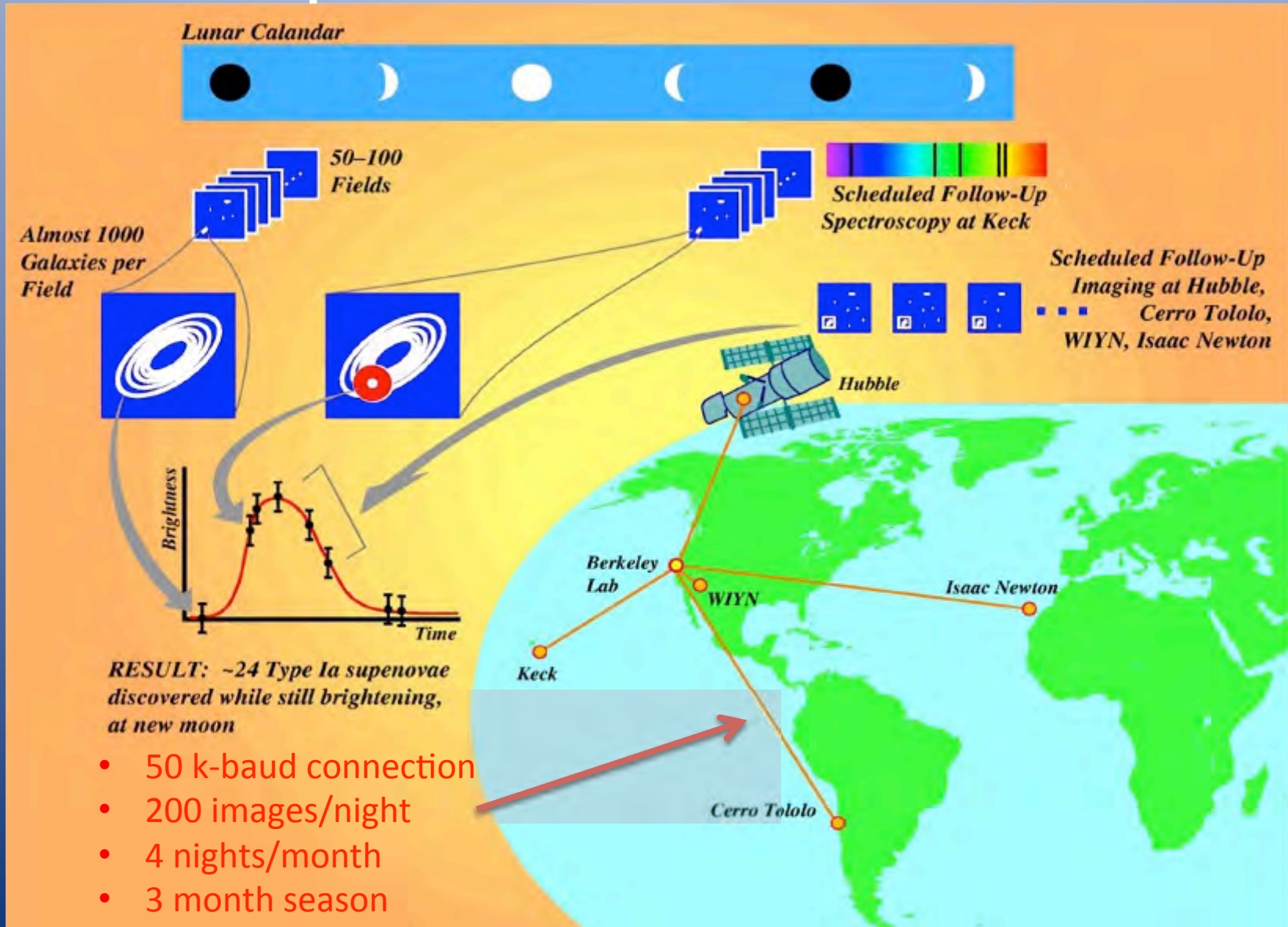


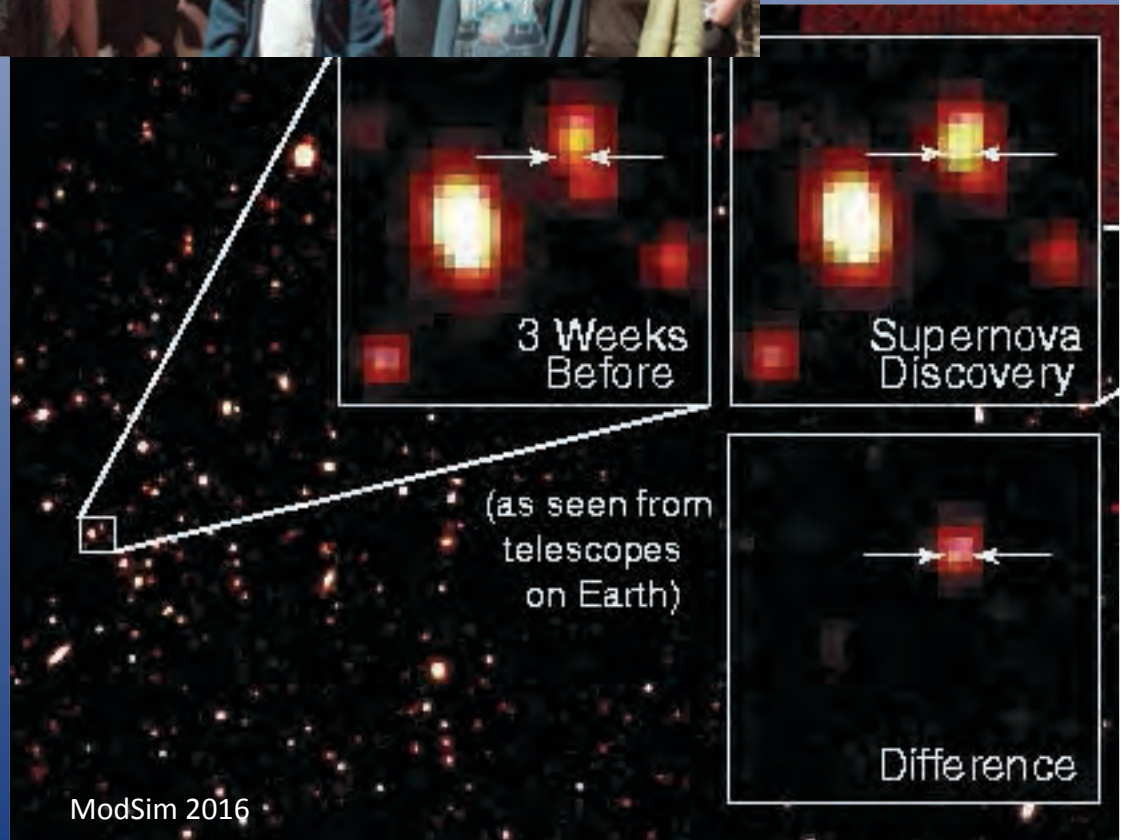
ModSim Challenges for a Scientist's Use of Workflows

Peter Nugent (LBNL/UCB)

Supernovae circa 1995



Supernovae circa 1995

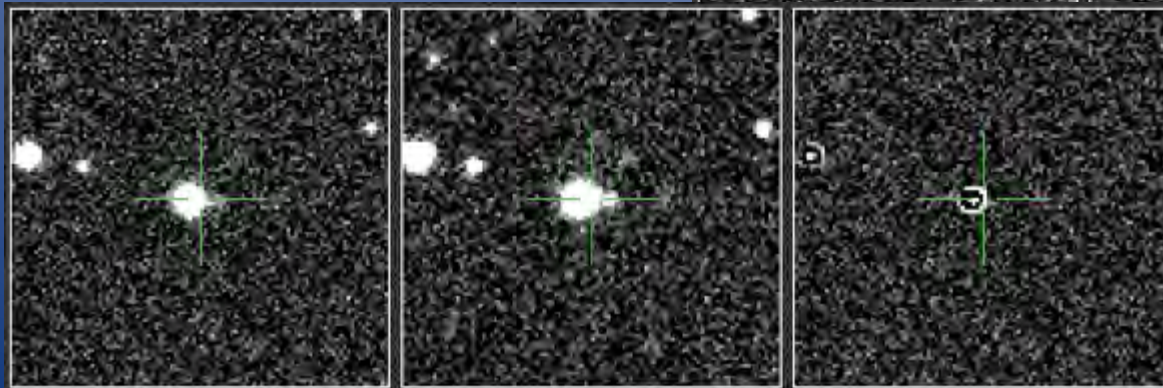
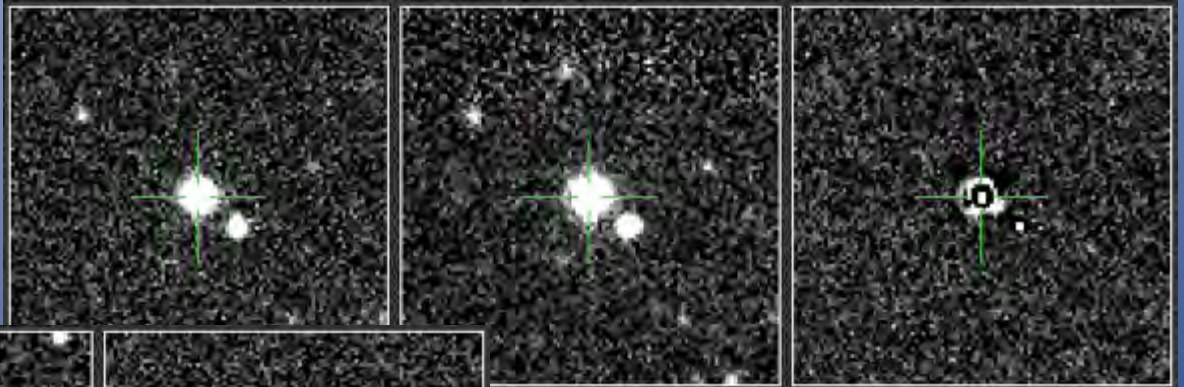


Supernovae circa 1998



Per image we would have ~ 200 $5\text{-}\sigma$ detections. We would require 2 independent detections.

