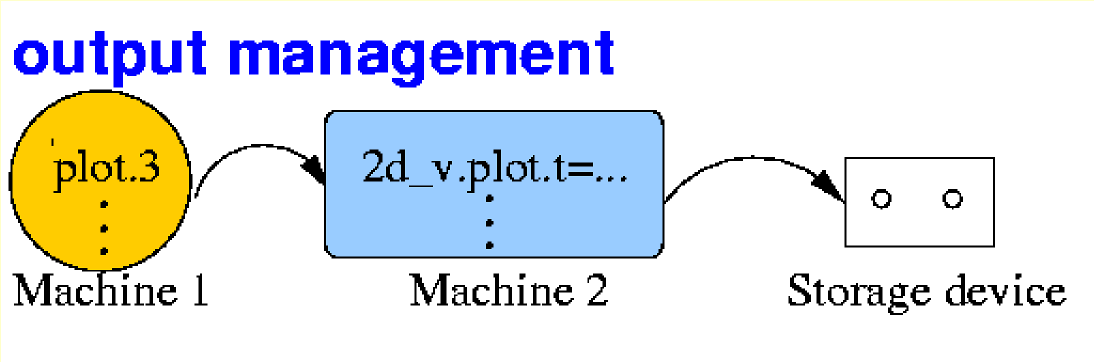


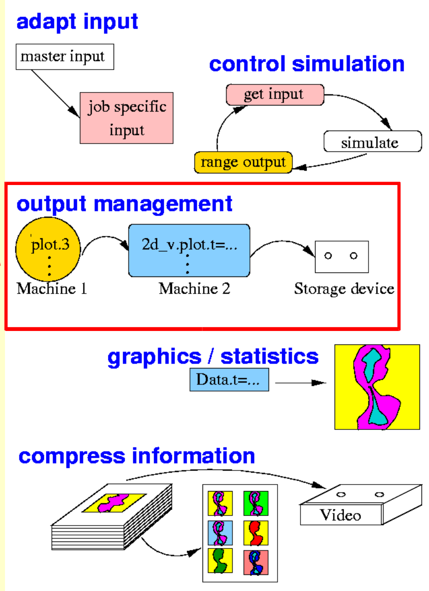


your time-dependent simulation may now run weeks or months:
a script controls the entire simulation

- submits the different part- jobs
- stops (and sends mail) if job dies
- manages the data- output
- does rudimentary graphics and data- analysis for controlling

Data- management 3



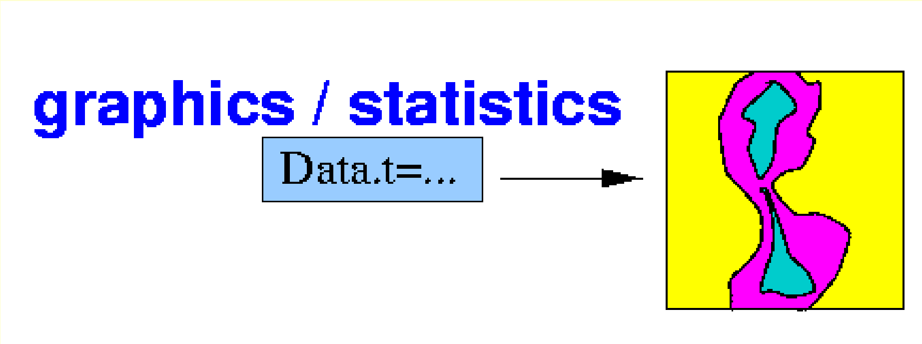


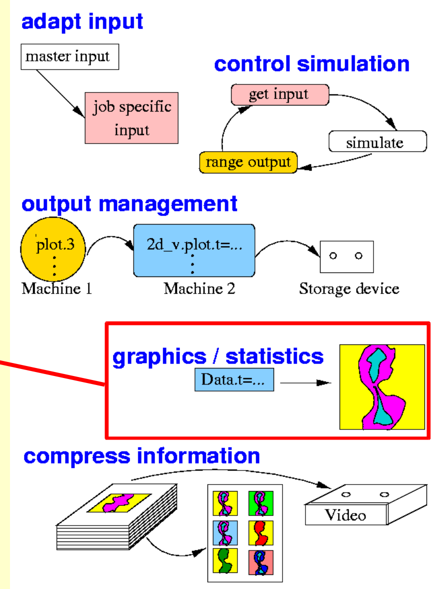
Now and then you have to get your data from the machine where your simulation runs and to store the data somewhere:

script (e.g. csh)

