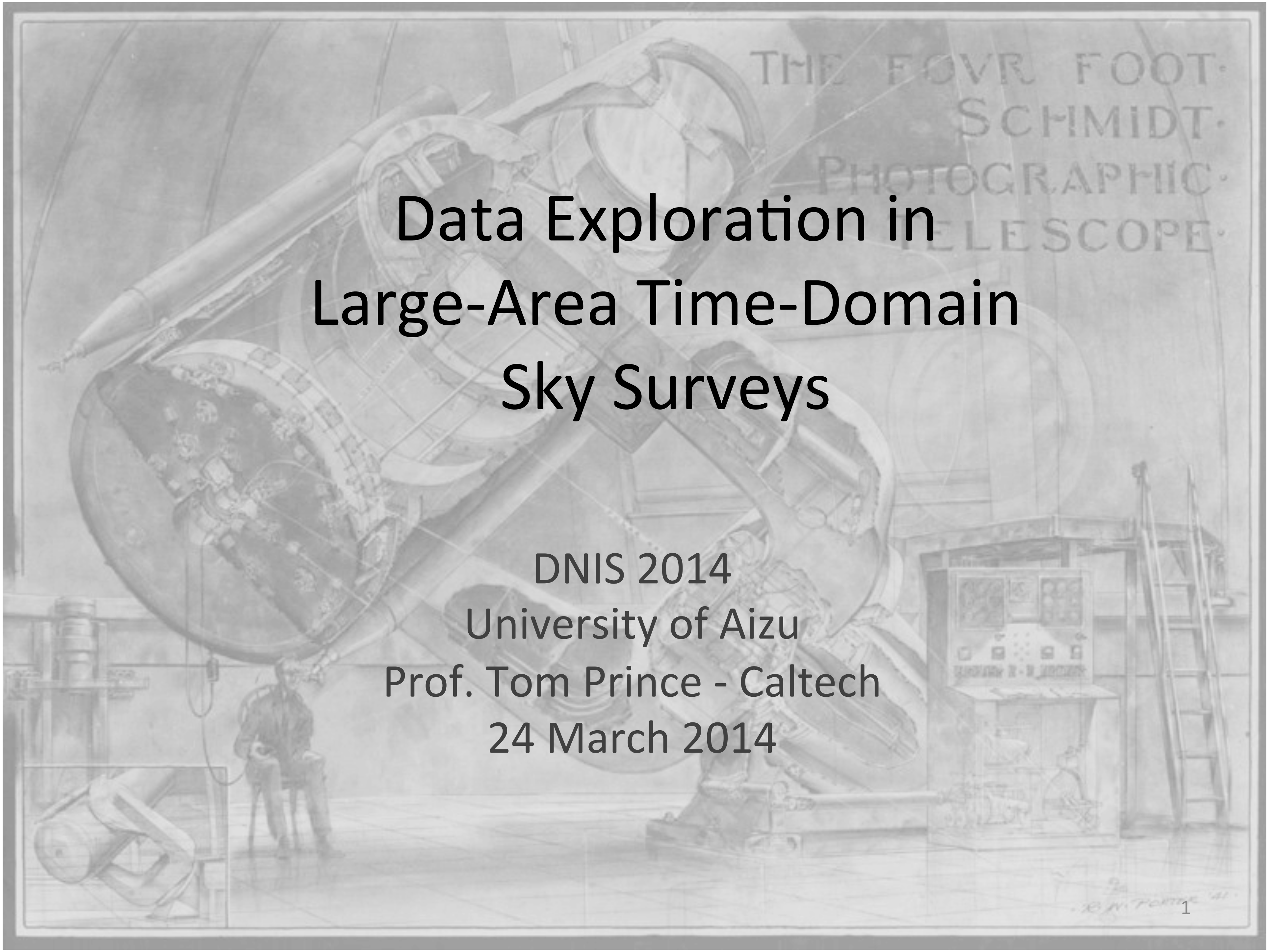
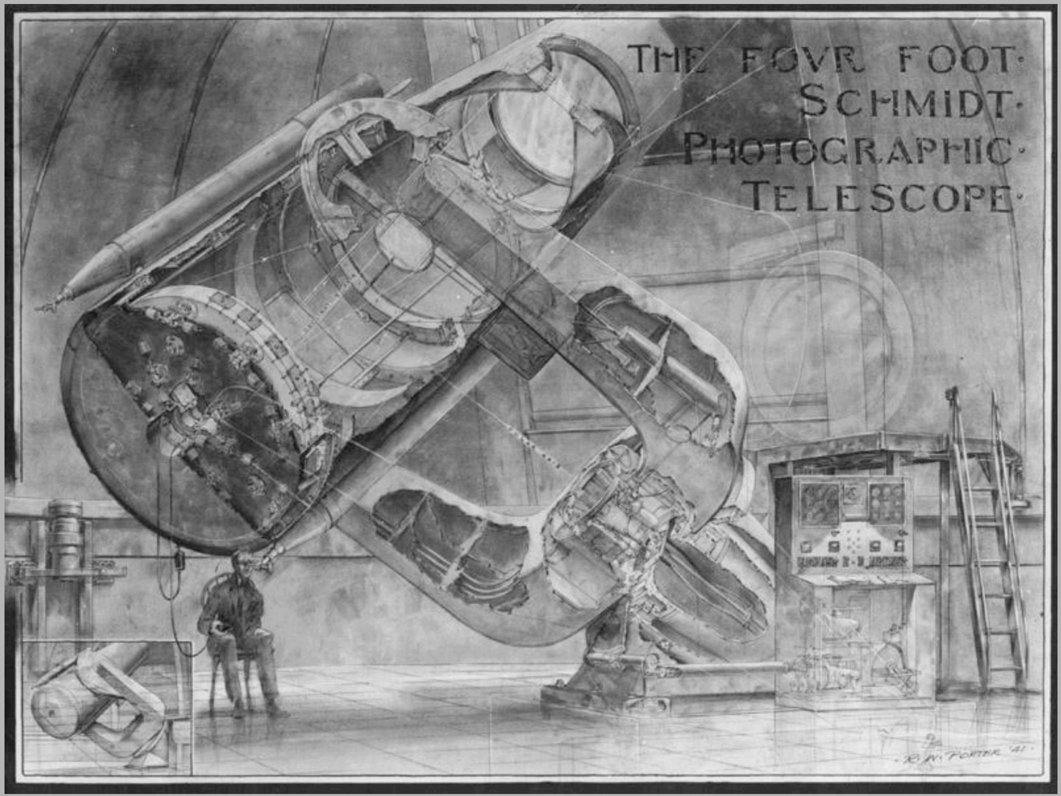


Data Exploration in Large-Area Time-Domain Sky Surveys

DNIS 2014
University of Aizu
Prof. Tom Prince - Caltech
24 March 2014



Part I:

A HISTORICAL INTRODUCTION: 1933-2021

Early Time-Domain Astronomy:
Fritz Zwicky (1898-1974)
Caltech Astronomer

Zwicky at the 18-inch Schmidt Telescope

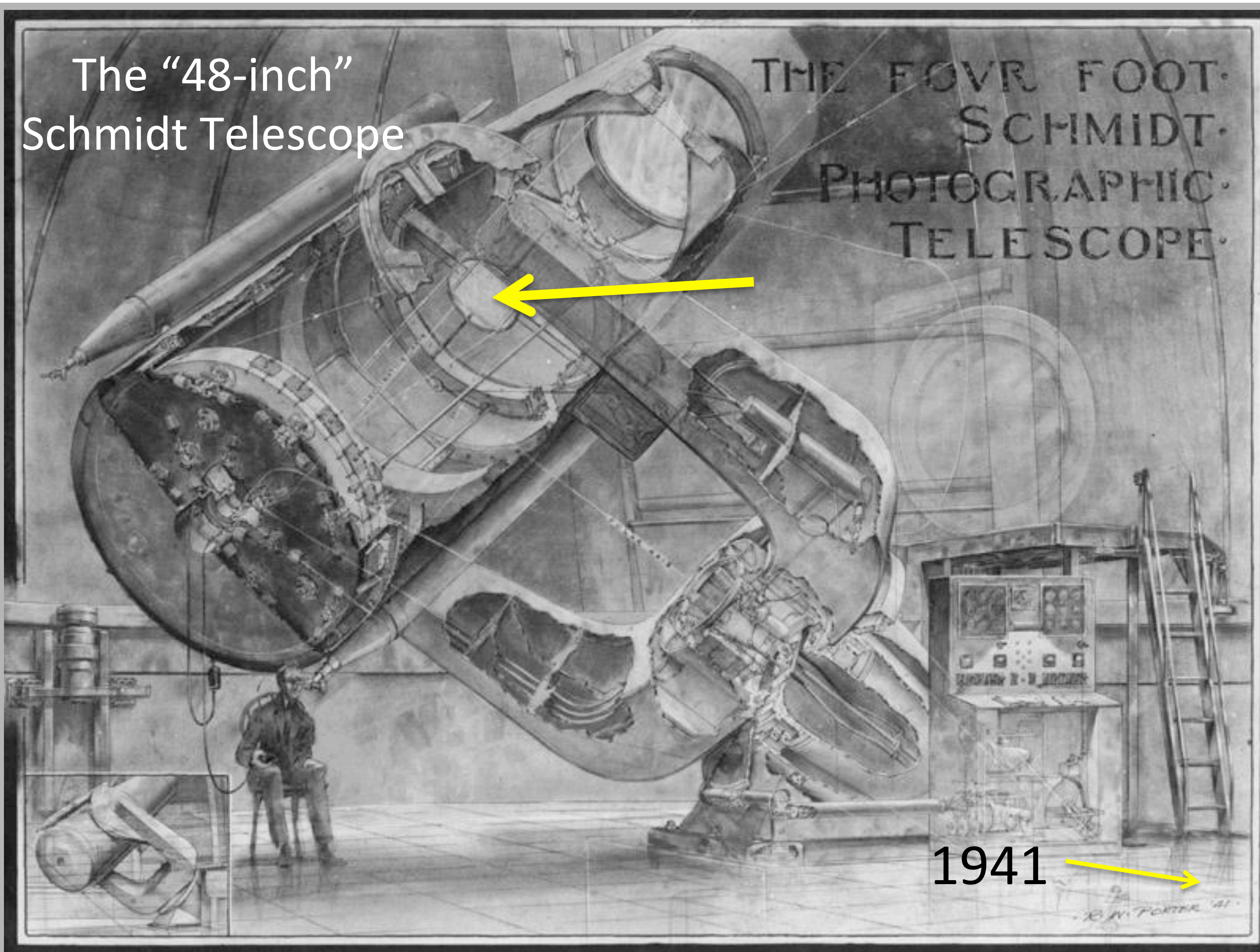
Supernovae

SN1937A

"the most unrecognized genius of twentieth century astronomy"

