IPython and the Scientific Python Ecosystem

Fernando Pérez

http://fperez.org, @fperez_org

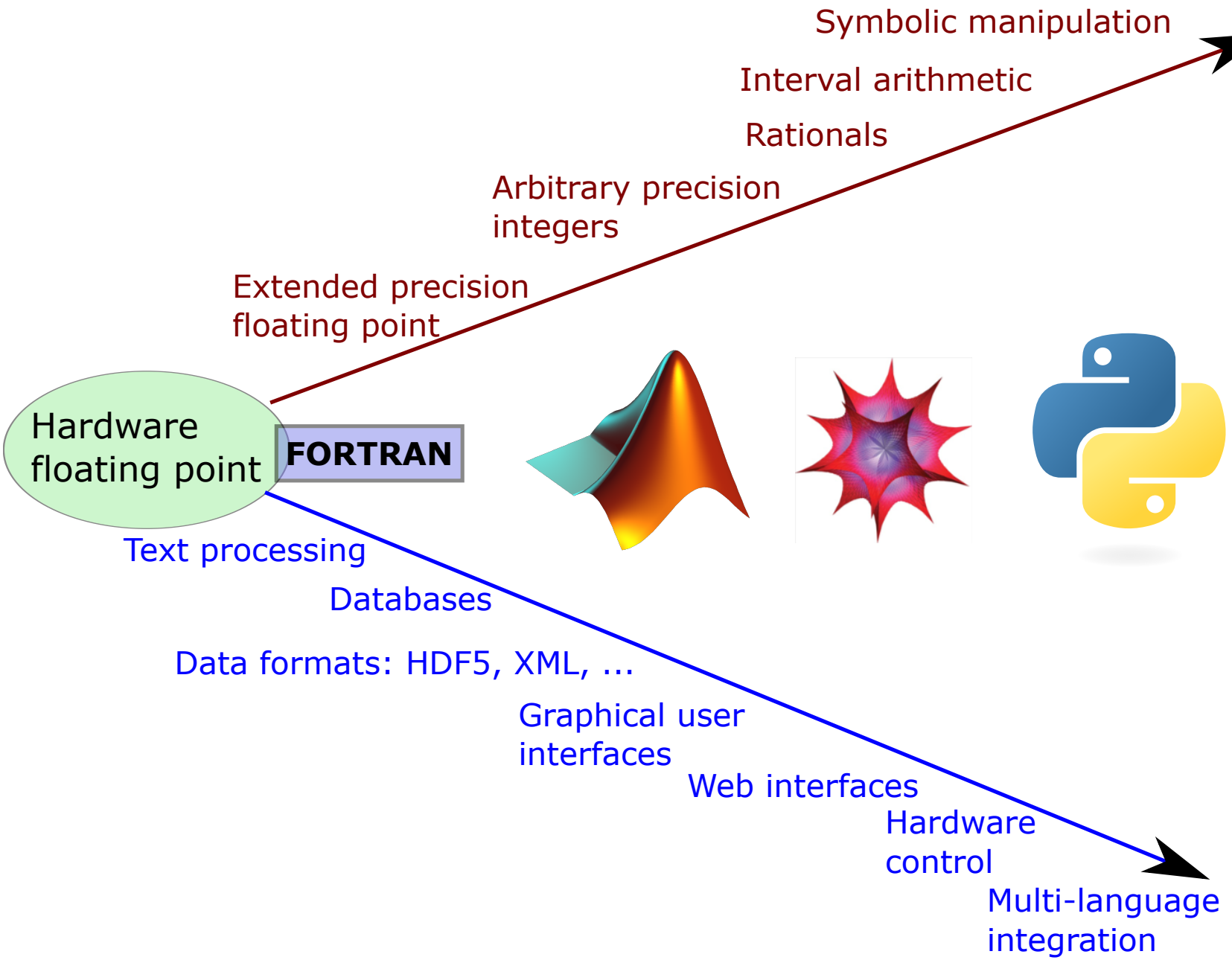
Fernando.Perez@berkeley.edu

Helen Wills Neuroscience Institute, UC Berkeley

Geilo Winter School on Reproducibility
Jan 21, 2013

Outline

Beyond (Floating Point) Number Crunching



NumPy: the Foundation. Modern Array Processing

- A flexible, efficient, multidimensional array object.
- Convenient syntax: `c = a+b`.
- Math library that operates on arrays: `y = sin(k*t)`.
- Basic scientific functionality:
  - Linear algebra
  - FFTs
  - Random number generation