- transfers data from computing to a front- end machine (e.g. tcsh)
- annotes data (perl)
- transfers data to a mass- storage device (data- silo, tape)

Data- management 4



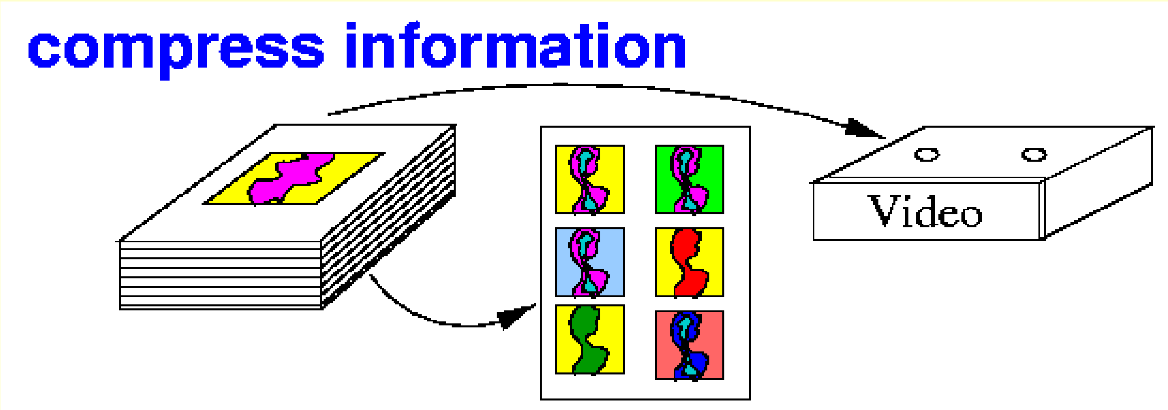


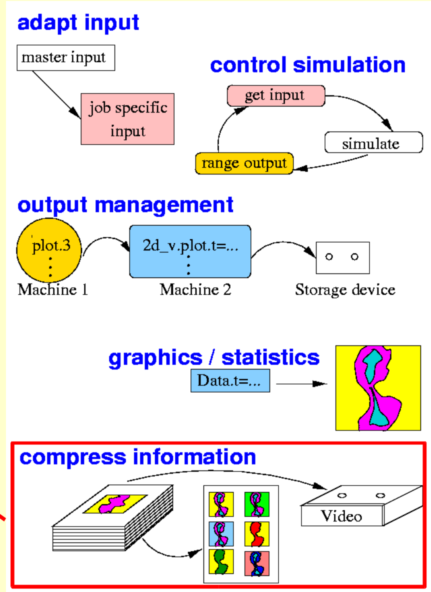
you finally will find some 100 GB of data on your storage device...
data- analyis starts now:

script to
- copy data to a graphics machine (e.g. csh)
- possibly run (part of) the graphics code (e.g. csh)
- possibly copy data to yet another machine and do statistics etc.
- range / annotate the output produced (perl)

Note: this post- processing of the data may again need an amount of CPU- time comparable tothe original computation !

Data- management 5





- Large amount of data asks for some form of 'data- compression' (for better understanding and publication)
- Time- dependent data asks for time- dependent analysis

Video is one possibility

(time dependent means or streamlines are some other possibilities)

Visualization

Visualization in A- MAZE

Problems we encounter:

- large amount of data
- large dynamical range (e.g. already a $1/r^2$ can be tedious)
- time dependence ('how does it flow?')
- various data formats, among them
  - block- structured meshes
  - sets of 3D rays with arbirarily spaced points

NCAR; AVS; IDL; Matlab; ....; What product to take?

