Copperhead: Data-Parallel Python

Bryan Catanzaro, UC Berkeley
Michael Garland, NVIDIA Research

Yunsup Lee, Mark Murphy, Armando Fox
Kurt Keutzer, UC Berkeley

(catanzar@cs.berkeley.edu)
(mgarland@nvidia.com)

(fox@cs.berkeley.edu)
This intrinsically data-parallel procedure

```python
def saxpy(a, x, y):
    return [a*xi + yi for xi,yi in zip(x,y)]
```

- is valid Python, AND
- compilable to parallel platforms: CUDA, OpenCL, TBB, ...

- Copperhead is a restricted subset of Python

- Program in data-parallel style
  - map, reduce, scan, split, join, scatter, gather...
Selective, Embedded, Just in Time Specialization

- Copperhead procedures live in normal Python modules
  - delineated by @cu function decorator
- Runtime intercepts & specializes function when called
  - infers type, shape, etc. using runtime data
  - generates C code
  - moves data as needed
  - runs in parallel
- Early experiments indicate promising efficiency
from copperhead import *

@cu
def saxpy(a, x, y):
    return [a*xi + yi for xi,yi in zip(x,y)]

x = [1.0, 1.0, 1.0, 1.0]  # can use NumPy or
y = [0.0, 1.0, 2.0, 3.0]  # CuArrays, too

result = saxpy(2.0, x, y)