Typically only 50-200 images taken per night - 4 sq. deg. of sky.



Cuts were made based on shape, motion, etc., and a scanner would have to look at ~ 5 candidates per image.

Supernovae circa 2000

Pain begins....

NEAT Search Facilities

Site:	Haleakala	Palomar I	Palomar II
Aperture:	1.2m	1.2m	1.2m
Nights/Month:	18 dark/gray	18 dark/gray	18 dark/gray
Imager Format:	4k × 4k	16k × 16k	16k × 24k
Imager Scale:	1.33"/pixel	0.50"/pixel	0.50"/pixel
Field of View:	1.1° × 3.4°	1.1° × 3.4°	2.3° × 4.0°
Filters:	open	open	4 fixed filters
Exposures:	3 × 60 sec	3 × 60 sec	TBD
Readout:	20 sec	20 sec	TBD
Night Area:	600□°	800□°	(2000 □°)
Start:	Mar 2000	Feb 2001	~Dec 2001
Data (Compressed):	12 Gbyte/night	17 Gbyte/night	(28 Gbyte/night)

**~1000 sq. deg. - 250 X increase in scale per night
EVERY NIGHT !!!**

Supernovae circa 2000



FEDEX Networking: Do not underestimate the bandwidth of a station wagon filled with DAT tapes... achieved 200 kB/s



Intelligent Optical Network Infrastructure

PTF (2009-2012), iPTF(2013-2016)

- CFH12k camera on the Palomar Oschin Schmidt telescope
 - 7.8 sq deg field of view, 1" pixels
 - 60s exposures with 15-20s readout in r, g and H-alpha
 - First light Nov. 24, 2008.
 - First useful science images on Jan 13th, 2009.
- 2 Cadences (Mar. - Nov.) 2009-2011
 - Nightly (35% of time) on nearby galaxies and clusters (g/r)
 - Every 3 nights (65% of time) on SDSS fields with minimum coverage of 2500 sq deg. (r) to 20th mag 10-sigma
 - H-alpha during bright time (full +/-2 days)

Nov-Feb, minute cadences on select fields.

Supernovae circa 2009

Discovery and Follow-up



P48:
Discovery Engine

P60:
Followup

Instrumentation, system
design, first results

Law, Kulkarni, Dekany et al. 2009 PASP 121 1395L

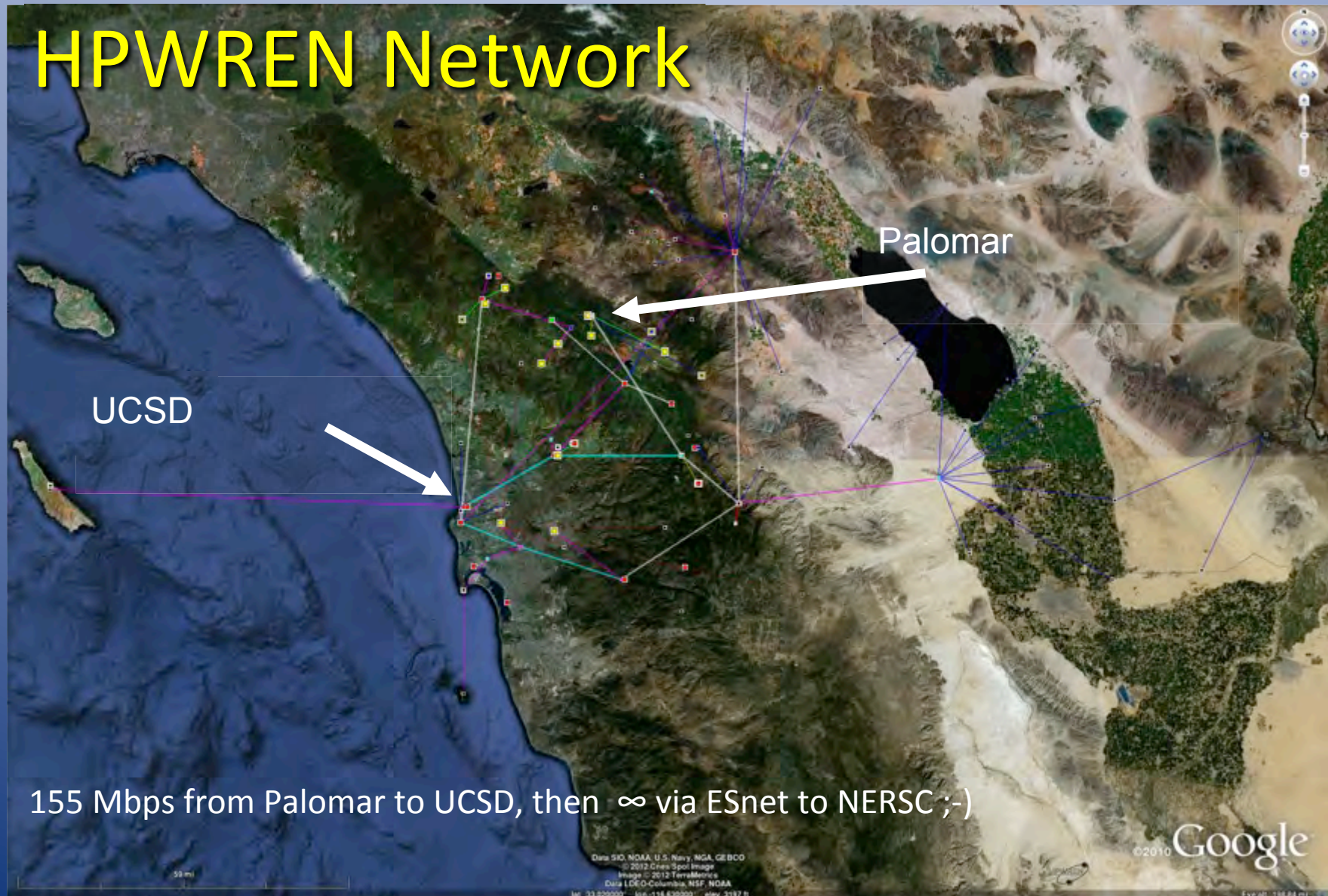
Science plans

Rau, Kulkarni, Law et al. 2009 PASP 121 1334R

2010 survey status

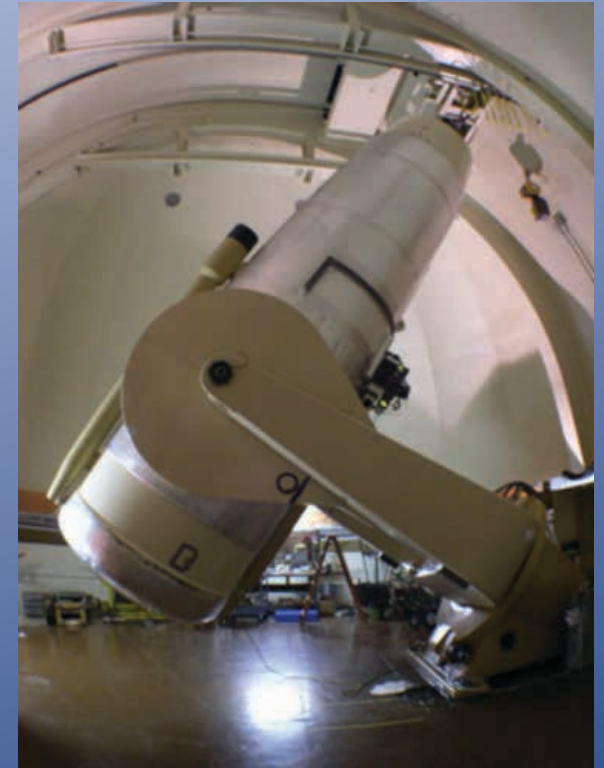
Law et al. 2010 SPIE 7735
ModSim 2016

HPWREN Network



155 Mbps from Palomar to UCSD, then ∞ via ESnet to NERSC ;-)

PTF Camera



92 Mpixels, 1" resolution, R=21 in 60s

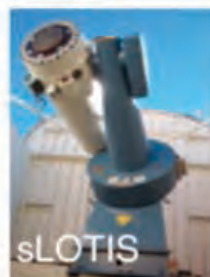
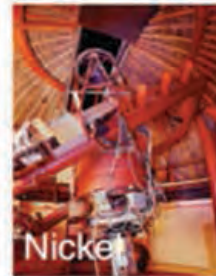
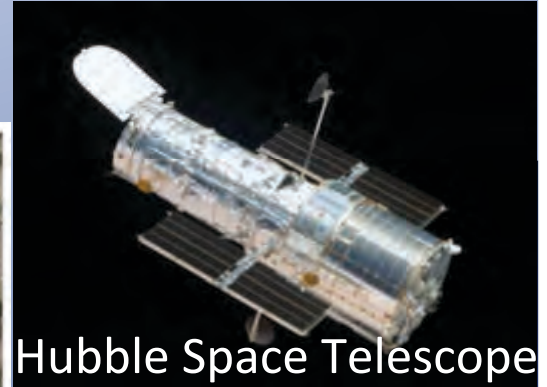
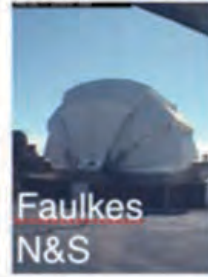
PTF Science