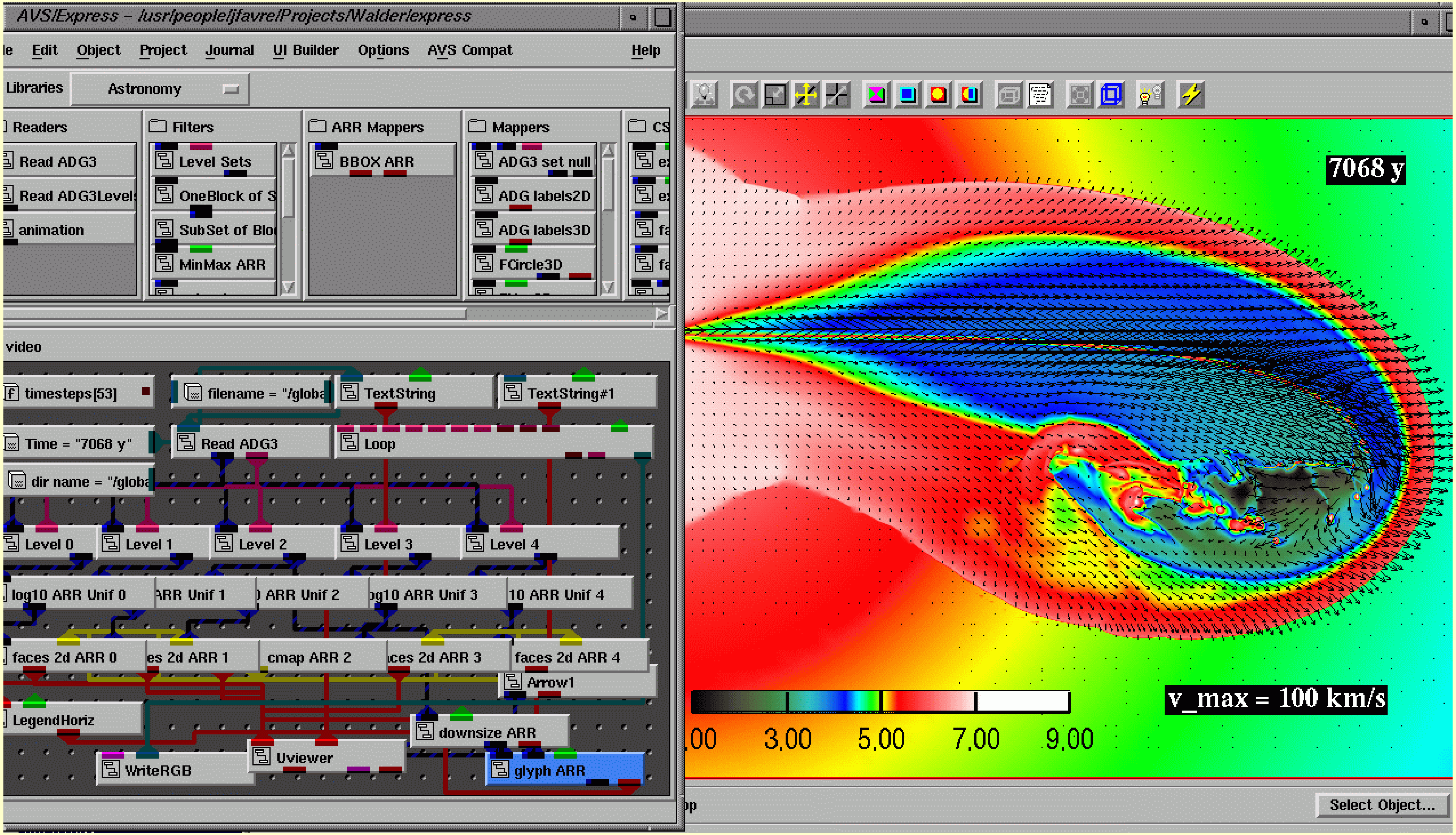
NCAR: (we use it for 2D contour plots)
Pros: easy to code yourself (fortran code), runs in background,
moderate machine ok, for free (http:// www.scd.ucar.edu/vets)
Cons: not interactive, moderate user interface and only for viewing

AVS: (we use it for 2D and 3D color plots and movies)
Pros: interactive, multi- color- tables, user interface
Cons: 'clumpsy', good machine required, difficult to code yourself,
have to adapt data file to requirements of AVS, expensive

IDL: (we use it for 1D and 2D plots)
Pros: kind of interactive, easy to code
Cons: 2D color plots give huge .ps files, often slow, expensive

Matlab: (we use it for 2D plots in D3NEBEL)
Pros: copes with some strange data formates (2D set of rays)
kind of interactive, easy to code
Cons: not made for graphics, slow, expensive

