
the open-source
sky survey
David W. Hogg (NYU)
http://astrometry.net/

non-text searching

- need to search things that *aren't text*, with queries that *aren't text*
- even "image search" in Google requires accurate text meta-data
- multi-billion-dollar question: "Here's a picture, what is it a picture of?"
  - we have answered this in *one tiny domain*

people

- Jon Barron (NYU, Toronto)
- David W. Hogg (NYU) – *astro PI*
- **Dustin Lang (Toronto)**
- Keir Mierle (Toronto, Google)
- Sam Roweis (Toronto, Google) – *comp sci PI*
- (with help from Blanton, Finkbeiner, Stumm)

blind calibration

- easy parts:
  - the sky is just a set of points in 2-d
  - excellent catalogs exist (*esp* USNO-B)

blind calibration

- hard parts:
  - the sky is big; astronomical images are small
  - bandpasses and sensitivities of images do not match those of the astrometric catalogs
  - we don't necessarily know *anything* about the images we see

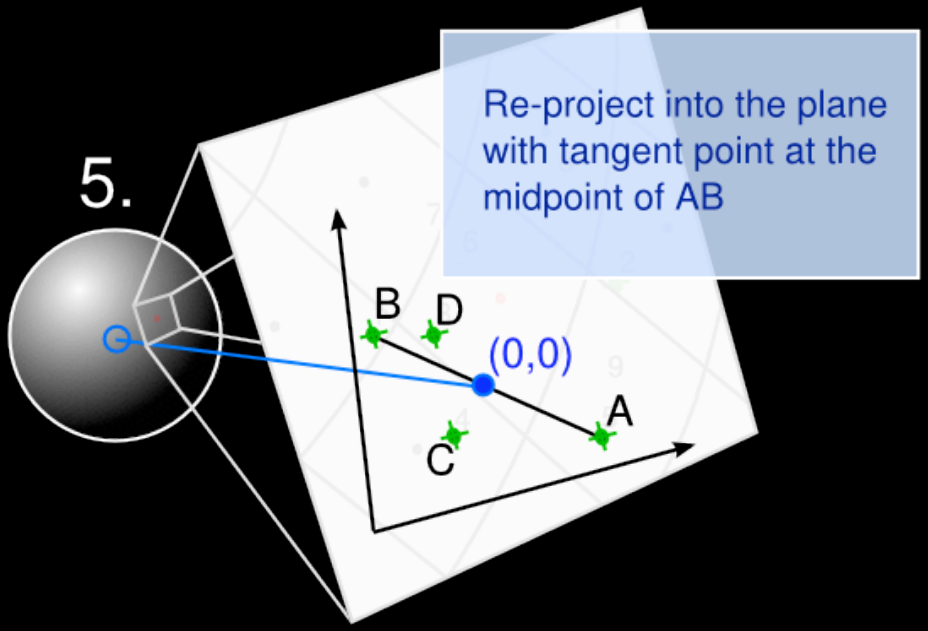
demo

- [web demo]