1933 - Zwicky & Baade:

(1) massive stars end their lives in explosions which blow them apart (supernovae)
(2) such explosions produce cosmic rays
(3) they leave behind a collapsed star made of densely-packed neutrons (neutron stars).

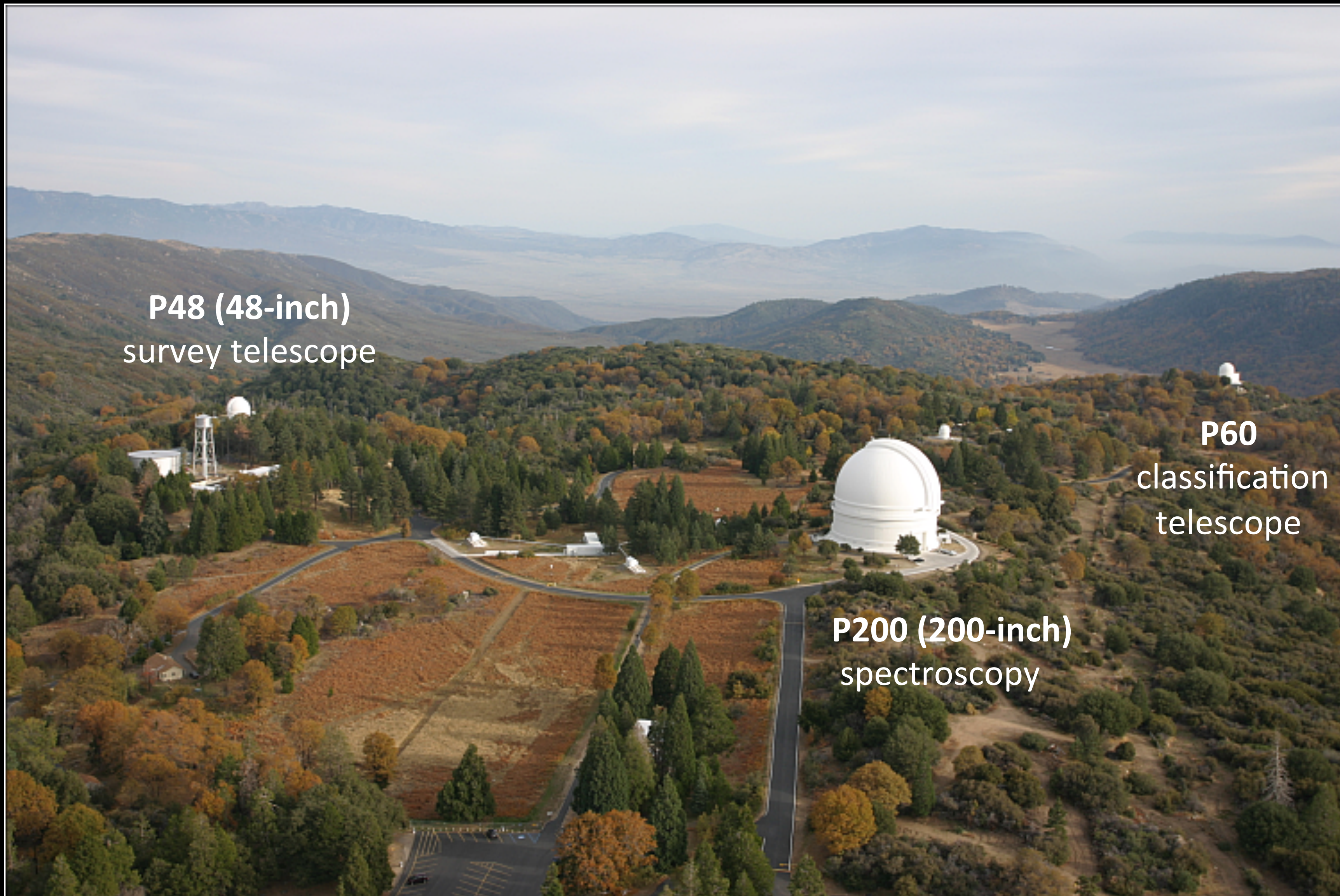


The Next Big Step

1940's

Best time-domain telescope until 2021!

Four feet = 48 inches
= 1.2 meter



The Caltech Palomar Observatory

Palomar Sky Survey Plate:
36 cm x 36 cm ~ 1GByte

Hubble at the 48-inch Schmidt
~ 45 square degree field of view

Human Eye: Retina->Brain ~ 10Mbit/s

Palomar Sky Survey Plates

1^o x 1^o image of Coma Cluster

(x 40 for full Palomar Sky Survey plate)

Image Understanding in the Time-Domain

Which are stars and which are galaxies?

Which are asteroids and which are stars?

Which are real objects and which are ghost reflections?

Which of the objects has varied in brightness by 1%?

(x 640 for full survey plate)
(x 640,000 for full sky)

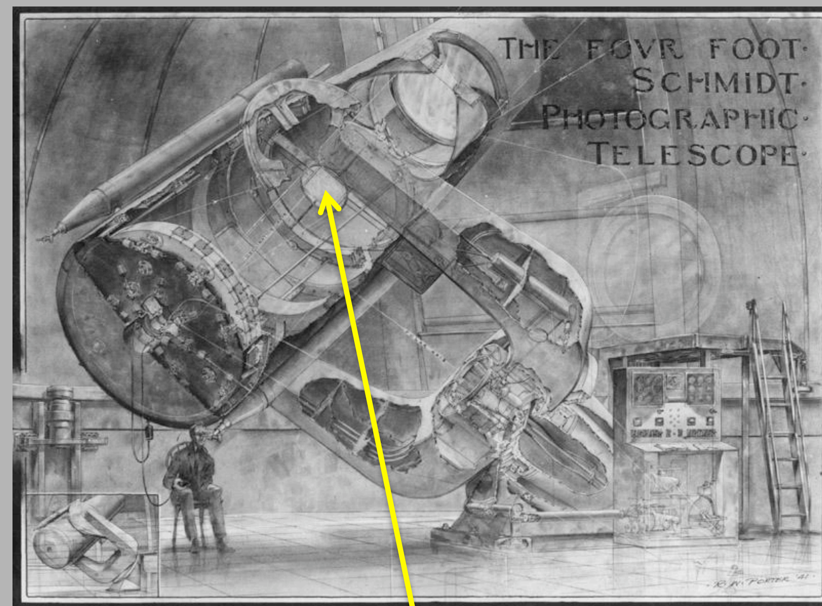
(0.25^{o} x 0.25^{0})

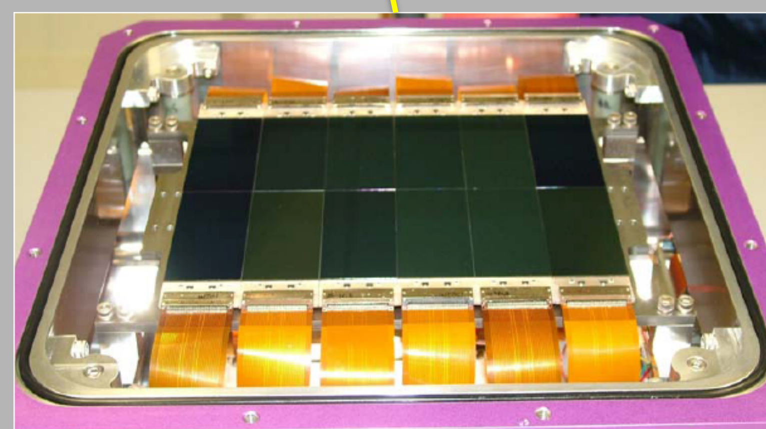
Scalable?