PTF Key Projects	
Various SNe	Dwarf novae
Transients in nearby galaxies	Core collapse SNe
RR Lyrae	Solar system objects
CVs	AGN
AM CVn	Blazars
Galactic dynamics	LIGO & Neutrino transients
Flare stars	Hostless transients
Nearby star kinematics	Orphan GRB afterglows
Type Ia Supernovae	Eclipsing stars and planets
Tidal events	H-alpha $\frac{1}{2}$ sky survey

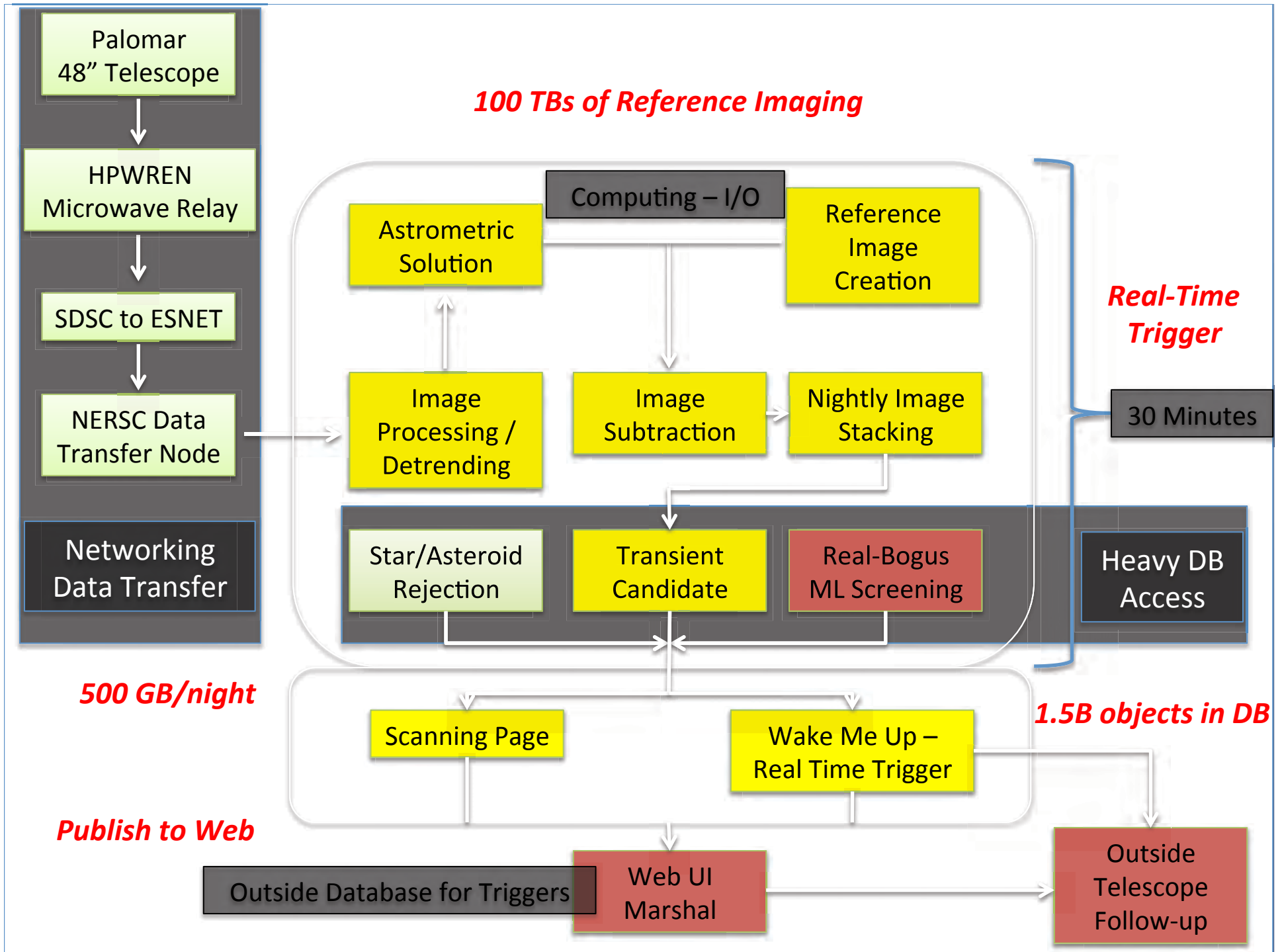
The power of PTF resides in its diverse science goals and follow-up.

PTF Science

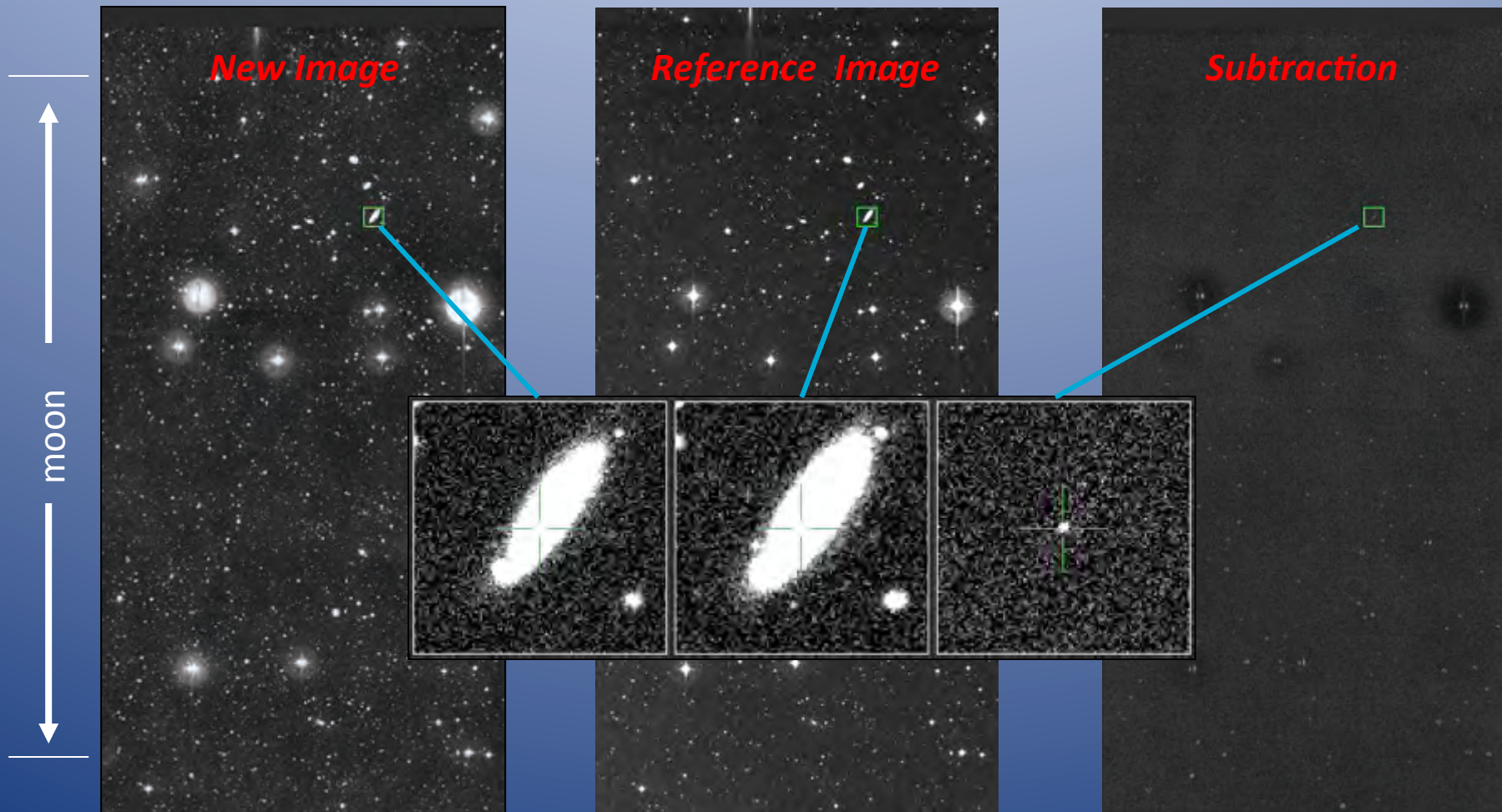
▼► Detected transients will be followed up using a wide variety of optical and IR, photometric and spectroscopic followup facilities.



The power of PTF resides in its diverse science goals and follow-up.



Real or Bogus – Machine Learning Analysis



4096 X 2048 CCD images - over 3000 per night – producing 1.5M bogus detections, 50k known astrophysical objects and only 1-2 new astrophysical transients of interest every night. Machine learning is used to wade through this sea of garbage.