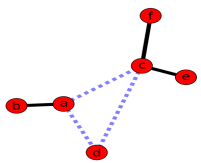
Scientific Python: a Rich Ecosystem

IPython

SymPy

machine learning in Python

scikits-image
image processing in python



NetworkX

Outline

Computing is not the 'third branch' of science...

It is now the backbone of theory *and* experiment!

Computing in science **must** improve drastically before we can really call it scientific.

Computing is not the 'third branch' of science...

It is now the backbone of theory *and* experiment!

Computing in science **must** improve drastically before we can really call it scientific.

The Lifecycle of a Scientific Idea (schematically)

1. Individual exploratory work
2. Collaborative development
3. Production work (HPC, cloud, parallel)
4. Publication (with reproducible results!)
5. Education
6. Goto 1.

The Problem with most tools

Barriers and discontinuities in workflow in between all the steps

The Lifecycle of a Scientific Idea (schematically)

1. Individual exploratory work
2. Collaborative development
3. Production work (HPC, cloud, parallel)
4. Publication (with reproducible results!)
5. Education
6. Goto 1.

The Problem with most tools

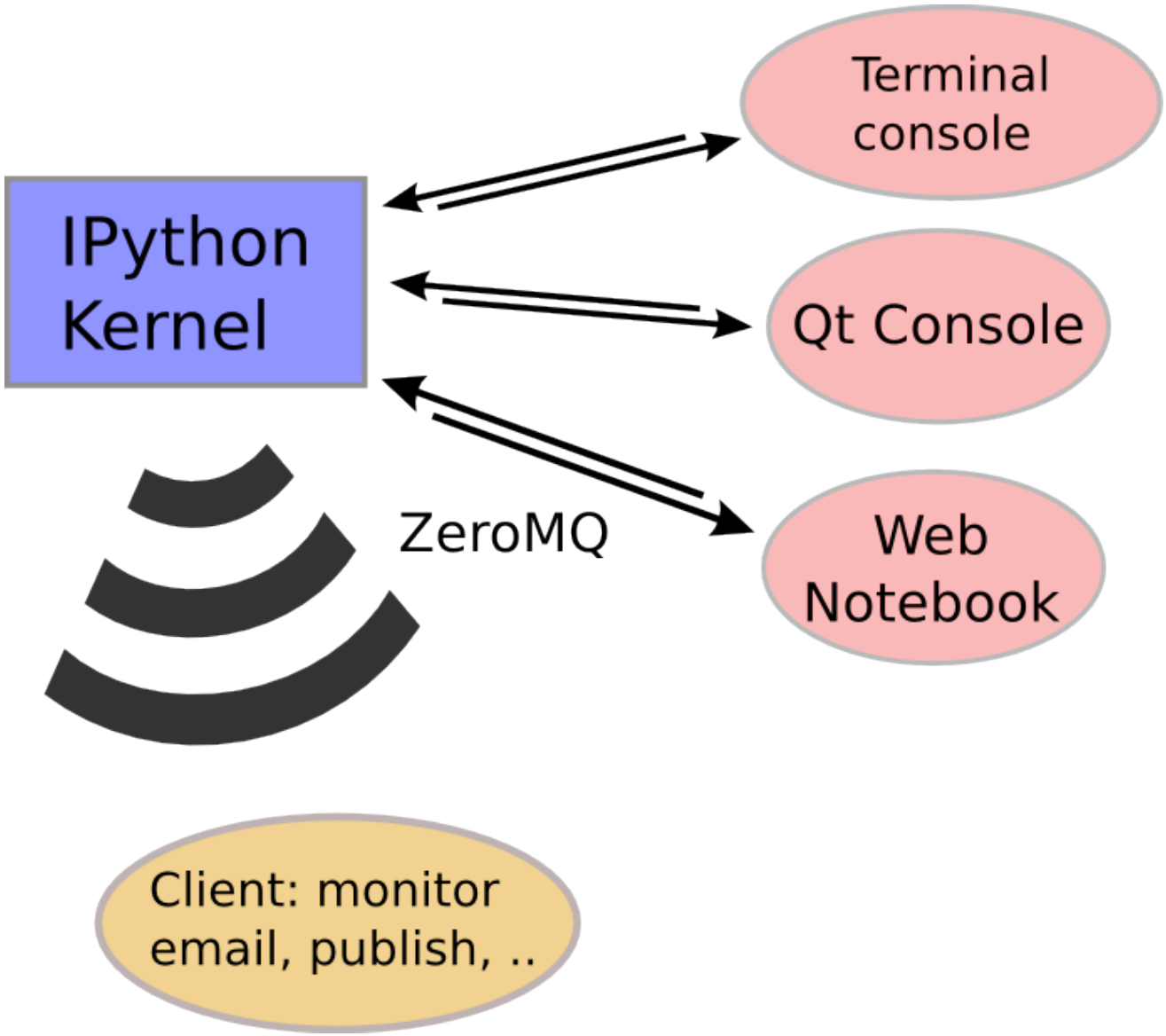
Barriers and discontinuities in workflow in between all the steps