Graphical user interface of AVS

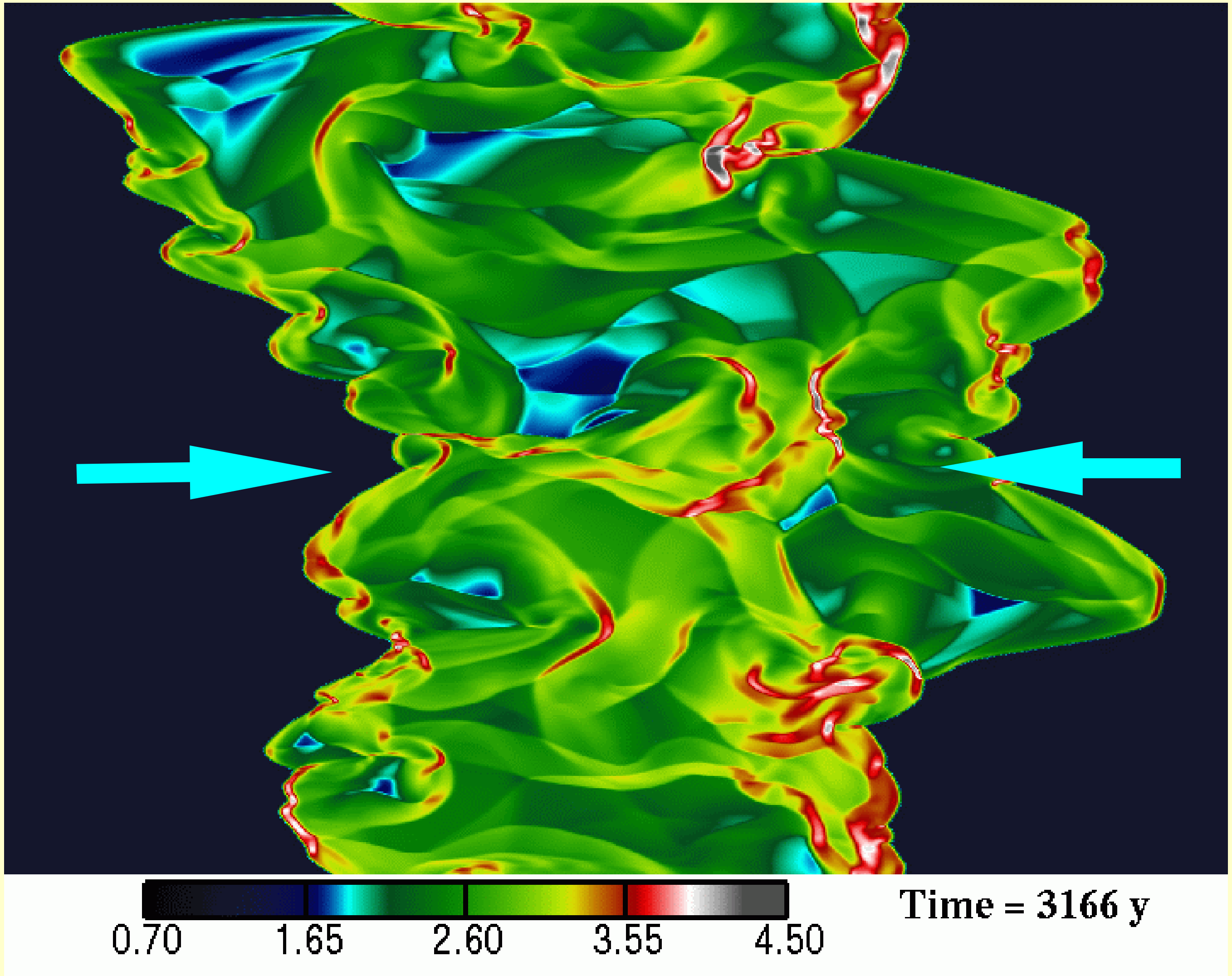


Multi- color color- table: in AVS

In many applications it is compelling to use a multi- color color- table.

Otherwise, you will not capture the whole range of interesting features.

