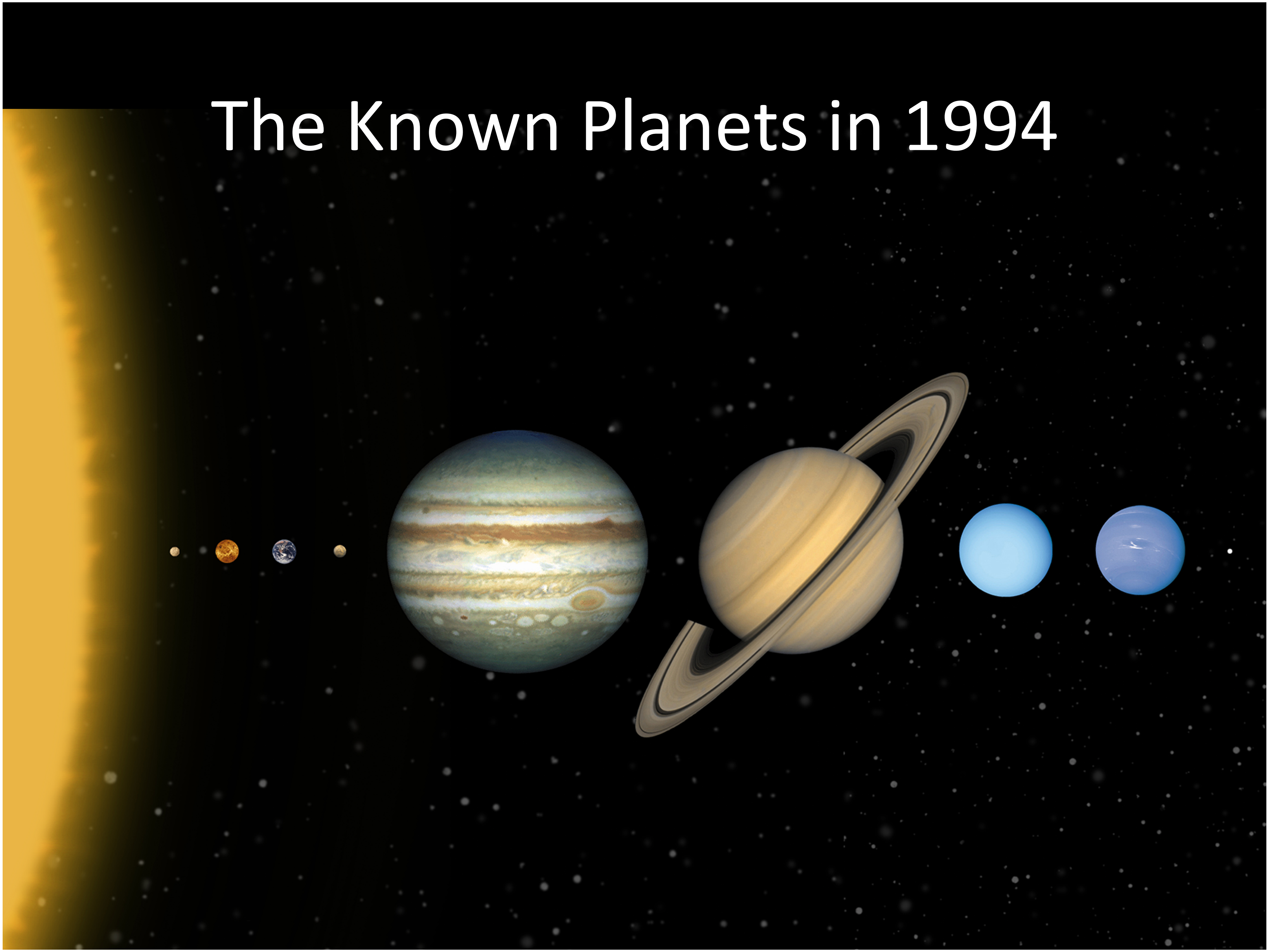
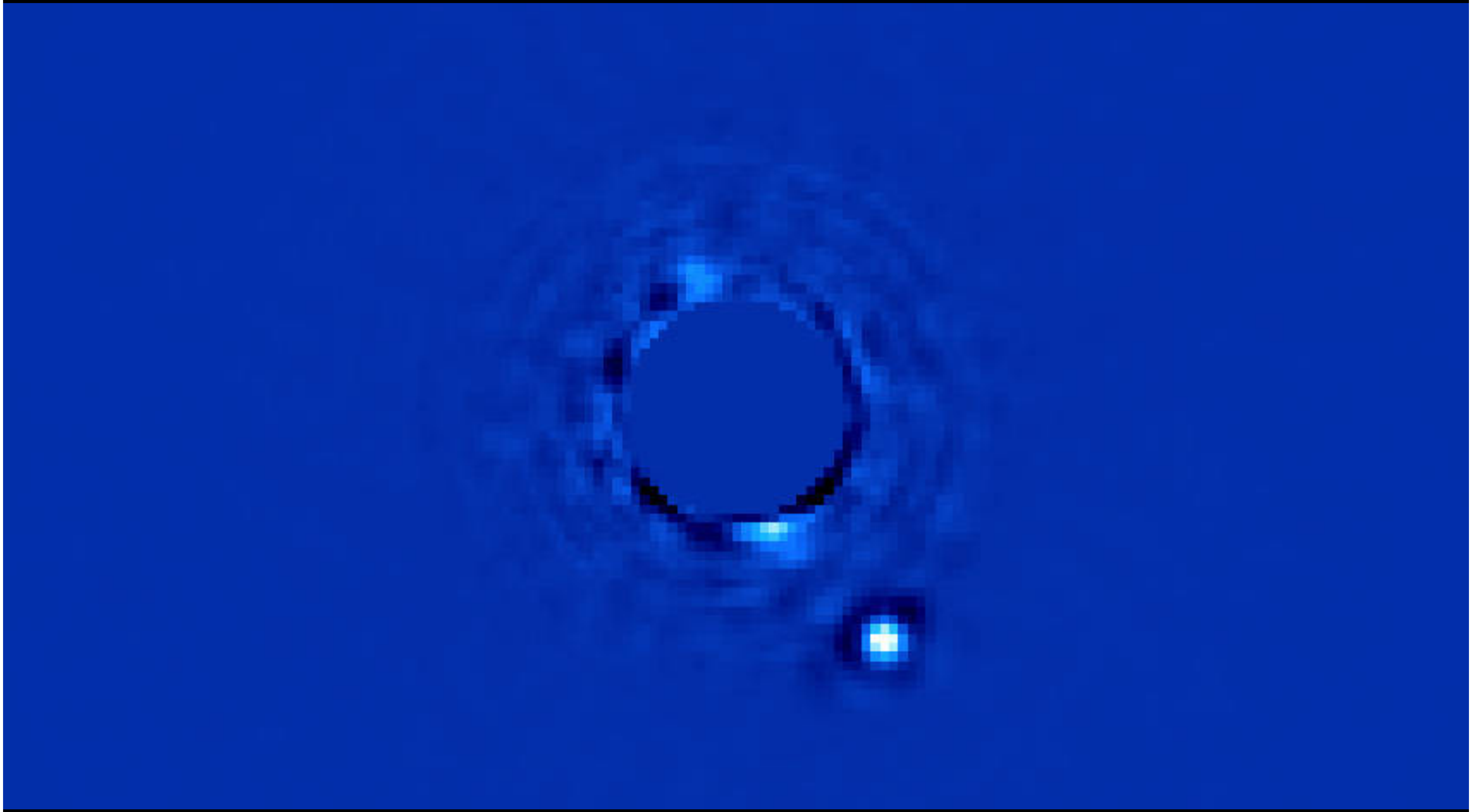
From Pixels to Planets

