# Cosmic Variability –

## Mass Data Management and Information Technology Challenges of Astronomical Databases

Wen-Ping Chen（陳文屏）

Institute of Astronomy
National Central University, Taiwan

2006 October 24 @ CODATA Beijing

中央大學

## Challenges of Astron. Observations

- **Sensitivity** --- farther, fainter, older
- **Resolution** --- clearer (angular), finer (spectral), …

## Next Frontier in Astrophysics

- Celestial objects vary in brightness (minor planets, stars, AGNs, gravitational lenses, etc).
- **Time domain** has not been much exploited.

## Lesson on Project Management

- Software cost should not be sneezed at, especially if the data are to be publicly available
- ***Rule of Thumb*** --- $1 hardware; $1 software; $3-10 informatics (i.e., databases)

The **BATC** (Beijing-Arizona-Taipei-Connecticut) project, a multi-wavelength sky survey which involves institutes in China, US and Taiwan, was initiated in early 1990s by a group of Chinese astronomers including Fang Lizhi, then exiled in the US.

The BATC Schmidt telescope in Beijing Obs.

It takes mutual trust to build up a collaboration!

The project, after the initial hardware, management, and communication challenges, has collected a total of 500-600 GB worth of imaging data, and now enters its scientific production peak (totaling > 100 SCI papers) after more than a decade of operations.

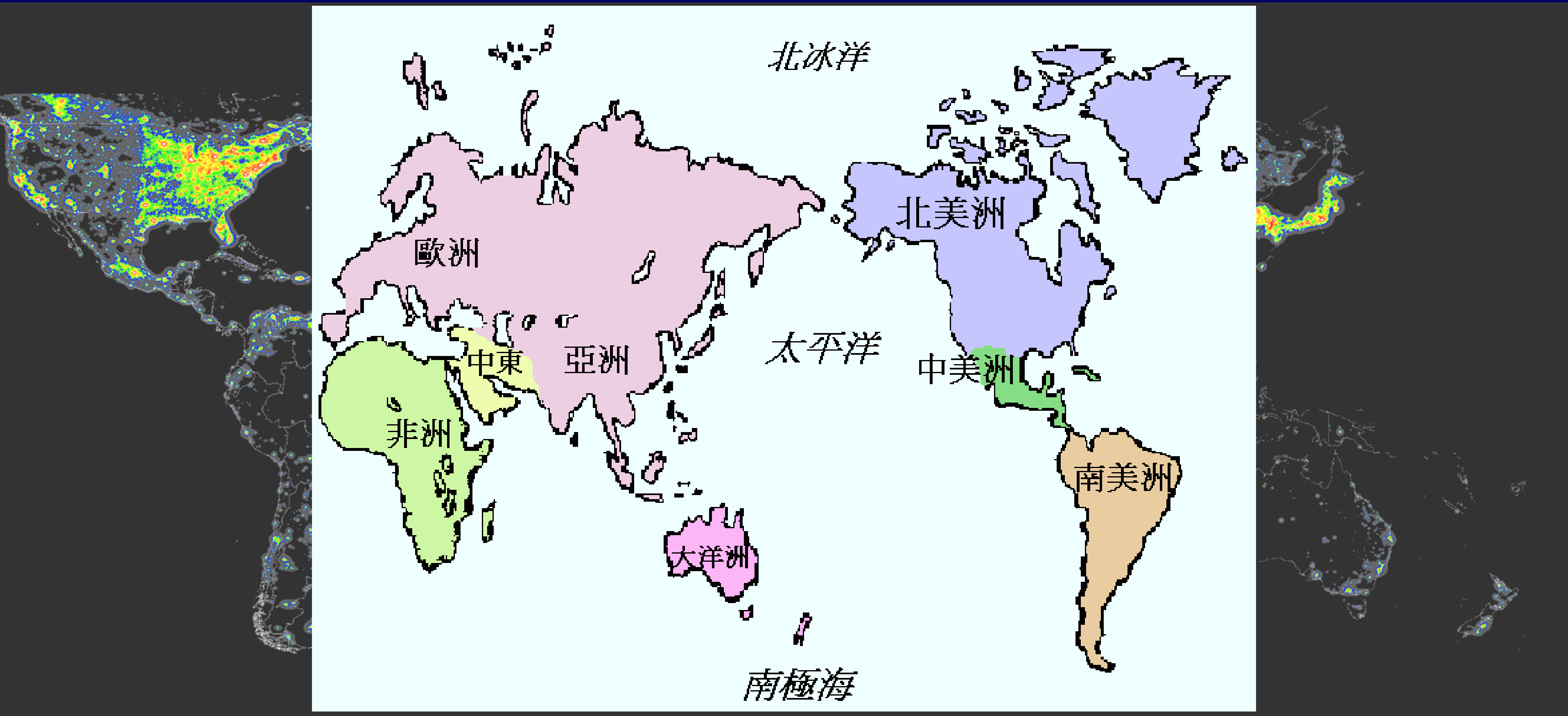
# Outline

Fully operational; in Taiwan

**TAOS**
Taiwan-America Occultation Survey

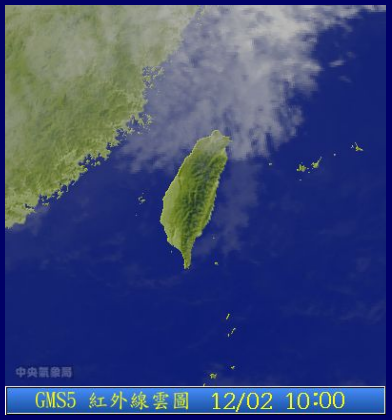
**Pan-STARRS**
Panoramic Survey
Telescope
and
Rapid Response System

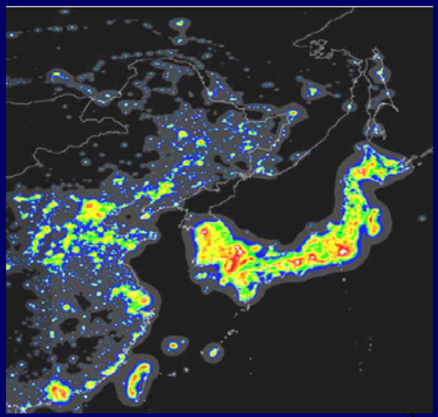
Being constructed;
in Hawaii, USA




Geographical Vantage:
- Many high mountains
- West Pacific
- Low latitude

➔ **time-domain astrophysics**



# LULIN OBSERVATORY
# 鹿林天文台

**Elevated to 2862m; above the inversion layer**

**… seen from Yusan (Jade Mt; 玉山) 4000- m**

LELIS


SLT

LOT


TAOS

# Lulin Observatory

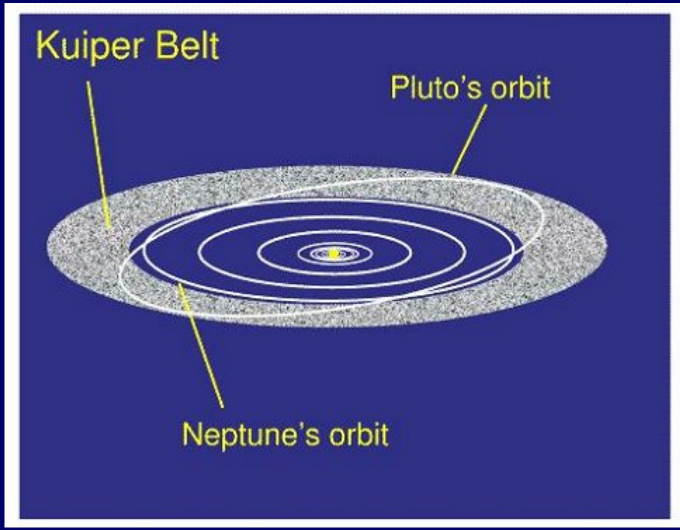
# Scientific Activities at Lulin

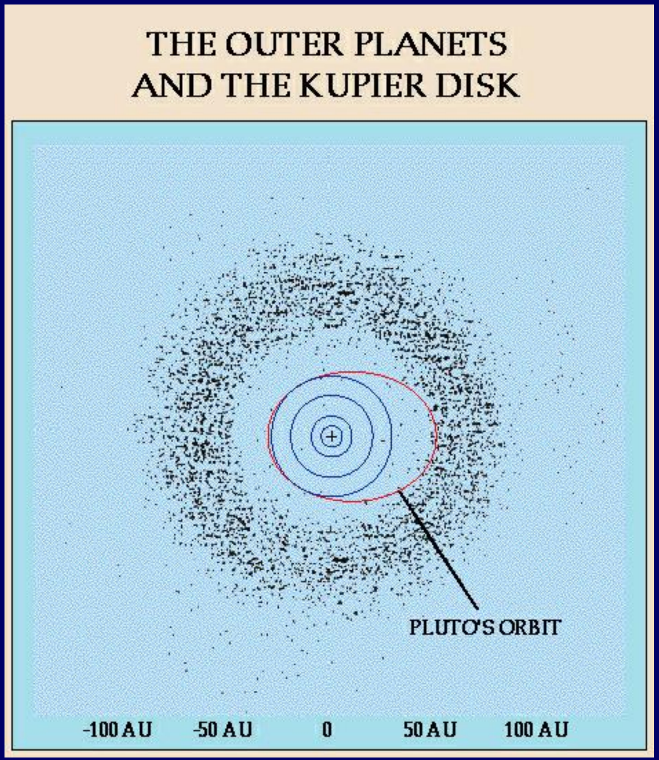
- **Time Variability**
  - Part of global network
  - Mass data processing
  - Data mining
- **TAOS** (Taiwan, USA, Korea)
- **LELIS** (NCU)
- **Taiwan Oscillation Network** (NTHU)
- **Sprite** (NCKU)
- **Taiwan Earth-Shine Network** (NTHU)
- **Atmospheric Experiments**