Today: Palomar Transient Factory

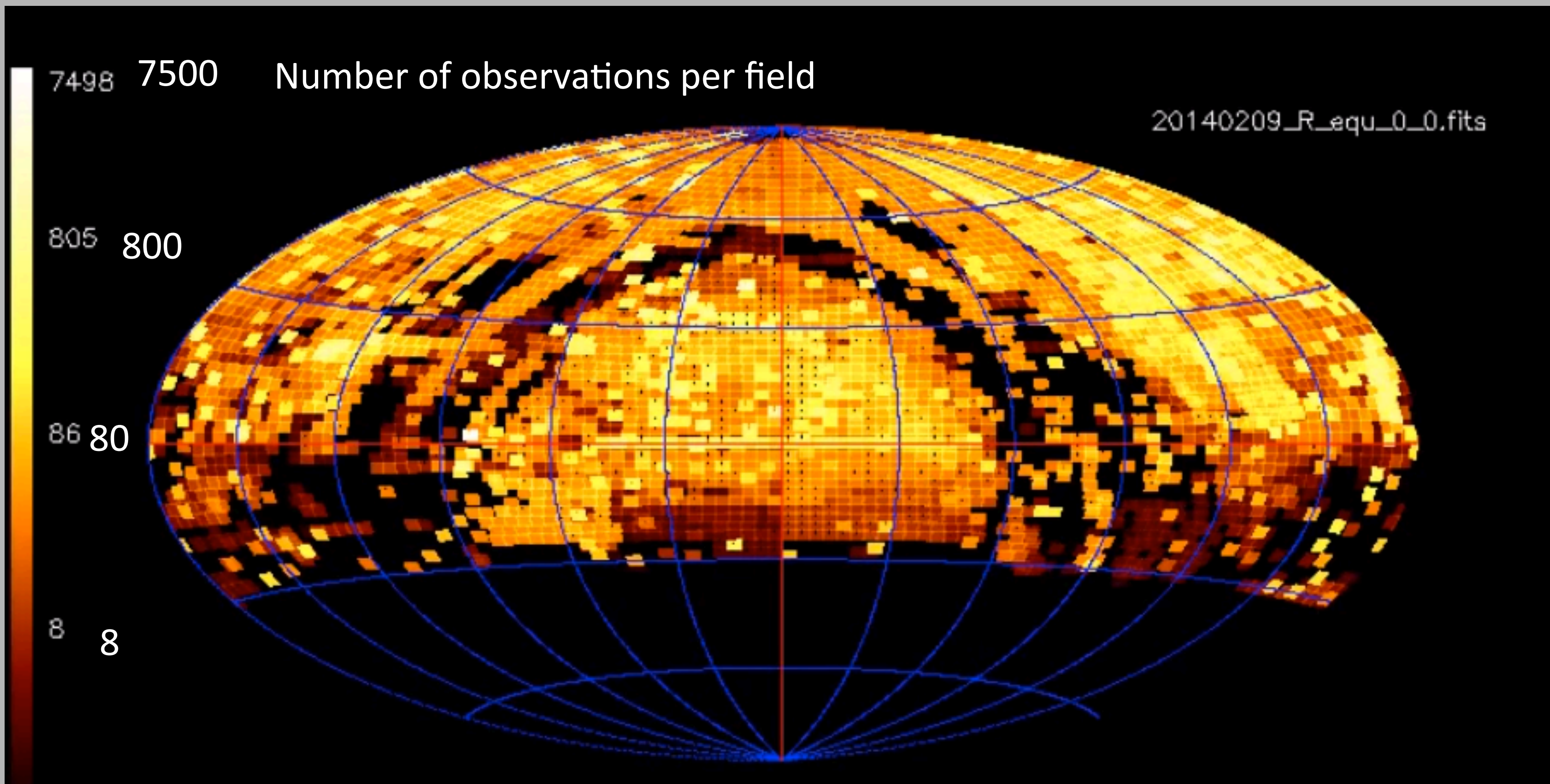
A 65 year old telescope with a modern CCD array & digital processing

- 12k x 8k CCD array ~100 Mpixel
- 1 minute exposures
- 7.25 sq degree field of view
- ½ billion objects in Northern Sky observed 10-10,00 times



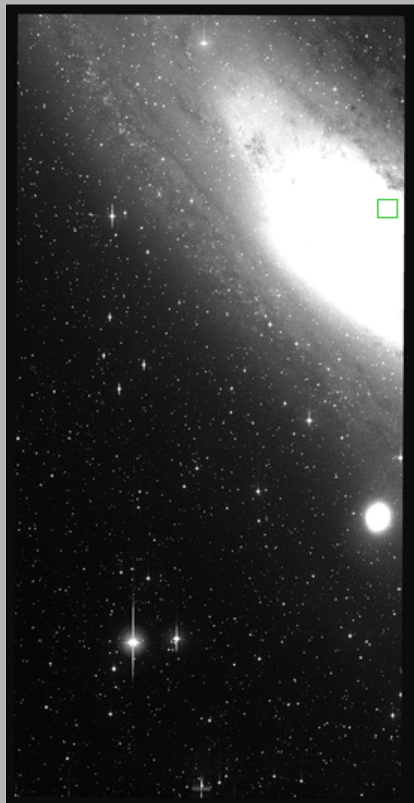


PTF Coverage of Northern Sky



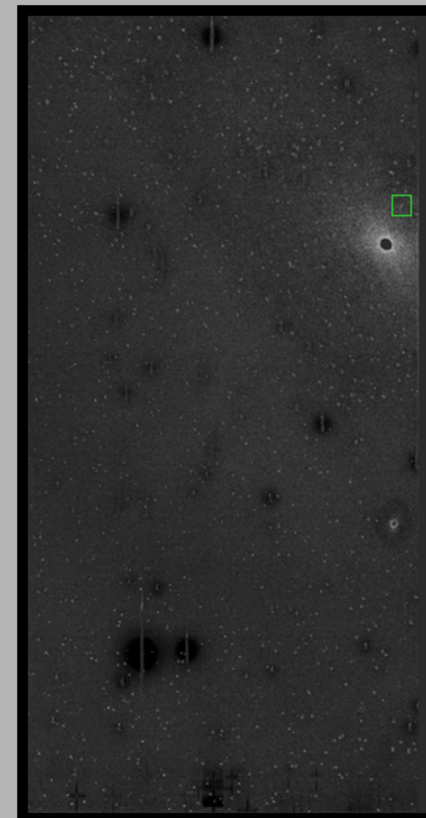
~1/2 Petabyte of Image Data

Reference



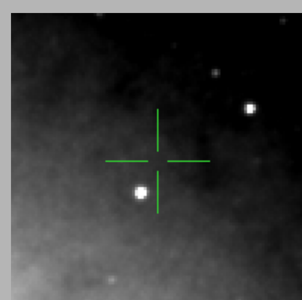
New Image

Subtraction

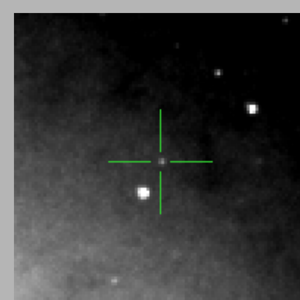


1 of 12
CCD chips

Example:
Nova in M31
(Andromeda)



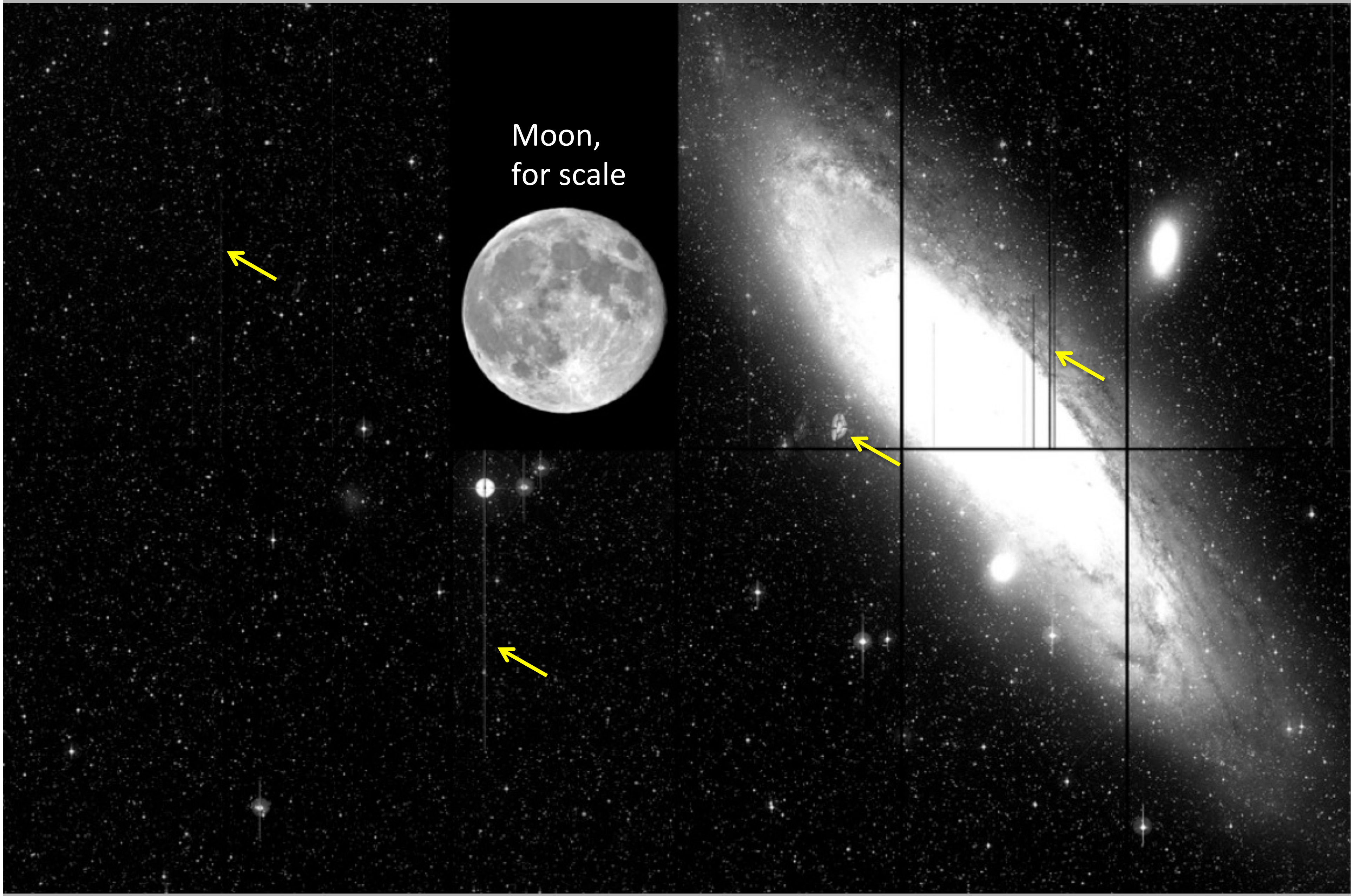
Reference



New Image

(see talk by Quimby)

Real?
or Bogus?