Lee Brownston, SGT, Inc.
Jon M. Jenkins, NASA

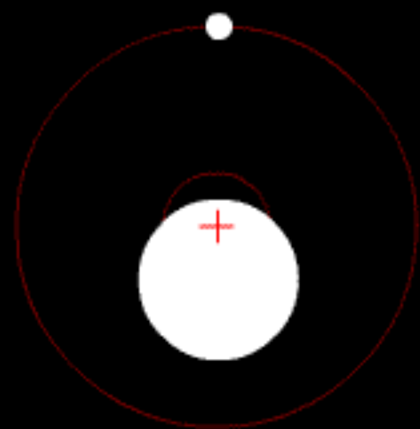
The Known Planets in 1994



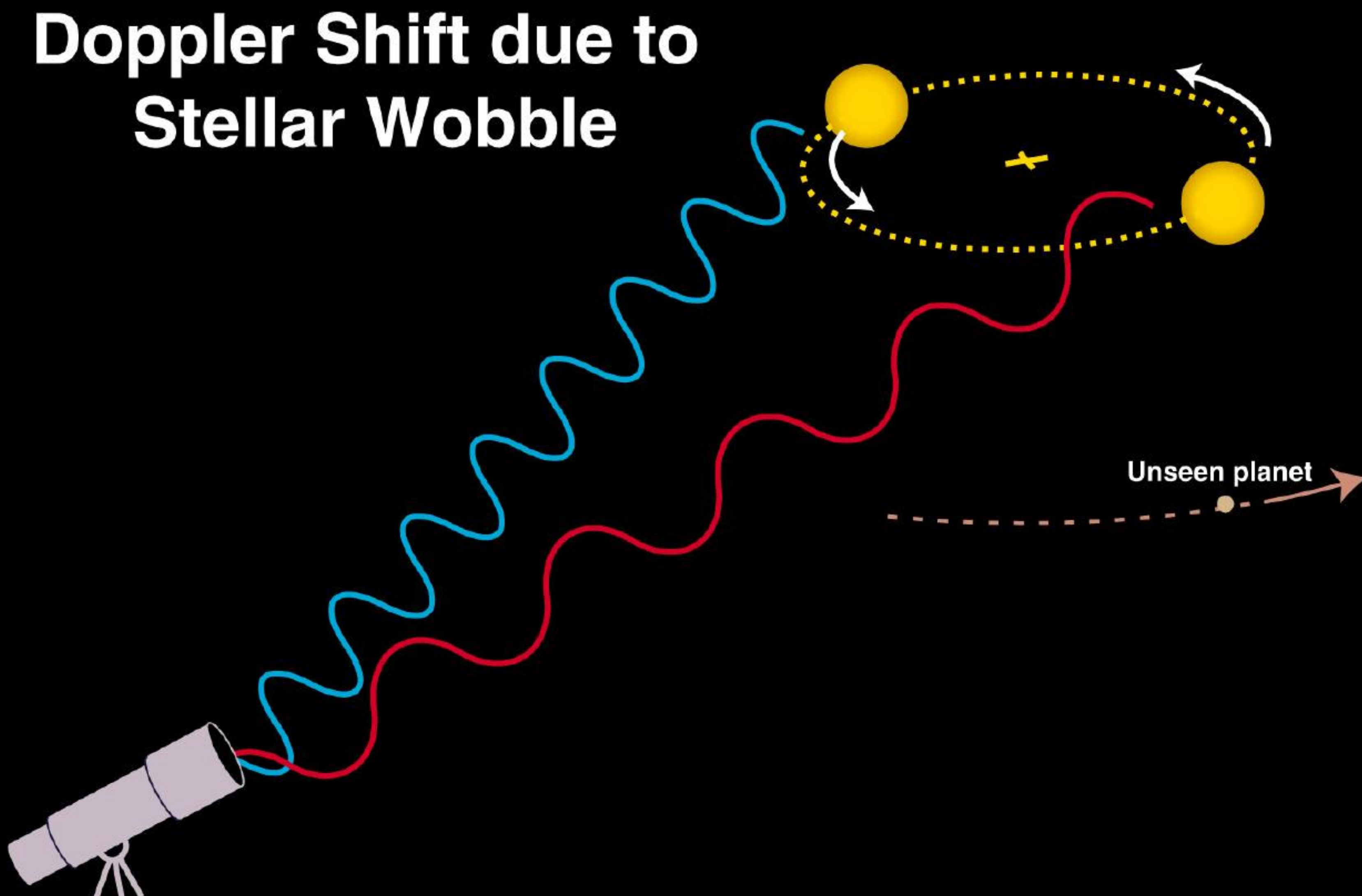
Direct Imaging: Beta Pictoris b



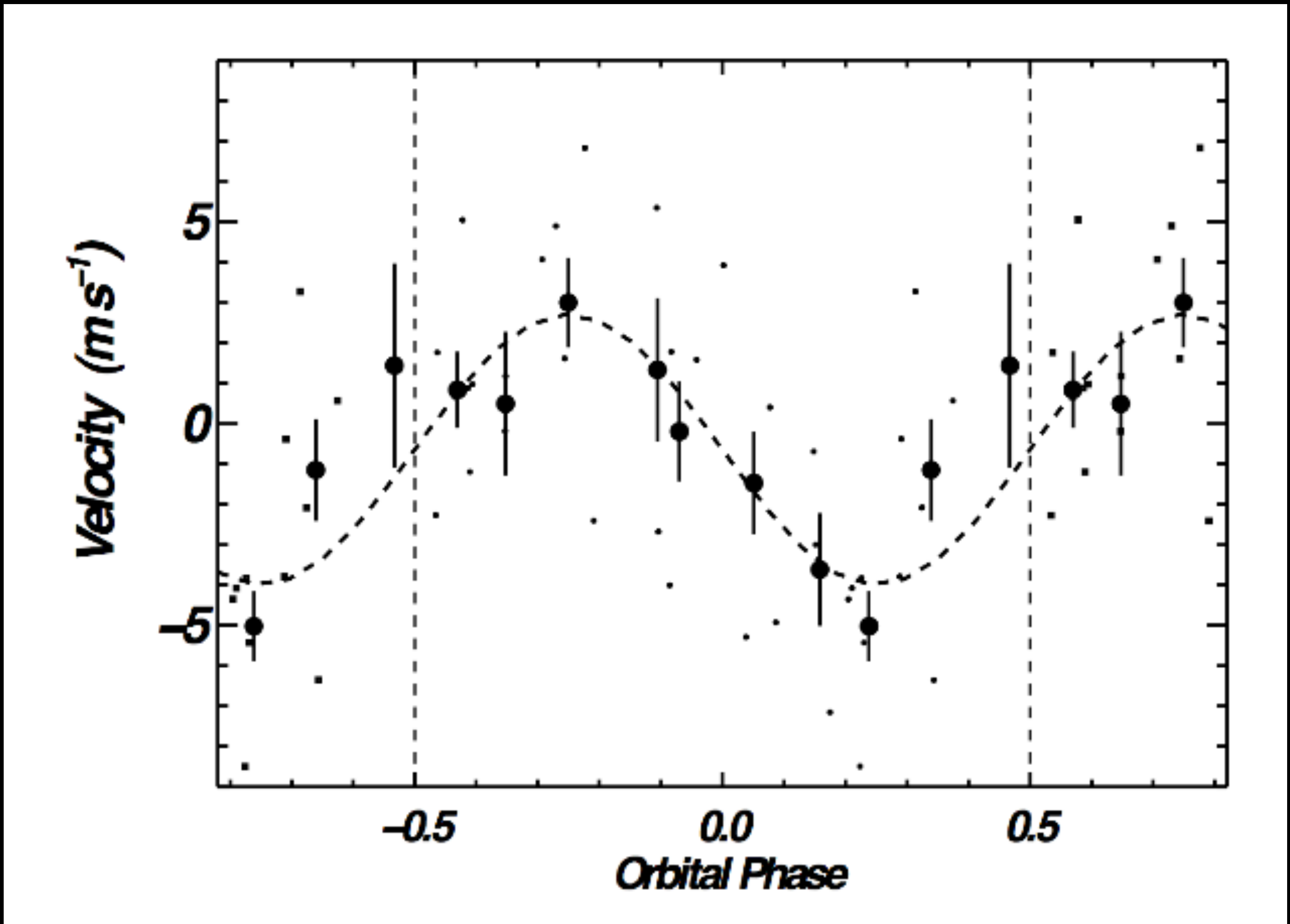
Planetary System Center of Mass



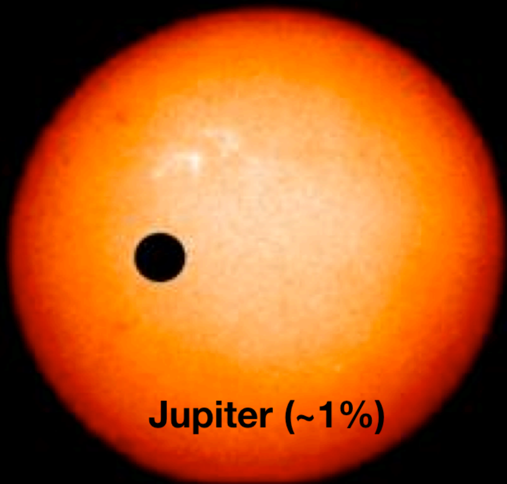
Radial Velocity

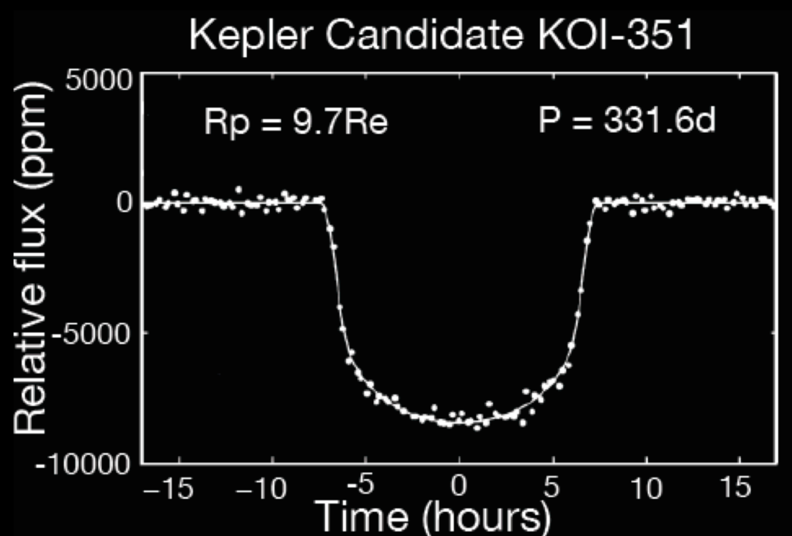


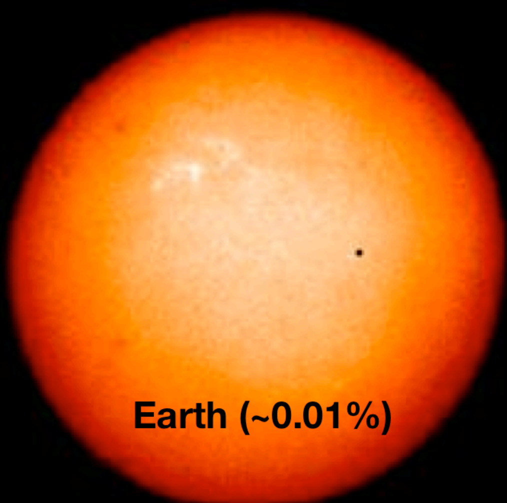
Doppler Measurements

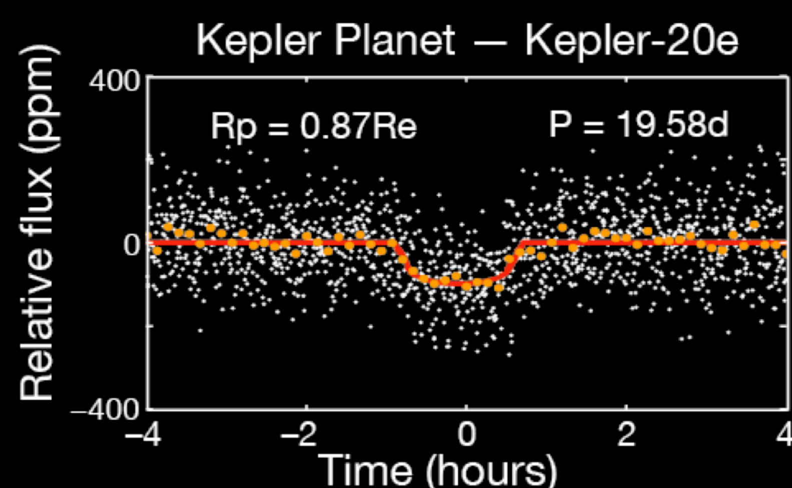