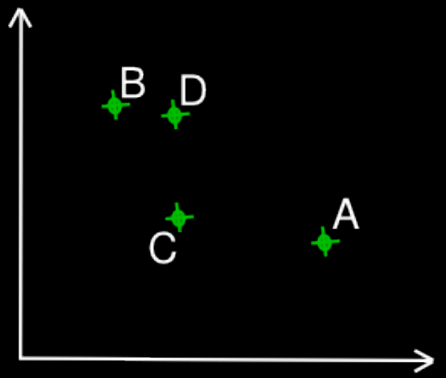
how it works

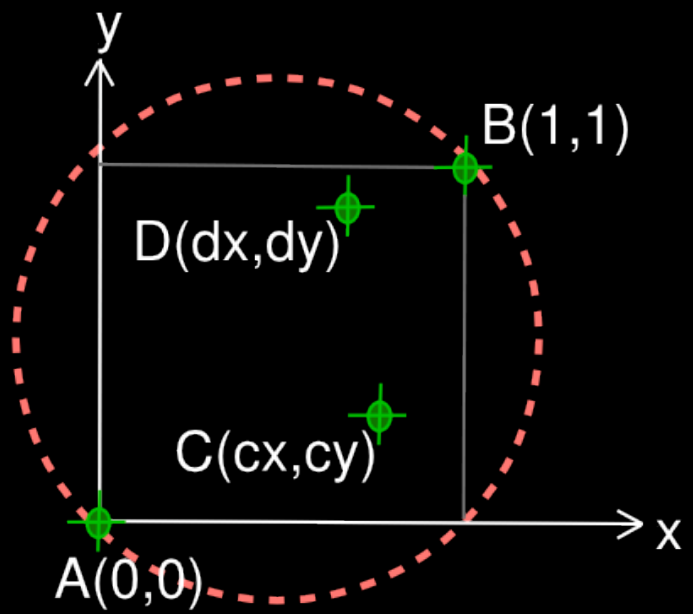
- use quads of stars to identify **hypotheses**
- test explanatory power of each hypothesis to **verify**
- typically try thousands of hypotheses per image
  - that's a lot, but a lot less than brute-force search
  - verify is fast

5.



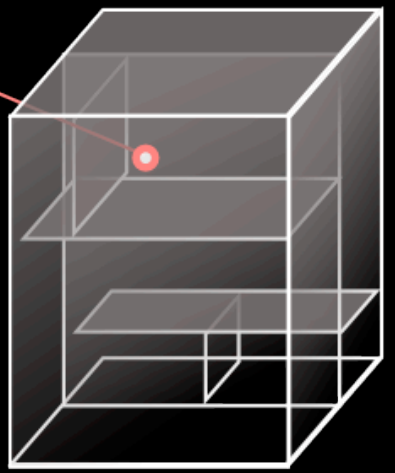
6.