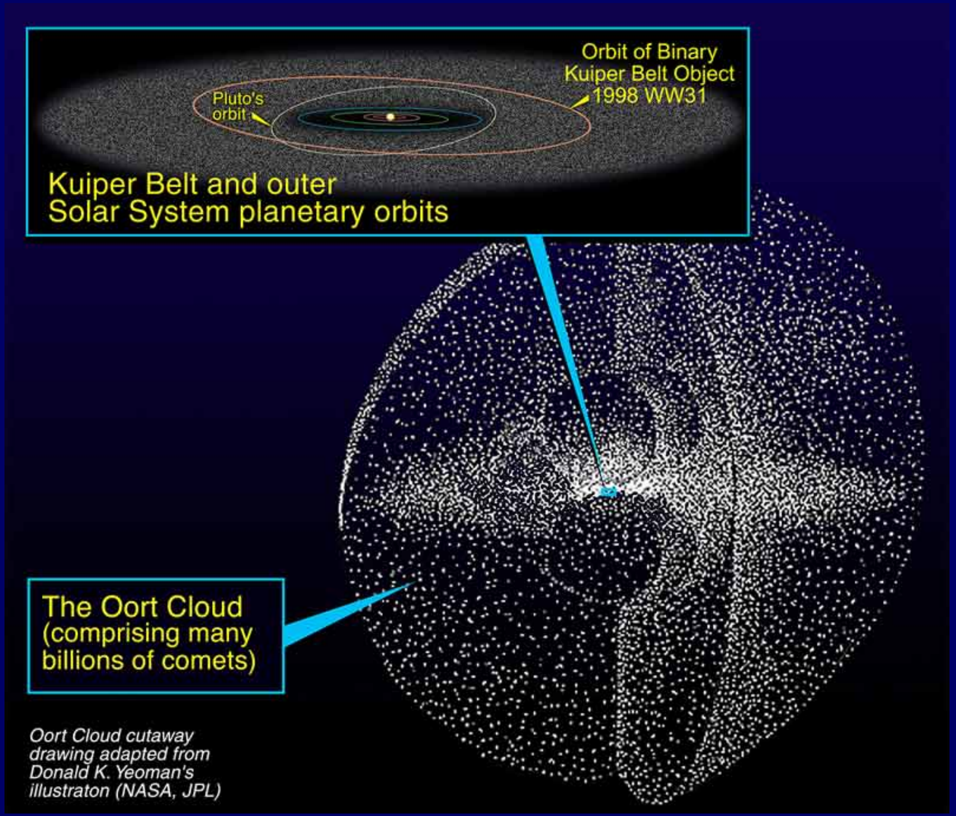
NCU/Lulin Observatory from Yusan North Peak

So far, discoveries of
10+ supernovae,
150+ asteroids
…

So far there are more than 1000 objects found beyond the orbit of Neptune, Pluto being one of the largest. These are seen by reflected sunlight, so only the largest can be detected by large telescopes

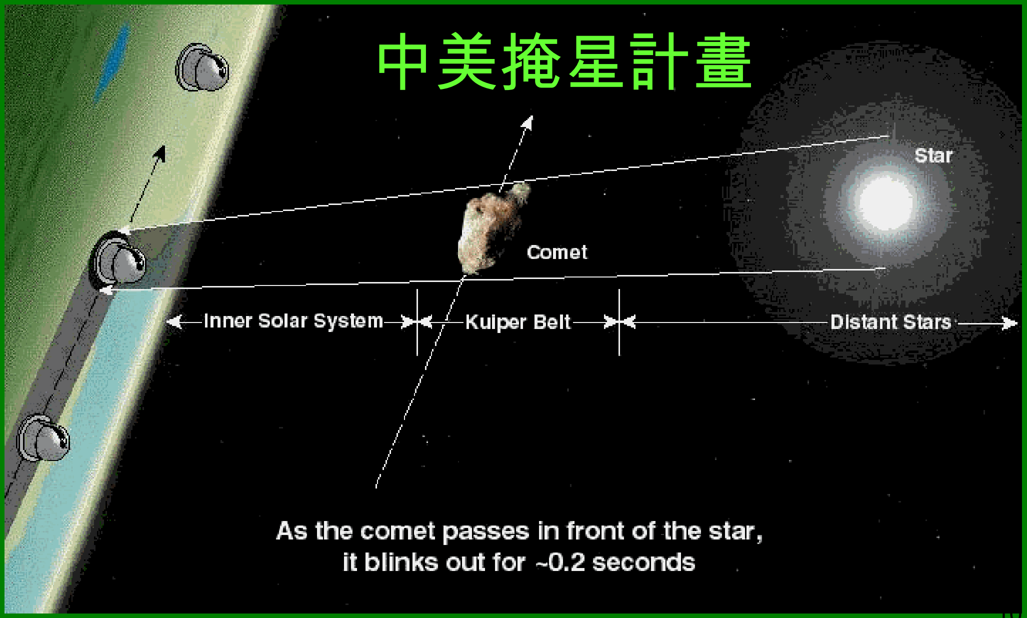


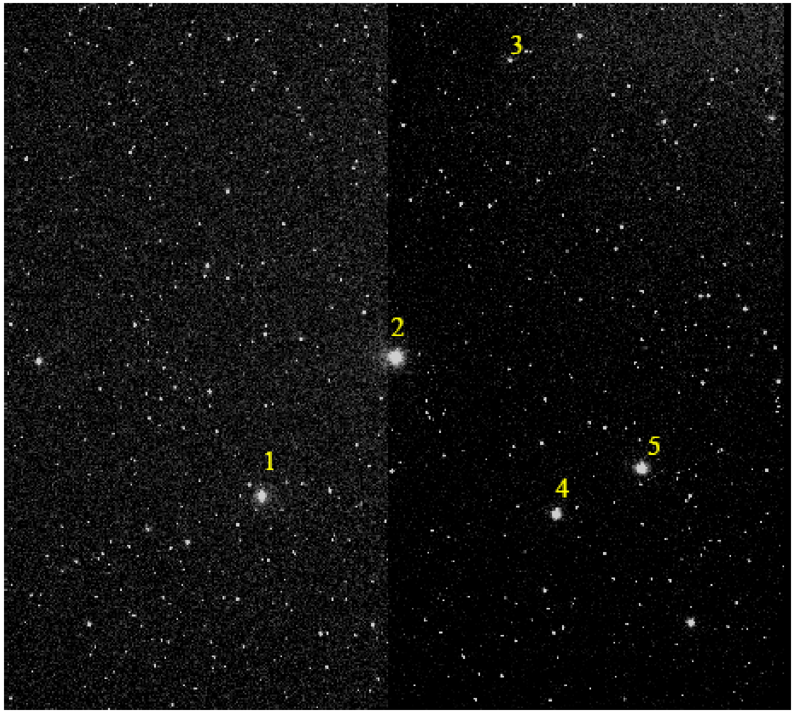
The **TAOS** (Taiwan-America Occultation Survey) project, a novel telescope array set up by groups from Taiwan, US and Korea, began routine observations in early 2005 and has the potential to give clue to the formation and evolution of our Solar System.

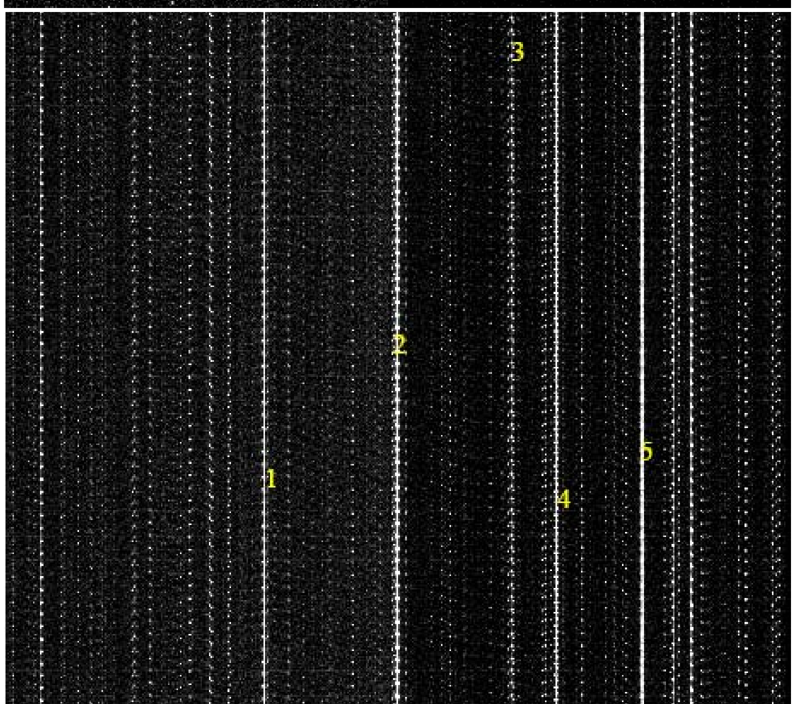
Comet nuclei too faint to be detected by direct imaging may be “seen” when they move in front of a background star --- a stellar occultation event.