IPython's goal:

Fluid transitions in all these steps

Demo

Pillar #1: An architecture for interactive computing



Pillar #2: the Notebook Format

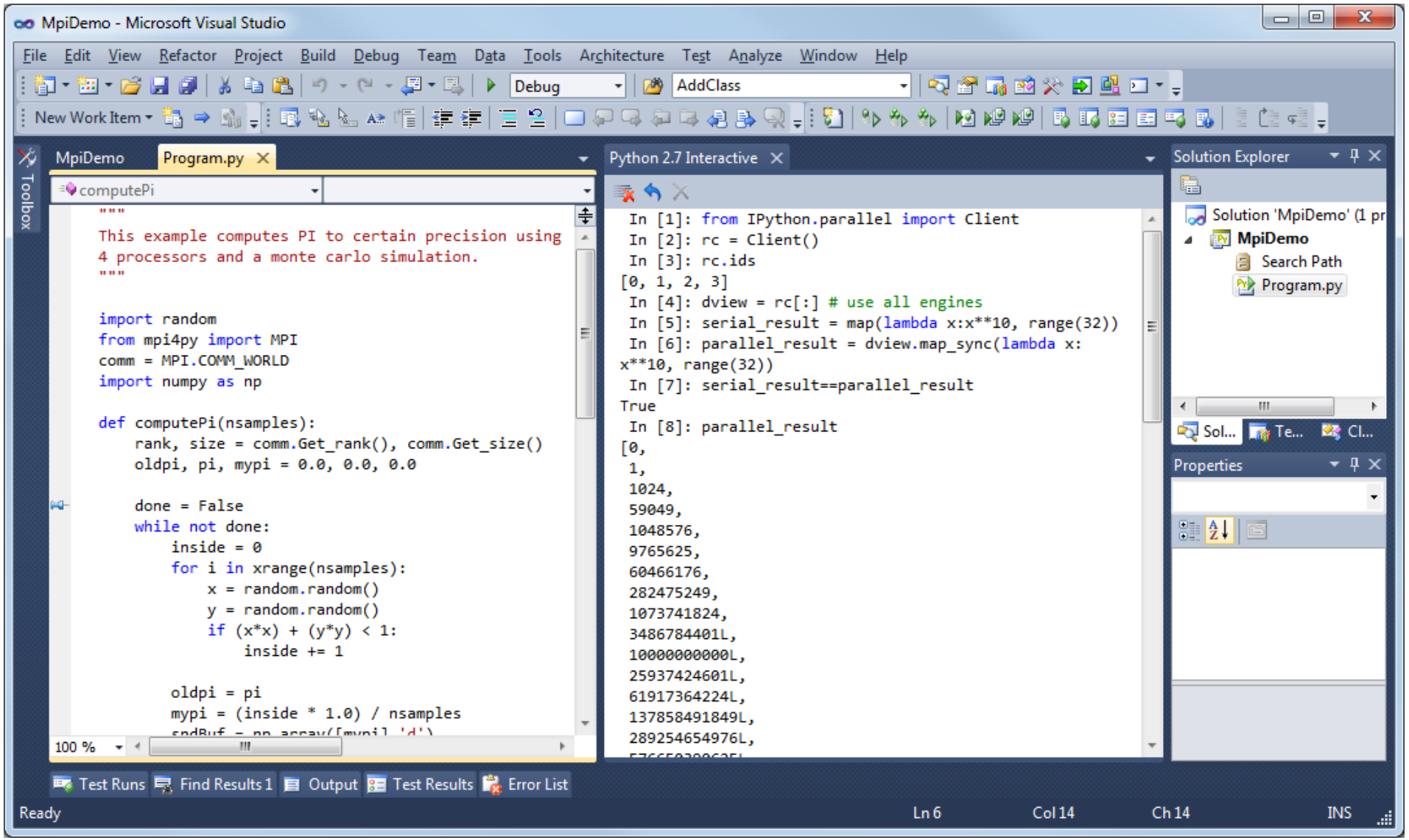
- **JSON** but version control-friendly
- Easy for machine processing, **fixable by hand if need be**.
- Lots of hooks for metadata
- **Not** Python-specific (R and Ruby notebooks exist, Julia planned)
- Produce Markdown, reST, LaTeX, HTML, etc...