Earth-size Planets: Detection Method

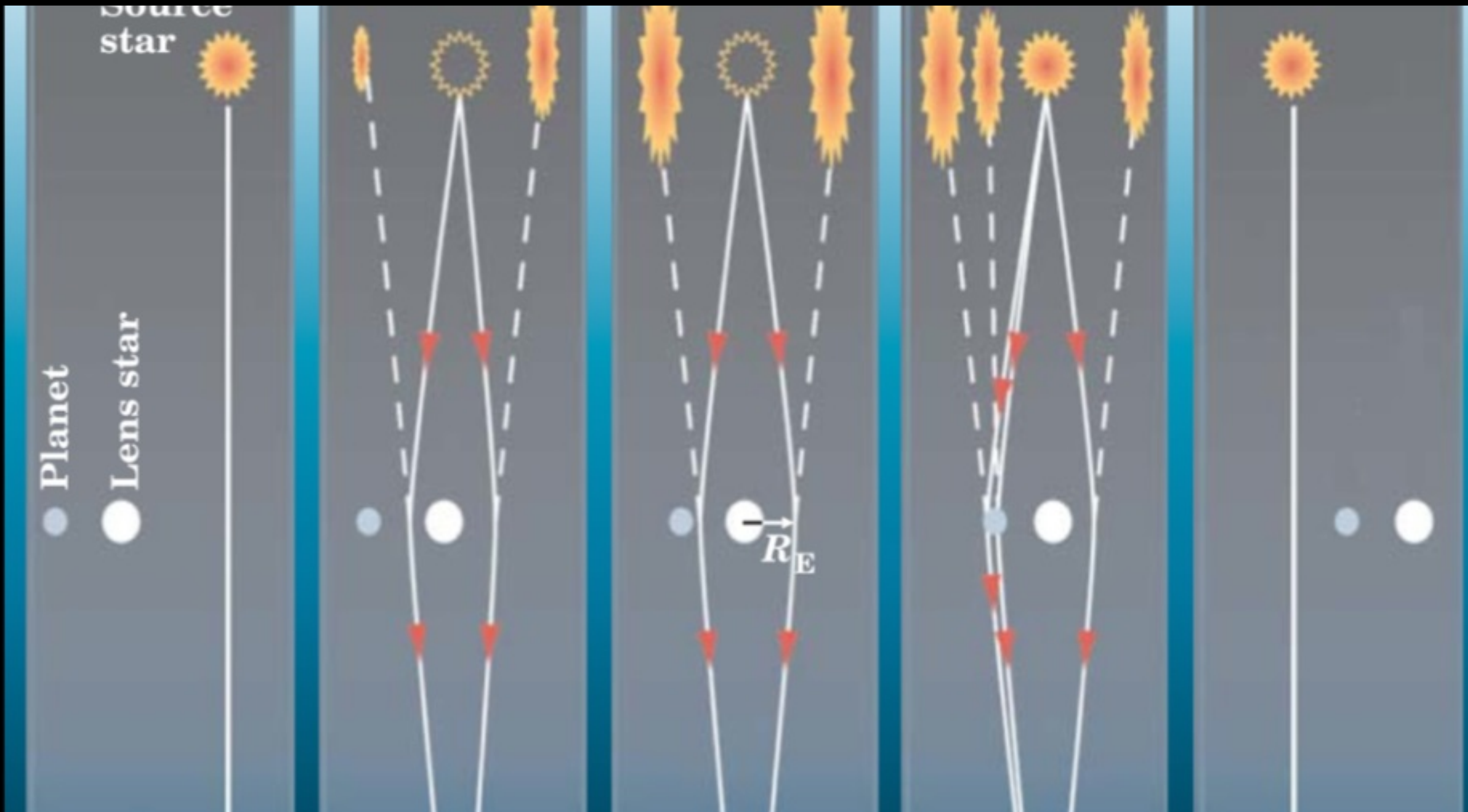




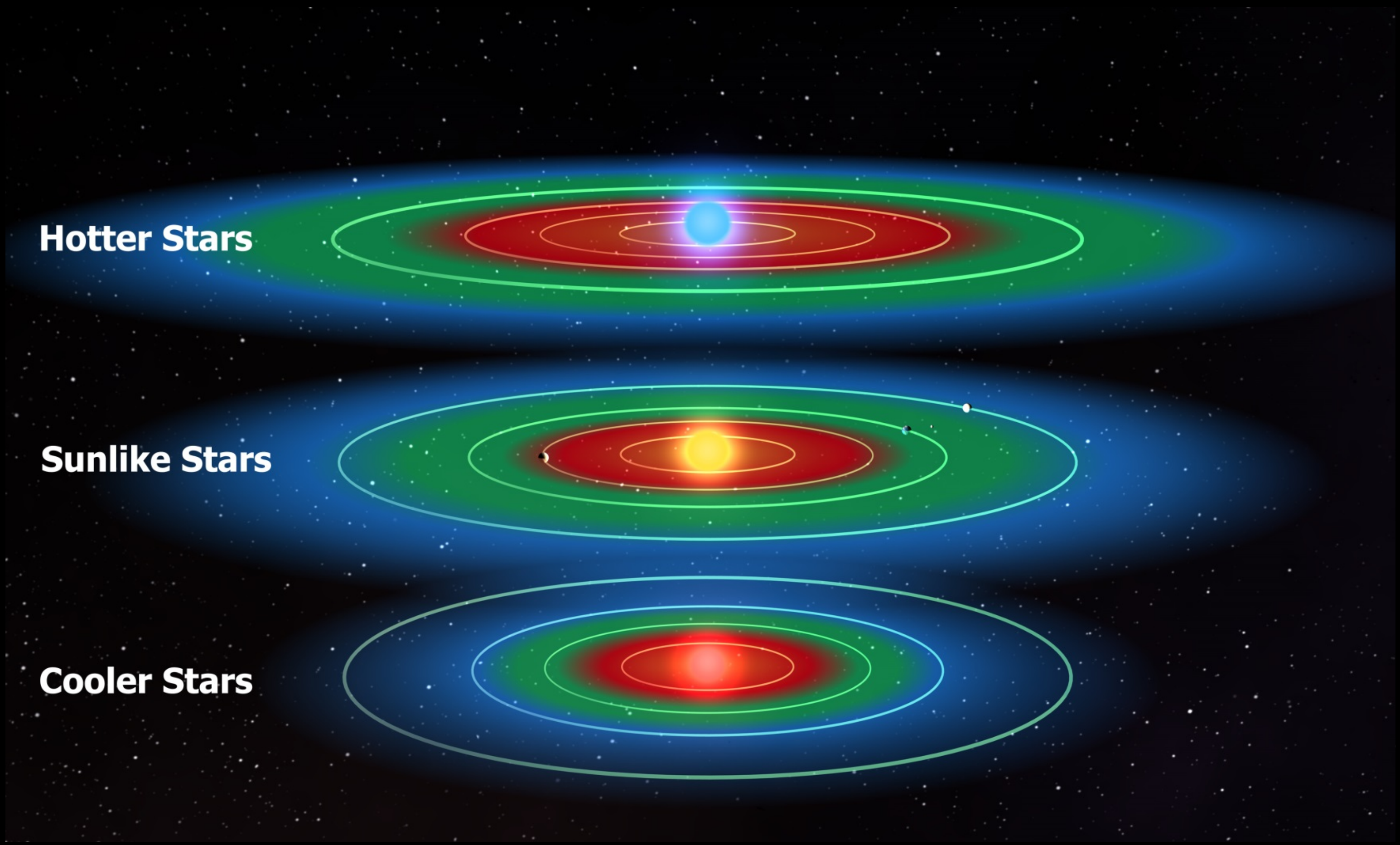


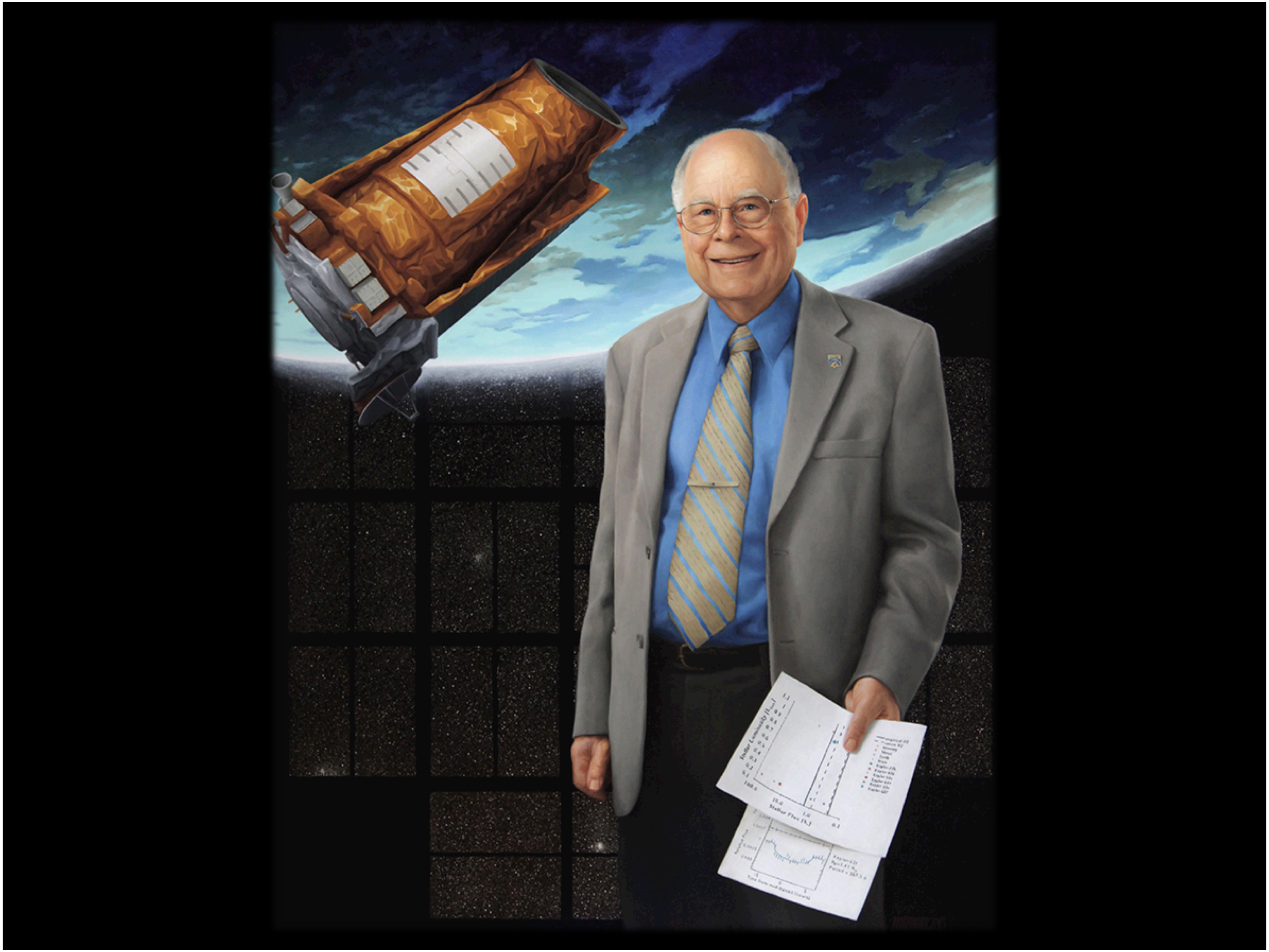


Gravitational Microlensing

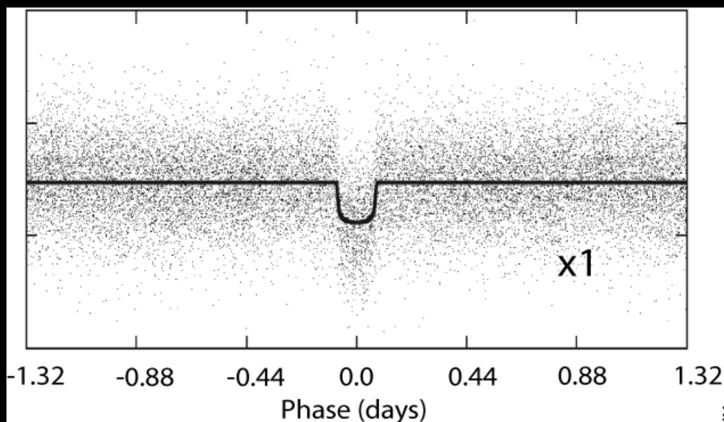


The Habitable Zone

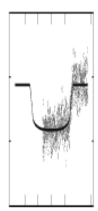



100.0
10.0
1.0
0.1
Stellar Flux

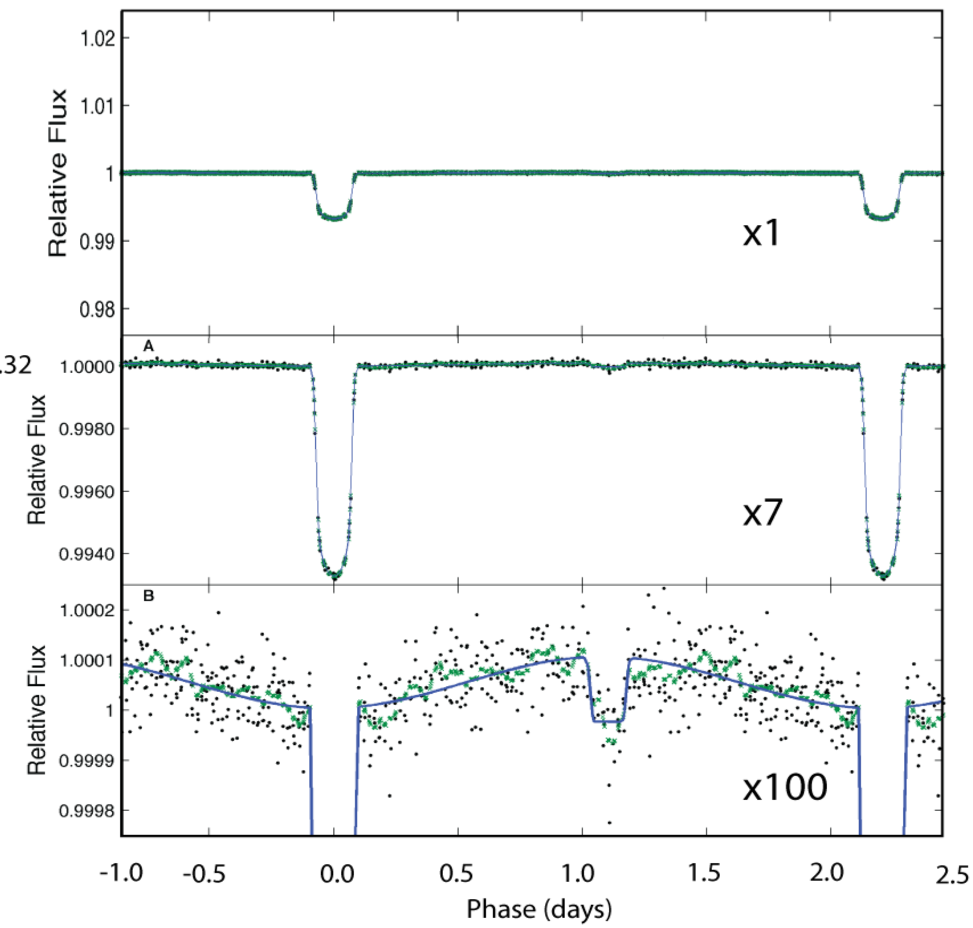
HAT-P-7b Ground vs. Space



16,620 HATNet data points (57.7 days of data)