PTF Database

	R-band	g-band
images	1.82M	305k
subtractions	1.52M	146k
references	29.2k	6.3k
Candidates	890M	197M
Transients	42945	3120

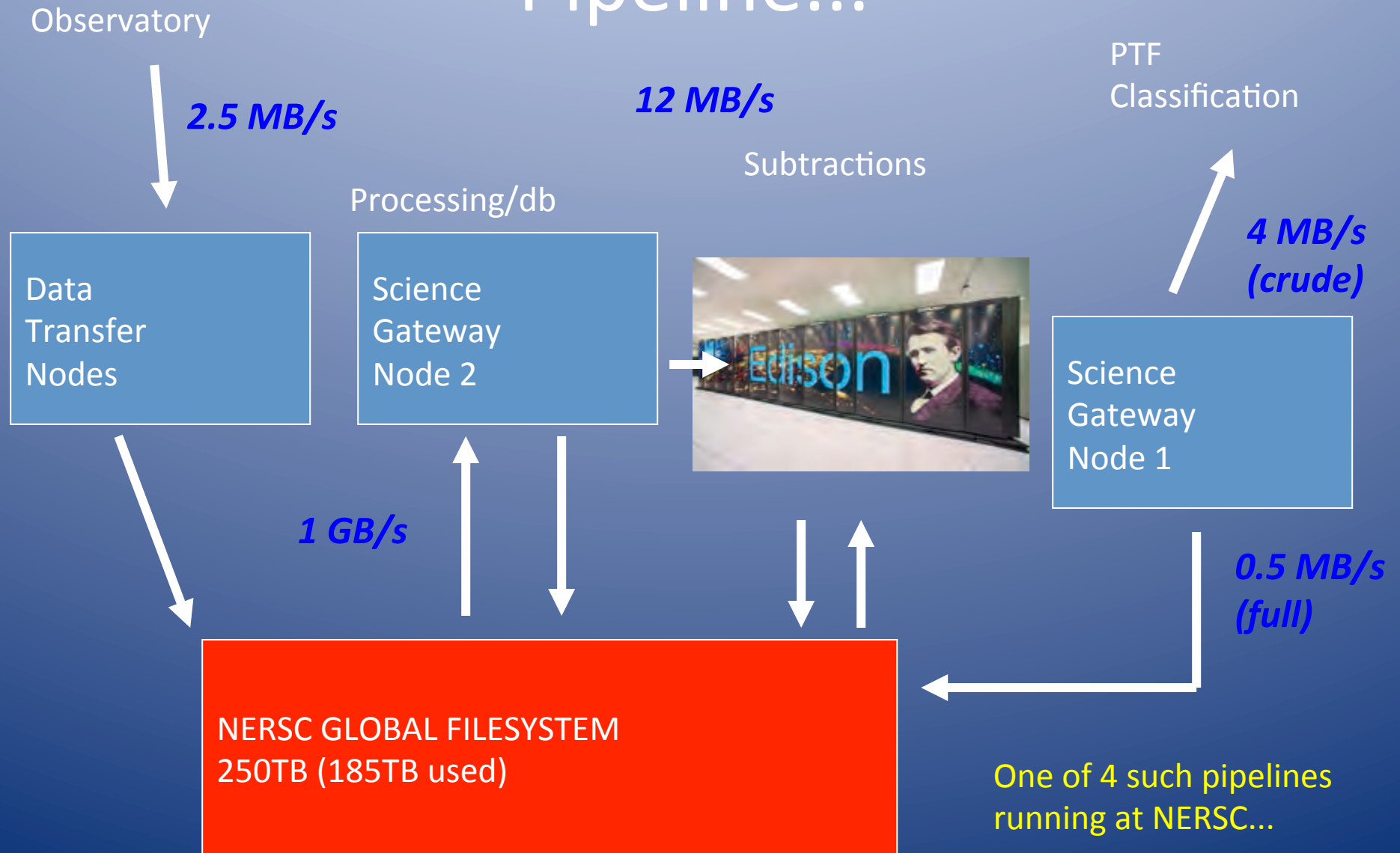
All in 851 nights from 2009-2012.

An image is an individual chip (~0.7 sq. deg.)

The database is now 1 TB.

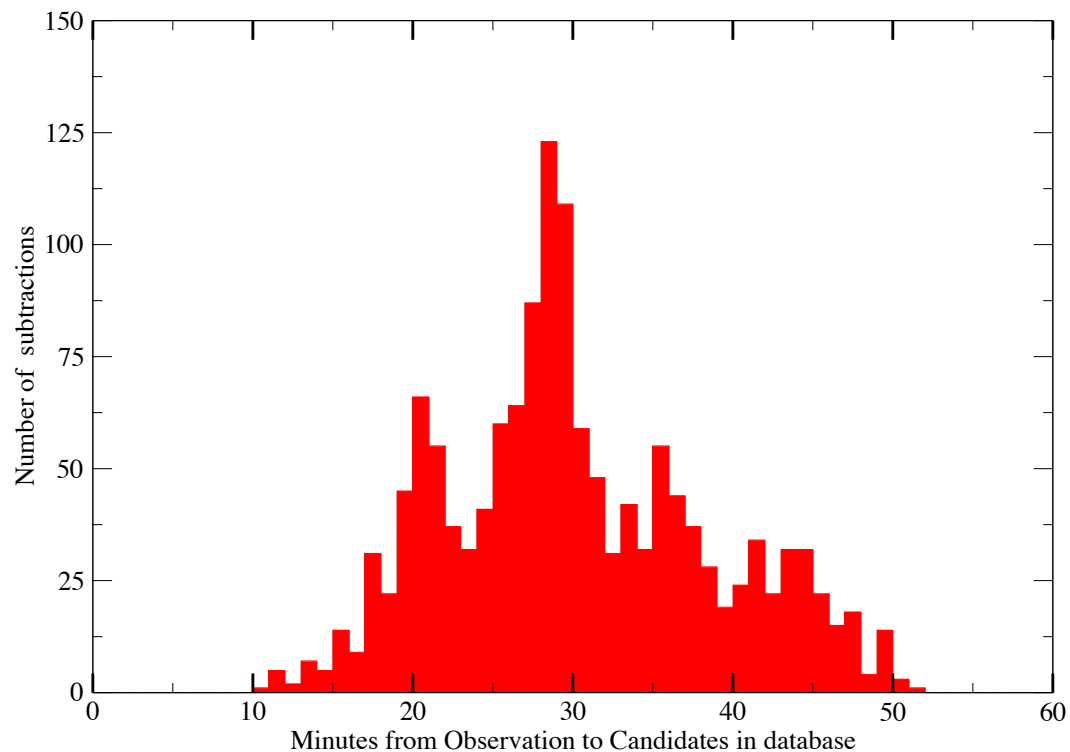
Now doubled in size from iPTF from 2013-2016.

Pipeline...



PTF Turn-around

Typical night: 2012-07-06

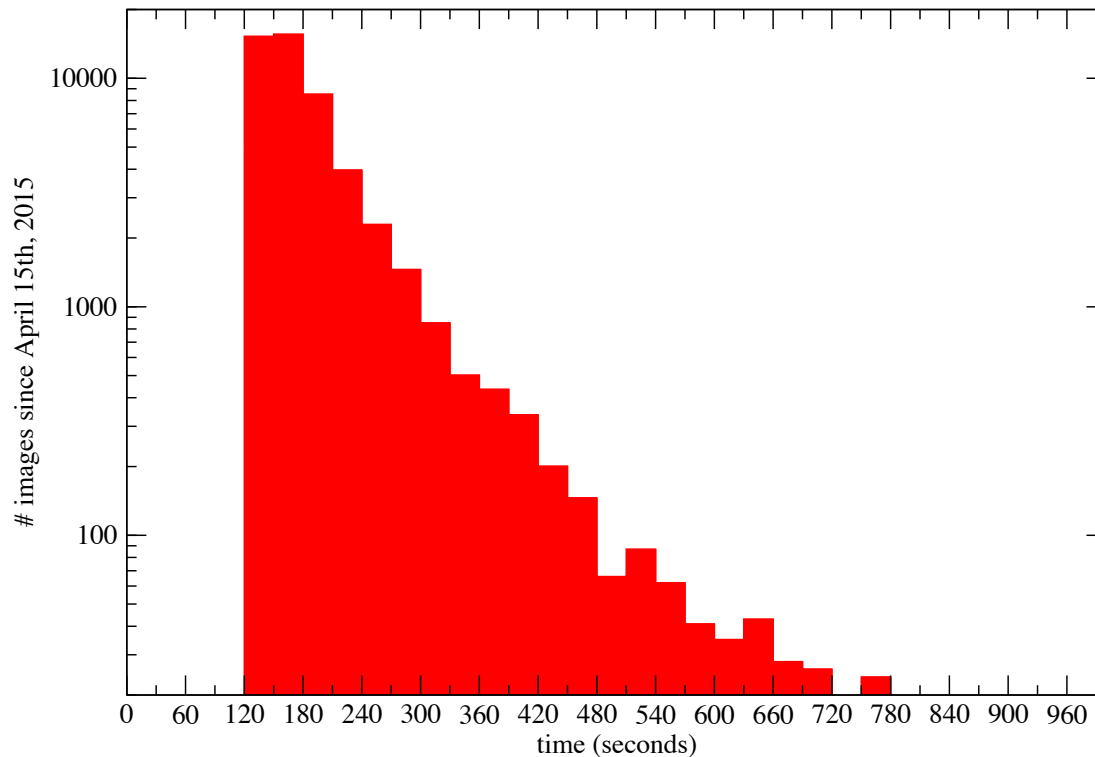


What does “real-time” subtractions really mean?

For 95% of the nights all images are processed, subtractions are run, candidates are put into the database and the local universe script is run in < 1hr after observation.

Median turn-around is 30m.