An open format for sharing, publishing and archiving executable computational work

Documented protocols and formats: a growing ecosystem around IPython

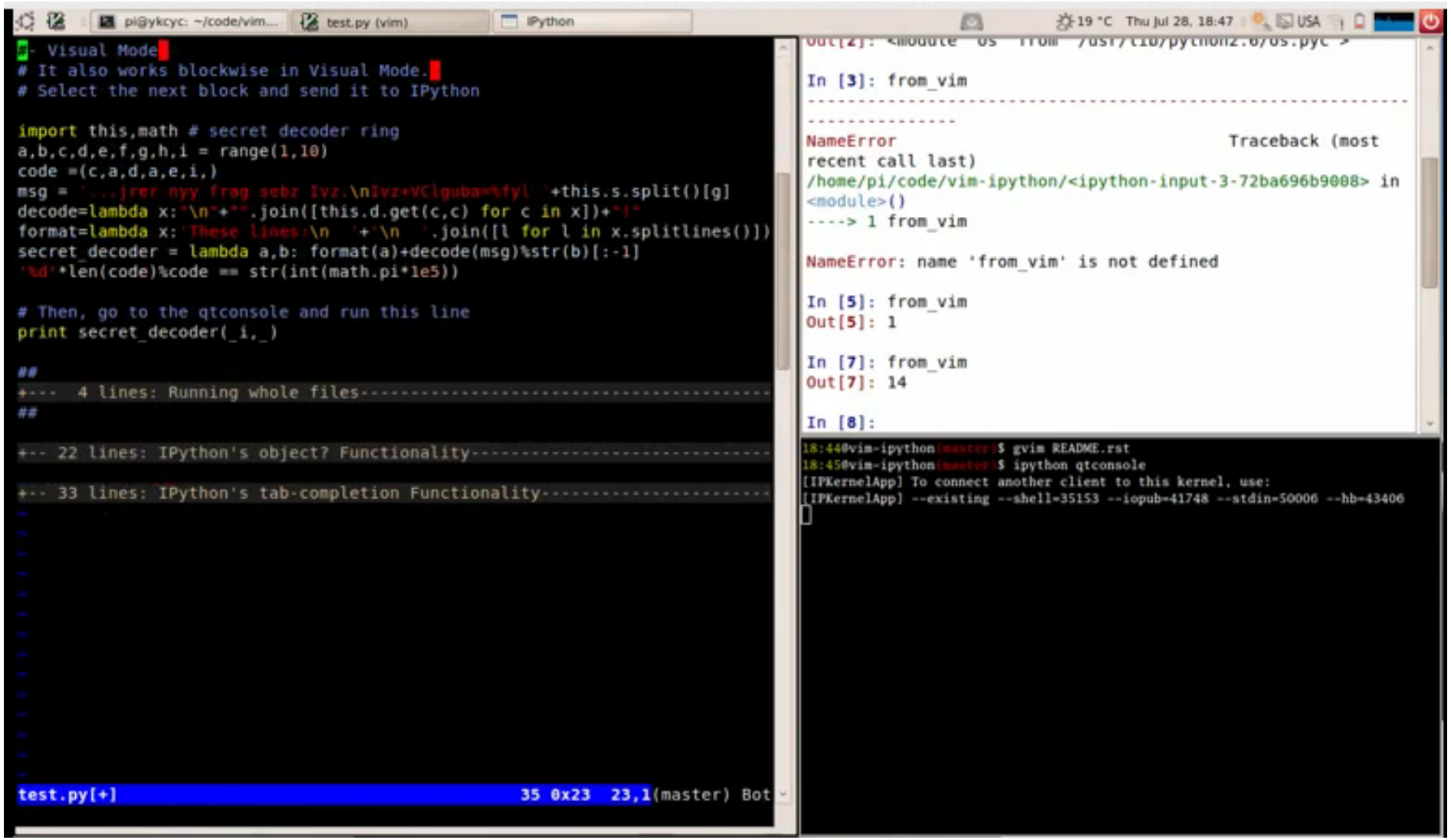
Microsoft Visual Studio 2010 integrated console