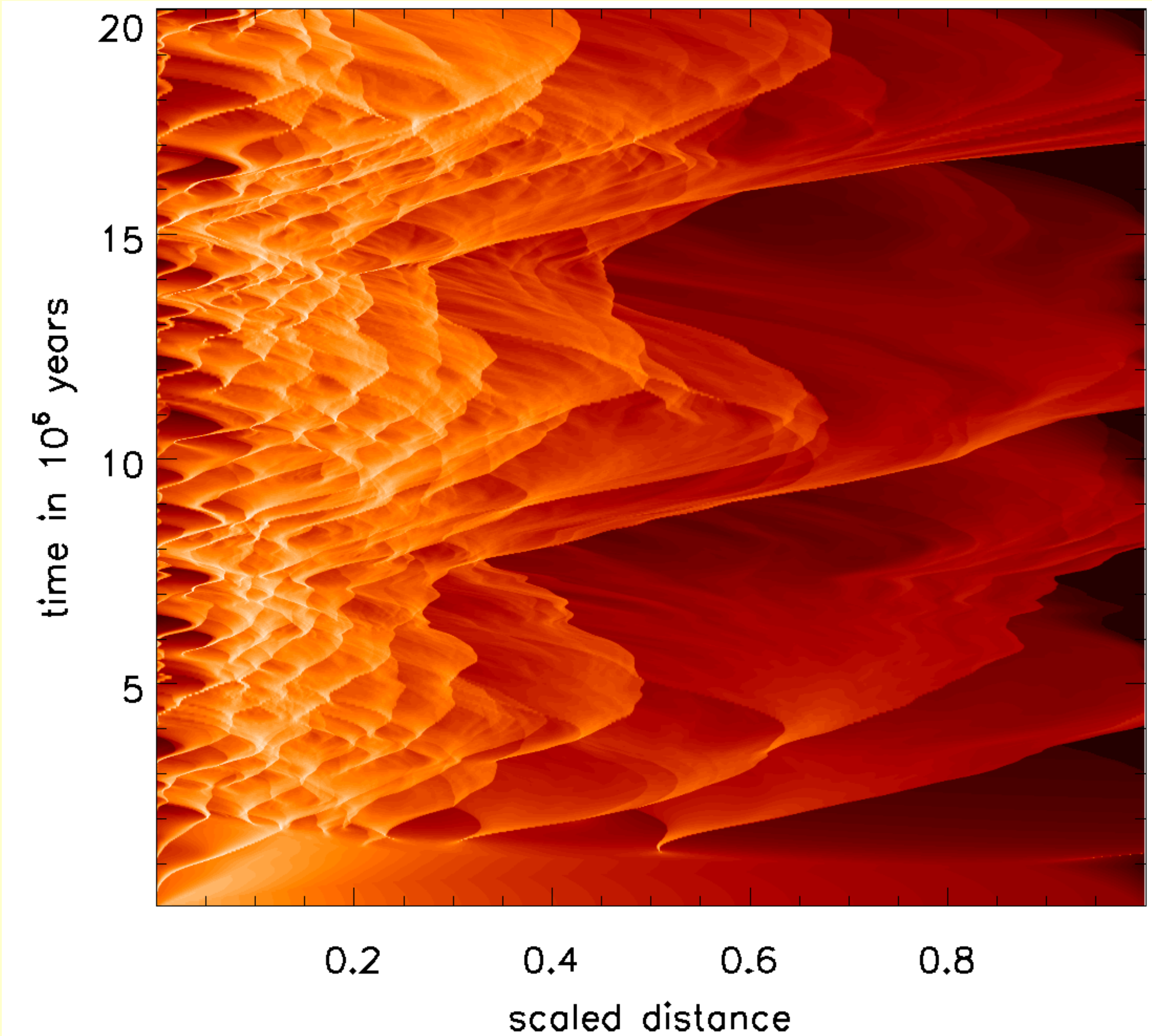
Note that not all graphics packages allow you to store the color- table you have designed.



Fully developed turbulent shell: density

IDL: 2D color contour plots give huge post scritp files

Alternatively, and what you see here, make a gif- file



Visualization software for each and every problem...

http:// seesar.lbl.gov/ANAG/chombo/index.html

http:// zeus.ncsa.uiuc.edu/~miksa/LCAVision.html

http:// public.kitware.com/Paraview/HTML/Index.html

http:// www.astroscu.unam.mx/scu/Varios/supermongo.html

http:// seesar.lbl.gov/AMR/index.html

Movie production: what for?

Depending on where you want to show the movie and on what you want to show with it, you may choose a different format:

For your web page?
Need small file, but moderate quality may be ok
(e.g. mpeg)

For a work station?
File size is unimportant, good qualtity and portability desirable (e.g. quicktime)

For a conventional video tape?
Produce singel *.gif (or equivalent files)
Then you will need help... equipment very expensive

Movie production: how to make it?

We currently use the program movie_convert, which forms part of standard Silicon Graphics Software.

Free- ware on the web includes:

http:// www.mpeg.org/MPEG/MPEG-content.html
http:// www.apple.com/quicktime
http:// www.schirmacher.de/arne/kino/index_e.html
http:// www.heroinewarrior.com/index.php3
http:// www.virtualdub.org
http:// mjpeg.sourceforge.net
http:// www.mainconcept.com

Summary

✭ having a code is not sufficient for doing numerical simulations

✭ you need specialized hardware
(machines for computing, graphics, storage devices)

✭ you need skills with regard to scripting
('by hand': too time consuming and too error prone)

✭ you need visualization software, possibly video software

✭ you need people to help you with these issues
(system administrators, computer scientists, graphics people)