Palomar Transient Factory (PTF) CCD image of Andromeda Galaxy (M31) field

Time Domain Astronomy

The data-centric problem statement:

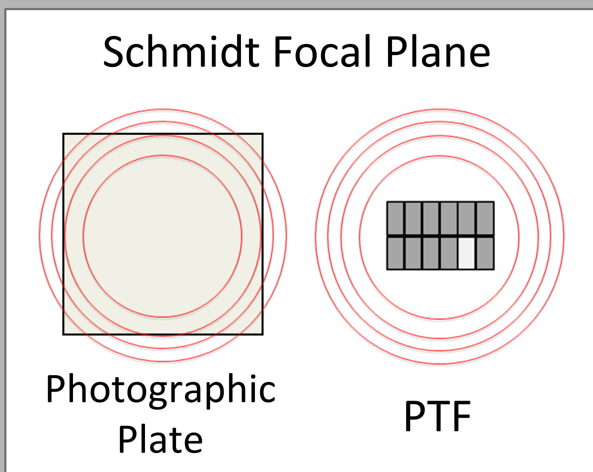
– Databases: Given ~1 petabyte of image data, extract sources and create a database of ~1 billion objects.

– Machine learning: In near real time, analyze ~Gbytes/minute of image data to identify which of millions of "candidates per night" are true astronomical transients. Similarly, identify fast-moving near earth asteroids.

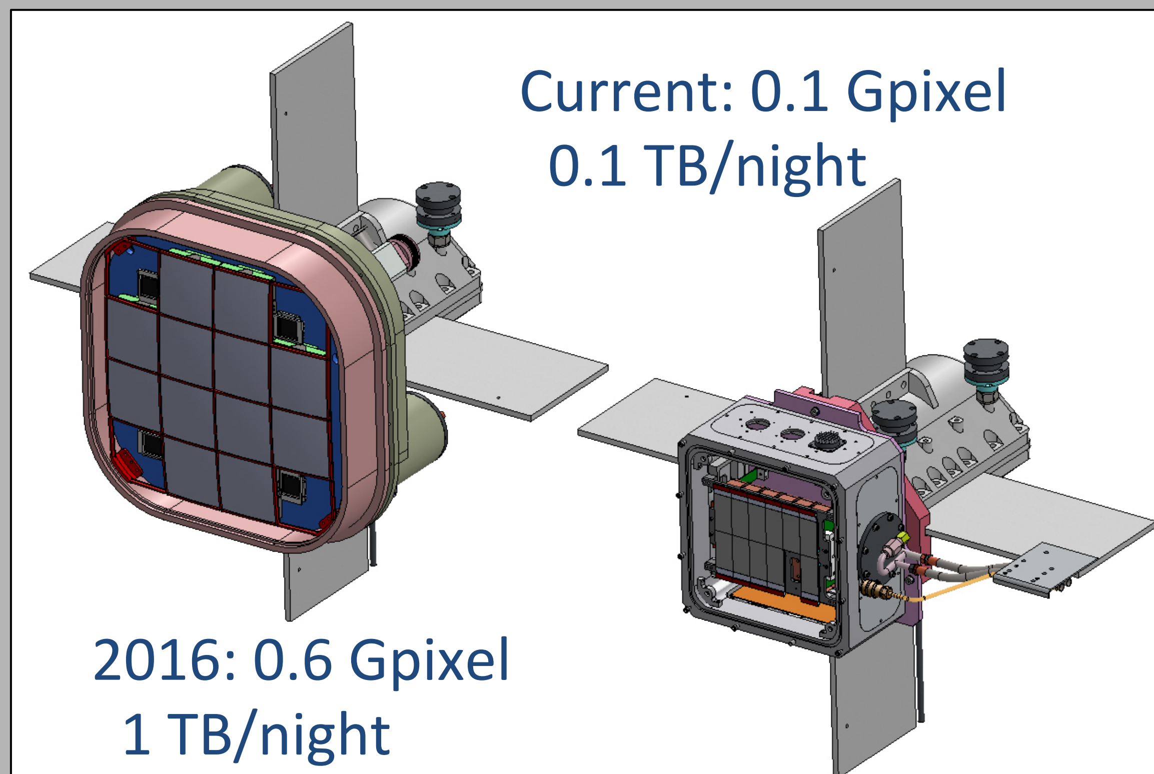
– Time-frequency algorithms: Using efficient queries on the database, determine which of the ~10's of millions of variable sources are "interesting sources"

Requires highly automated data exploration
Reduce false positives to manageable level

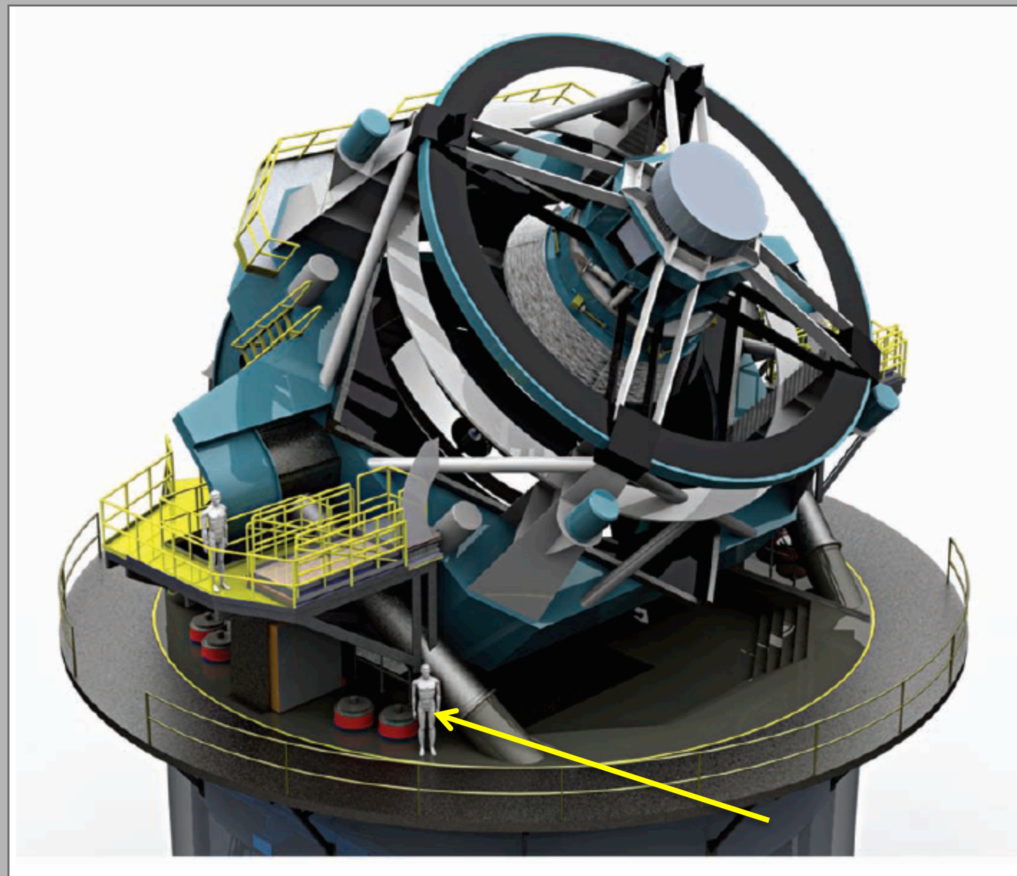
The Next Big Step (2016): *Zwicky* Transient Facility (ZTF)