Single night at 1.2 m FLWO with Kepler Cam

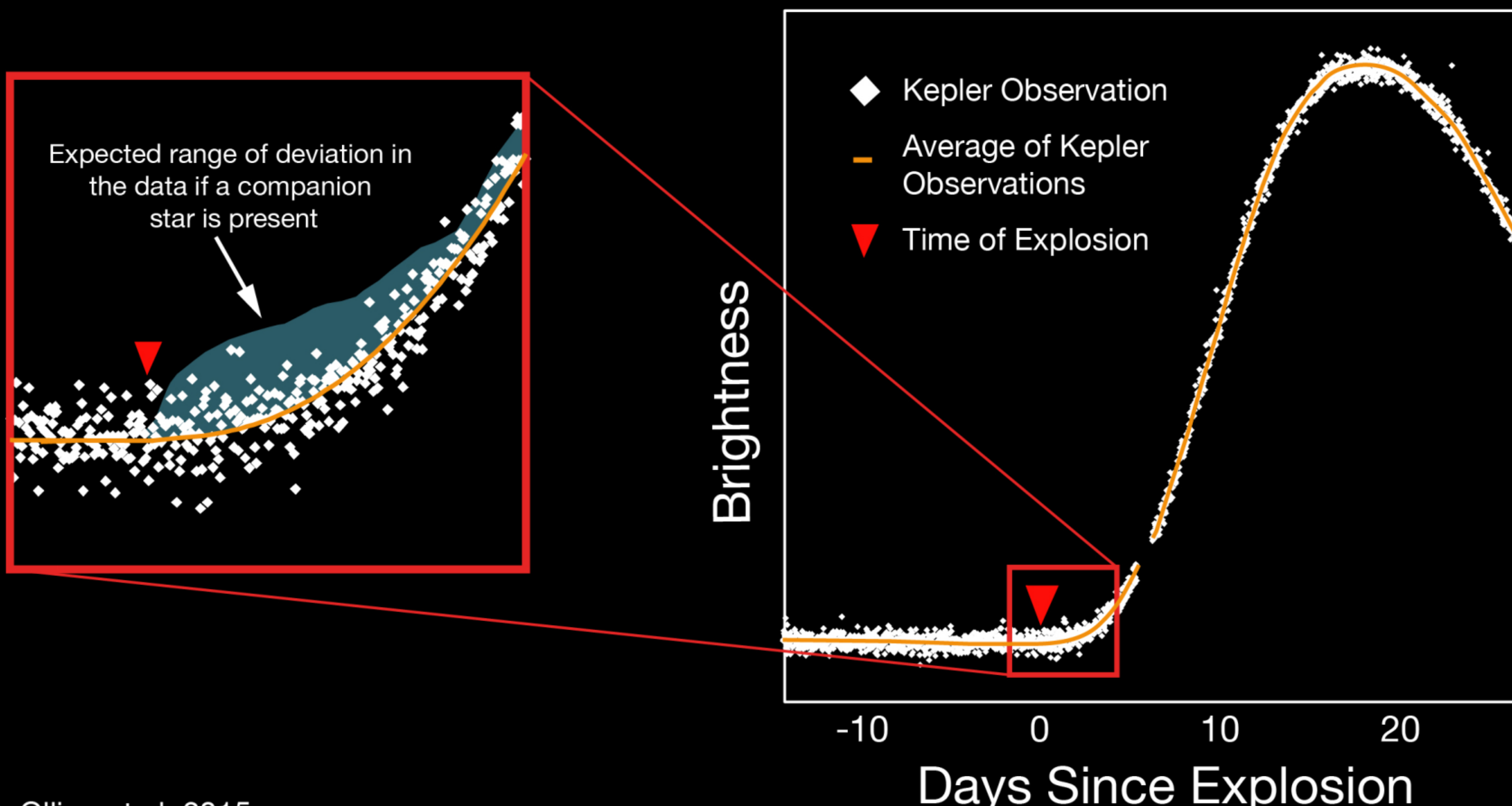


HAT-P-7b data from the ground
A. Pal et al., 2008

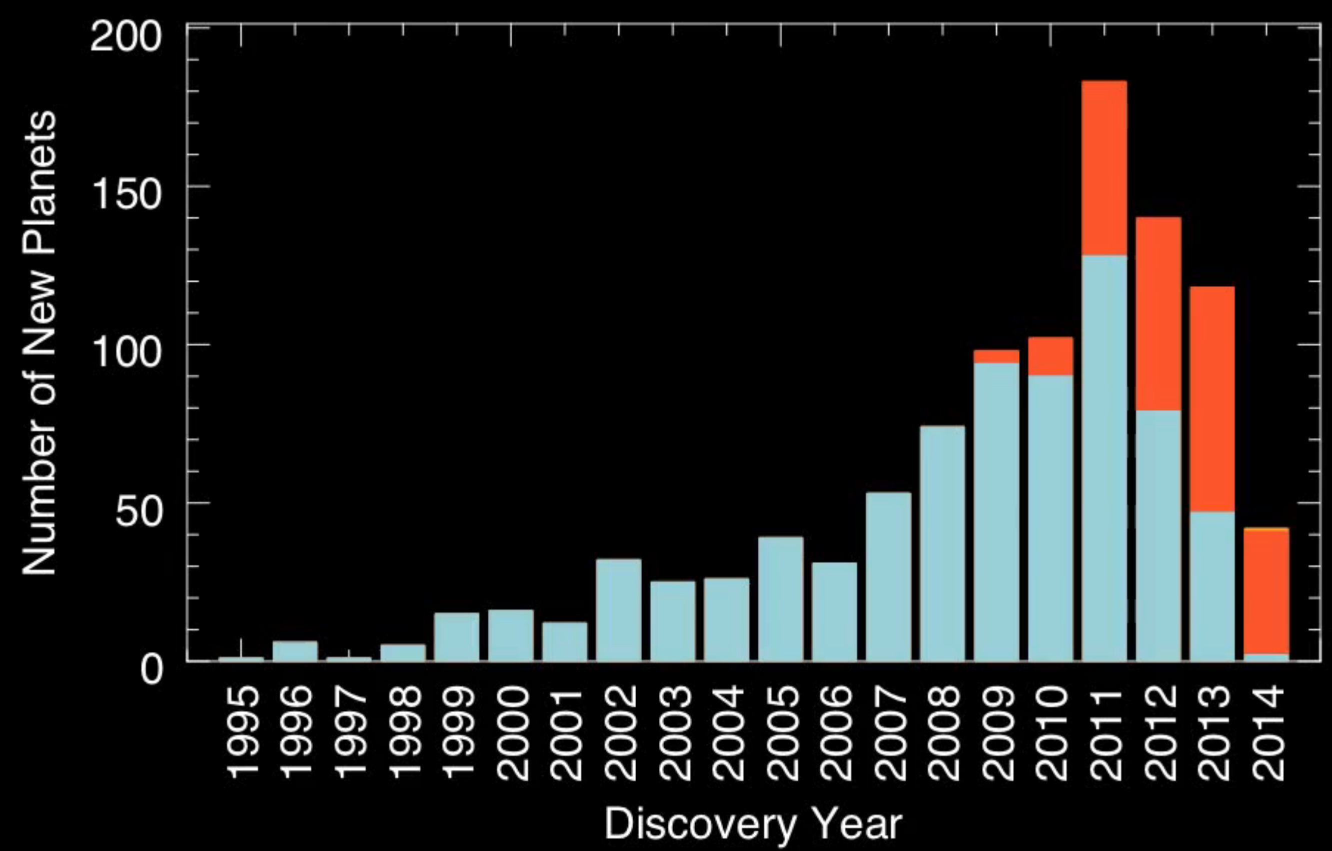
Kepler Commissioning data (10 days)
W. Borucki et al., 2009