Dino Viehland and Shahrokh Mortazavi (Microsoft); http://pytools.codeplex.com



A vim client to control an IPython kernel/console

Paul Ivanov (Berkeley), https://github.com/ivanov/vim-ipython



Notebooks on Windows Azure Cloud

Shahrokh Mortazavi (Microsoft), B.G., F.P.: http://bit.ly/JQeojD.

Python - IPython Notebook

https://www.windowsazure.com/en-us/develop/python/tutorials/ipython-notebook/

Windows Azure

PORTAL

HOME PRICING DEVELOP MANAGE COMMUNITY SUPPORT ACCOUNT

overview .net node.js php java python downloads

Free trial

WINDOWS AZURE INTRO

TUTORIALS
Web with Django
Web with Django + MySQL
Django with Visual Studio
IPython Notebook
show all

HOW TO GUIDES
Blob Service
Table Service
Queue Service
Service Bus Queues
Service Bus Topics
Command Line Tools
show all

COMMON TASKS
Install Python
SQL Database Management
Custom DNS
Enable Remote Desktop
Enable SSL
CDN
Staging Deployment
show all

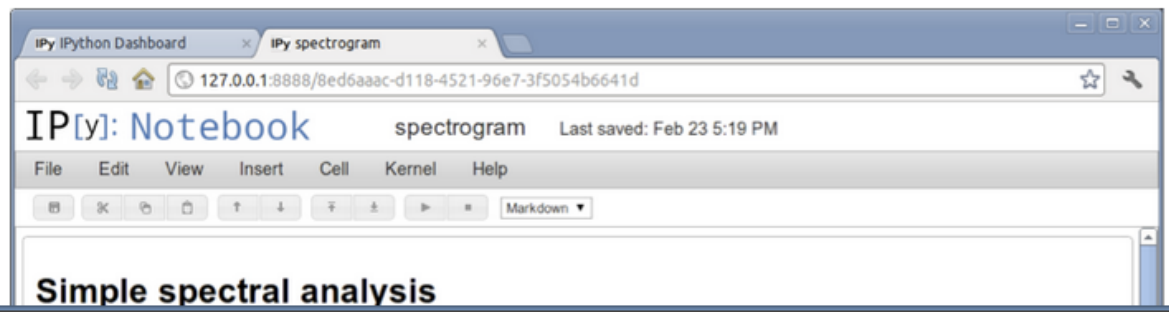
BEST PRACTICES
Troubleshooting
Security

IPython Notebook on Windows Azure

For a quick overview of installation and IPython, please watch:

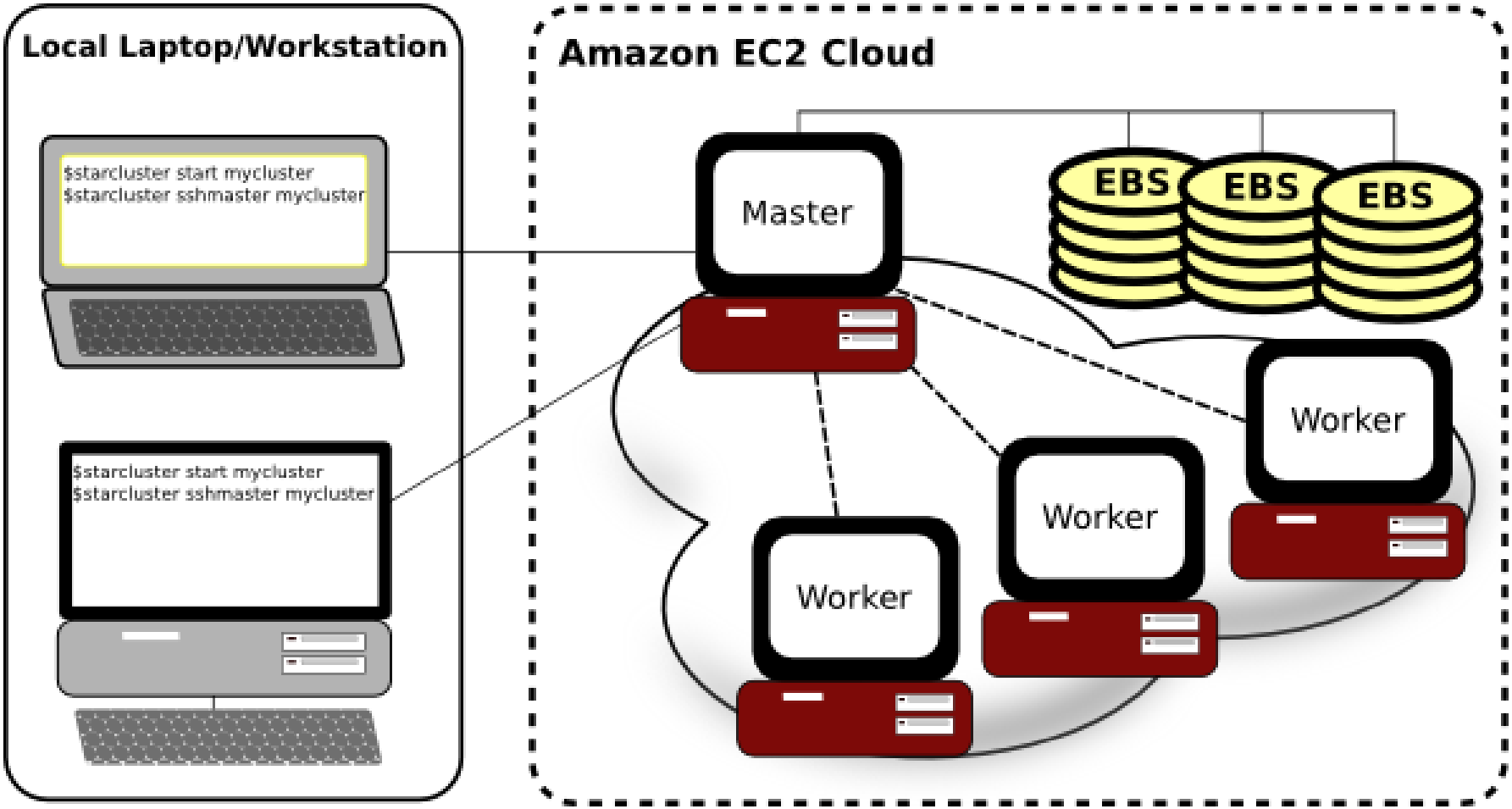
The IPython project provides a collection of tools for scientific computing that include powerful interactive shells, high-performance and easy to use parallel libraries and a web-based environment called the IPython Notebook. The Notebook provides a working environment for interactive computing that combines code execution with the creation of a live computational document. These notebook files can contain arbitrary text, mathematical formulas, input code, results, graphics, videos and any other kind of media that a modern web browser is capable of displaying.