# Project Overview

- A census of the small objects in the outer solar-system
- An array of wide-field telescopes (D=50 cm, f/1.9, FOV=3 sq. deg) to monitor brightness changes of ~1,000 stars at 5 Hz rate
- Looking for a 'blink' of starlight (occultation) when an object (> 2 km) moves in front of a distant star
  **Frequency of events ➔ population of "interveners"**
- Data rate a few 100 GB per night; only "interesting" data downloaded via the dedicated microwave E1 connection
- Real-time data analysis (light curves, rank statistics)
- Requiring coincidence detection of the same event by all telescopes to guard against false positives

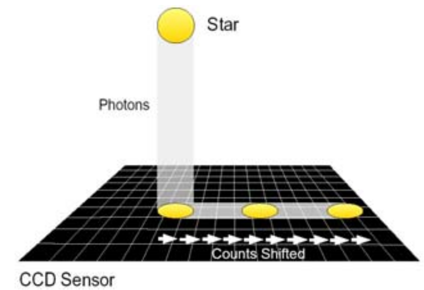



# Data Acquisition

## Typical CCD imaging

*Every star, together with surrounding skies, get exposure at the same time*




## TAOS data

*Integrate for 200 ms and then read out 32 rows of pixels, with the shutter remains open*

*The sequence continues, so each star appears as a series of dots 'zipper'*

➔ *'Fake' neighboring stars and skies!*

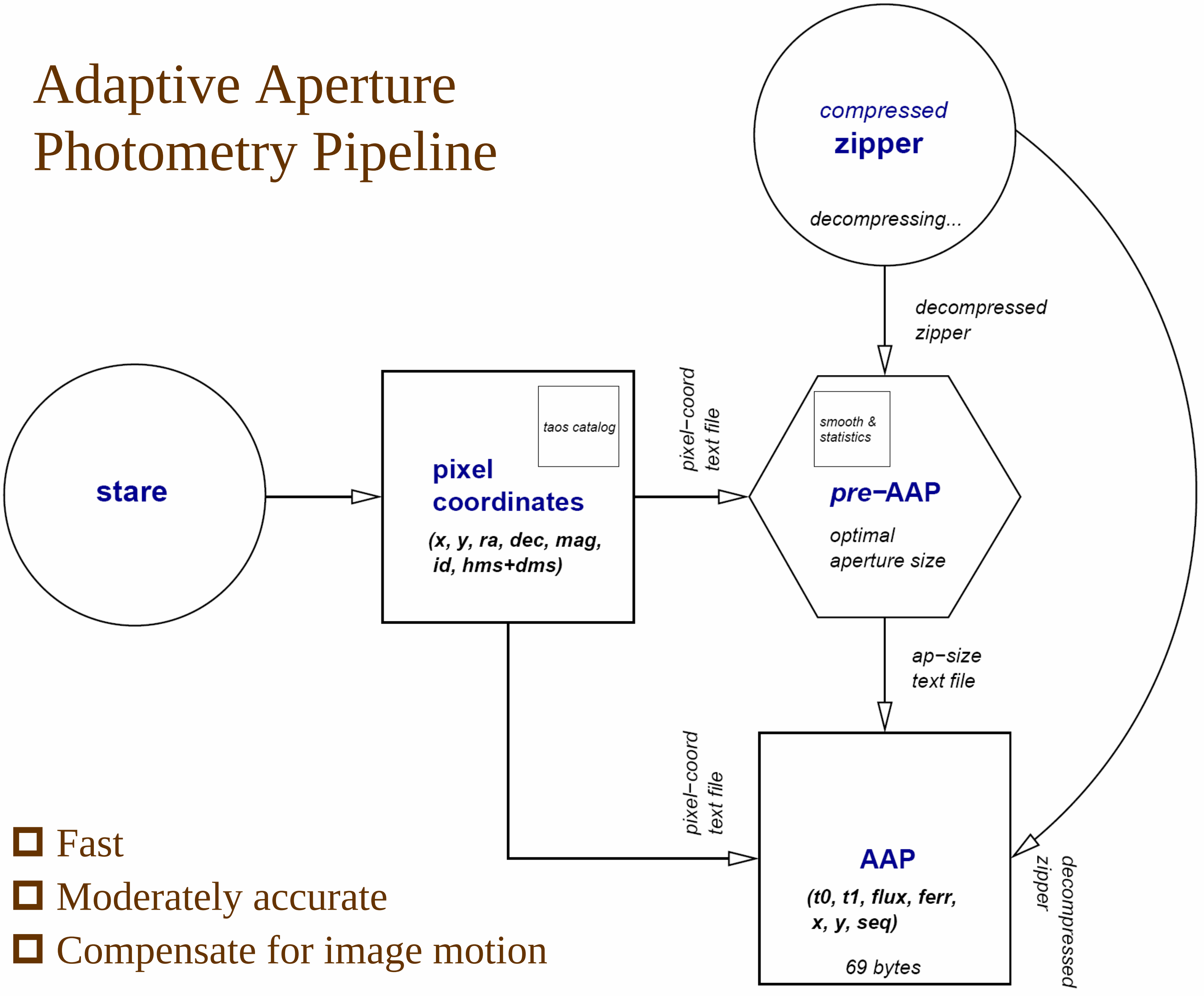
# TAOS Telescopes

Lulin Observatory
Central Taiwan
altitude=2862 m

TAOS is the only one of its kind in the world to conduct a census of small (1-2 km size) icy bodies at the outer reach of the solar system.

With a special data acquisition and a non-parametric statistical analysis scheme

# Adaptive Aperture Photometry Pipeline



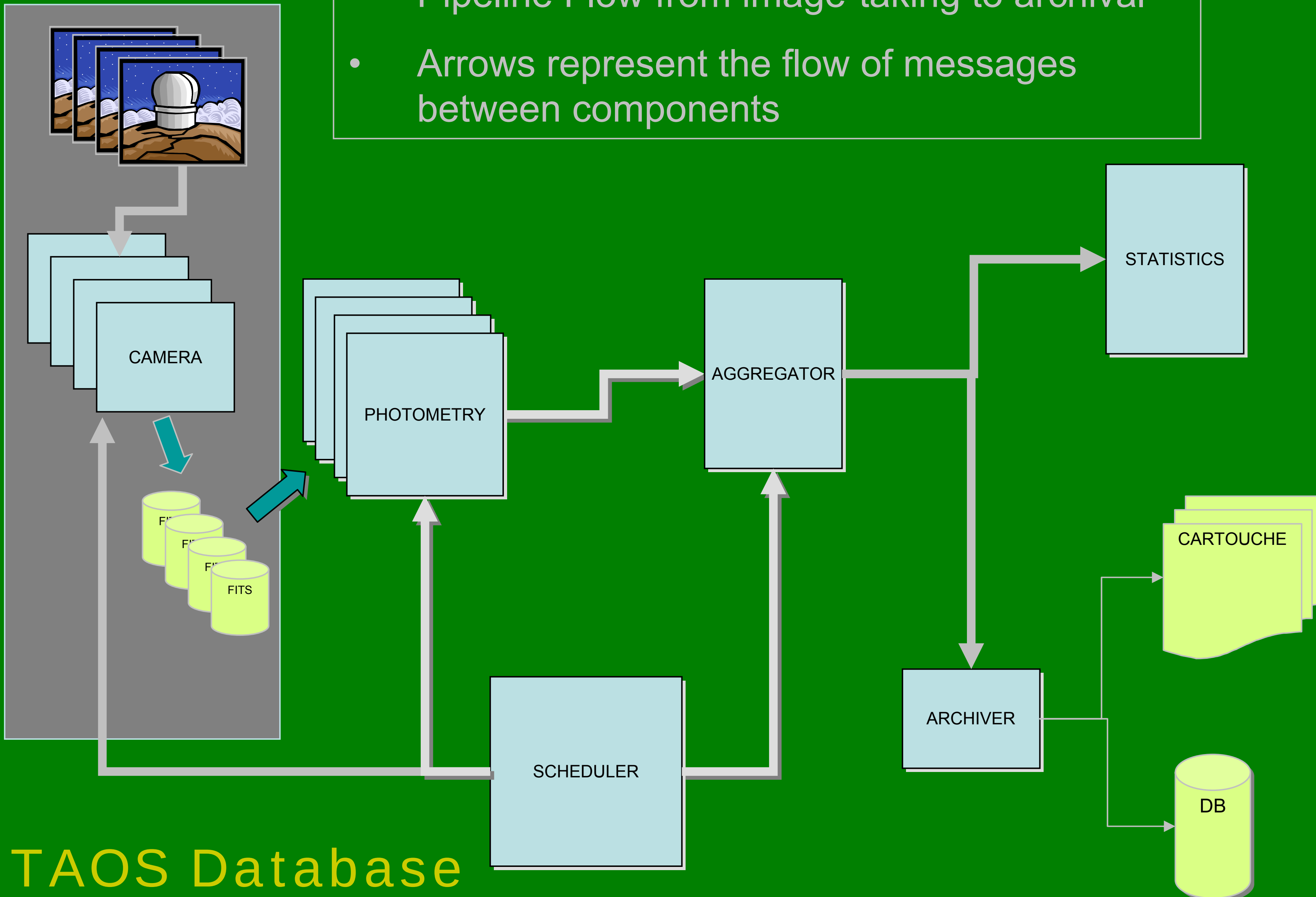

- Fast
- Moderately accurate
- Compensate for image motion

# Sample Output of the TAOS Photometric Light Curves

```
# Produced by KIWI on 2005-12-27+085846 (rectphot8)
# square aperture size (in diameter) : 4.8
#
```