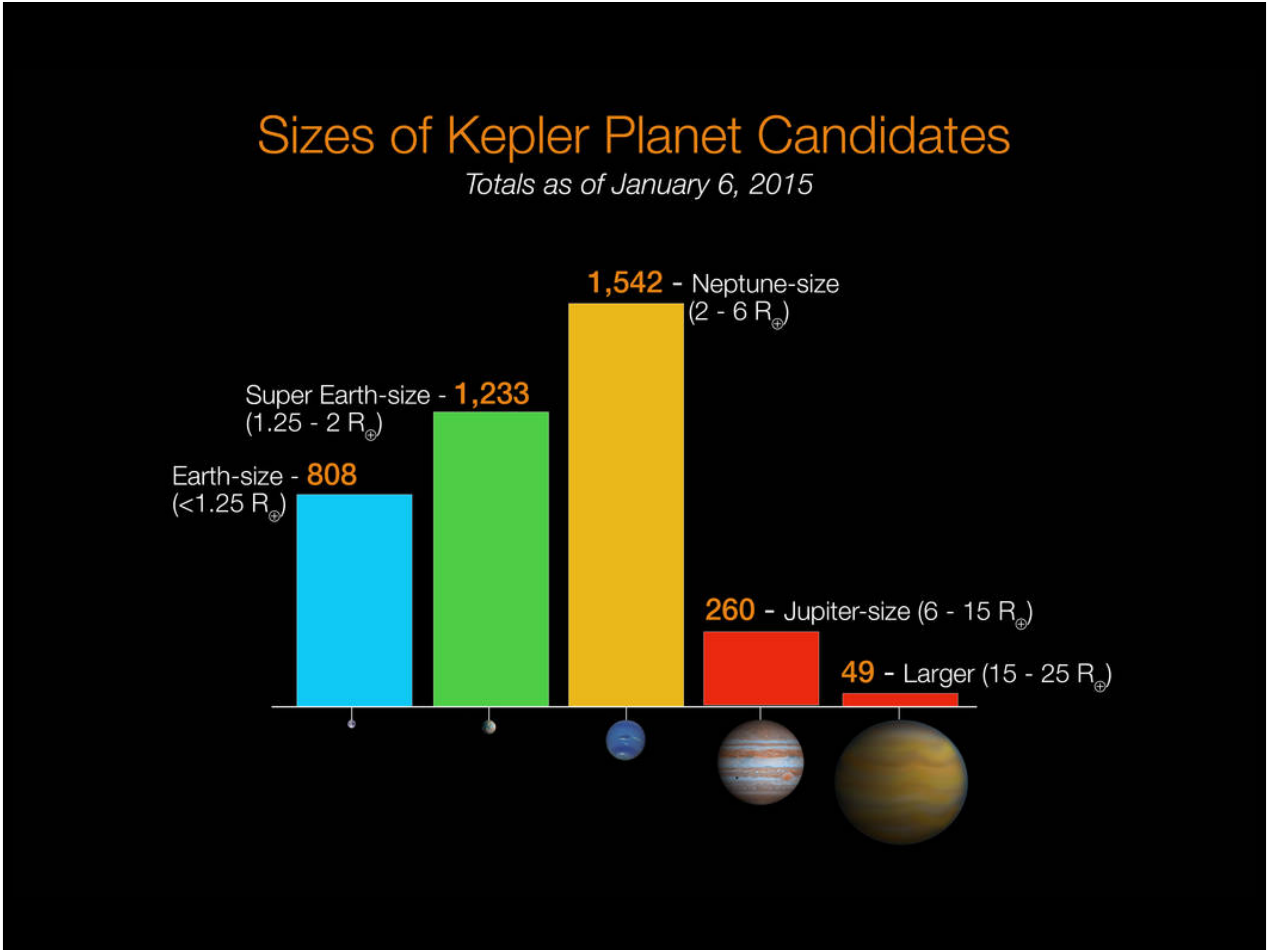
Kepler's Field of View

Kepler Observations of Supernova KSN 2011b



KEGS: Olling et al. 2015


200
150
100
50
0
Number of New Planets
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
Discovery Year

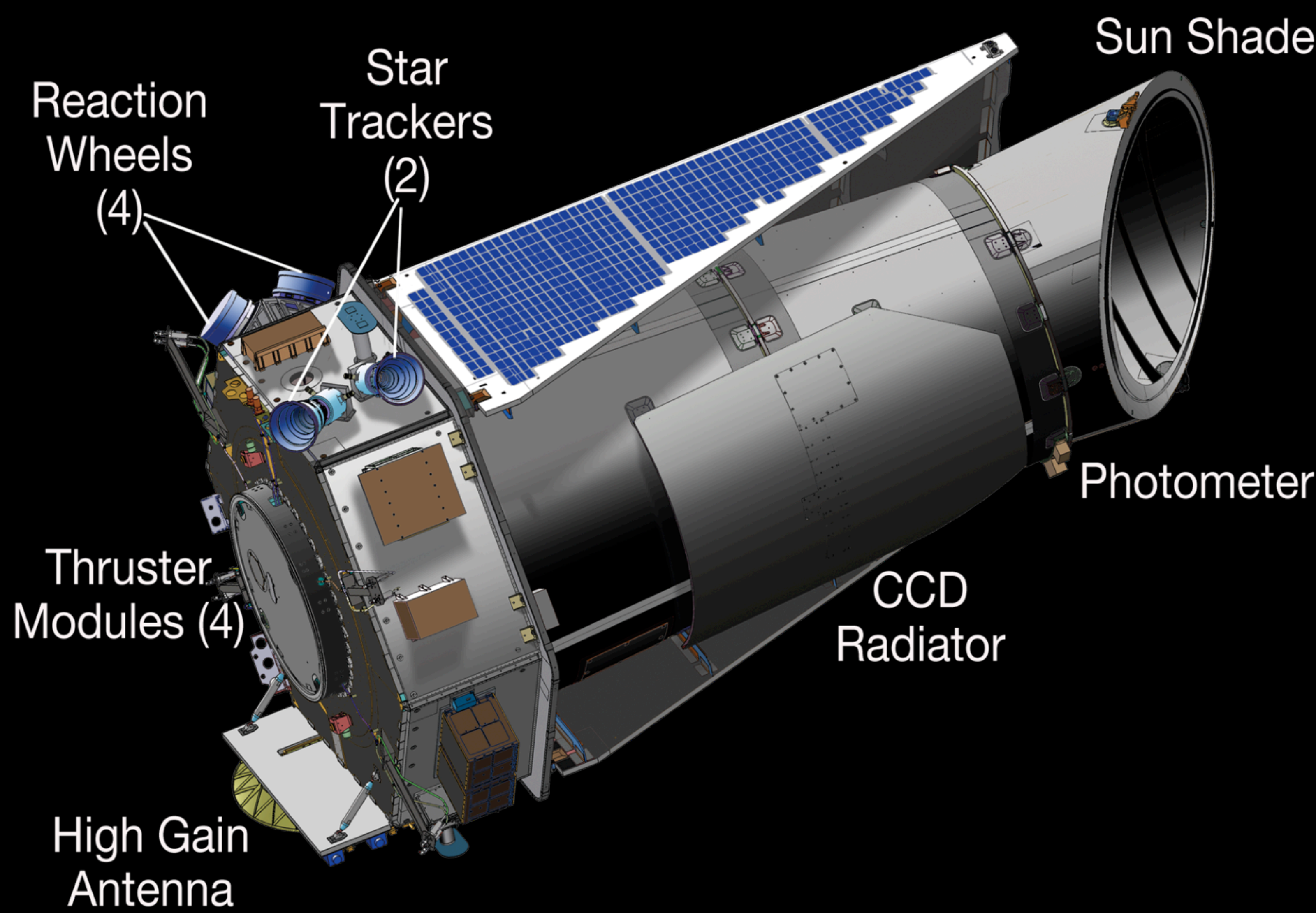

Sizes of Kepler Planet Candidates
Totals as of January 6, 2015
1,542 - Neptune-size
(2 - 6 $R_\oplus$)
Super Earth-size - 1,233
(1.25 - 2 $R_\oplus$)
Earth-size - 808
(<1.25 $R_\oplus$)
260 - Jupiter-size (6 - 15 $R_\oplus$)
49 - Larger (15 - 25 $R_\oplus$)


NASA Kepler's Hall of Fame:
Small Habitable Zone Planets
As of January 2015
NEW
Kepler-438b
Kepler-442b
Kepler-440b
Kepler-186f
Kepler-62f
Kepler-62e
Kepler-296e
Kepler-296f
Earth
ARTIST'S CONCEPTS