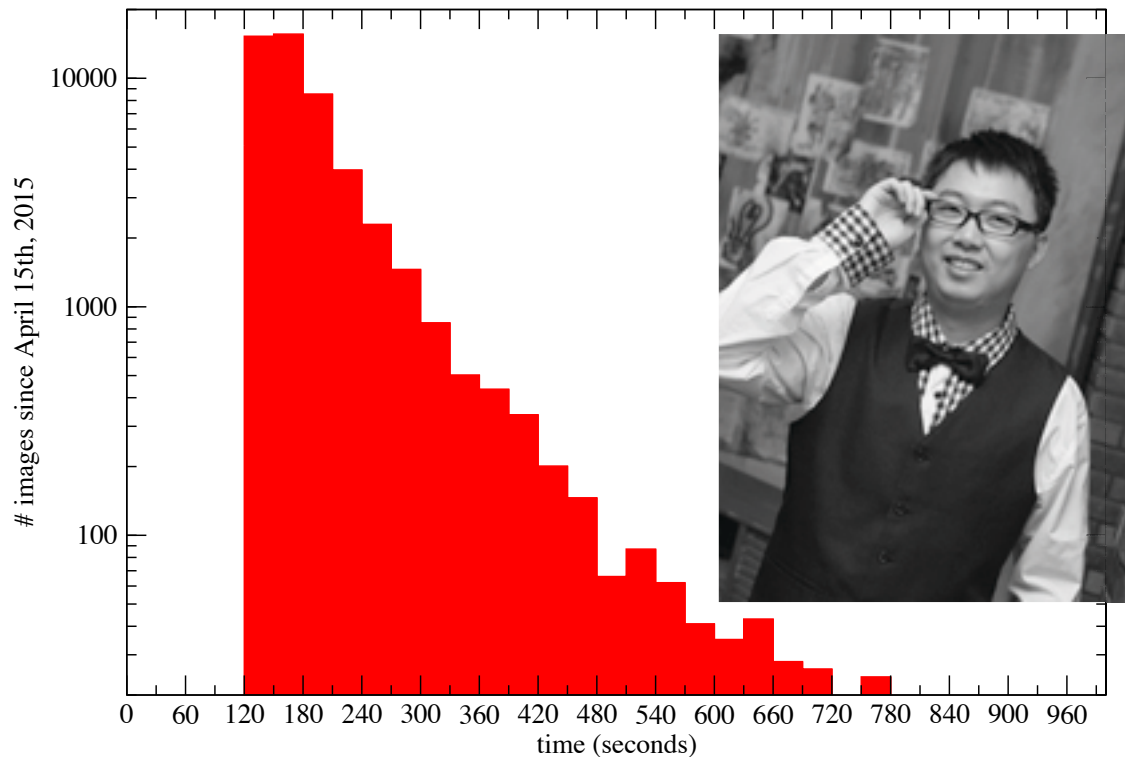
iPTF turn-around



Due to the X-SWAP project (Extreme-Scale Scientific Workflow Analysis and Prediction), funded through the ASCR LAB-1088 call (Analytical Modeling for Extreme-Scale Computing Environments), we have been able to understand and eliminate a lot of our inefficiencies and decrease the turn-around by an order of magnitude!

Better use of the Lustre filesystem (for everything), better use of OpenMP in all codes, reserved nodes, etc.

iPTF turn-around



We made major changes to the old pipeline.

- Pipeline completely instrumented for timings.
- Identified and fixed python load time on Edison (15min to 5 sec).
- Moved all I/O in processing to Lustre /scratch filesystem
- Now optimizing db access

Yi Cao's Caltech thesis May 3, 2016. This fall will become an eScience Postdoctoral Fellow at University of Washington.

ModSim 2016

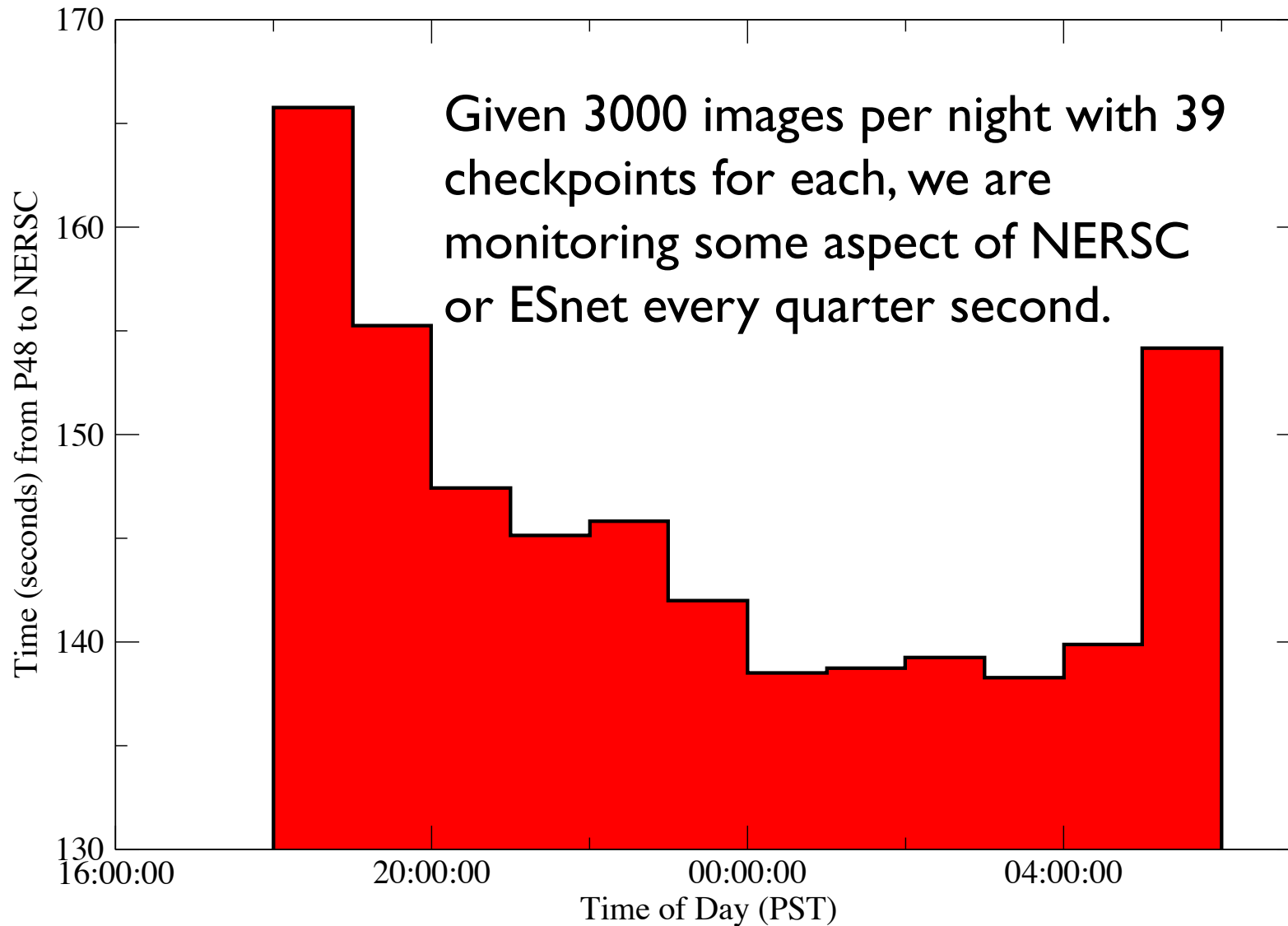
Typical turnaround is now < 5 minutes for 95% of the data!

Instrumented Pipeline with 39 Checkpoints

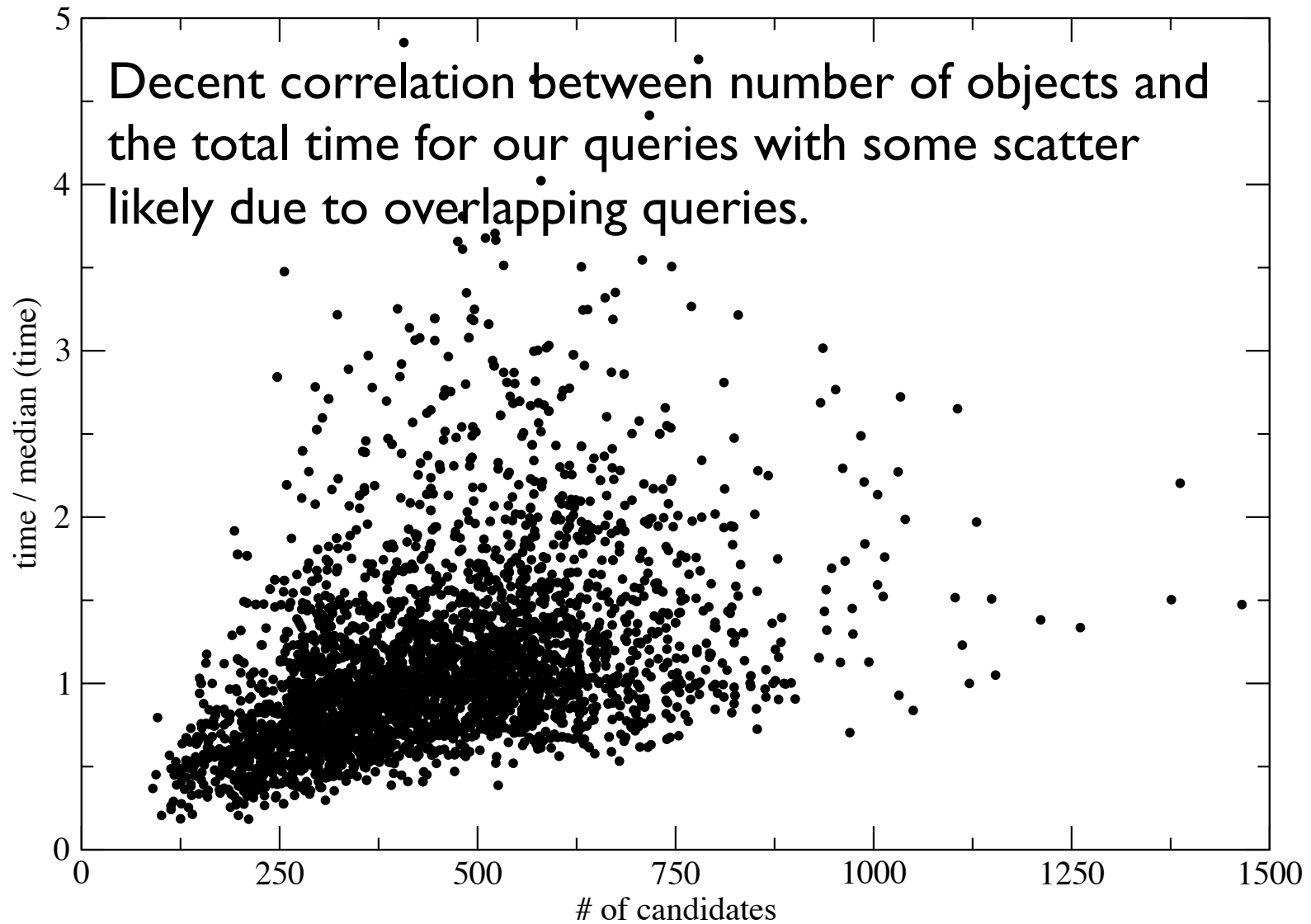
Covers everything from:

- Pulling the data from the telescope
- I/O on scratch
- Subtraction software
- Running ML algorithms
- Loading the db with discoveries
- Performing difficult geometric queries to match with known stars, asteroids, previous discoveries, etc.
- Copying data from scratch to project

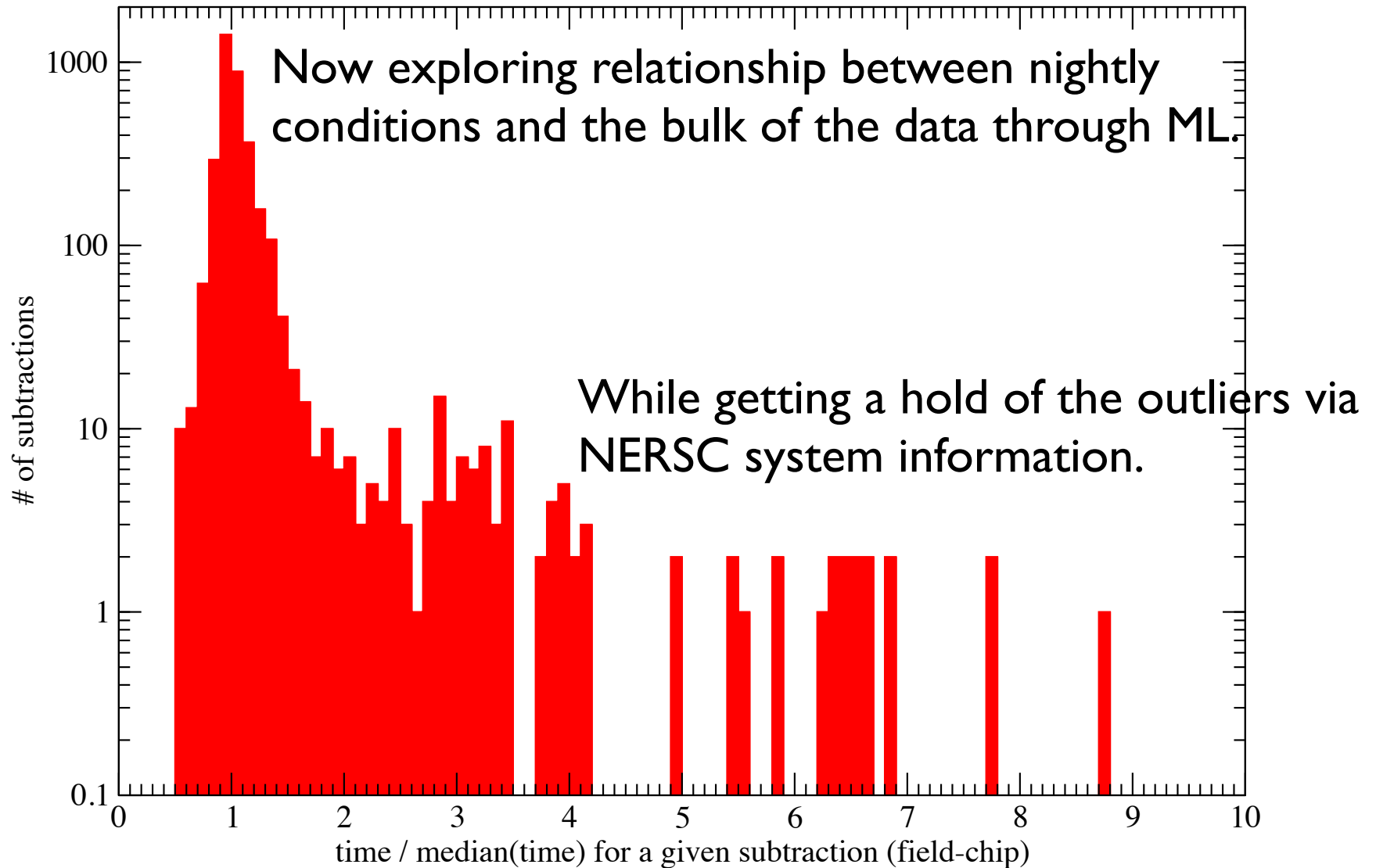
For 8 hours every night, we now know more about the NERSC center than they do in real-time.



DB Access



I/O time on Cori

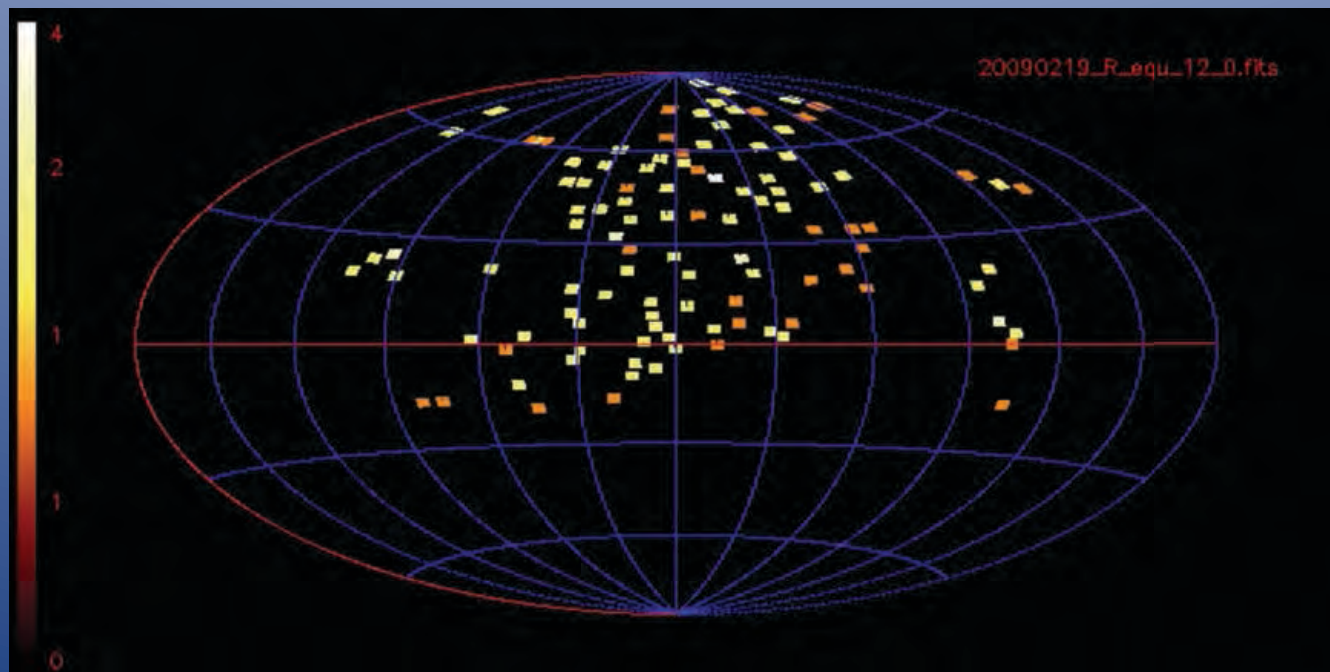


PTF Sky Coverage

To date:

- 2338 Spectroscopically typed supernovae
- 10^6 Galactic Transients
- 10^4 Transients in M31