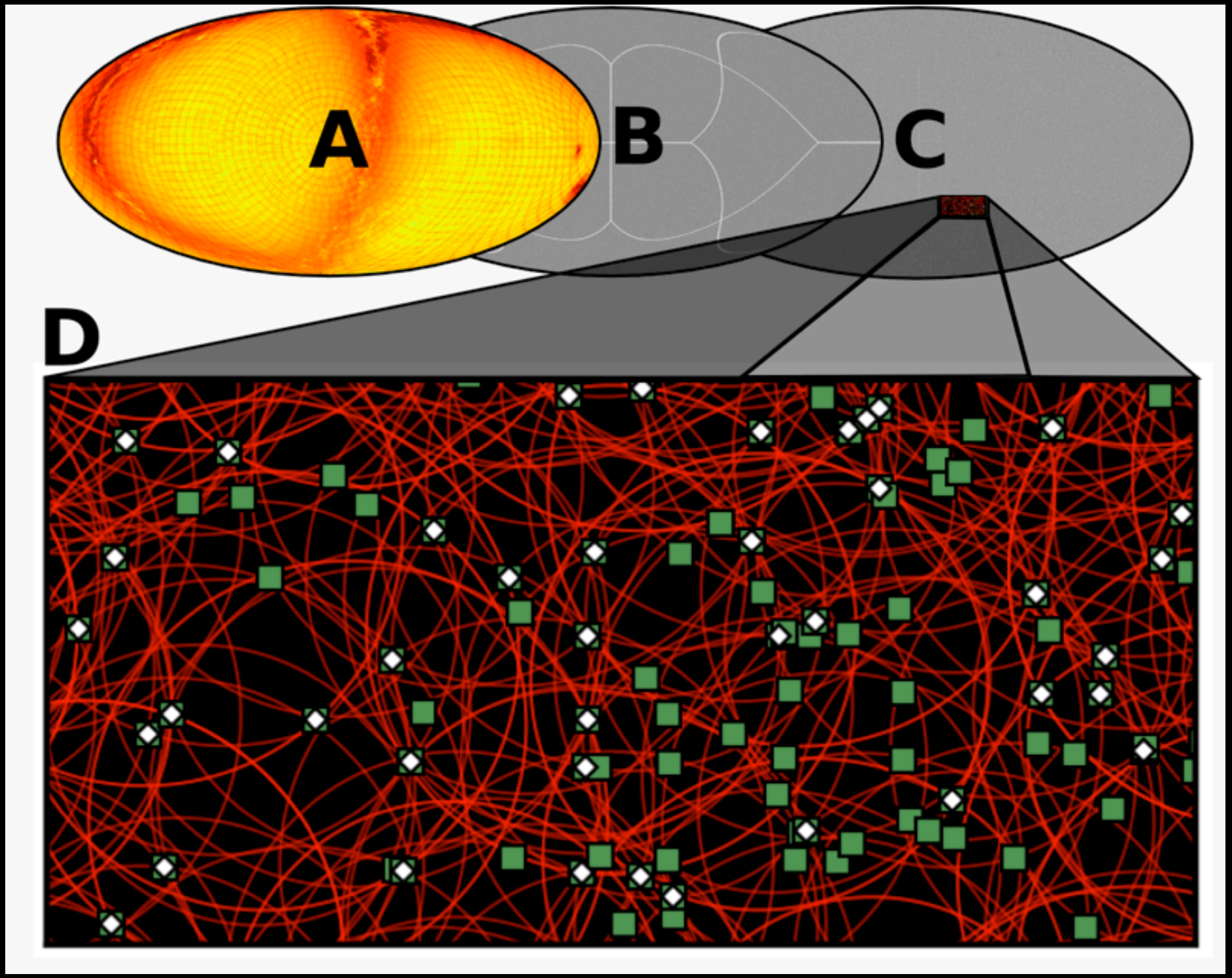