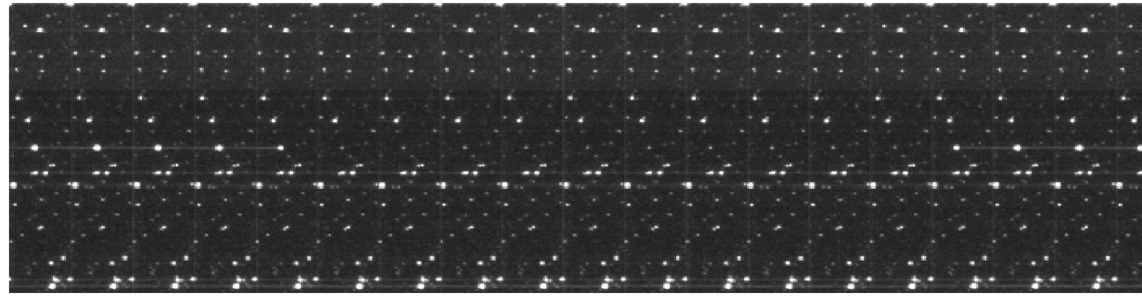
| beg. time | end time | count | ct err | x | y | rb |
|---|---|---|---|---|---|---|
| 66293.663414 | 66293.818818 | 26656.42 | 141.08 | 478.73 | 49.79 | 2 |
| 66293.913422 | 66294.068817 | 26787.13 | 137.69 | 478.86 | 125.79 | 2 |
| 66294.163444 | 66294.318814 | 26650.18 | 142.46 | 478.72 | 201.90 | 2 |
| 66294.413452 | 66294.568816 | 26329.72 | 138.72 | 478.86 | 277.74 | 2 |
| 66294.663449 | 66294.818813 | 26454.31 | 140.94 | 478.79 | 353.83 | 2 |
| 66294.913452 | 66295.068814 | 27486.91 | 145.05 | 478.93 | 429.89 | 2 |
| 66295.163458 | 66295.318812 | 27318.67 | 140.27 | 478.77 | 505.80 | 2 |
| 66295.413483 | 66295.568820 | 26646.11 | 140.48 | 478.61 | 581.82 | 2 |
| 66295.664491 | 66295.818816 | 25601.26 | 146.22 | 478.41 | 657.70 | 2 |
| 66295.913628 | 66296.068814 | 26560.96 | 143.56 | 478.57 | 733.74 | 2 |
| 66296.163496 | 66296.318812 | 26050.15 | 145.18 | 478.49 | 809.87 | 2 |
| 66296.413508 | 66296.568819 | 26172.10 | 148.19 | 478.33 | 885.90 | 2 |
| 66296.663500 | 66296.818816 | 25504.24 | 149.49 | 478.31 | 961.89 | 2 |
| 66296.913505 | 66297.068813 | 25484.46 | 139.52 | 478.32 | 1037.92 | 2 |
| 66297.163527 | 66297.318812 | 26705.86 | 146.23 | 478.26 | 1114.00 | 2 |
| 66297.413534 | 66297.568816 | 25456.52 | 153.24 | 478.25 | 1190.05 | 2 |
| 66297.663539 | 66297.818815 | 24907.12 | 144.34 | 478.47 | 1266.08 | 2 |
| 66297.913528 | 66298.068814 | 25864.63 | 140.92 | 478.66 | 1342.22 | 2 |

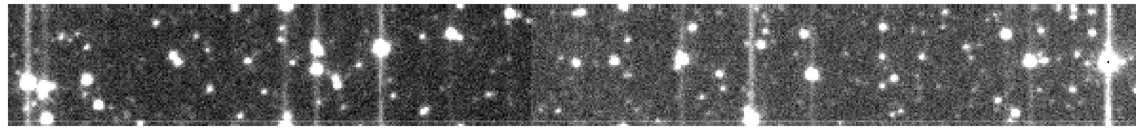
- Pipeline Flow from image-taking to archival
- Arrows represent the flow of messages between components

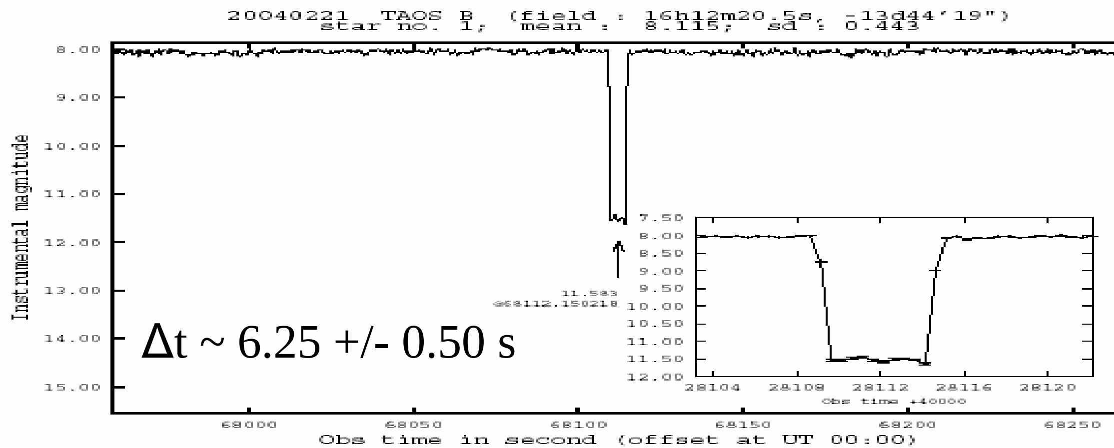



# TAOS Database

# TEST DRIVE





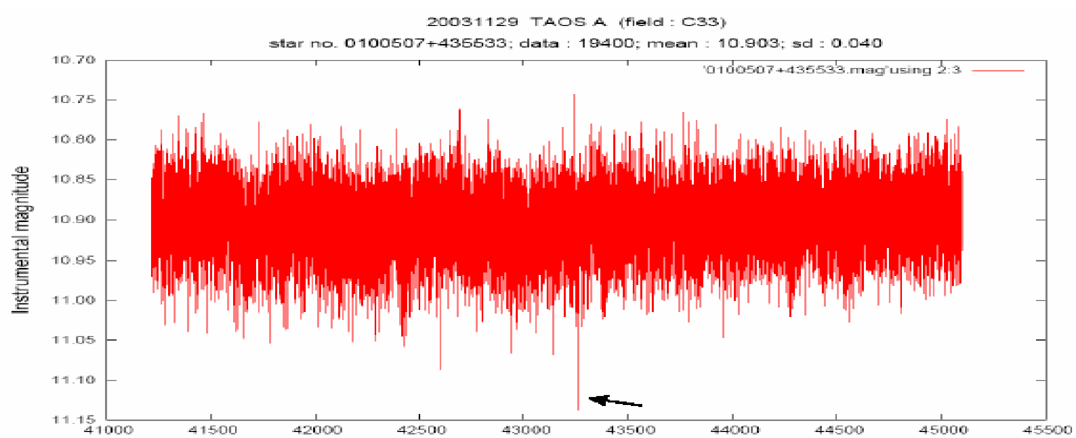



**2004 February 21**

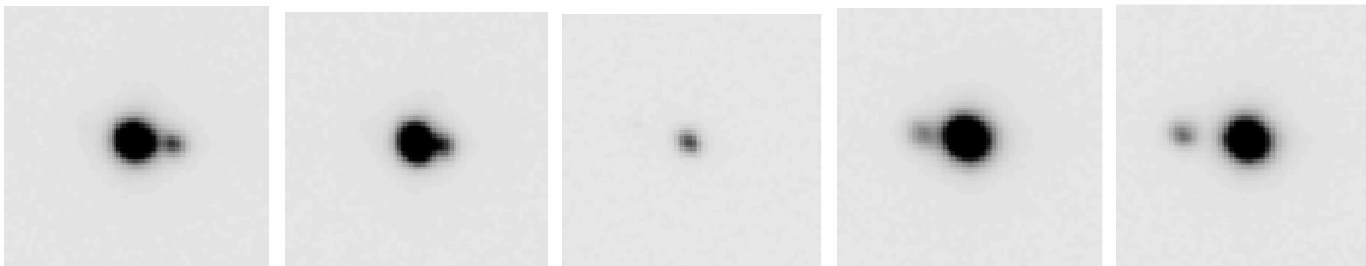
TAOS detected the occultation event of HIP 079407, $m_V$=8.8 mag) by **(51) Nemausa** ($m_V$=11.9)

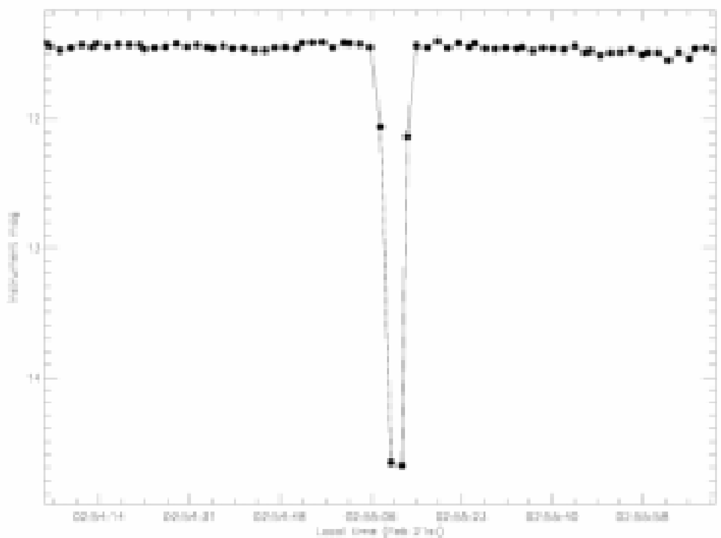
Prediction by Isao Sato (左藤勳)

