Python and Science: the great mystery

Python isn't known for its baffling performance

Scientific computing <3 performance (think Fortran)

Yet they love each other, which paves the way for accelerating solutions
THE PYTHRAN COMPILER, 7 YEARS LATER

Serge « sans paille » Guelton

Compiler Engineer / Wood Craft Lover / RedHat employee

FOSDEM'20 — 1st February 2020
IN A NUTSHELL

A *non-intrusive* compiler for scientific kernels written in *Python*

```python
import numpy

#pythran export log_likelihood(float64[], float64, float64)
def log_likelihood(data, mean, sigma):
    s = (data - mean)**2 / (2 * (sigma**2))
pdfs = numpy.exp(-s)
pdfs /= numpy.sqrt(2 * numpy.pi) * sigma
return numpy.log(pdfs).sum()

# native parallel vectorized module
pythran log_likelihood.py -fopenmp
    -march=native -DUSE_XSIMD
```
WHAT ABOUT... CYTHON?

- Cython introduces an (optional) new dialect
- Cython allows mixed mode execution

Upside:

- Huge user base (numpy, scipy, scikit-*)
- Graduate conversion
- Nice error reporting
- Nice profiling support through (cython -a)

- Low 3rd party requirements (cc)
WHAT ABOUT... PyPy?

- PyPy is a full (jitting) interpreter
- PyPy has limited Numpy acceleration
- Strong *theoretical* and *engineering* foundations

Upside:

- *Full* language support
WHAT ABOUT... NUMBA?

- Numba favors explicit JIT compilation
- Numba favors an explicit, low-level style
- Based on the LLVM compiler infrastructure

Upside:

- Powered by the conda package manager
- Good adoption in the community
- Easy integration (@numba.jit)
4 COMPILERS, 4 STORIES

Enough speak about others,
Let's rollback Pythran history :-)

extra game: guess-the-compiler-initial-commit
Move the pypy trunk into its own top level directory, so the path names stay constant.
commit 54ab11e7765f92682f033ea61d6916d2eed0256e
Author: William Stein <wstein@gmail.com>
Date:    Thu Oct 19 21:28:18 2006 -0700

Official Pyrex version 0.9.4.1
2012

Presented at Pyconfr

Python (python - object) =
OpenMP + C++

*Upside*: met Victor Stinner

*Downside*: no Numpy support => no interest
commit 52370f5564fa90916ce6f38cf4e63b8fa1761643
Author: Travis E. Oliphant <teoliphant@gmail.com>
Date:  Thu Mar 8 05:15:08 2012 -0600

first commit
Presented at ScyPy 2013:

**Pythran**: Enabling Static Optimization of Scientific Python Programs

*Upside*: Numpy compatibility, great conference and energy, creation of [https://github.com/numfocus/python-benchmarks](https://github.com/numfocus/python-benchmarks)

*Downside*: in spite of lower-level DSL, Numba is the thing
2014

Benchmarks-driven engineering:

https://github.com/serge-sans-paille/numpy-benchmarks/

commit 6bf225ac9ade74044ad63be89fd37f6b46a184d1
Author: serge-sans-paille <sguelton@quarkslab.com>
Date: Fri Mar 14 10:38:36 2014 +0100

Initial commit
2014

commit be01e422c4314e6cc6f470df6573e674784da573
Author: Kevin Modzelewski <kevmod@gmail.com>
Date:   Wed Apr 2 19:26:58 2014 -0700

Initial commit
2014

Finally some online validation

commit ebb9aecd396ebf1cf01d6809b78e8e78b8ee7357
Author: serge-sans-paille <serge.guelton@tb>
Date:   Sun Sep 7 15:55:39 2014 +0200

Setup Travis CI

In addition to the various test scenarios stored in .travis.yml:

- make CLI documentation compiler dependent, using
  pythran-config --compiler new flag
- fix a typo in the DEVGUIDE
- split test_numpy in three parts otherwise travis times out. It was too big anyway
- fix libs flags
- remove test_simd as travis does the job for us
2014

commit 0da9c09ace57efcbdf237f347a909f991125e0f7
Author: Joel Akeret <jakeret@phys.ethz.ch>
Date:   Thu Oct 16 11:13:46 2014 +0200

first commit
2015

Start of the OpenDreamKit project

OpenDreamKit also supports open source research codes directly by investing into structural improvements and new features to not only connect all of these tools but also enrich them, and make them more sustainable.
2016

**GAST**: Python 2 and Python 3 compatibility layer

Started as a pun using the breton language,

Ended up as a package downloaded 3M times a month

```
commit 5deeef2b8077ca620f81b5c2c156b6e6afed597d0
Author: serge-sans-paille <serge.guelton@tb>
Date:   Wed Jun 22 15:42:47 2016 +0200

Gaea's touch
```
2016

Last commit in Hope’s master branch
2017

First use of tags in Pythran

```bash
> git show `git tag | head -n1`
commit 295ea3154d8891b9fc55b252cc6002003c48eda5
Author: serge-sans-paille <serge.guelton@tb>
Date:   Tue Jul 4 21:10:32 2017 +0200

0.8.1 - GDR Calcul
```
2017

Last commit in the Pyston's master branch
2018

Published *Pythran: Crossing the Python Frontier*

Article in *Computing in Science and Engineering*,
March 2018

Good for visibility of the project!
2018

Managing external dependencies, always a pleasure :-)  

commit d07f5f91c1ef5d4d52f197b08bc3a21b6cb9948  
Author: serge-sans-paille <serge.guelton@tb>  
Date:   Fri Aug 10 17:02:55 2018 +0200  

Moving to xsimd  

Use xsimd instead of boost::simd for vector instructions. This relies on a patch to xsimd to provide the scalar version of each mathematical function.
2019

So far, so good for diversity

commit 3ebb085d486e0b6d6520927879e14772597cfaac
Author: serge-sans-paille <serge.guelton@tb>
Date: Mon Oct 14 15:20:17 2019 +0200

Support clang-cl.exe as a windows compiler

- Test on AppVeyor with that compiler, but don't use it as default
- Fix interaction with numpy.int type on Win32
Pain always strikes twice

commit 97ea22f7126742f824a4bf7b347c820829b9fa12
Author: serge-sans-paille <serge.guelton@tb>
Date:   Thu Jan 2 18:23:06 2020 +0100

Remove all reference to py2 code and behavior from pythran

Basically:

- __builtin__ -> builtins
- take into account behavior change of map/zip/range
- also take into account the now invalid map(None, ... pattern
- modified division behavior
- StandardError is no longer available, so is cmp etc
- reduce has moved from __builtin__ to functools

This is a huge change! But also a welcome one, -1000 lines of code :-}
2020

Acknowledgments

people = {
    'Liyun He', 'Jean Laroche',
    'Pierrick Brunet', 'Logilab',
    'Mehdi Amini', 'Sylvain Corlay',
    'Stefan Behnel', 'OpenDreamKit',
    'Nicolas M. Thiéry', 'Yann Diorcet',
    'Pierre Augier', 'Joël Falcou',
    'Miro Hrončok',
    'HPC Project', 'Adrien Guinet',
    'Lucie G.', 'Laëtitia G.'} + others

random.shuffle(people)
print(people)
**SOLDEVI EXCAVATIONS**

- Users matters:
  - Portability across Python versions
  - Portability across OS and arches

- Not being an industrial project is fine

- Make it fun:
  - Write articles (and not only code)
  - Share ideas
  - Meet people

- [https://github.com/fluiddyn/transonic](https://github.com/fluiddyn/transonic)