Whether you're absolutely new to Python and want to learn it in a fun, interactive environment or do some serious parallel/technical computing, the IPython Notebook is a great choice. As an illustration of its capabilities, the following screenshot shows the IPython Notebook being used, in combination with the SciPy and matplotlib packages, to analyze the structure of a sound recording:



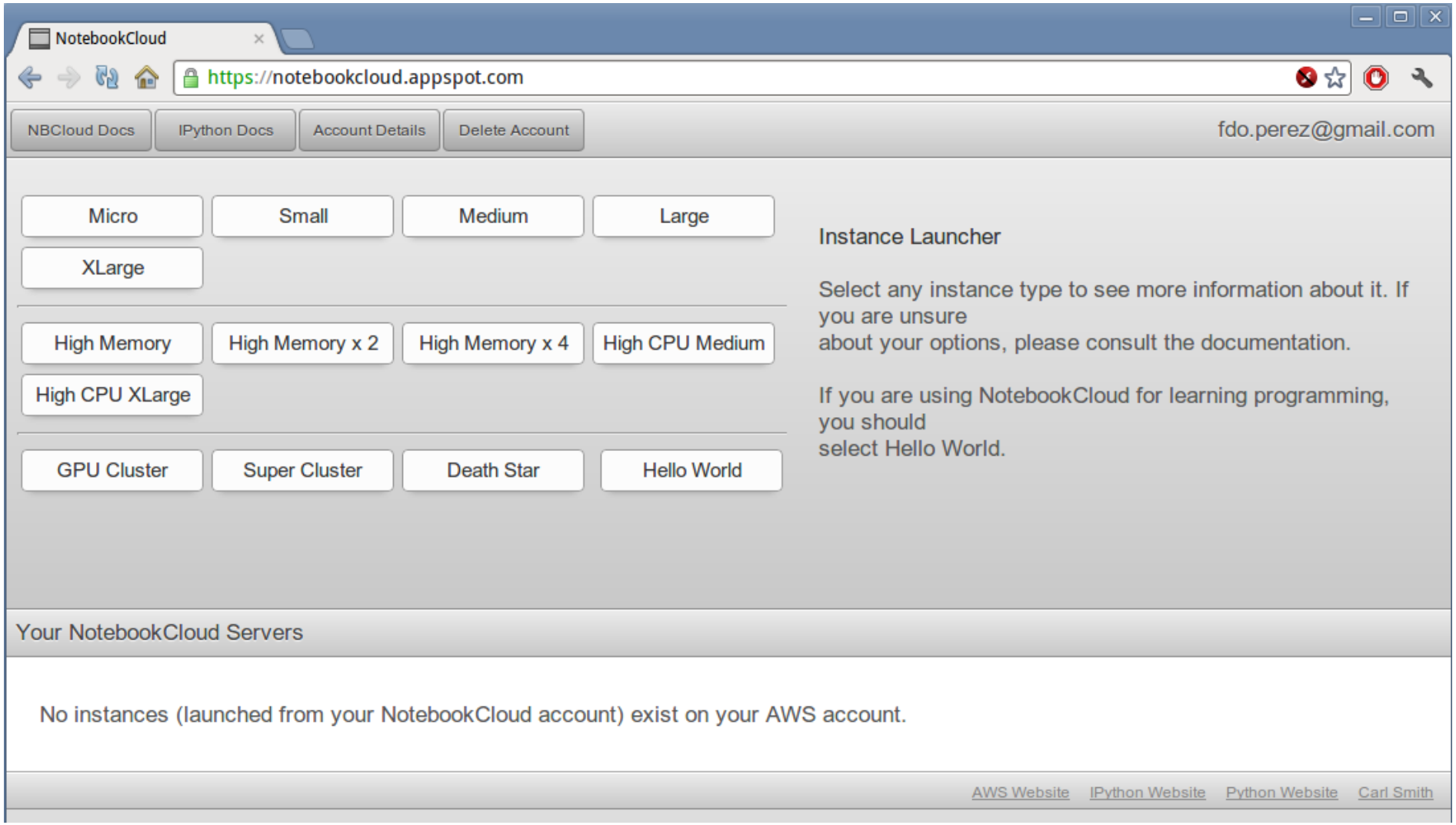
Star Cluster: IPython parallel+Notebook on Amazon EC2

Justin Riley (MIT): `http://web.mit.edu/star/cluster`



One-click single notebook on Amazon EC2

Carl Smith (UK): `https://notebookcloud.appspot.com`.