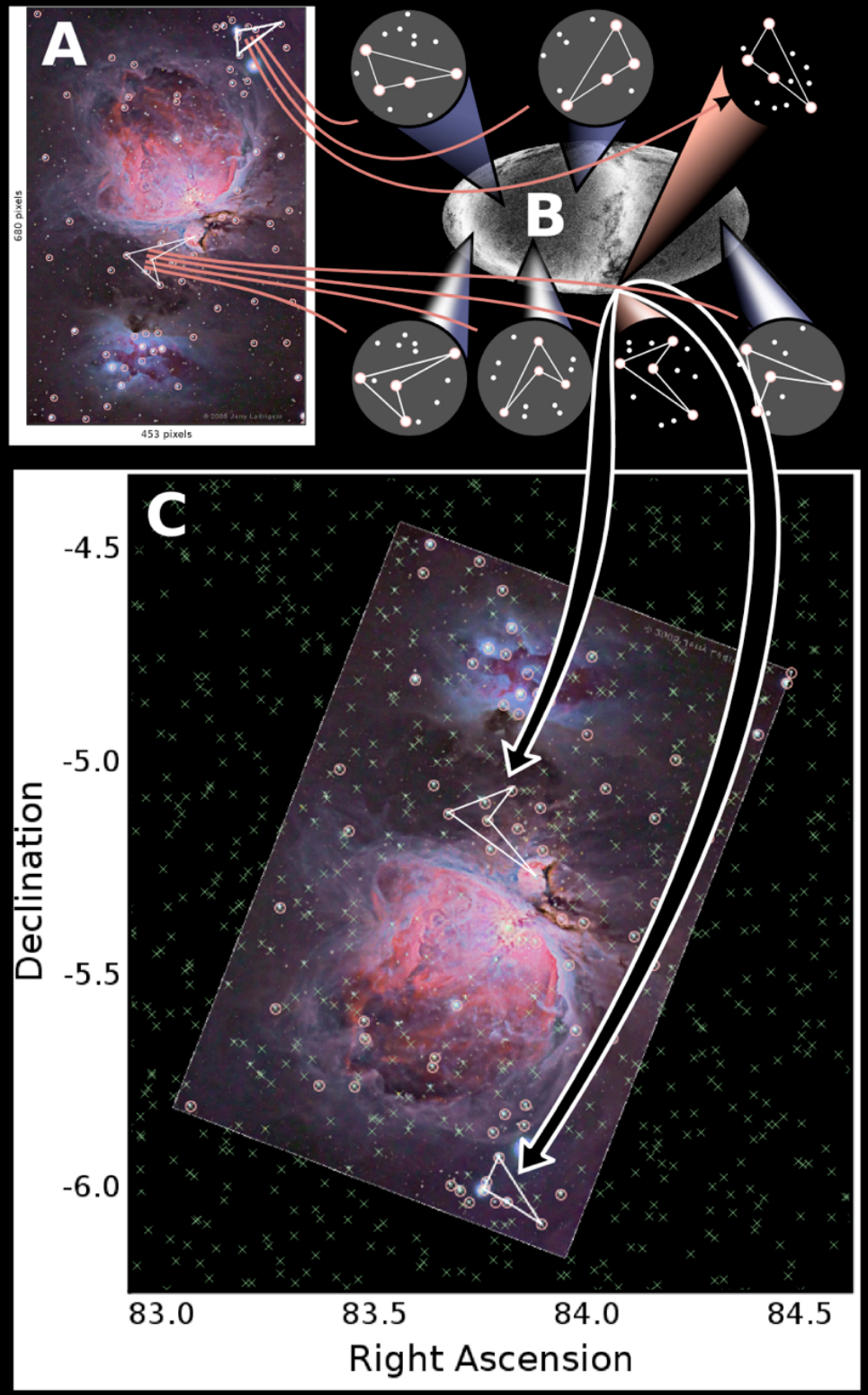
7.