ZTF: *10 x's the throughput of PTF!*



Into the Future: LSST

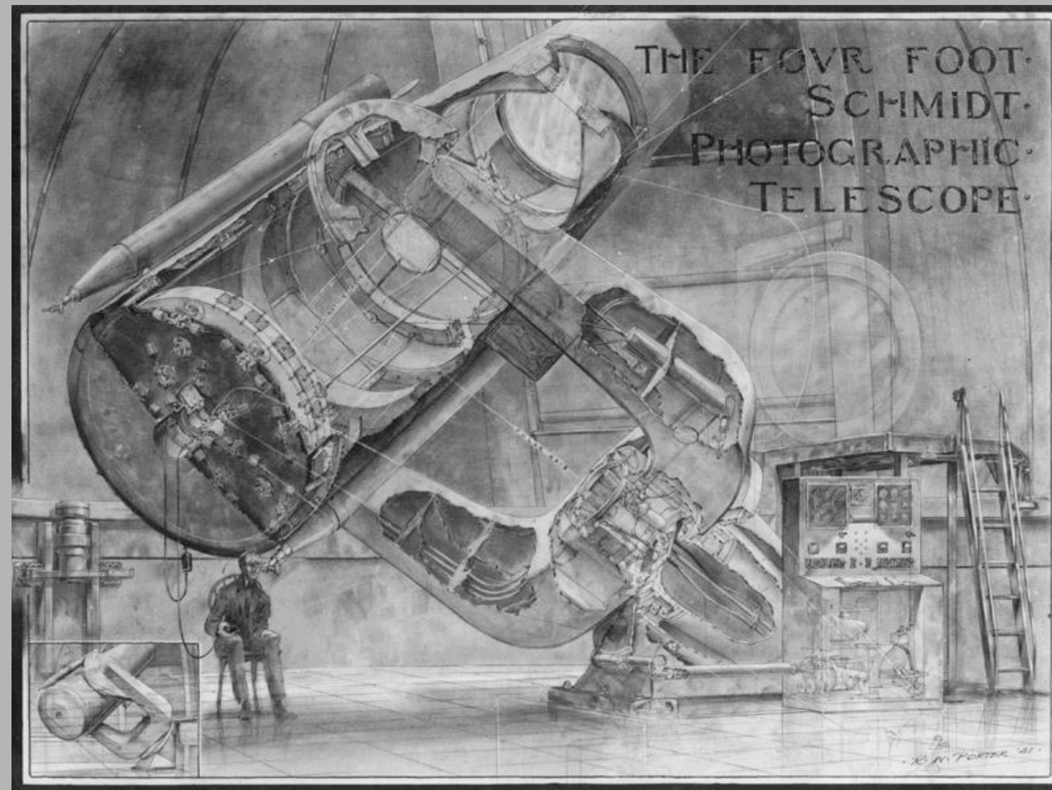


LSST: 10 x's the throughput of ZTF. 100 x's PTF

~2021

Large Synoptic Survey Telescope

8.4 meter, 3.2 Gigapixel, 30 Tbyte/night



Part II:

TIME-DOMAIN *ASTRONOMY*
-- SOME REMARKS

Astronomy "101"
(Optical/Near-Infrared)

Photometry = measurement of brightness (flux)

Astrometry = measurement of position

Transient: object that suddenly appears, or large change in flux

Variable: object usually detected, but flux varies

Magnitudes (logarithmic scale for flux)

- m~20.5: PTF limit (single observation)
- m~24.5: LSST limit
- Δm=2.5 => can see 10x's fainter; Δm=5 => 100x's fainter

Angular units

- 1 arcminute = (1/60) degree.
- 1 arcsecond = (1/3600) degree. Typical limit due to "seeing".

Non-ideal conditions!

- Weather (clouds)
- Moon
- Satellites, airplanes, …

Many existing optical time-domain surveys

(+ SDSS Stripe 82,ASAS,...)
(+ many more to come)

Not just PTF!

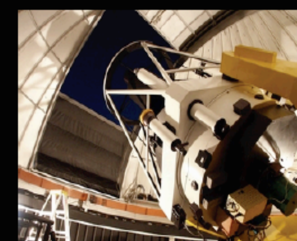
SkyMapper

PanStarrs

La Silla Schmidt

Palomar Schmidt

OGLE IV



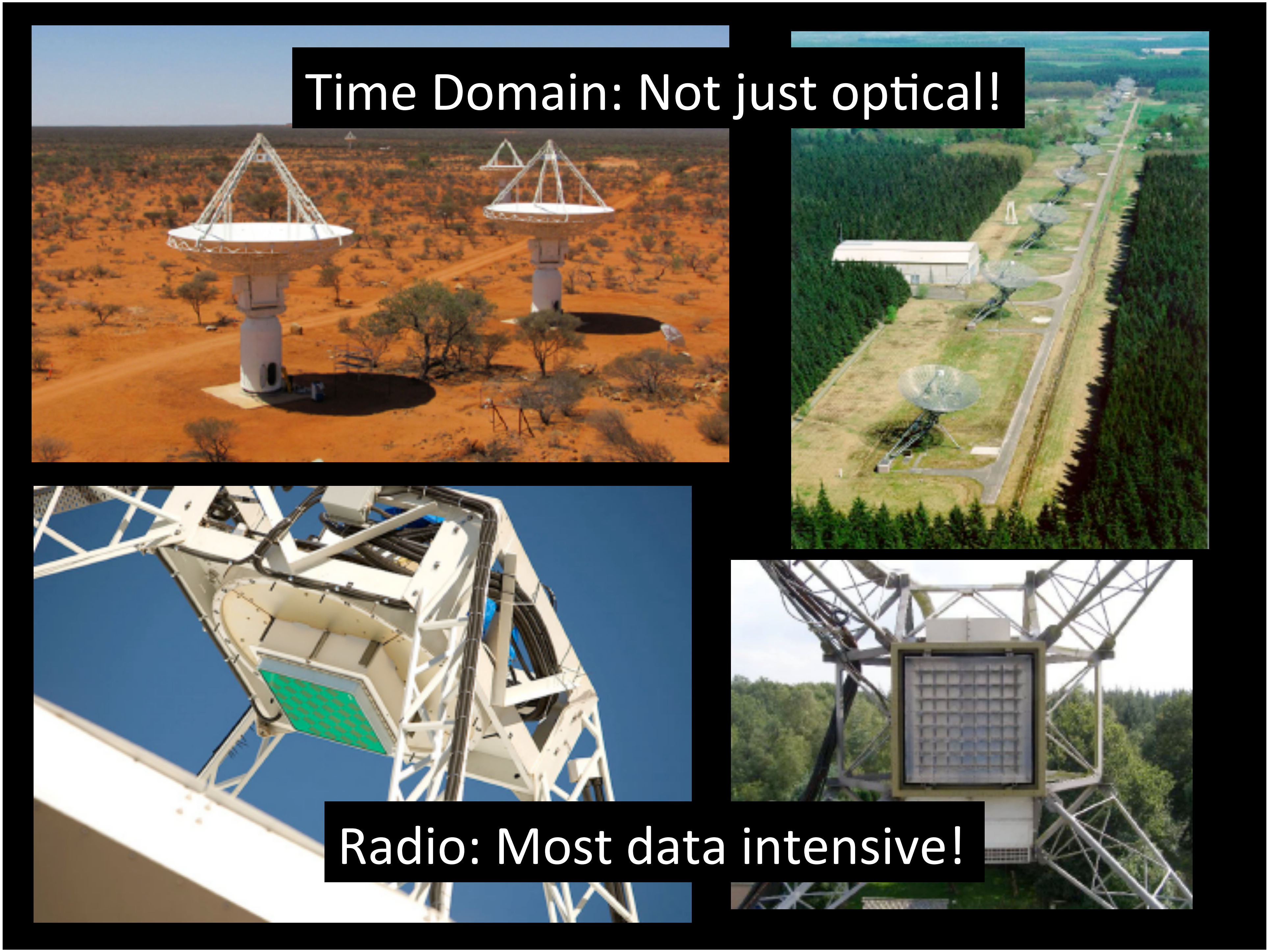
Uppsala Schmidt

Catalina Schmidt

VST

CTIO

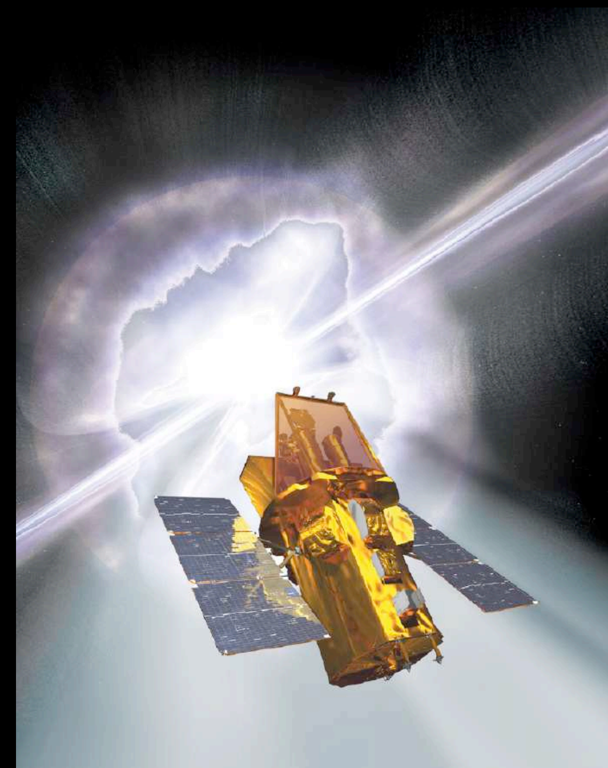
(from B. Schmidt talk)


Time Domain: Not just optical!
Radio: Most data intensive!