Other projects using IPython

Scientific

- **Software Carpentry**
- **EPD:** Enthought Python Distribution.
- **Continuum:** Anaconda, Wakari.
- **Sage**: open source mathematics.
- **PyRAF**: Space Telescope Science Institute
- **CASA**: Nat. Radio Astronomy Observatory
- **Ganga**: CERN
- **PyMAD**: neutron spectrom., Laue Langevin
- **Sardana**: European Synchrotron Radiation
- **ASCEND**: eng. modeling (Carnegie Mellon).
- **JModelica**: dynamical systems.
- **DASH**: Denver Aerosol Sources and Health.
- **Trilinos**: Sandia National Lab.
- **DoD**: baseline configuration.
- **Mayavi**: 3d visualization, Enthought.
- **NiPype**: computational pipelines, MIT.

Web/Other

- **Visual Studio 2010**: MS.
- **Django**.
- **Turbo Gears**.
- **Pylons** web framework
- **Zope** and **Plone** CMS.
- Axon Shell, BBC **Kamaelia**.
- **Schevo** database.
- **Pitz**: distributed task/bug tracking.
- **iVR** (interactive Virtual Reality).
- **Movable Python** (portable Python environment).
- ...

(Incomplete) Cast of Characters

- **Brian Granger** - Physics, Cal State San Luis Obispo
- **Min Ragan-Kelley** - Nuclear Engineering, UC Berkeley
- **Matthias Bussonnier** - Physics, Institut Curie, Paris
- **Brad Froehle** - Mathematics, UC Berkeley
- **Paul Ivanov** - Neuroscience, UC Berkeley.
- **Robert Kern** - Enthought
- **Thomas Kluyver** - Biology, U. Sheffield
- **Jonathan March**- Enthought
- **Evan Patterson** - Physics, Caltech/Enthought
- **Jörgen Stenarson** - Elect. Engineering, Sweden.
- Stefan van der Walt - UC Berkeley
- John Hunter - TradeLink Securities, Chicago.
- Prabhu Ramachandran - Aerospace Engineering, IIT Bombay.
- Satra Ghosh- MIT Neuroscience
- Gaël Varoquaux - Neurospin (Orsay, France)
- Ville Vainio - CS, Tampere University of Technology, Finland
- Barry Wark - Neuroscience, U. Washington.
- Ondrej Certik - Physics, U Nevada Reno
- Darren Dale - Cornell
- Justin Riley - MIT
- Mark Voorhies - UC San Francisco
- Nicholas Rougier - INRIA Nancy Grand Est
- Thomas Spura - Fedora project
- Many more! (~150 commit authors)

Support
Thank you!

- **Enthought**, Austin, TX: **Lots!**
- **Microsoft**: WinHPC support, Visual Studio integration, Azure (thanks to Shahrokh Mortazavi).
- **DoD/DRC Inc**: funding through Sept. 2012 (thanks to Jose Unpingco and Chris Keees).
- NIH: via NiPy grant
- NSF: via Sage compmath grant
- Google: summer of code 2005, 2010.
- Tech-X Corp., Boulder, CO: Parallel/notebook (previous versions)

News!

ALFRED P. SLOAN FOUNDATION

2-year funding, core team (7 people)

Thanks!!

NumFOCUS: http://numfocus.org

Foundation to promote Open Source Scientific Computing

- Support the development of multiple projects.
- Community-created and driven.
- A neutral ground for industry, academia and government to support scientific open source.
- 501(c)3 - donations are tax-exempt in the USA