XENON IT

Python Zealotry on the Cosmic Frontier
OR: "How we became the best direct DM expt. in record time"
Disclaimer:
Everybody thinks their kid is better than other people's kids.

(Though Belle II seems quite nice)
ATLAS ≠ XENON1T

- Decade-scale lock-in
- $O(10000)$ people
- Can influence infrastructure
- Lots more data
- More final states

- New expt. ~ 3 yrs.
- $O(100)$ people
- Greenfield
- No dedicated soft. people
- Prototype $\rightarrow$ production
- Beggar infrastructure
- Influences soft. dev.
- PB/year tape
- Final analysis on "laptop"
Our REVOLUTIONARY idea
"Think like startup"
making Enterprise-ware

Root

Py Data
- Numba - Python DAQ and reconstruction (JIT) since can refactor better
- Scipy - better docs/examples of minimizers
- Numpy - fast and huge community
- Scikit-learn
- Ipy notebooks - data inspection
XENON IT notes

- Digitizer to NoSQL converter
- Python online/offline reconstruction
- Django DAQ control, data mgmt, data quality website
- Powerful web frameworks!
- Github, Travis CI, PR, "Slack", etc
my 3 points
Root is the NoSQL analysis format

event = \{ \text{time: int,} \\
\text{tracks: [\{ \text{px: float, ... \},} \\
\text{\ldots \} \}] \}

- HDFS hard w/o joins
- Elsewhere: Install MongoDB
- A refactored-out basic-typed Root file interesting.
  - Needed for our infrastructure
  - "pip install" so can build upon!
  - Competing formats not used as widely for exploratory analysis.
Python Fast as F...!!

- JIT is secret sauce
  - Tight loops
  - Could run 300MB/s live
- Refactoring and profiling easier in Python
  - "Premature optimization root of all evil"
- SWIG/Cython/Boost pain for sustainability

```
from numba import jit
from numpy import arange

# jit decorator tells Numba to c
# The argument types will be inf
@jit
def sum2d(arr):
    M, N = arr.shape
    result = 0.0
    for i in range(M):
        for j in range(N):
            result += arr[i,j]
    return result

a = arange(9).reshape(3,3)
print(sum2d(a))
```
Outside View

HEP software buried in legacy requirements and spread thin.

Nothing on software engineering process? Analysts copy-paste. Need it easier to do things well.

Tech invented wheels...

Where we strong? Where we different?

*my own! Provocative...
Viva la revolution!
1. Hit-finding in each channel
2. Peak classification is a separate step
3. Not satisfied? You can change it!