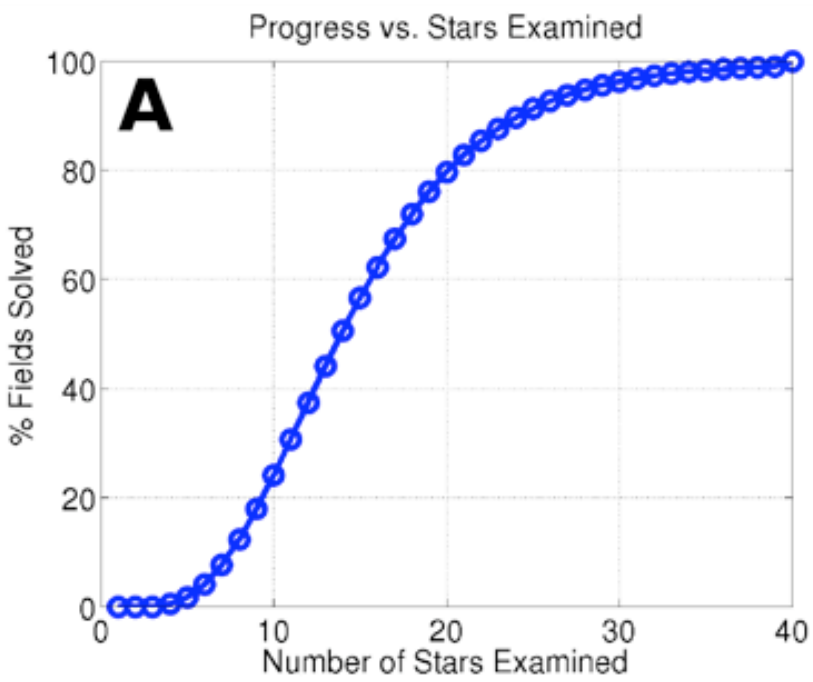
CODE:
(cx,cy,dx,dy)