D~150 km



# The 1 m telescope at Lulin also detected the same event with traditional CCD imaging.



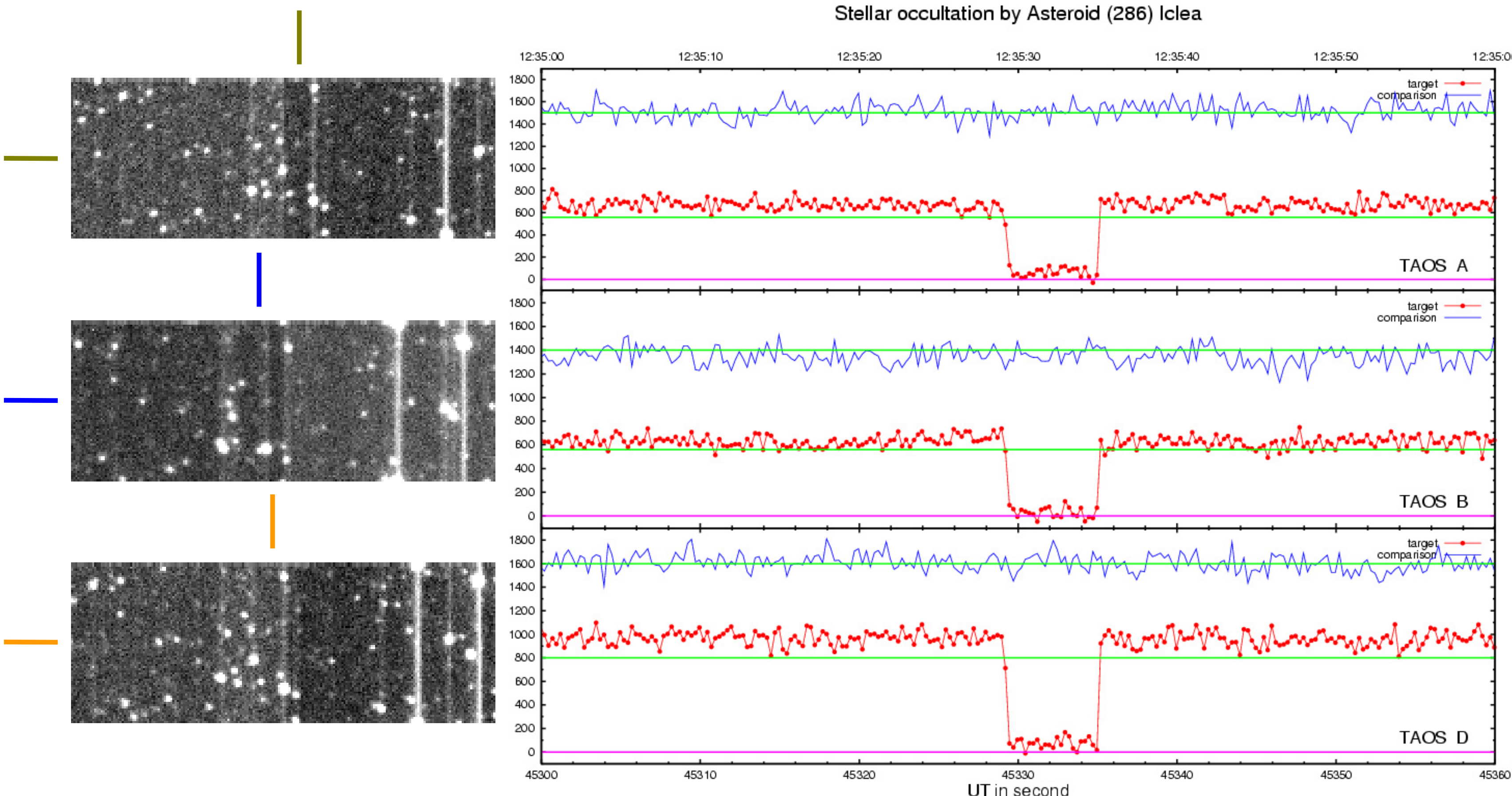


$t_{exp}$ =1 s
$t_{ro}$ < 2 s

By A. Chen

# TEST DRIVE II

**2006 Feb 06** three TAOS telescopes detected a suspected occultation of TYC 076200961 ($m_V$ ~ 11.83) by **(286) Iclea** ($m_V$ ~ 14.0 mag, D~ 97 km)



# Panoramic Survey Telescope And Rapid Response System

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SAIC Science Applications International Corporation
An Employee-Owned Company

INSTITUTE FOR ASTRONOMY
UNIVERSITY OF HAWAI'I

MAUI HIGH PERFORMANCE COMPUTING CENTER

AIR FORCE RESEARCH LABORATORY

國立中央大學

ACADEMIA SINICA

# Project Overview

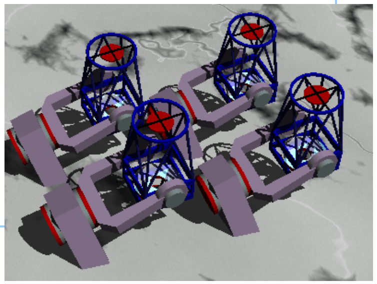
- All-sky survey (3π)
- Frequent revisit (cadence 4-7 days)

Wide-Field Imaging
Short Duty Cycle
Efficient Operations

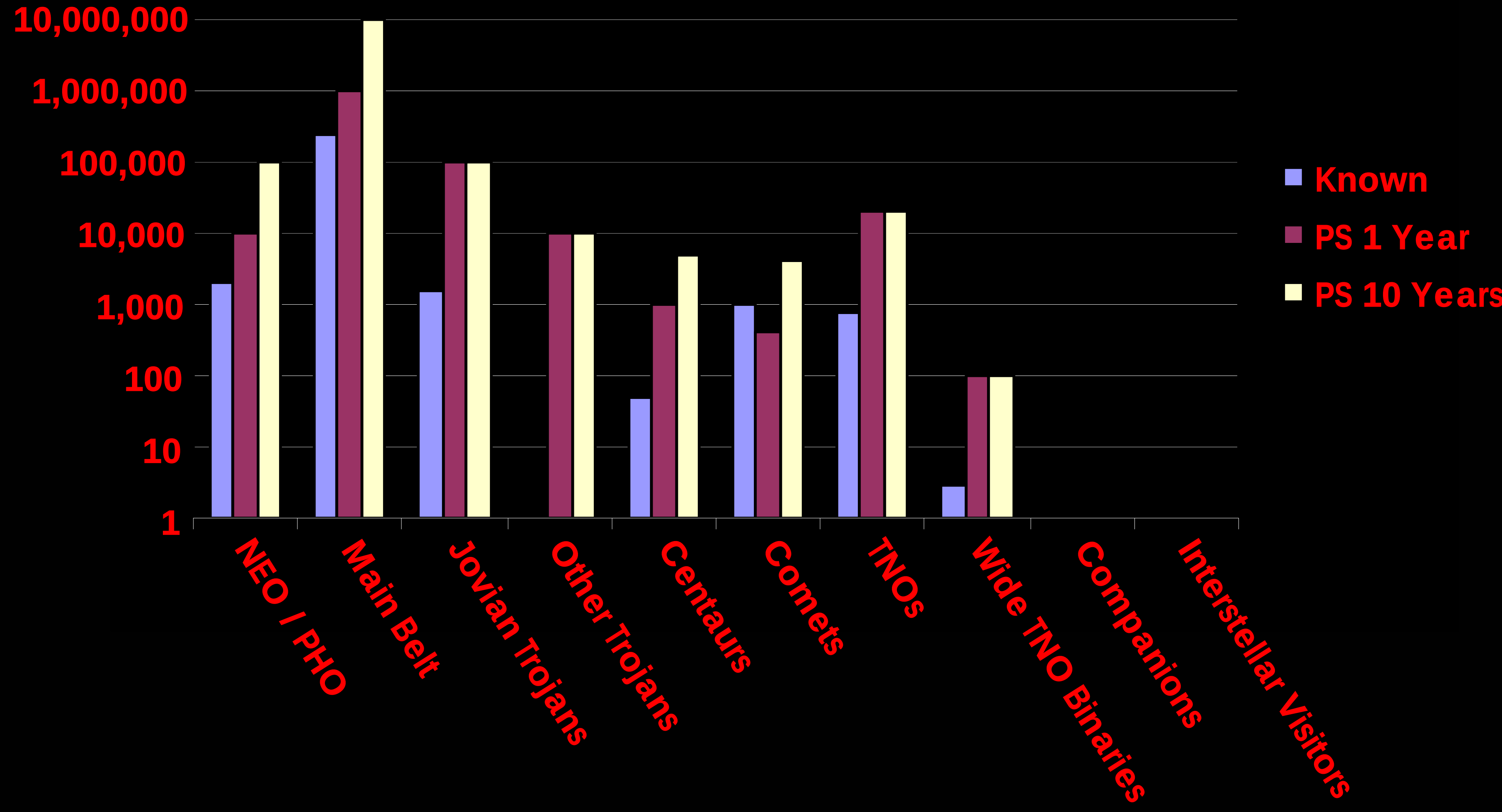
- An array of 4 telescopes, located in Hawaii, each of D=1.8 m, equipped with a 1.4 gigapixel camera of an Orthogonal Transfer Array CCD detector (=40 cm square focal plane) ➔ 7 square-degree FOV with 0.26” pixels

- Detection of moving, transient, and variable celestial objects down to very faint limits
- Cumulate very deep sky images