Time Domain: Not just long wavelength

MAXI Mission on ISS

(see talk by Kawai)

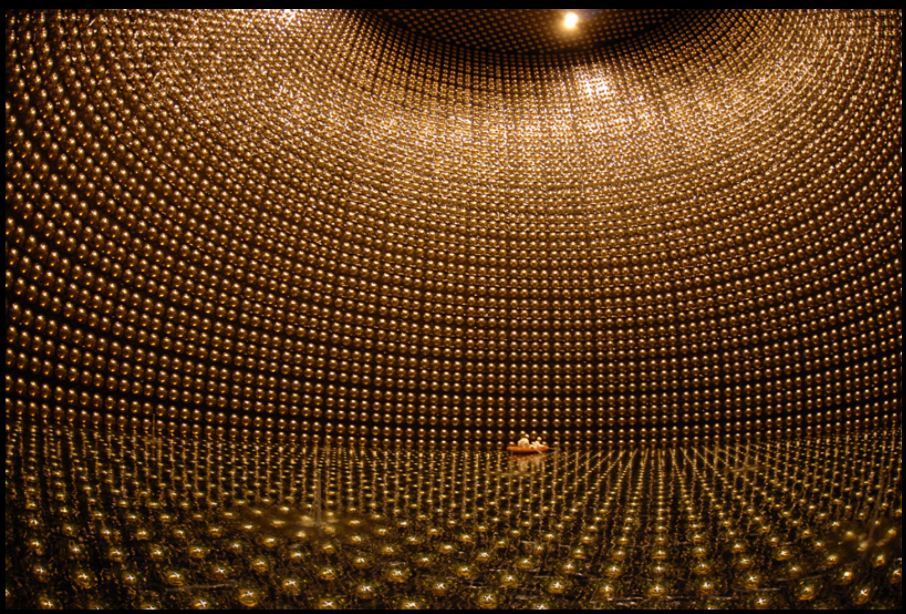


Swift Gamma-ray Burst Mission

X-rays and Gamma-rays

Time Domain: Not just photons!

Neutrinos



Super-Kamiokande

Gravitational Waves

Kagra Project: Large-scale
Cryogenic Gravitational-wave Telescope

Many existing optical time-domain surveys

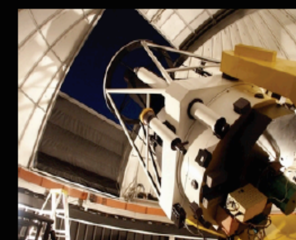
(+ SDSS Stripe 82,ASAS,…)
(+ many more to come)

SkyMapper

PanStarrs

MOA telescope

OGLE IV



La Silla Schmidt

Palomar Schmidt

Catalina Schmidt

Uppsala Schmidt

VST

CTIO

(from B. Schmidt talk)

Comparison of Surveys

Survey Camera	D (m)	$\Omega_{\rm fov}$ (deg^2)	Etendue (m^2 deg^2)	Pixels ($''$)	$t_{\rm exp}$ (sec)	$t_{\rm OH}$ (sec)	$m_{\rm lim}$ (5σ)	$\dot{V}_{-19}$ (Mpc^3/s)	$f_{\rm spec}$	FoM	$N_{\rm obs}$ (yr^{-1})
SNLS	3.6	1.0	10.2	0.2	300	40	22.8	1.3×10^3	0.1	0.06	0.8
DECam	4.0	3.0	37.7	0.3	50	17	23.7	4.7×10^4	0.06	0.9	12
HSC	8.2	1.7	89.8	0.2	60	20	24.6	5.1×10^4	0.02	0.4	6
CRTS[41]	0.7	8.0	3.1	2.5	30	18	19.5	1.5×10^3	1.0	0.6	46
PTF	1.2	7.3	8.2	1.0	60	46	20.7	2.8×10^3	1.0	≡1.0	10
PS1 3π[44]	1.8	7.0	17.8	0.3	30	10	21.8	2.5×10^4	0.4	3.7	48
CRTS-2*	0.7	19.0	7.3	1.5	30	12	19.5	1.5×10^3	1.0	1.5	125
BlackGEM*[45]	0.6(4×)	2(4×)	11.3	0.6	30	5	20.7	9.3×10^3	1.0	3.3	63
ATLAS*[46,47]	0.5(2×)	30/60	11.8	1.9	30	5	19.8	2.3×10^4	1.0	8.4	473
LSST*	8.4	9.6	319.0	0.2	30	11	24.7	5.1×10^5	0.02	4.1	54
ZTF	**1.2**	**47**	**53.1**	**1.0**	**30**	**15**	**20.4**	$\mathbf{3.0 \times 10^4}$	**1.0**	**10.6**	**288**

Diameter

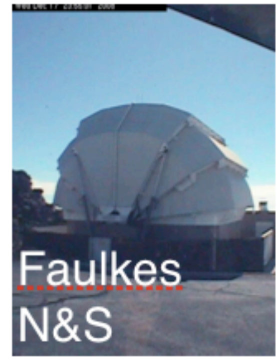
Field of View

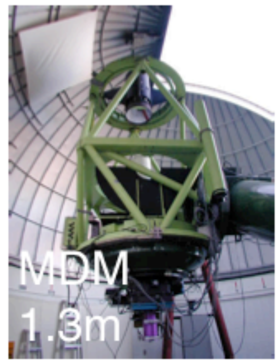
Limiting Magnitude

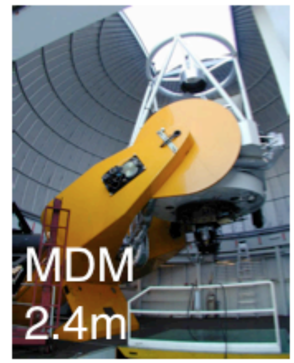
Figure of Merit

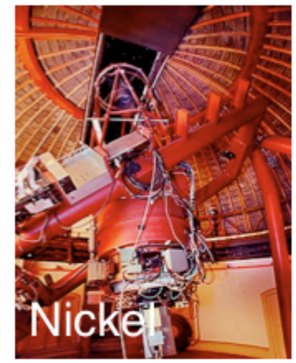
Survey evaluation requires Figure of Merit (FOM) specialized for specific science goal

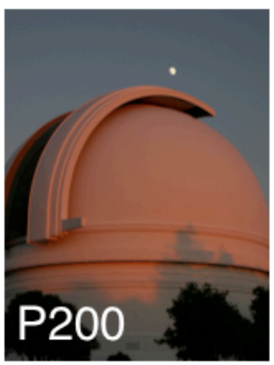
PTF follow-up telescopes

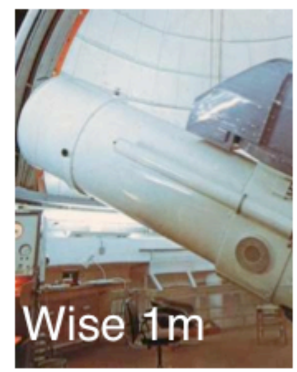












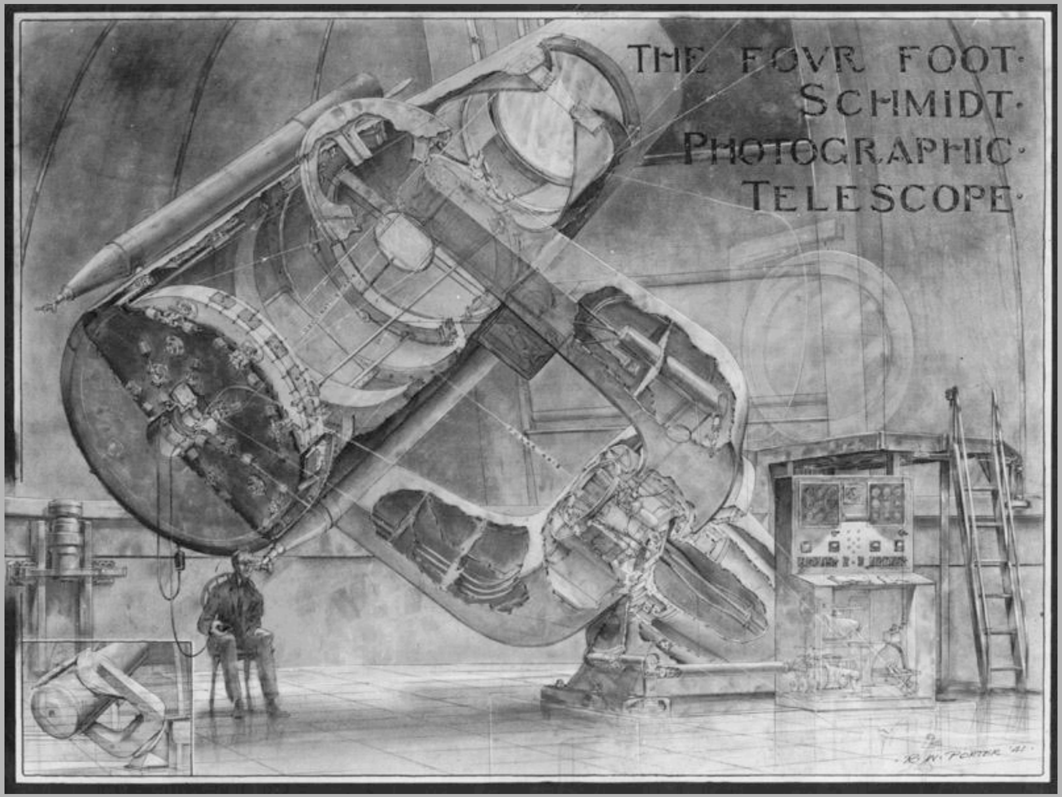
Time Domain Astronomy: Not just survey telescopes

Science of Time-Domain Surveys

- Supernovae
- Gamma-ray bursts
- Gravitational wave counterparts
- Microlensing
- Compact binaries
- Black hole novae
- Flare stars
- Stellar mergers
- Active galactic nuclei