139 publications, 6 in *Nature* and 2 in *Science* since late 2009

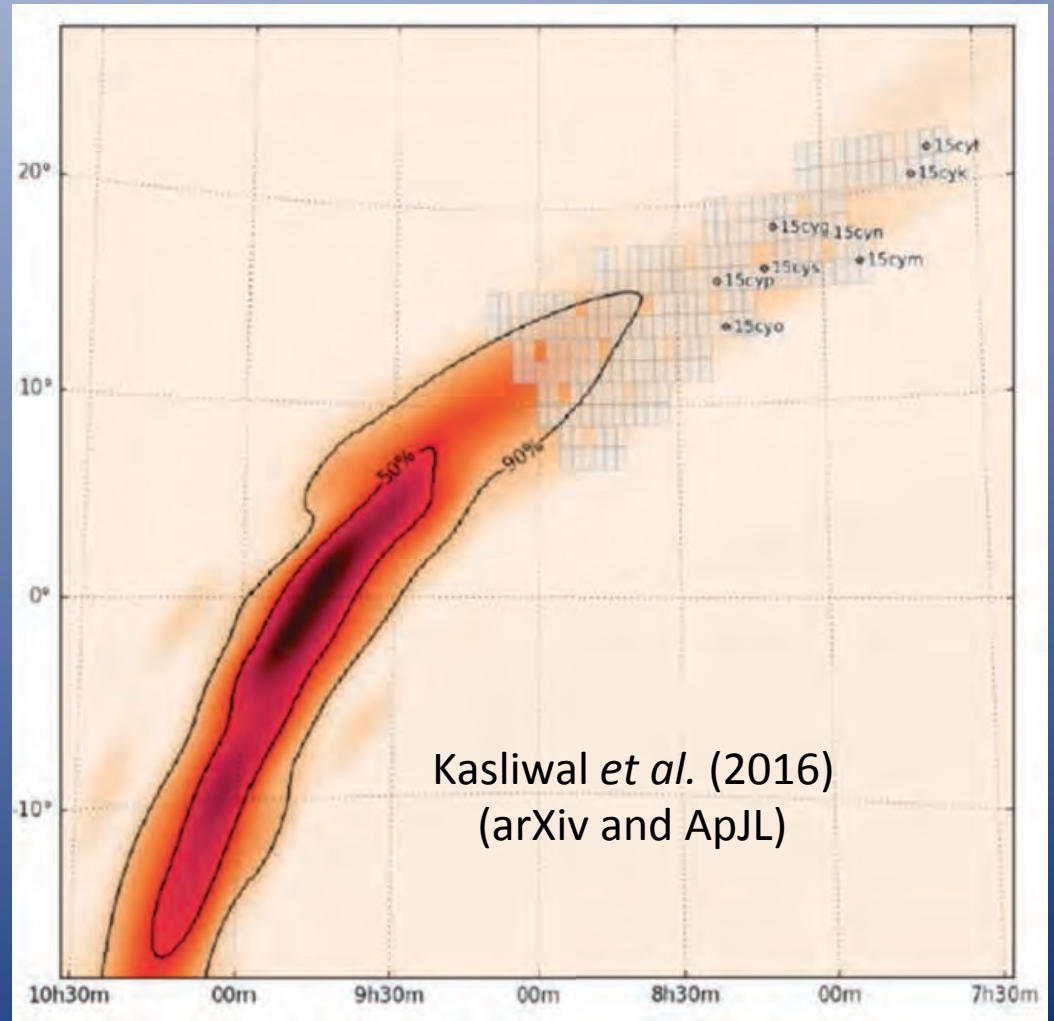
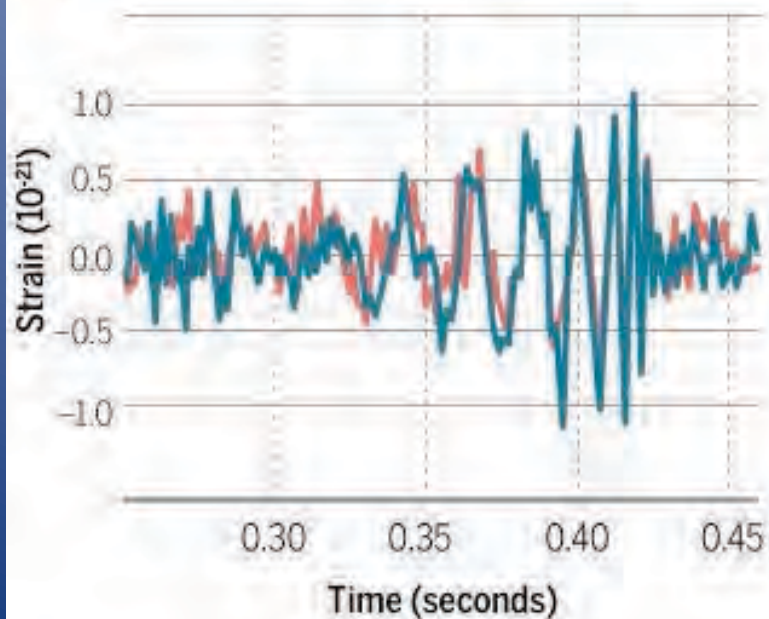


Gravitational Wave Trigger GW150914

Signals in synchrony

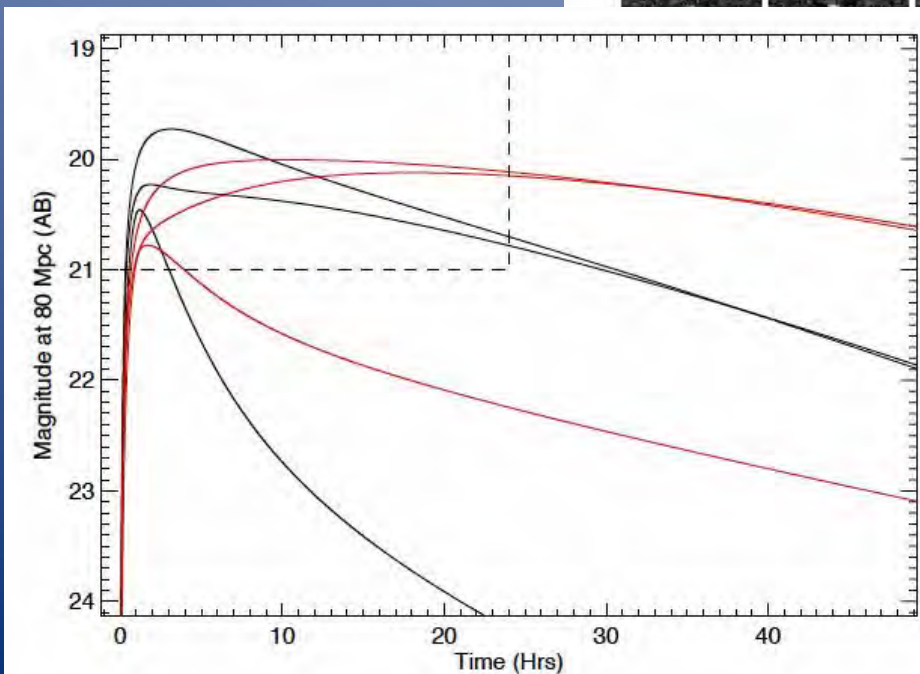
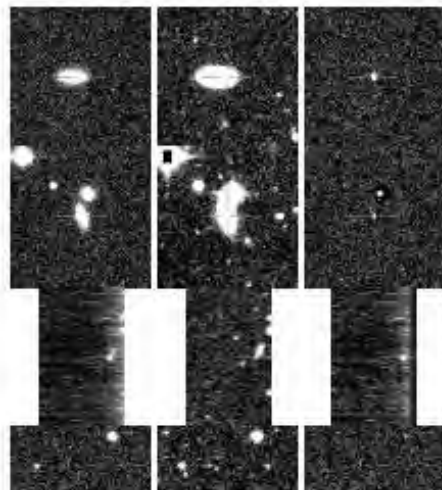
When shifted by 0.007 seconds, the signal from LIGO's observatory in Washington (red) neatly matches the signal from the one in Louisiana (blue).

● LIGO Hanford data (shifted) ● LIGO Livingston data



GW150914

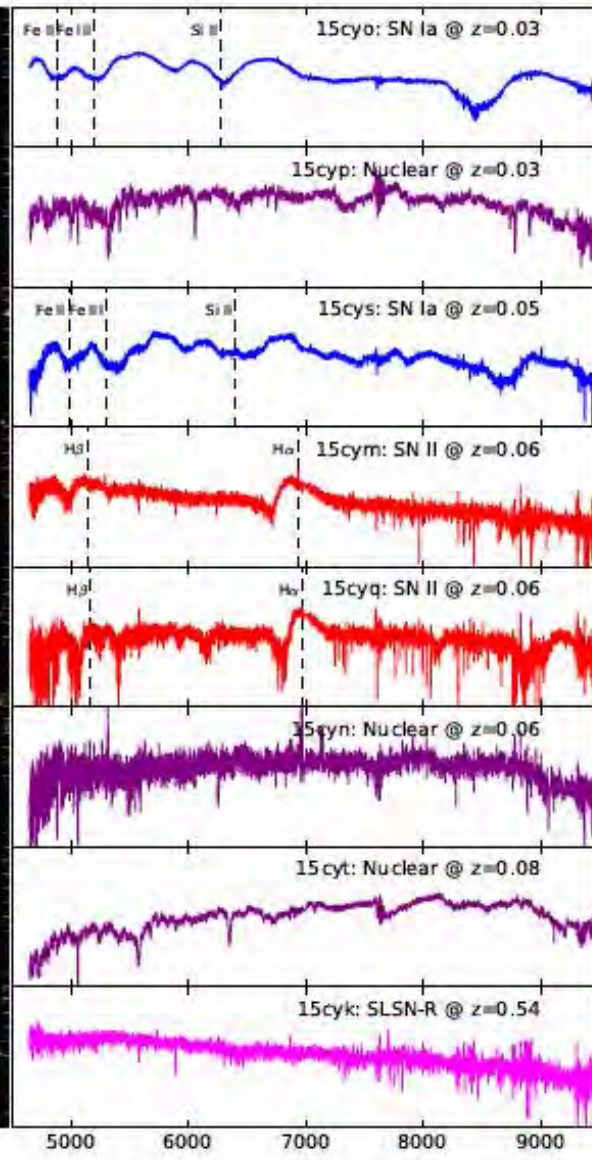
Going to have to be able to sift through a lot of stuff, and react quickly with follow-up, to get on the optical companion for a GW trigger.