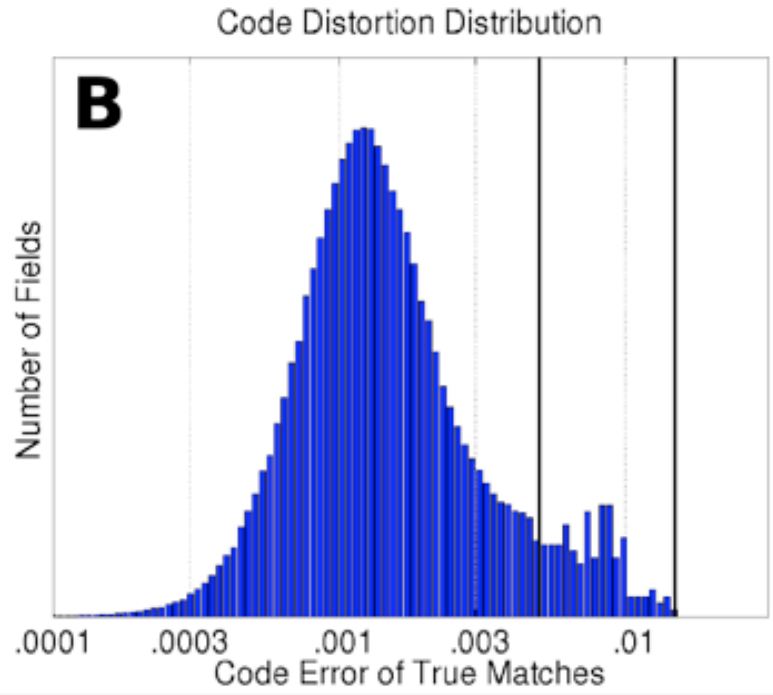


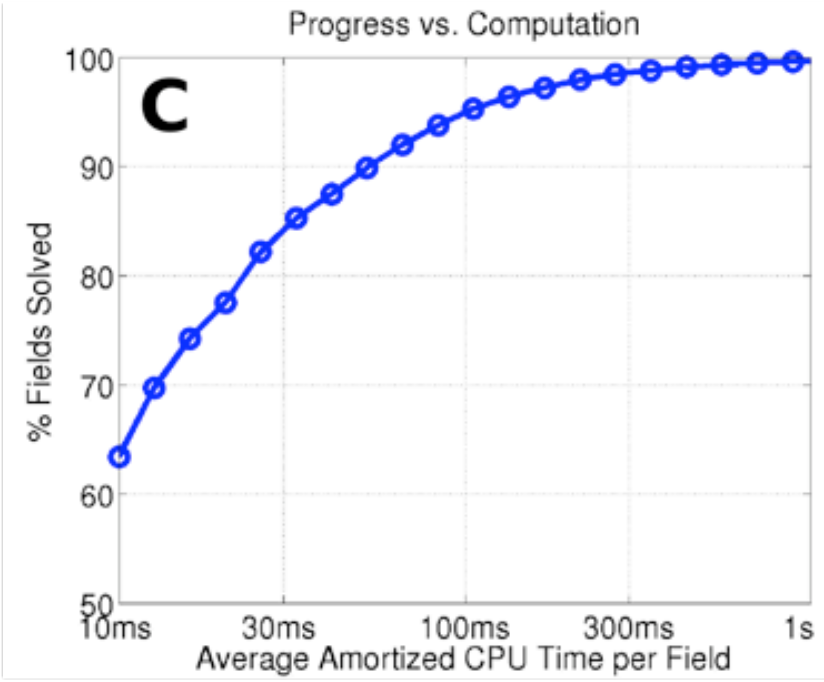
8. 4D kdtree

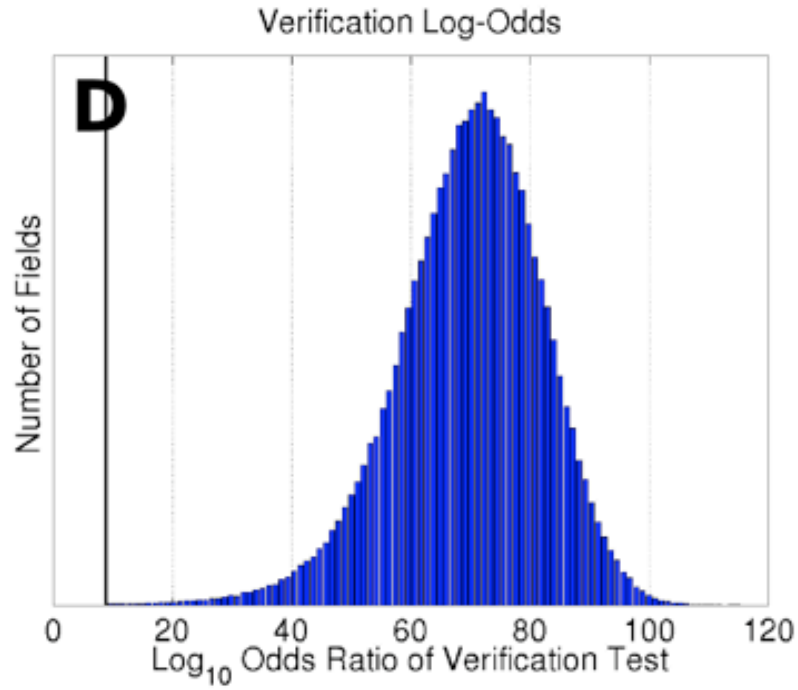

A
B
C
D


A
680 pixels
453 pixels
B
C
-4.5
-5.0
-5.5
-6.0
Declination
83.0
83.5
84.0
84.5
Right Ascension


Progress vs. Stars Examined
A
% Fields Solved
Number of Stars Examined
0
20
40
60
80
100
0
10
20
30
40


Code Distortion Distribution
B
Number of Fields
Code Error of True Matches
.0001
.0003
.001
.003
.01


Progress vs. Computation
C
% Fields Solved
Average Amortized CPU Time per Field
50
60
70
80
90
100
10ms
30ms
100ms
300ms
1s


Verification Log-Odds
D
Number of Fields
Log10 Odds Ratio of Verification Test
0
20
40
60
80
100
120

project status

- we are currently alpha (invitation only)
- go beta this spring?
- *all code is open source* (vanilla c)
  - runs on Linux and Mac
  - (Windows if you have skills)

blind calibration works

- for astrometric WCS
  - limited by USNO-B at present
- for date
  - precision of years with pms; better with variables?
- for bandpass and photometric zeropoint
  - rough bandpass: UBVRIJK