…

PTF Key Projects	
Transients in nearby galaxies	Search for eLIGO/neutrino EM counterpart
Thermonuclear SNe	Core Collapse SNe
Blazars/AGN	Tidal Disruption Flares
H-alpha Sky Survey	Orphan GRB afterglow
AM CVn	CVs
Galactic dynamics	RR Lyrae
Flare stars	Rotation in clusters
Nearby Star Kinematics	Eclipsing stars and planets
Asteroids	KBOs

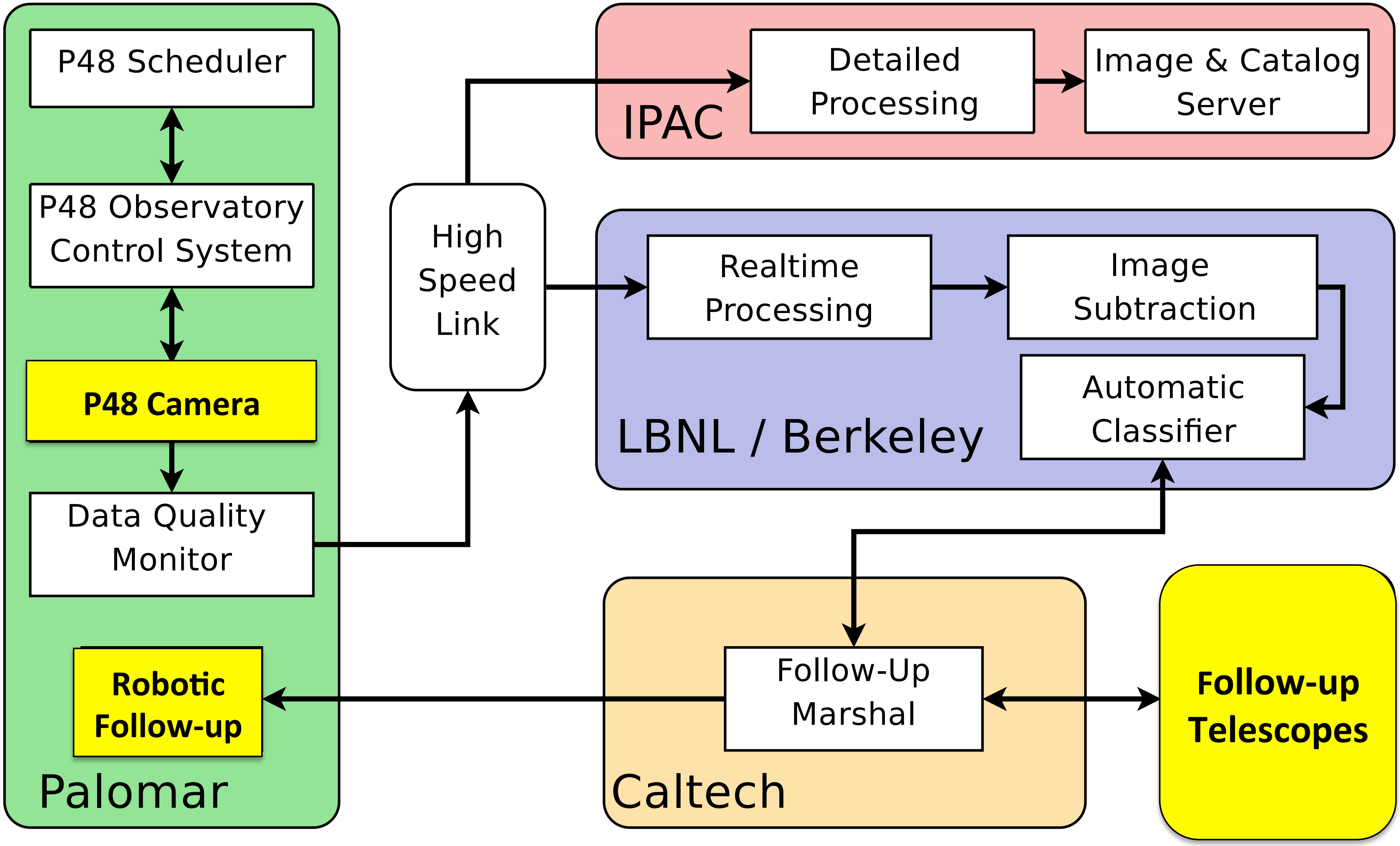


Part III:

DATA AND COMPUTATION IN ASTRONOMY

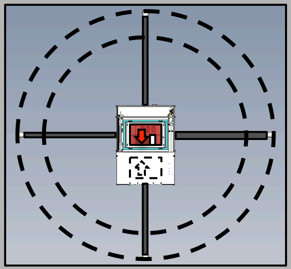
Three Issues: Machine-learning, Databases, Algorithms

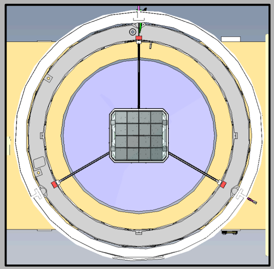
PTF/ZTF Data Flow

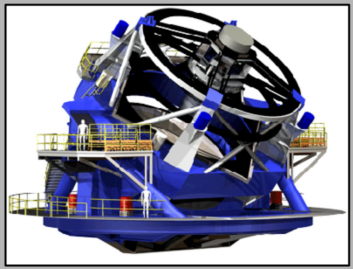


(Big) Data Challenge

	PTF (Now)	→ ZTF 2016	→ LSST 2021
Alerts per night	4×10^4	3×10^5	2×10^6
Image data rate	1 GB per 90s	3 GB per 45s	6 GB per 5s
Image archive	0.05 PB/yr	0.5 PB/yr	6 PB/yr



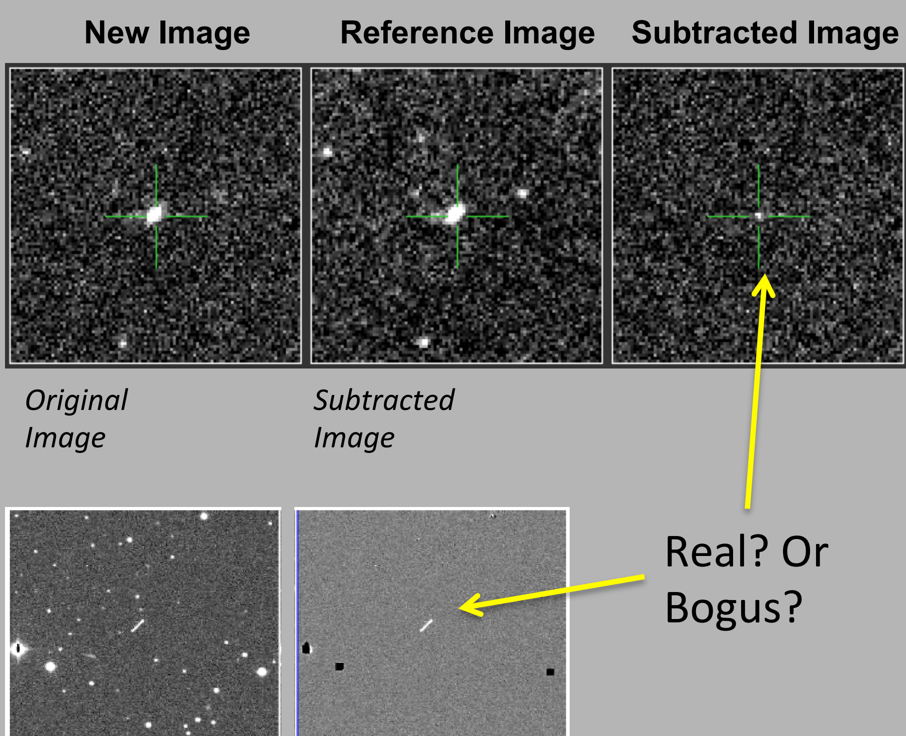


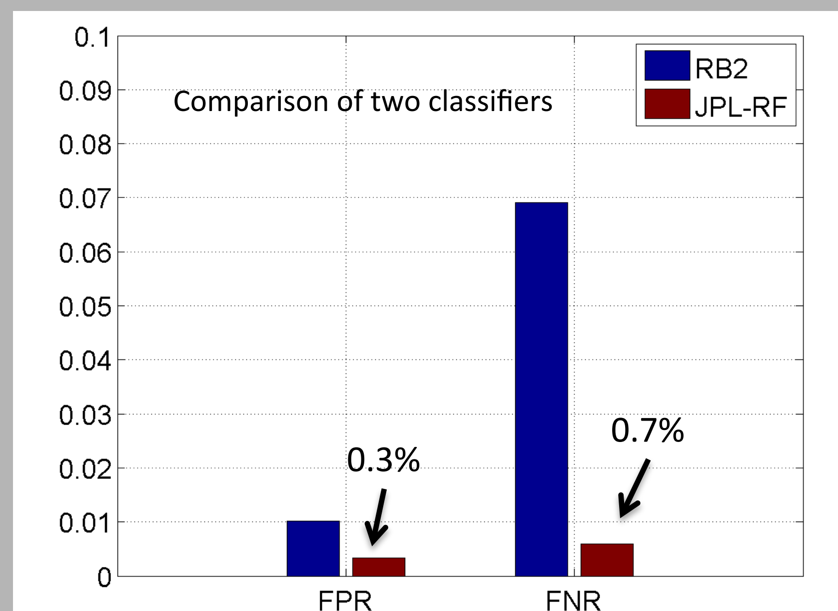


About x100 increase from PTF -> LSST

Machine Learning

- Data/Events Rates are too high for human analysis (~million)
- Machine-learning required to reduce event rate to manageable level for human vetting and telescope follow-up (~10's per night)
- Low false positive and low false negative ("Real-Bogus")