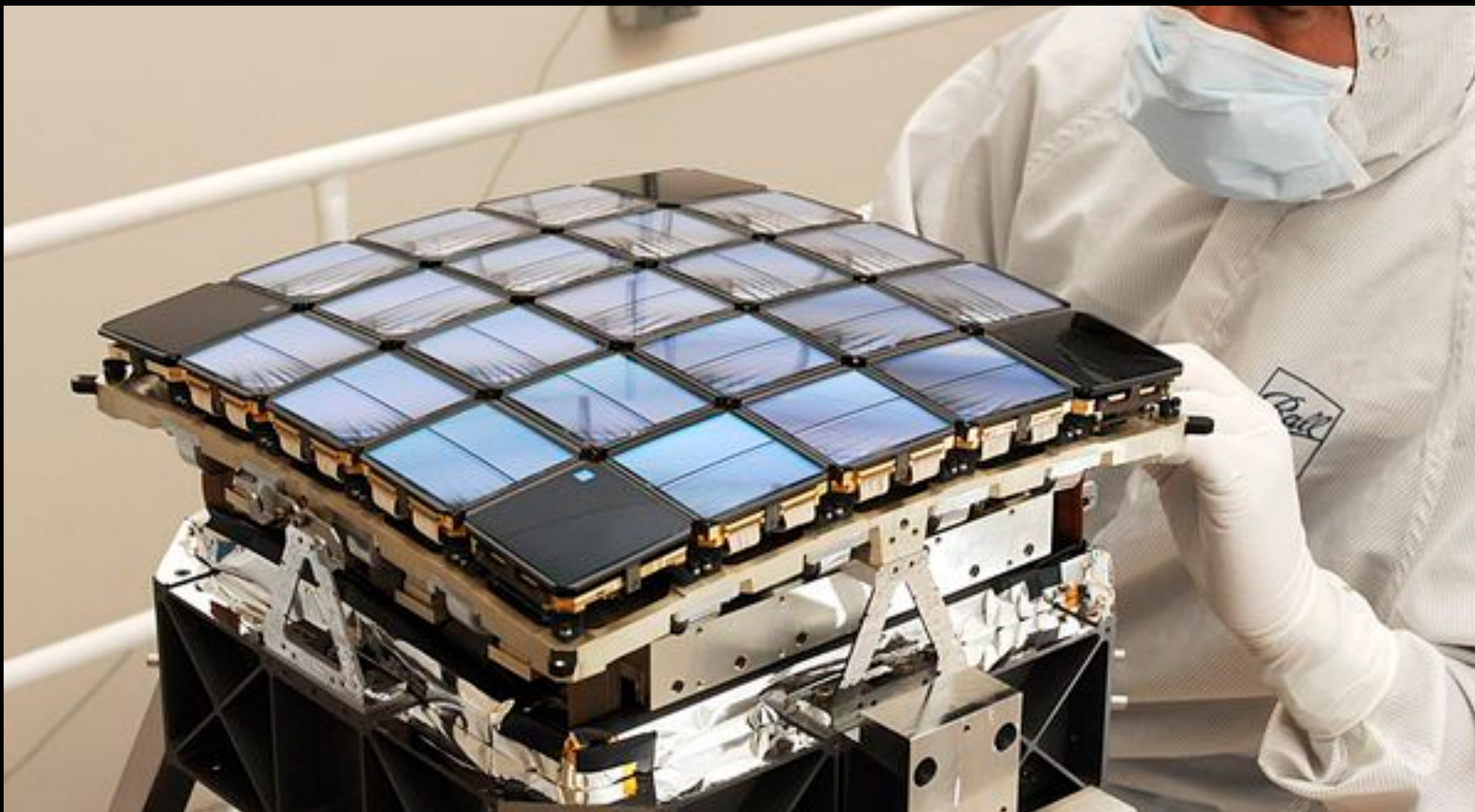
5253 Planet Transits in a Typical Week

- http://i.imgur.com/28LQQo2.gifv

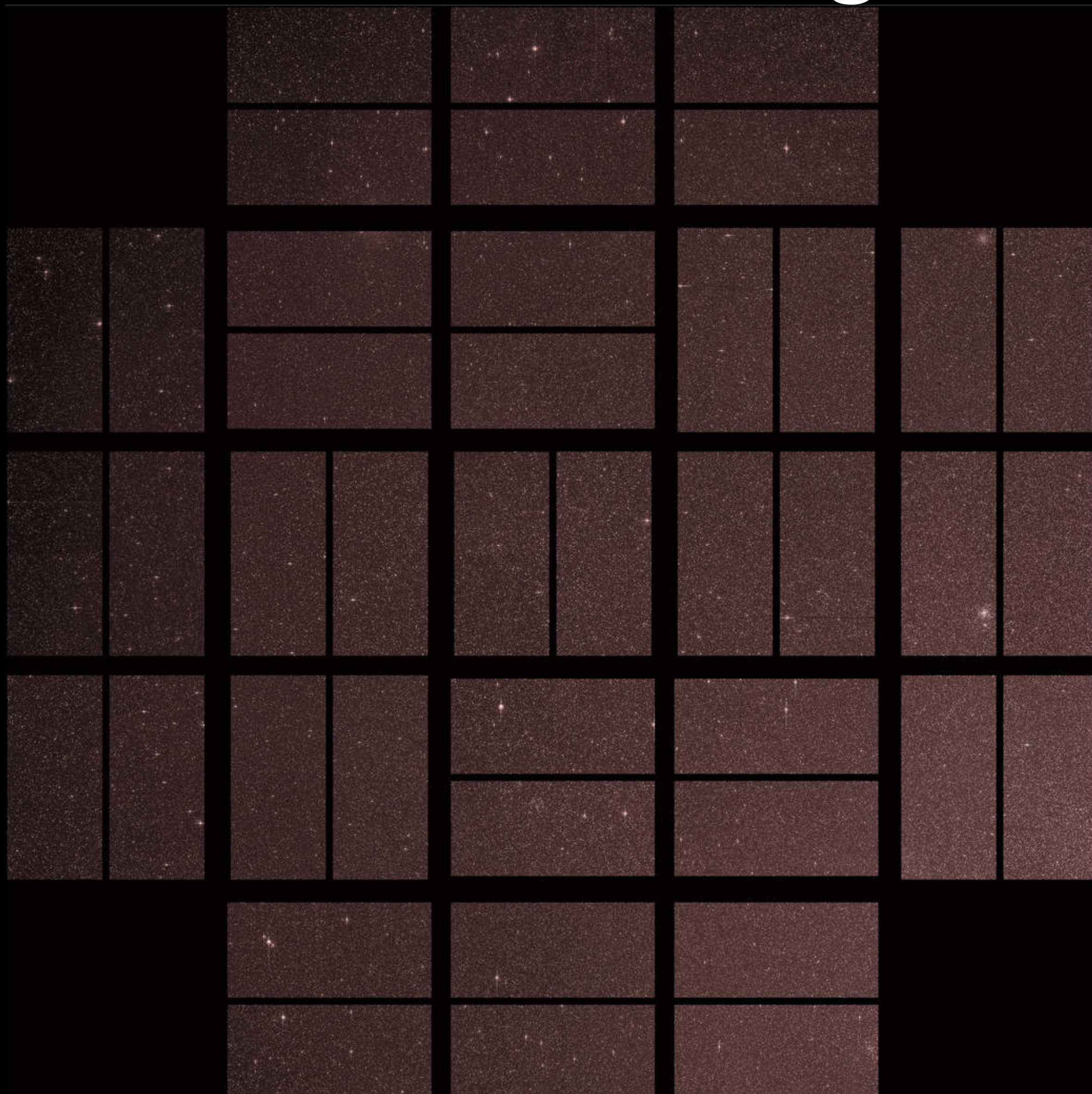
The Spacecraft



The Focal Plane

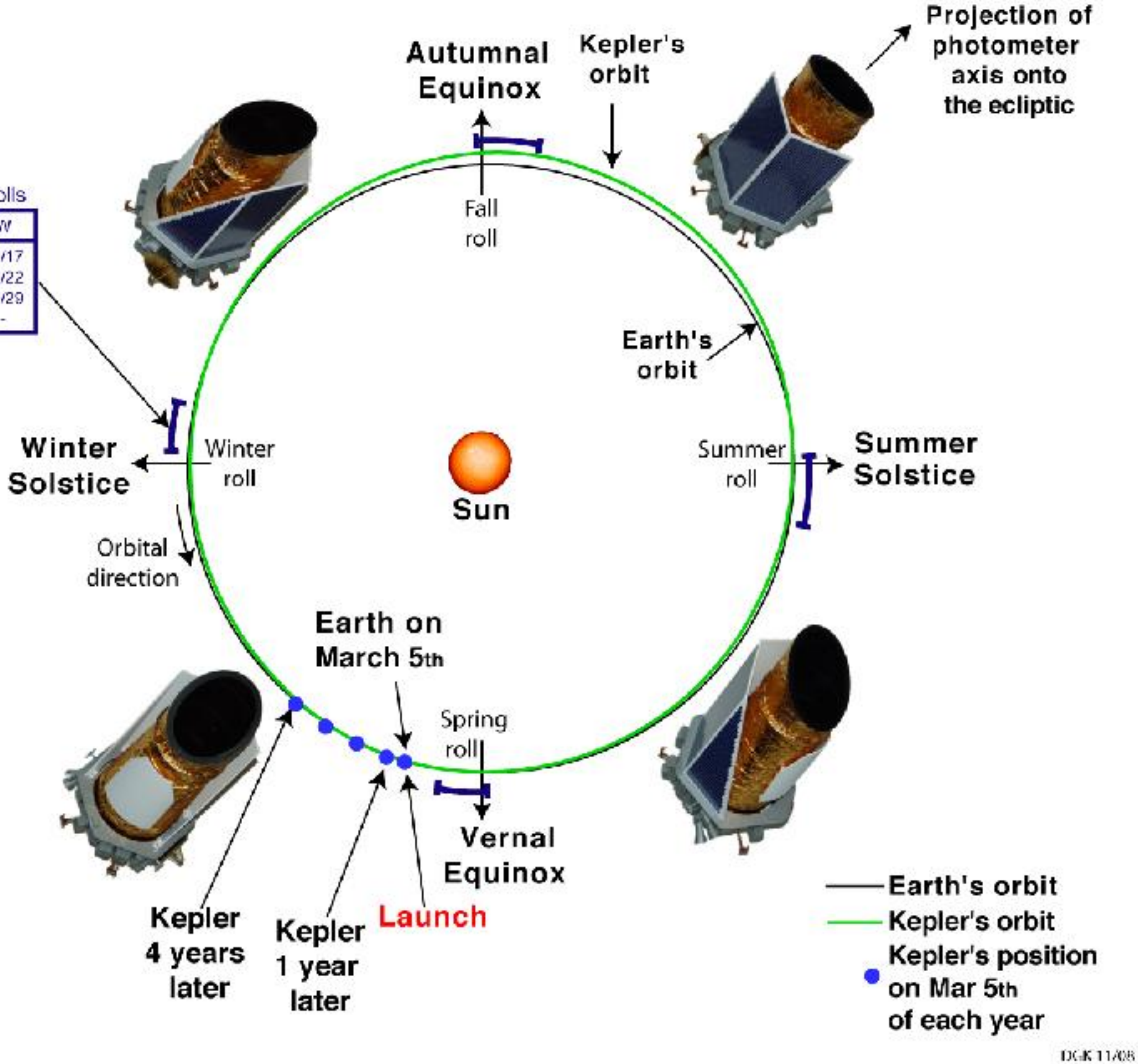


Full Frame Image



Planned Dates for Quarterly Rolls

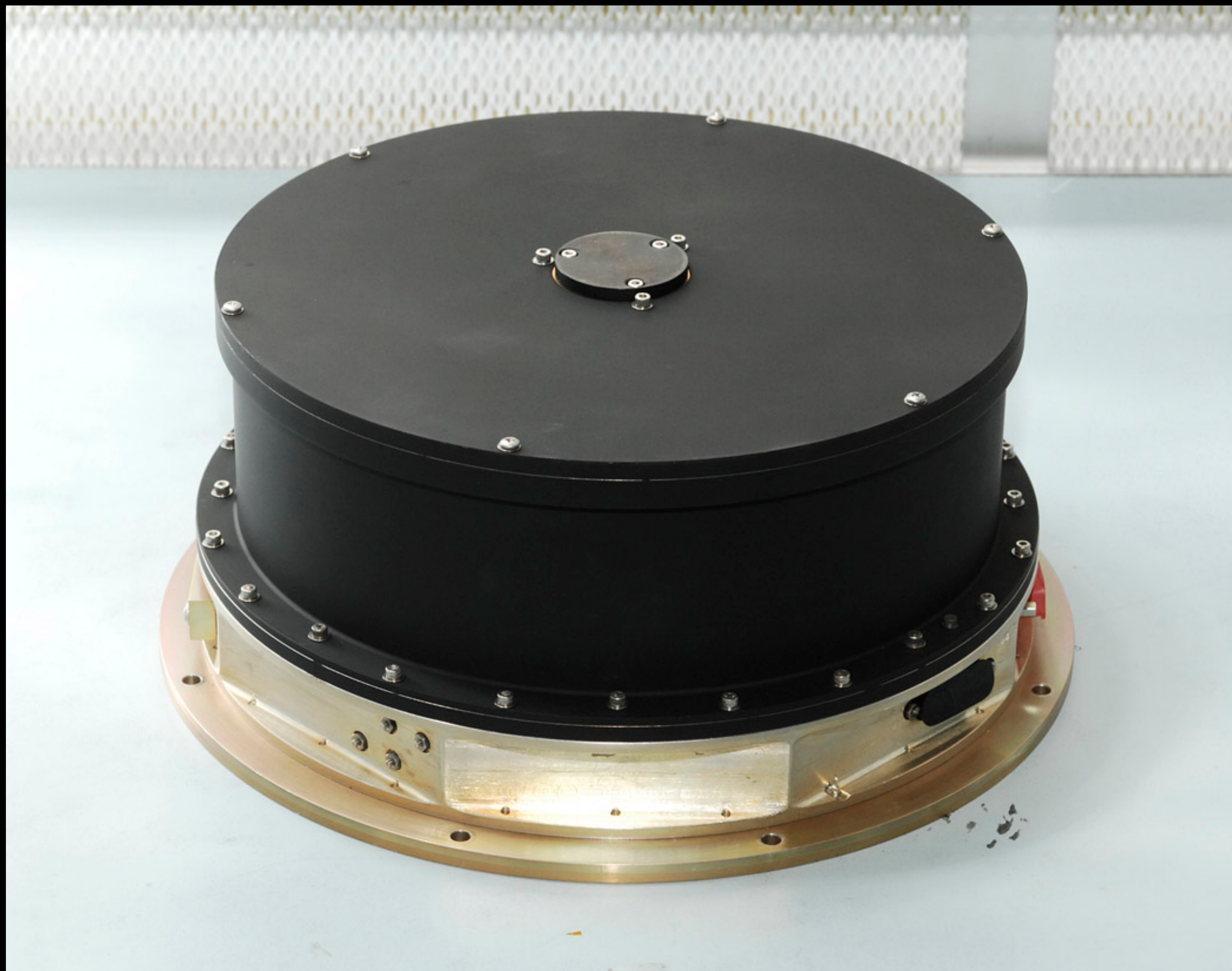
Year	Sp	Su	Fall	W
2009	3/17	6/18	9/17	12/17
2010	3/22	6/24	9/23	12/22
2011	3/28	6/27	9/29	12/29
2012	4/2	6/28	-	-