new Ref

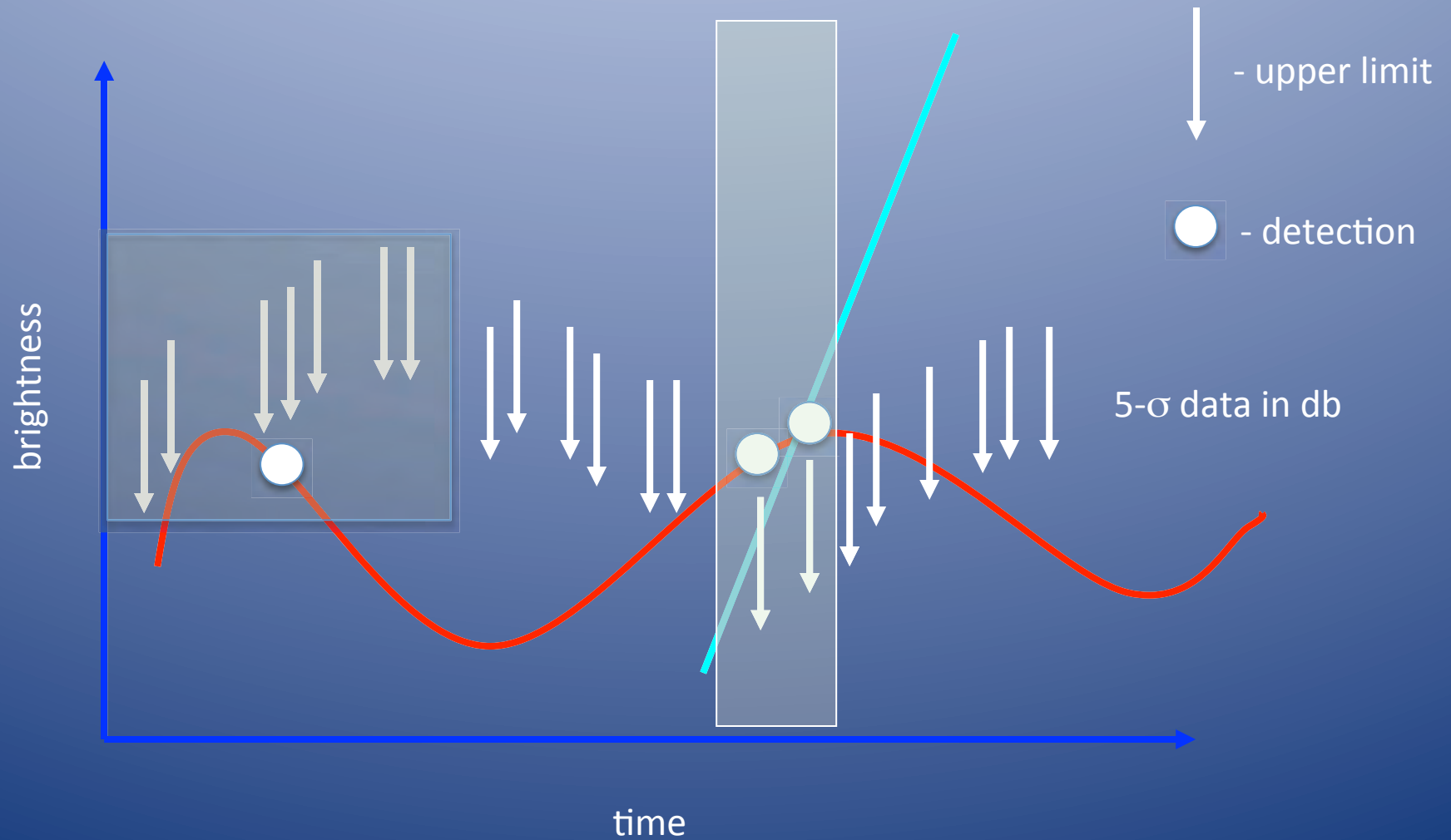
Sub

SDSS

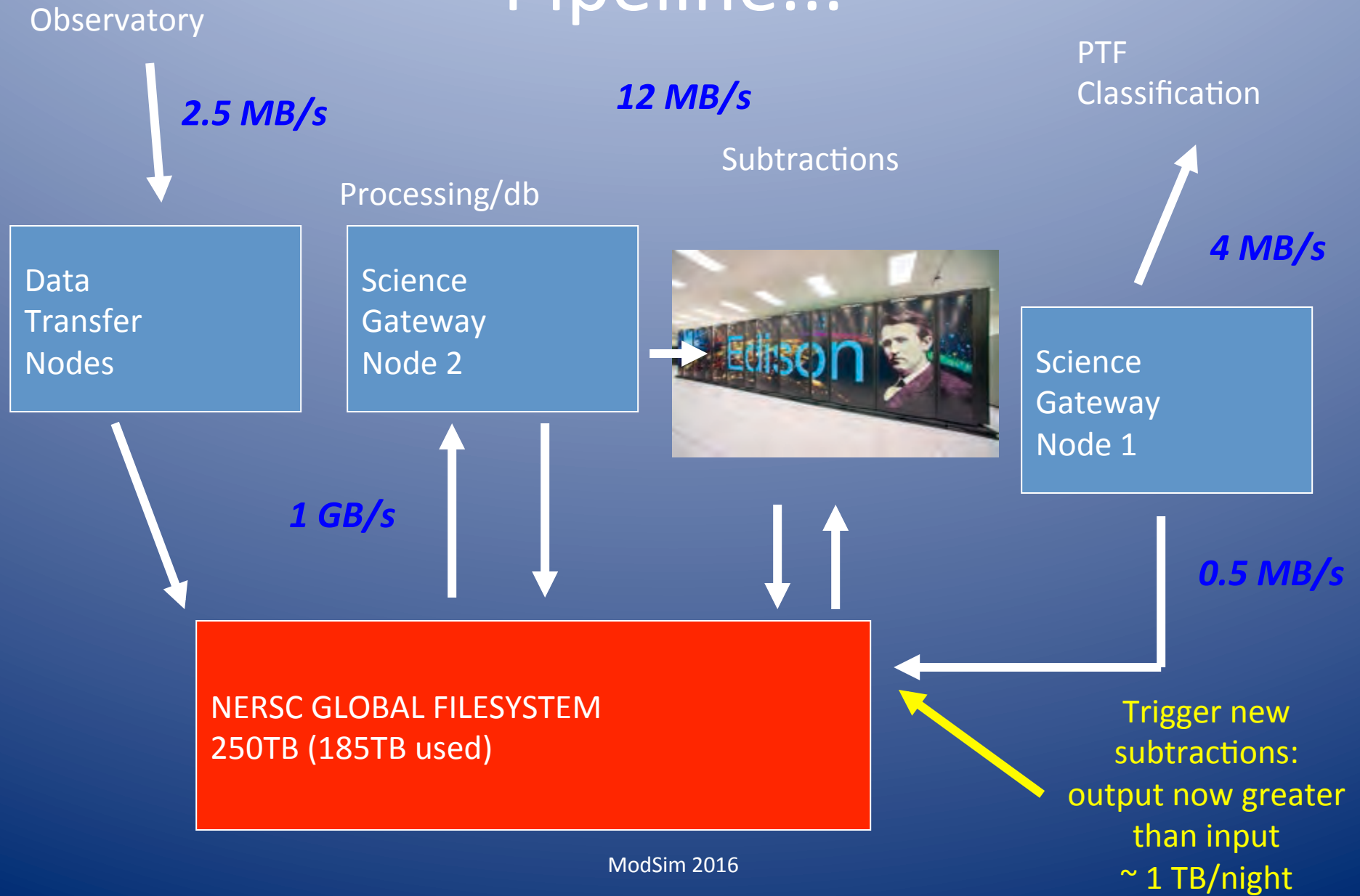


Observed Wavelength (\AA)

Bottlenecks...crude vs. real



Pipeline...



Zwicky Transient Facility



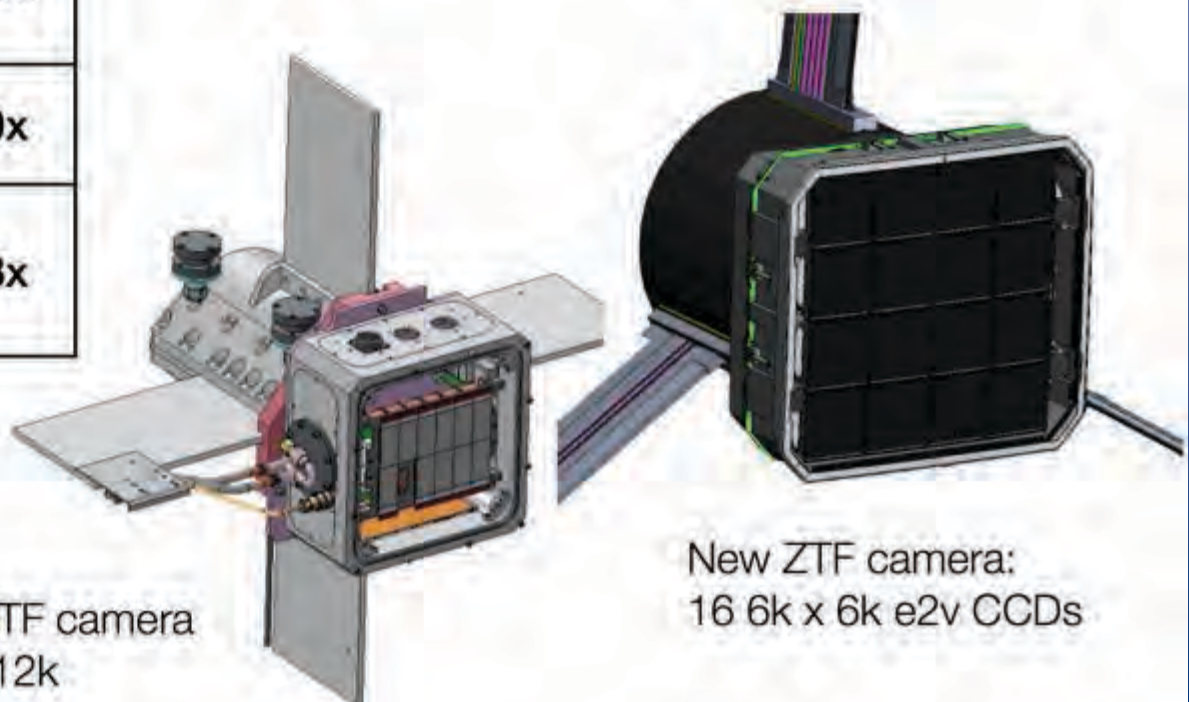
ZTF will survey an order of magnitude faster than PTF.

	PTF	ZTF
Active Area	7.26 deg ²	47 deg ²
Overhead Time	46 sec	<15 sec
Optimal Exposure Time	60 sec	30 sec
Relative Areal Survey Rate	1x	15.0x
Relative Volumetric Survey Rate	1x	12.3x

3750 deg²/hour

⇒ 3π survey in 8 hours

>250 observations/field/year
for uniform survey



Existing PTF camera
MOSAIC 12k

New ZTF camera:
16 6k x 6k e2v CCDs

Future



LSST - 15TB data/night
Only one 30-m telescope
How many triggers can we handle???