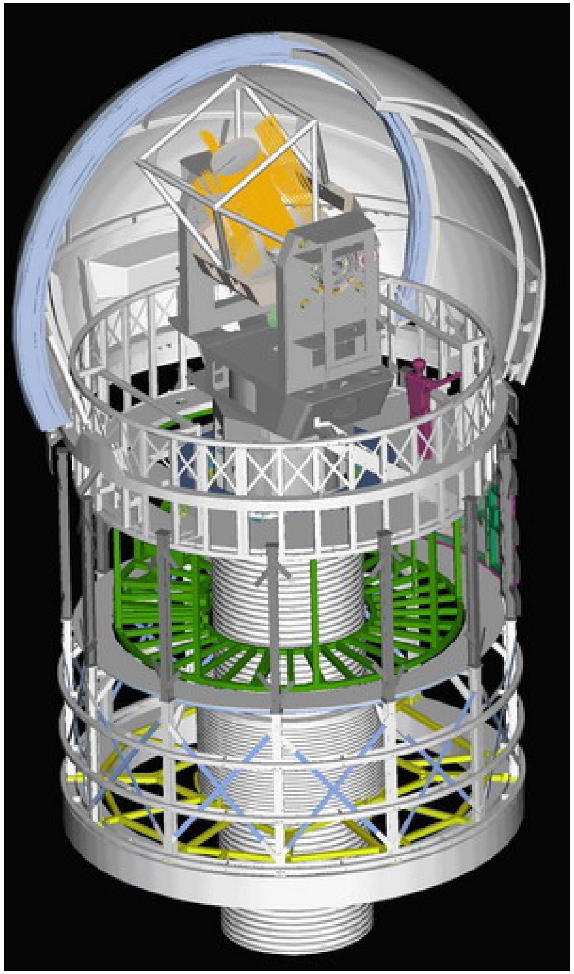
**several TB/night**

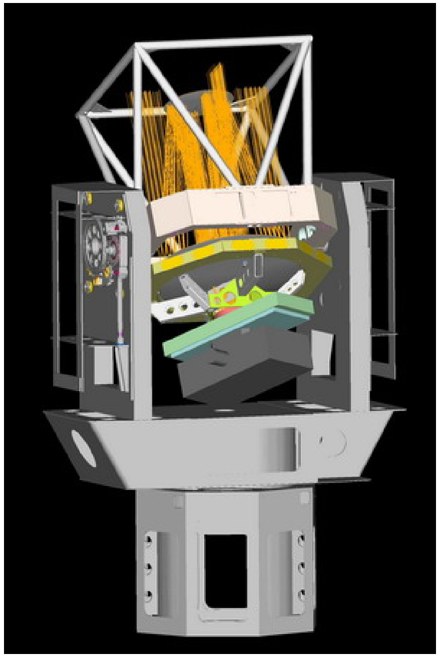


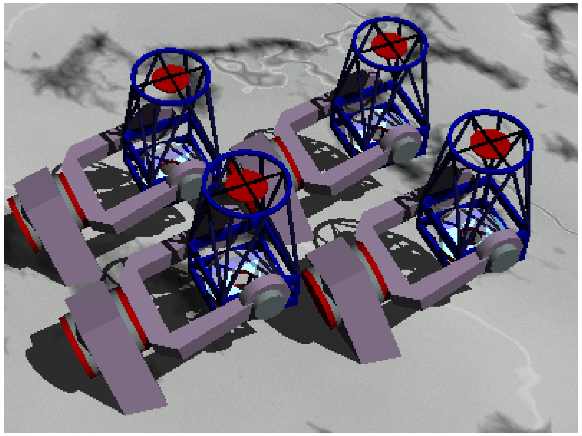
# Pan-STARRS Minor Planet Summary

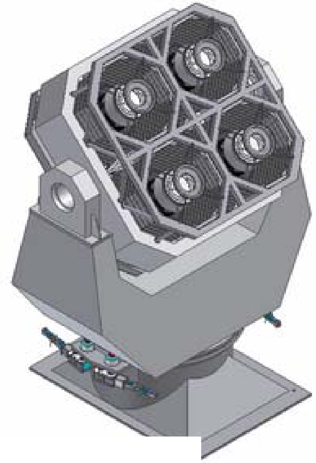


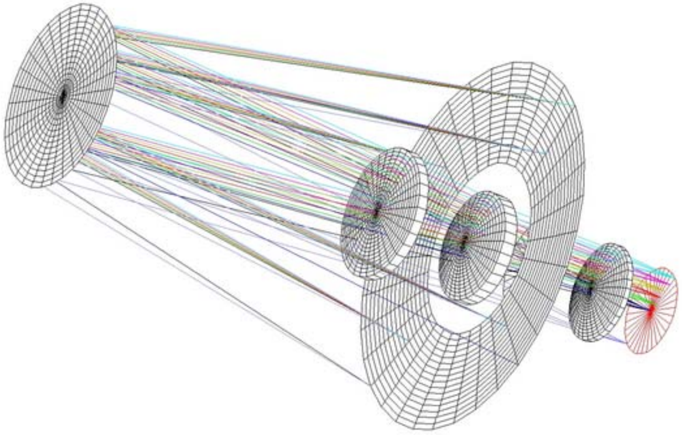
# The Telescopes



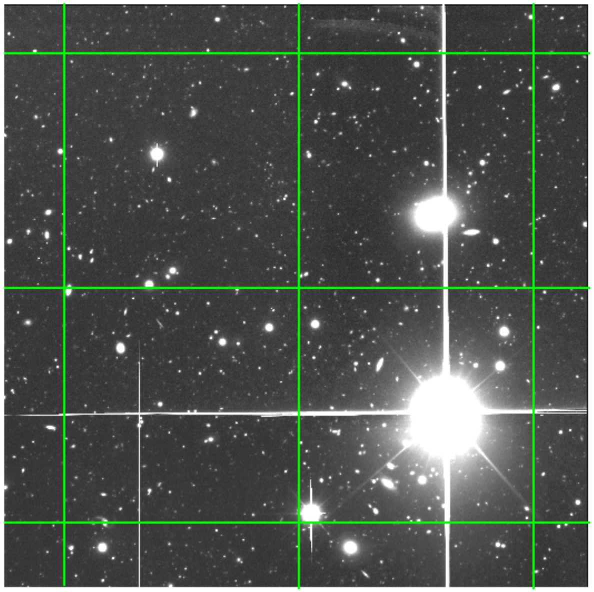


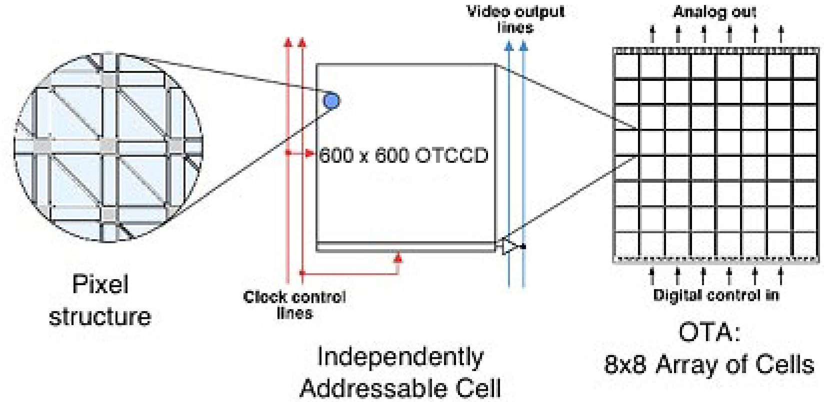






# The Detector

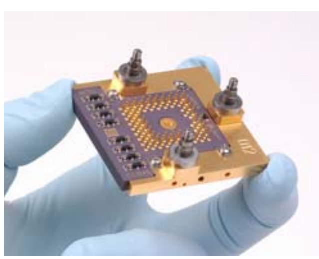


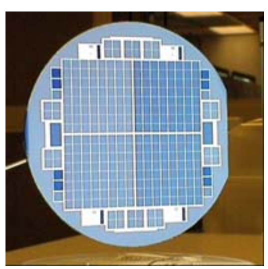


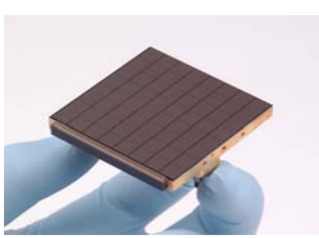

MIT Lincoln Laboratory

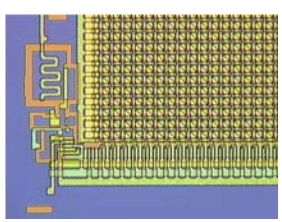
Independently addressable orthogonal transfer CCDs (cells)

- Reducing cost by increasing yield
- Fast readout: Gigapixels in 2 s
- On-Chip guiding
- Minimizing effects of bright stars
- Compensating for image motion