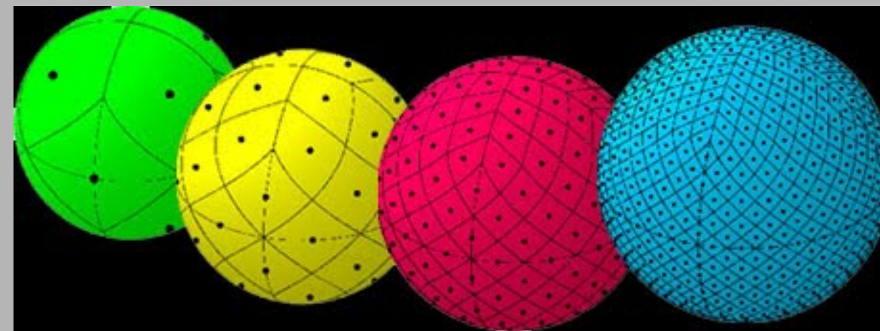


False Positive Rate False Negative Rate

(see talk by Wozniak)

Database Issues

Significant detailed technical issues of parallelization, efficiency, scalability, …



Healpix: Hierarchical indexing

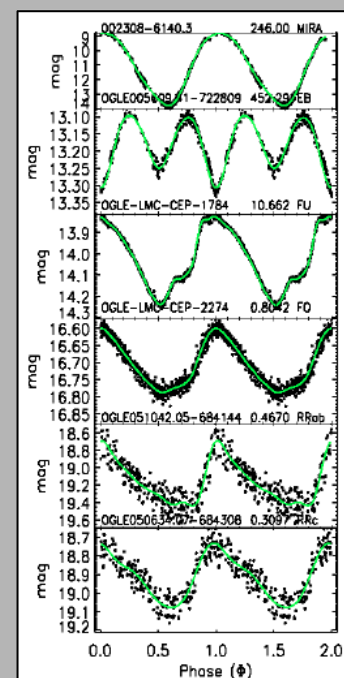
Sky Position Database Schema:

- Index by sky location => Healpix
- "What are the temporal characteristics of this list of objects with known locations"
- Alternative: Index by filter/field/chip

Time Domain Database Schema:

- Index by time-domain characteristic
- "Give me all objects with the following temporal characteristics"

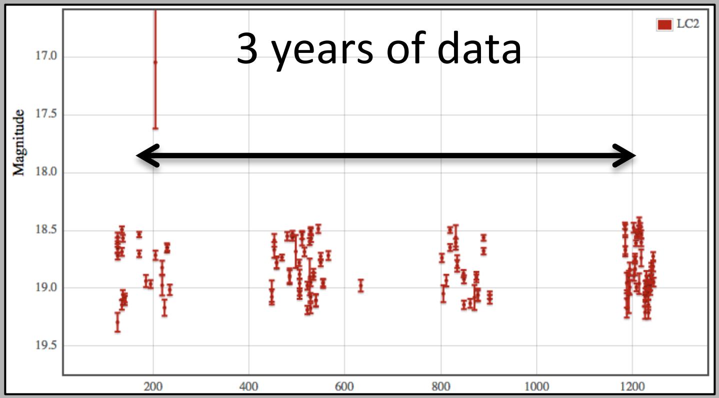
Light Curve Shape Characteristics



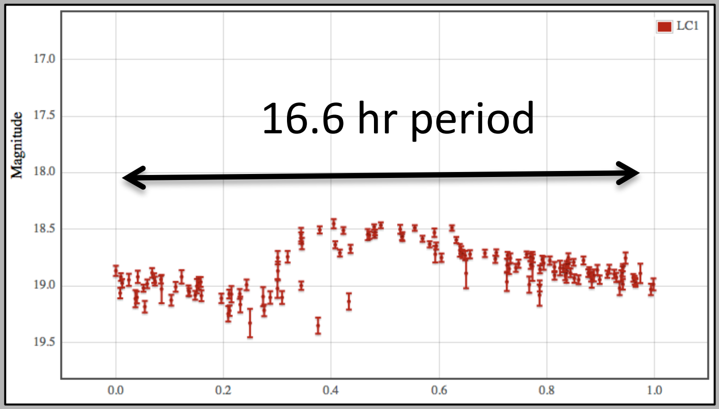
Time-Domain Algorithms

- Unevenly sampled data
  - Issues
    - Effects of strong sampling window functions?
    - Significance of detection of periodicity?
    - Best estimate of power spectrum for unevenly sampled data?
  - Still relatively immature algorithmically
  - Current techniques
    - Lomb-Scargle
    - Analysis of Variance (AoV)
    - Conditional Entropy

What are optimal, efficient algorithms for going from this:

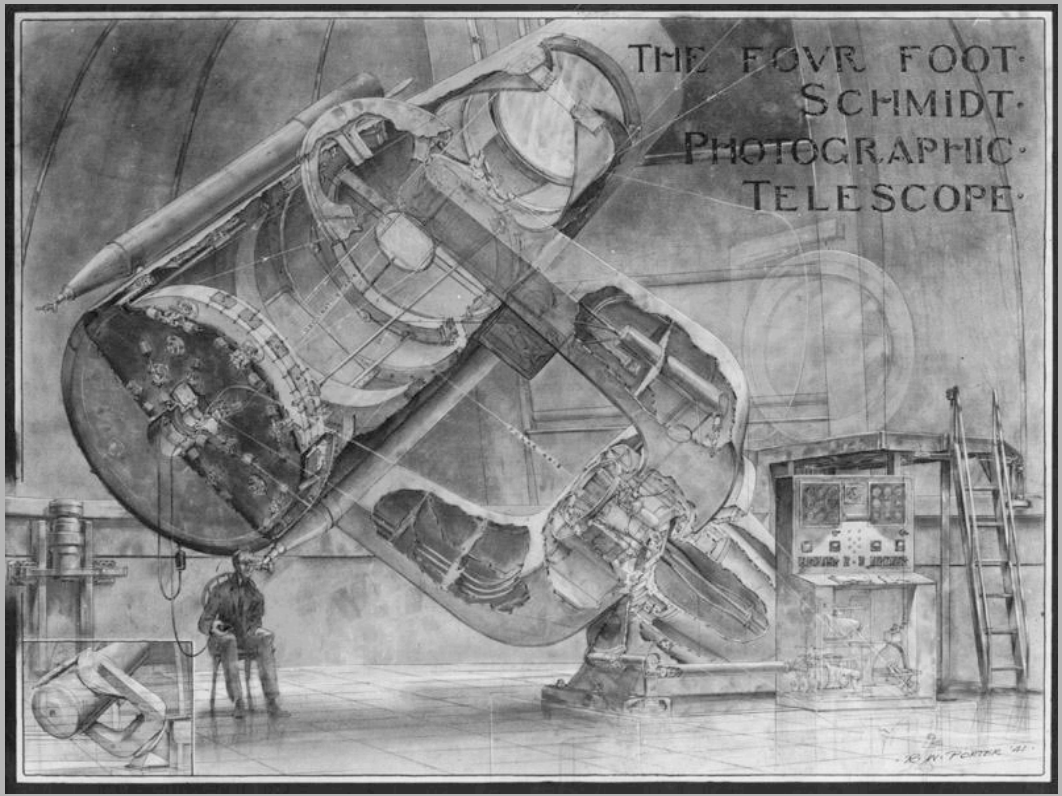


to this??



Virtual Observatories: Part of the astro-informatics landscape

- Standards for astronomical metadata and tools to allow efficient analysis of data across varied astronomical resources
- Formed in 2000-2003 timeframe
- International Virtual Observatory Alliance (IVOA)
  - 2 dozen international members
  - Japanese Virtual Observatory (JVO)
  - Virtual Astronomical Observatory (VAO-US)



CONCLUDING REMARKS

Summary

- Optical time-domain astronomy is on the verge of major growth in data rate & volume: x10 by 2016 and x100 by 2021.
- Highly efficient machine learning techniques will be required to eliminate false positives and to classify source types
- Major growth in number of detected sources will require efficient, well-designed databases
- Time-domain analysis algorithms will require significant further development

Astronomy Talks at this Workshop

- ***Implementing the Palomar Transient Factory Real-Time Detection Pipeline in GLADE: Results and Observations***
  Florin Rusu (classification, real-bogus)
- ***Astrophysical Image Modeling***
  Robert M. Quimby (image subtraction)
- ***Open Data from the Monitor of All-Sky X-ray***
  Dr. Nobuyuki Kawaii
- ***Big Data in Astronomy***
  S. R. Kulkarni (PTF and big data)
- ***Towards an Intelligent Astronomical Event Broker: Automated Transient Classification and Follow-up Optimization*** -
  Przemek Wozniak (Classification, machine learning)
- ***Machine-Learning Enabled Stellar Classification and the Prediction of Fundamental Atmospheric Parameters*** -
  Adam A. Miller (Classification, machine learning)
- ***Exploratory Analysis of Light Curves: A Case-Study in Astronomy Data Understanding*** -
  Dhriti Khanna and Vasudha Bhatnagar (time-domain algorithms)

Some References

"Automated discovery and classification of transients and variable Stars in the synoptic survey era", Bloom et al., PASP, **124**, 1175 (2012)

"Machine –assisted discovery of relationships in Astronomy", Graham et al., MNRAS, 431, 2371 (2013).

"Data mining and machine learning in astronomy", Ball & Brunner, arXiv:0906.2173v2 [astro-ph.IM].