View from the ecliptic North Pole

March 6, 2009

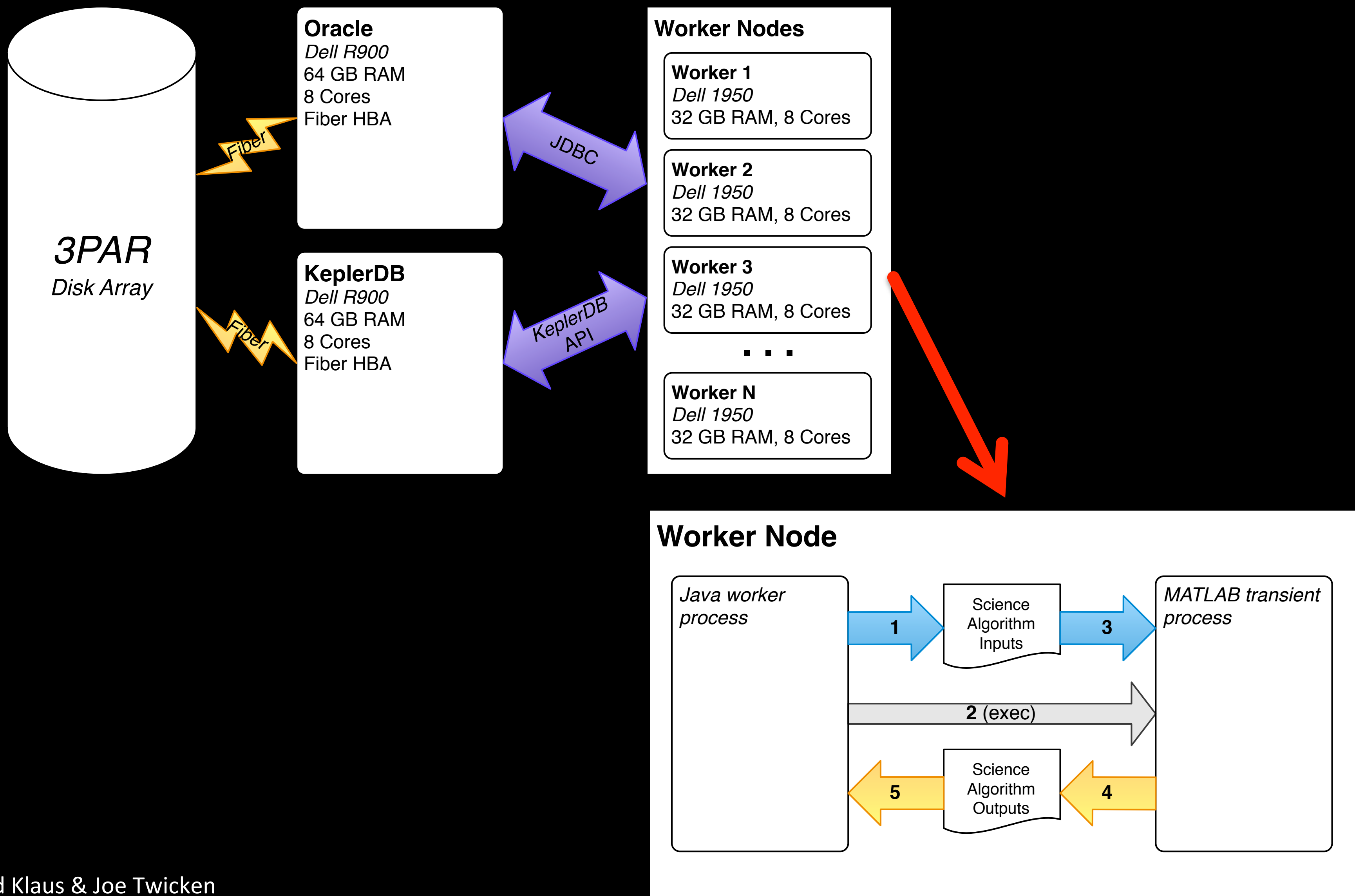
Reaction Wheel



K2 Captures Neptune and Moons

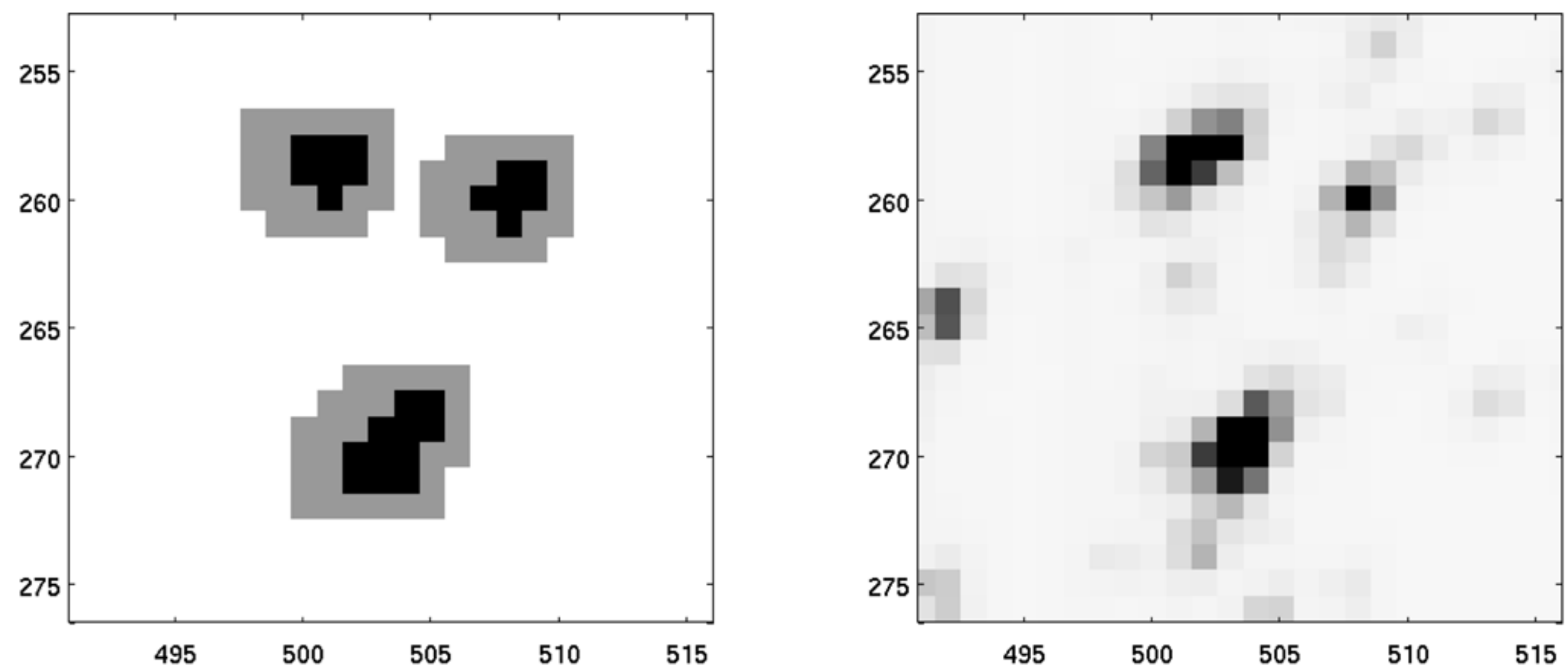
- https://youtu.be/Tw-q3uM_5_0

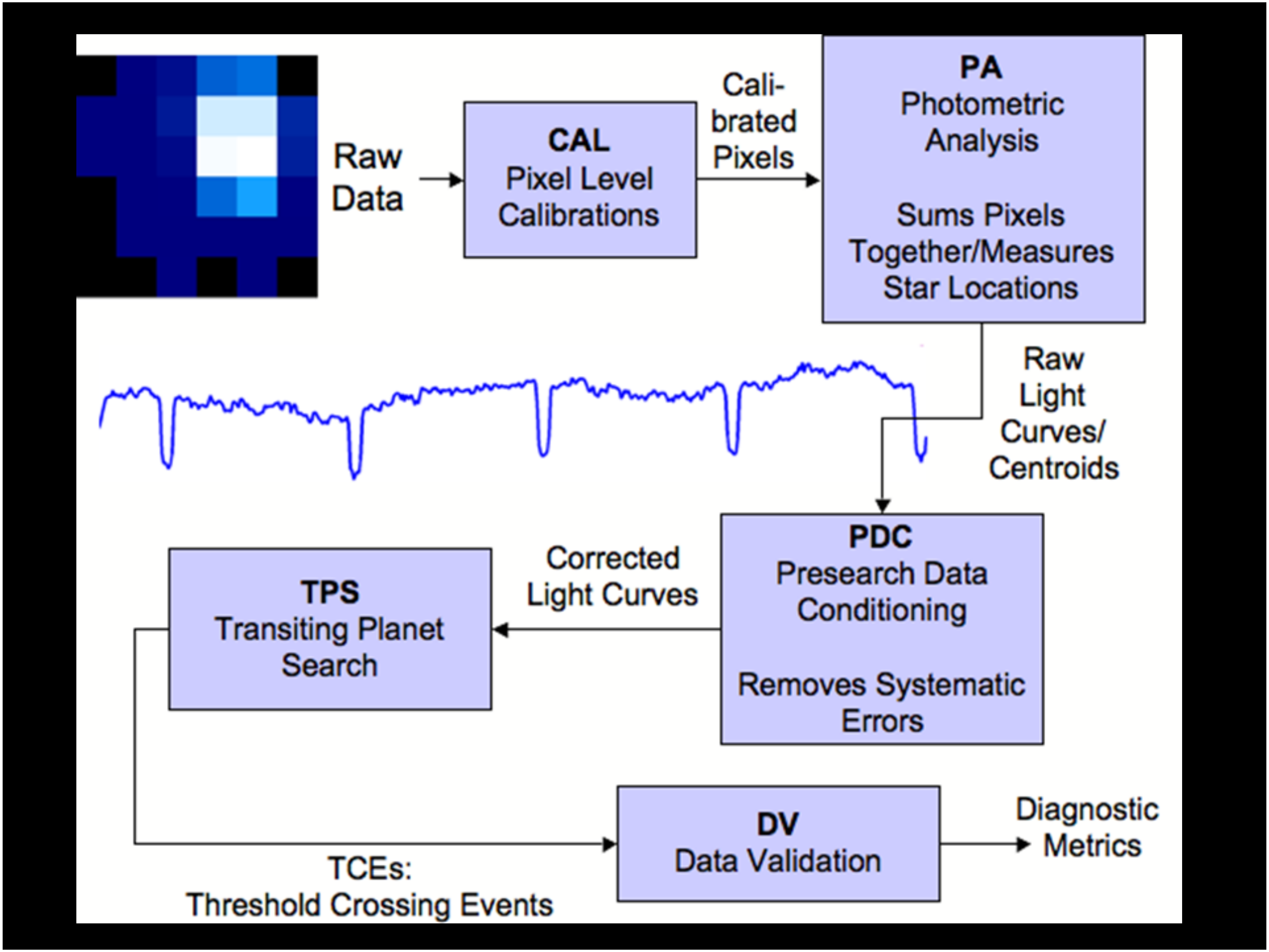
Baseline Architecture

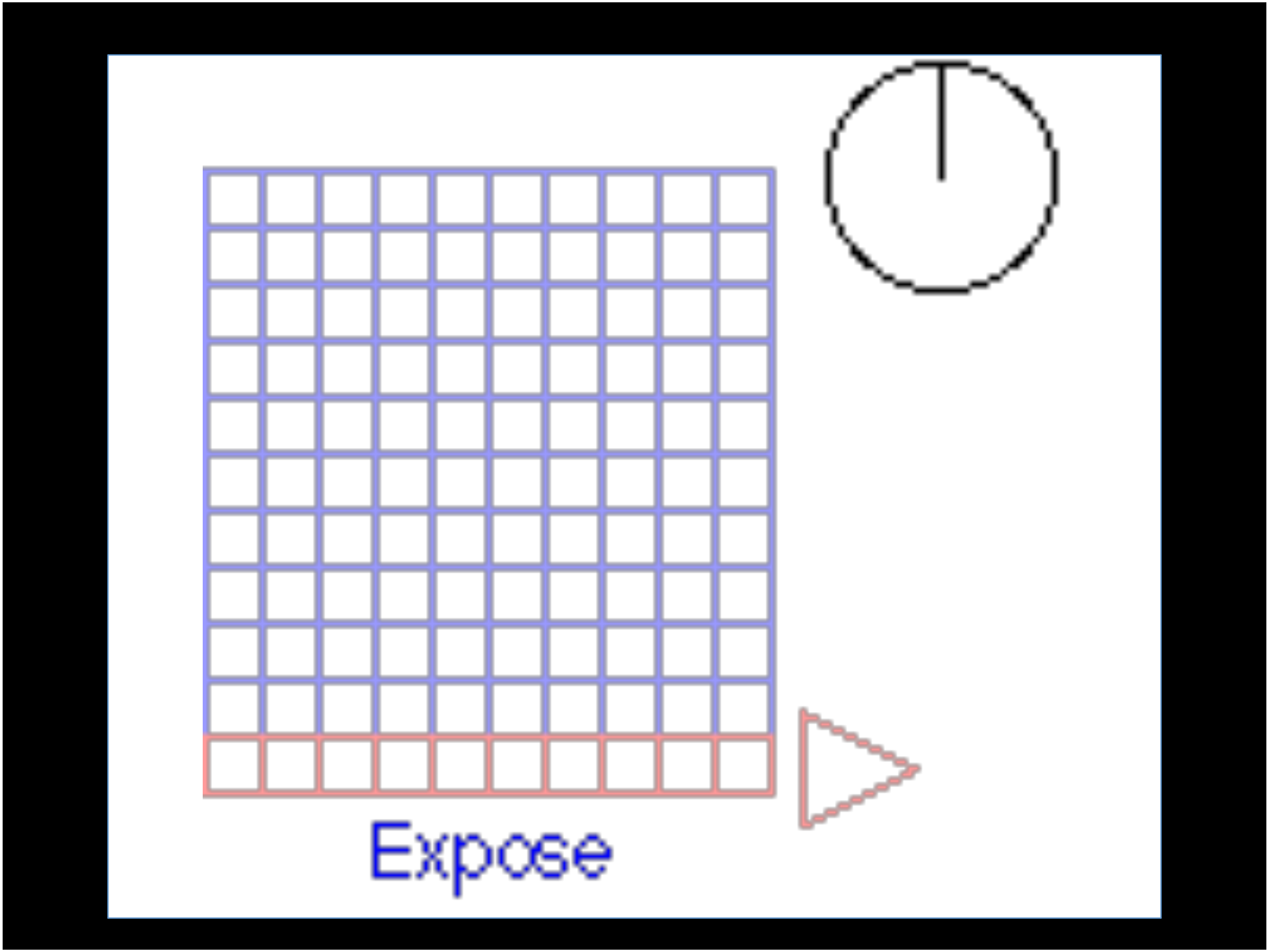


Todd Klaus & Joe Twicken

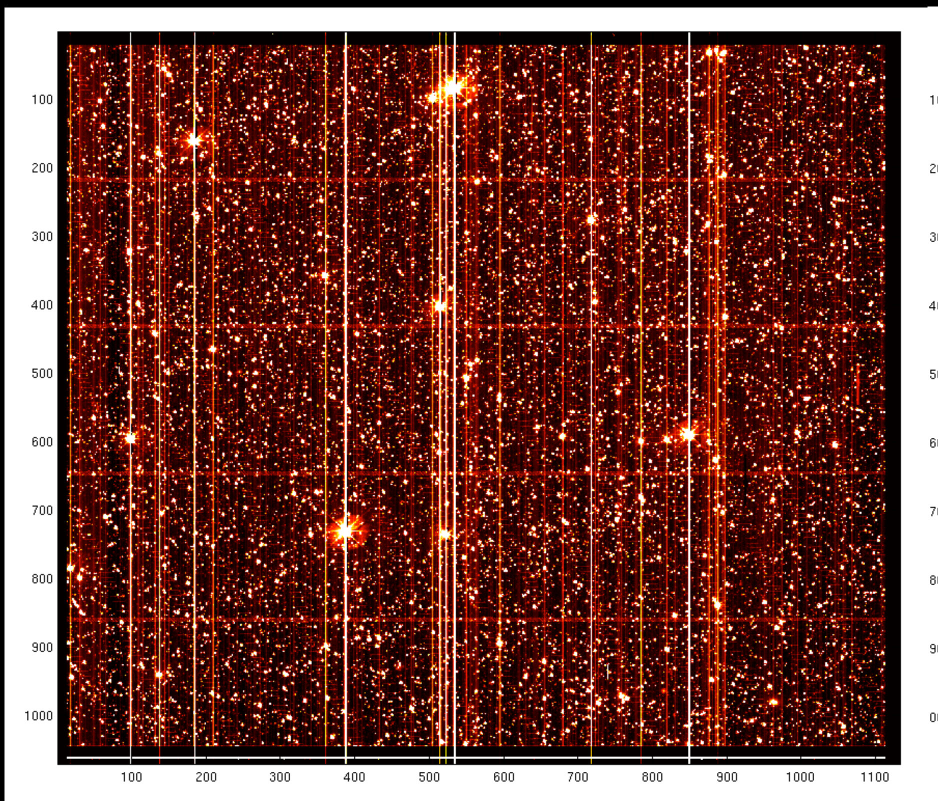
Target and Aperture Definition



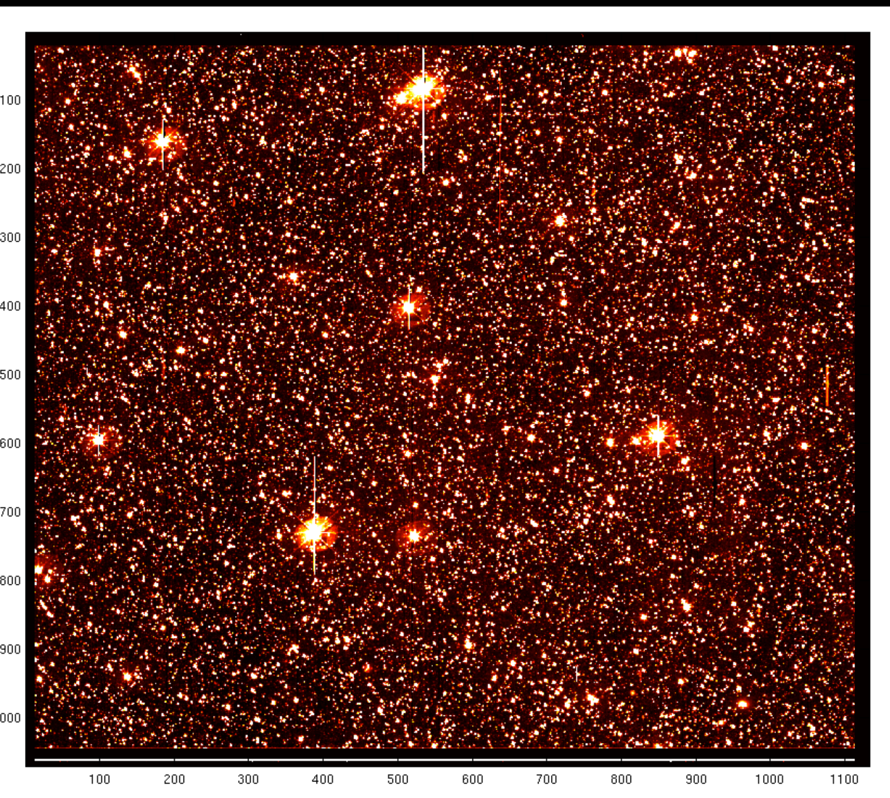