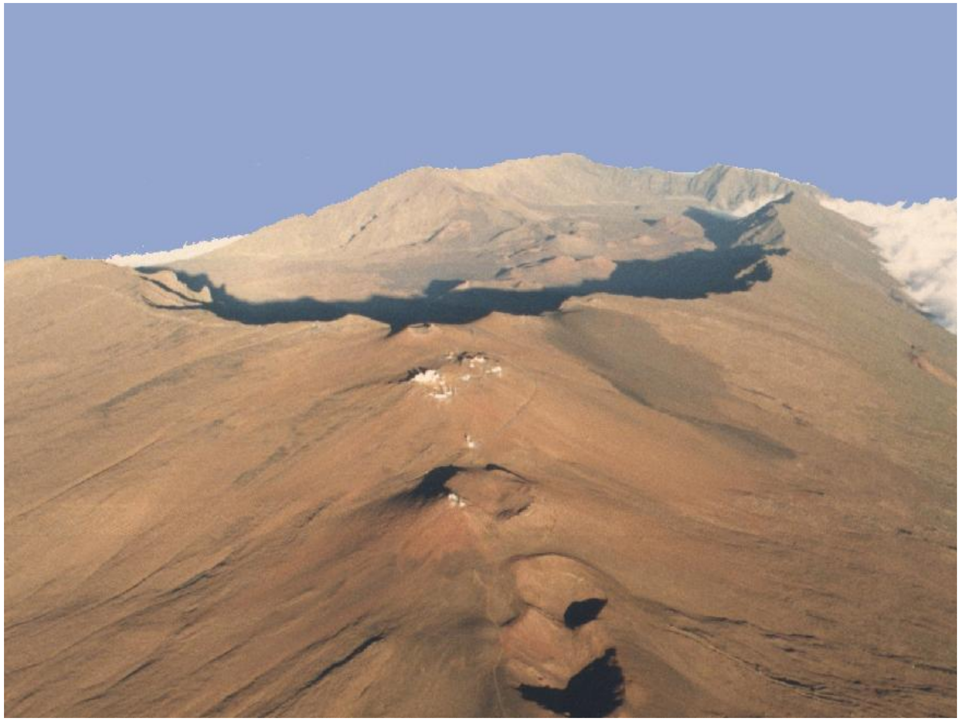


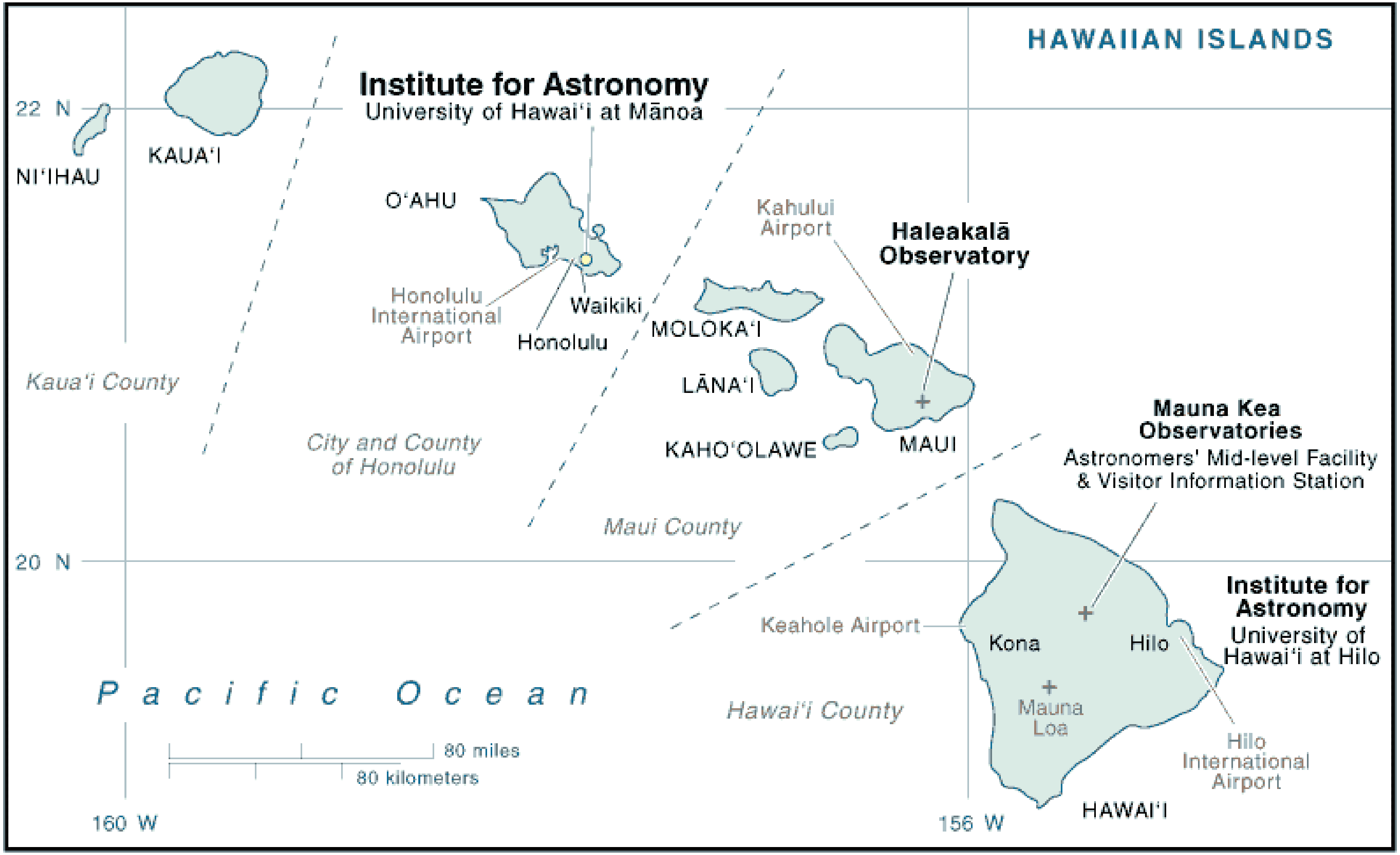




# The Site(s)

Site for Prototype telescope (PS1) ---
Haleakala High Altitude Observatory
(Maui)





HAWAIIAN ISLANDS
Institute for Astronomy
University of Hawai'i at Mānoa
22 N
20 N
160 W
156 W
NI'IHAU
KAUA'I
O'AHU
Honolulu International Airport
Waikiki
Honolulu
Kaua'i County
City and County of Honolulu
Kahului Airport
Haleakalā Observatory
MOLOKA'I
LĀNA'I
KAHO'OLAWE
MAUI
Maui County
Mauna Kea Observatories
Astronomers' Mid-level Facility & Visitor Information Station
Institute for Astronomy
University of Hawai'i at Hilo
Keahole Airport
Kona
Hilo
Mauna Loa
Hilo International Airport
HAWAI'I
Hawai'i County
Pacific Ocean
80 miles
80 kilometers

# Eventual Mauna Kea site for full Pan-STARRS

# IT Challenges

Each raw **image** from a **single** Pan-STARRS camera will contain **2 Gbytes** (2 bytes per pixel). In the full survey mode, typical exposures last **30 seconds**, so the raw data rate is several terabytes per night for the full telescope system. The amount of data produced by Pan-STARRS is so large that it will not be practical to archive every image. Software techniques are therefore being developed to extract the important information from the images, while allowing less crucial information to be discarded.