  - tens of percent precision given current catalogs
- for point-spread function

web 2.0

- user-generated content
  - blogs, moblogs, flogs, vlogs, wikis, "friend" sites
- file sharing
  - Flickr, YouTube, bittorrent (all with APIs)
- communities
  - tags, groups, feeds, comments, reviews, favorites
- *new technologies create new opportunities*

astrophotographers

- typical data processing:
  - read many FITS files from CCD in several bands
  - hand-select good seeing (“lucky imaging”)
  - hand-align and stack
  - turn into jpegs and post to the web
- science-grade data but...
  - hard to use for science
  - how do we find them?
  - there are often no (or hard-to-use) meta-data

AAVSOers

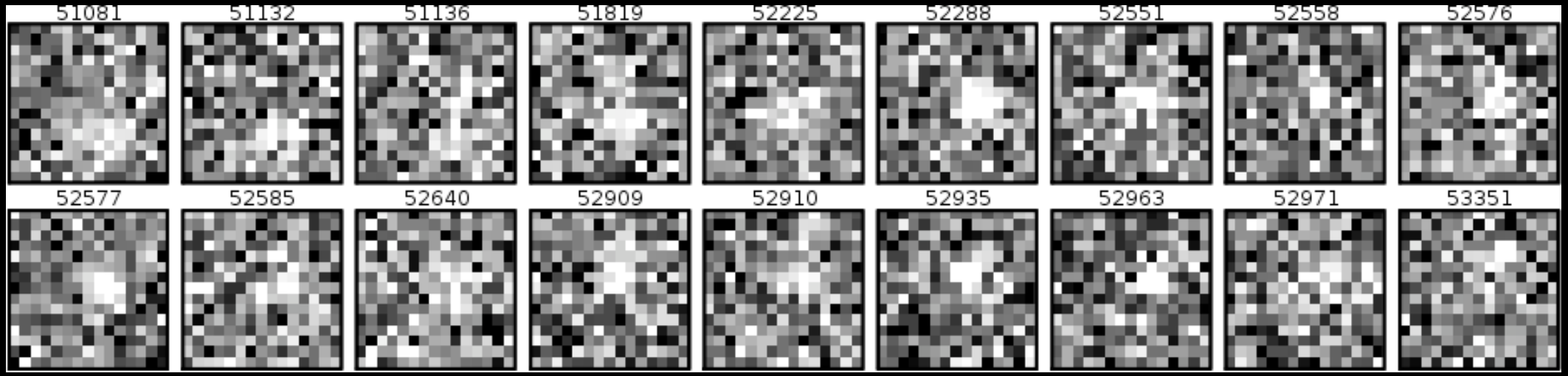
- typical data processing:
  - take many images with a CCD
  - flatfield, calibrate, measure one point source carefully in every image
  - submit magnitudes, put data in basement
- clearly science-grade data, but
  - worth so much more than just individual magnitudes
  - needs to be archived and as an *imaging database*