Data storage requirement ~0.5 Pb in year 1

# The Data Flow

**Subsystems**

✓TEL – Telescopes

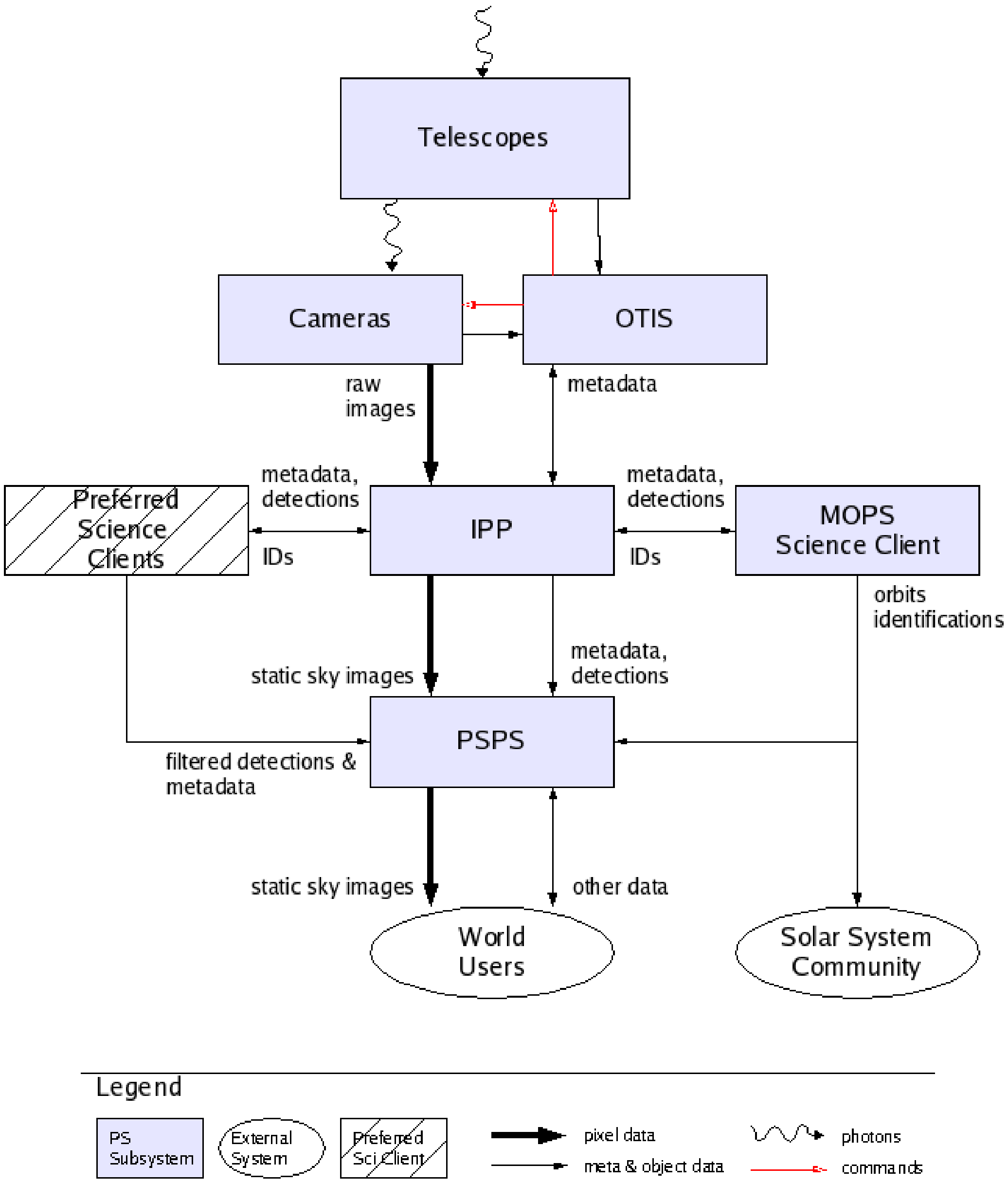
✓CAM – Cameras

✓OTIS – Observatory, Telescope & Instrument Software

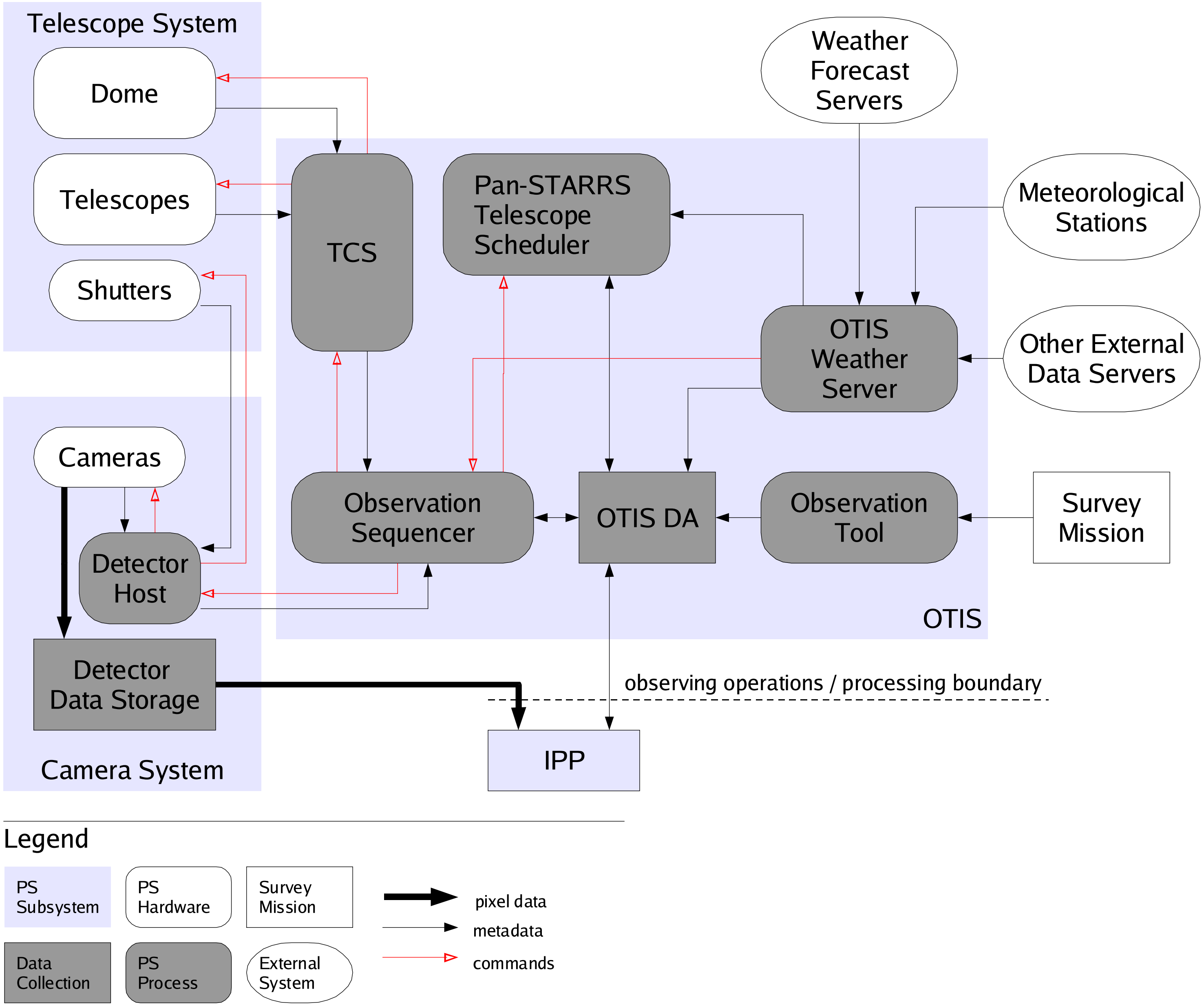
✓IPP - Image Processing Pipeline

✓MOPS – Moving Object Processing Software

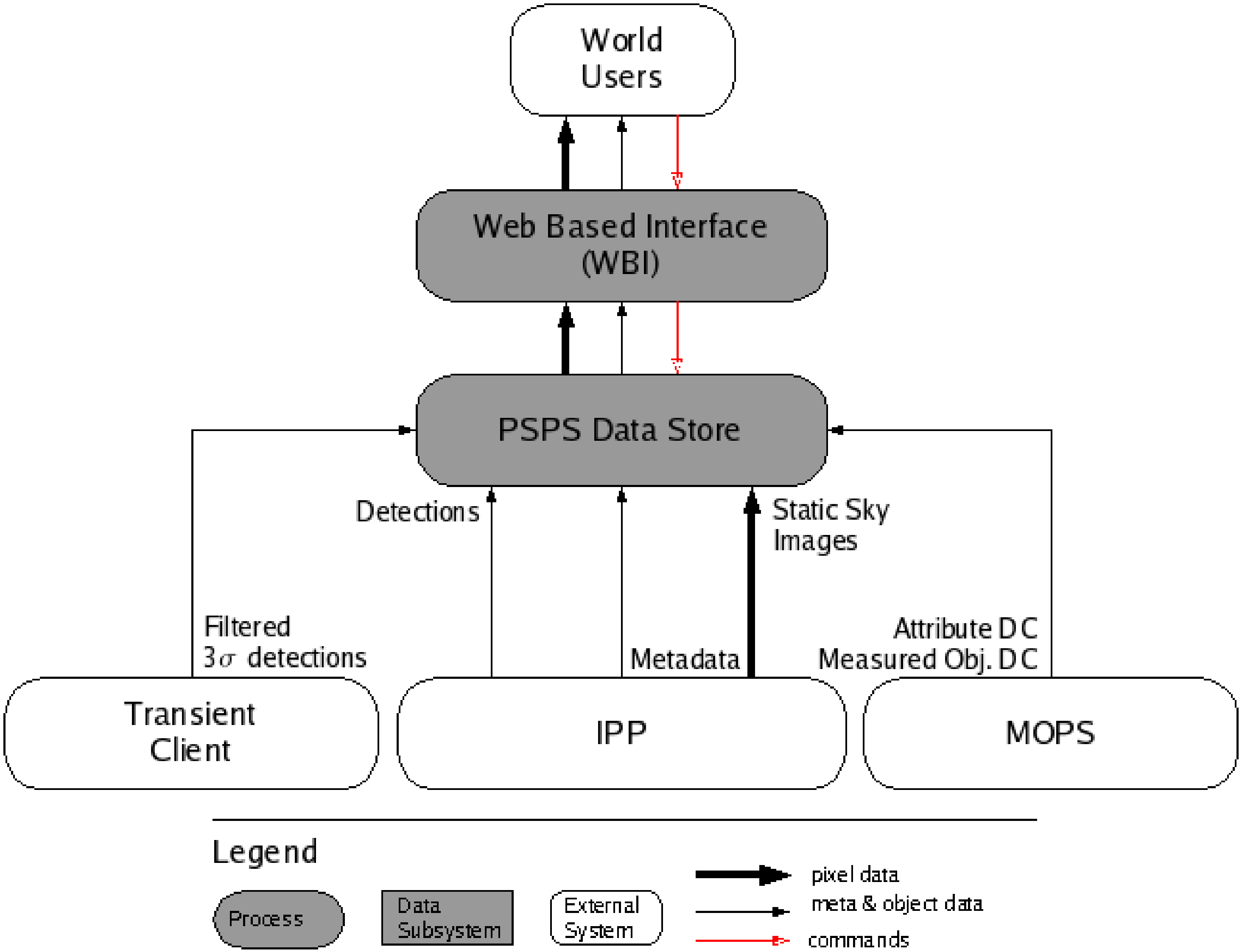
✓PSPS – Published Science Data Products



# Summit Process Flow Diagram



# PSPS Overview



# Database Sizing Justification

| Object DB | P2 | P4Σ | P4Δ |
|---|---|---|---|
| obj/deg | 2.7E+04 | 1.1E+05 | 2.0E+05 |
| deg/fpa | 7.00 | 7.00 | 7.00 |
| FPA/night | 3000.00 | 750.00 | 750.00 |
| nights/year | 250.00 | 250.00 | 250.00 |
| bytes/obj | 64.00 | 64.00 | 64.00 |
| DB OH | 4.00 | 4.00 | 4.00 |
| Years | 10.00 | 10.00 | 10.00 |
| PB | **0.36** | **0.36** | **0.67** |

| Cum. Sky Catalog | |
|---|---|
| deg | 3.0E+04 |
| obj/deg | 4.3E+05 |
| filters | 6.00 |
| bytes/obj | 300.00 |
| Compress | 1.00 |
| DB OH | 4.00 |
| Copies | 2.00 |
| PB | **0.19** |

| Static Sky Images | |
|---|---|
| deg | 3.0E+04 |
| pix/deg | 3.2E+08 |
| filters | 6.00 |
| bytes / pix | 7.20 |
| Compress | 0.40 |
| DB OH | 1.00 |
| Copies | 9.00 |
| PB | **1.51** |

| Data Product | Size (PB) |
|---|---|
| Static Sky Img | 1.51 |
| Object Data | 1.43 |
| Cum. Sky Cat. | 0.19 |
| Metadata | 0.04 |
| Postage Stamps | 0.01 |
| MOPS | 0.0021 |
| Filtered Trans. | 0.00001 |
| Total (PB) | **3.19** |

# 2006/02


# Conclusions

- Time-domain astrophysics as a niche science
- Astronomers demanding to push the IT forefronts

  Telescope/Detector technology
  → larger, finer observations

  Rapid cadence → huge data volume

- Data processing, analysis, storage, archival, distribution
  (*$1 hardware*, *$1 software*, *$3-10 DB*)
- Need to involve software engineers, IT managers, statisticians … from the **very beginning** of a project to design the experiment