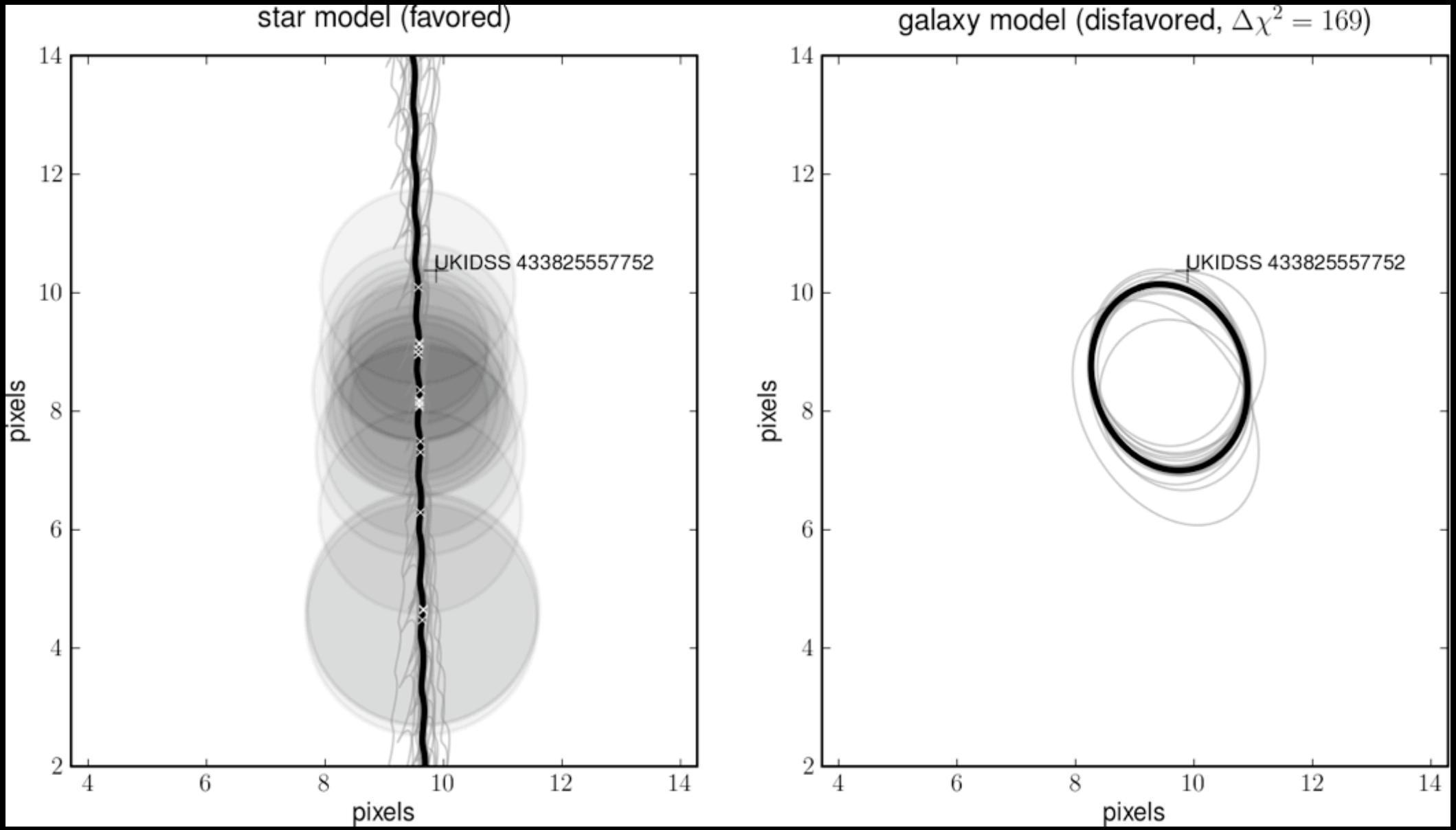
science with hobbyists

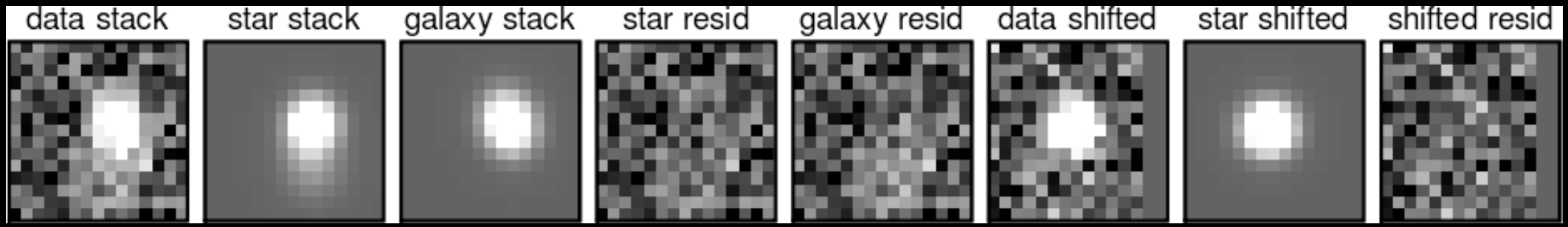
- rapid and high time-resolution response
  - GRBs, planetary microlensing, variable stars
  - near-earth object orbit determination
- pre-event imaging for transient events
- *ab initio* discovery
  - known classes, such as SNe, NEOs, transits
  - new classes, such as “gamma-free” GRBs
- deep, faint, and proper-motion science

going deep

- The combined aperture of all amateur telescopes exceeds the combined aperture of all professional
  - by far, but...
  - in the visible / optical
  - can we really get sqrt(N)?


51081
51132
51136
51819
52225
52288
52551
52558
52576
52577
52585
52640
52909
52910
52935
52963
52971
53351


star model (favored)
galaxy model (disfavored, $\Delta\chi^2 = 169$)
UKIDSS 433825557752
UKIDSS 433825557752
pixels
pixels
pixels
pixels
2
4
6
8
10
12
14


data stack
star stack
galaxy stack
star resid
galaxy resid
data shifted
star shifted
shifted resid

historical data

- the best astrometric catalog is USNO-B
  - one billion stars with positions and proper motions
  - less than *one percent* of the available historical data
- archives
  - contain *millions* of science-grade plates
  - Harvard archive alone has the sky *500 times over*
  - scanning is cheap but not done
  - meta-data are often more difficult than scanning

a new "observatory"

- automatically calibrate and archive all data
  - amateur, professional, historical; <1000 Tb
  - *data vetting* and interoperability
  - "opposite" of the Virtual Observatory
- create a global community of observers
  - information can flow both ways
  - think "astronomical" wikipedia or wikimapia
  - open-source sky survey
  - "if you like this part of the sky, you might also like..."

the end

- http://astrometry.net/