Raw Data
CAL
Pixel Level Calibrations
Cali-brated Pixels
PA
Photometric Analysis
Sums Pixels Together/Measures Star Locations
Raw Light Curves/ Centroids
PDC
Presearch Data Conditioning
Removes Systematic Errors
Corrected Light Curves
TPS
Transiting Planet Search
TCEs:
Threshold Crossing Events
DV
Data Validation
Diagnostic Metrics


Expose

CAL: Pixel Level Calibrations

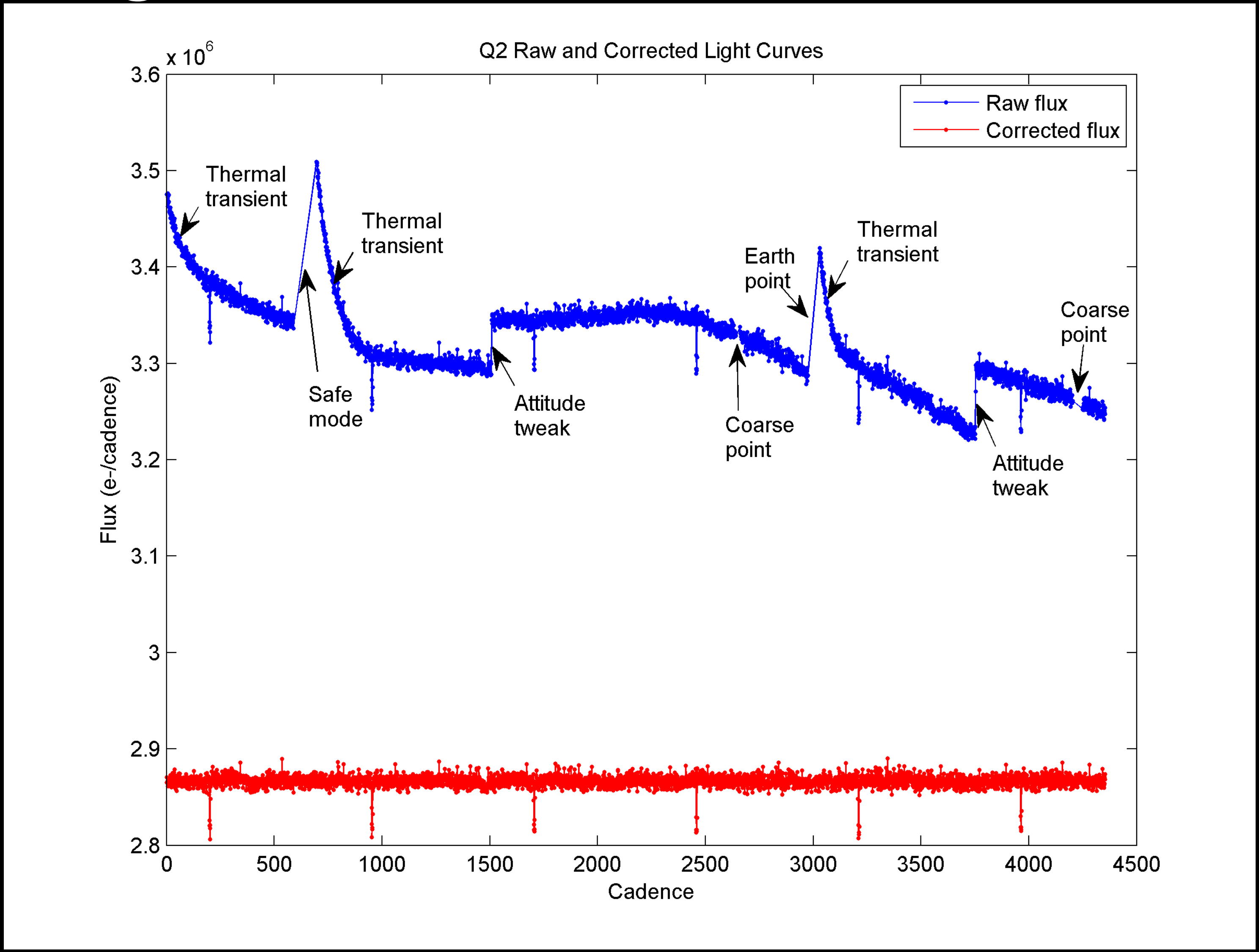


Raw FFI

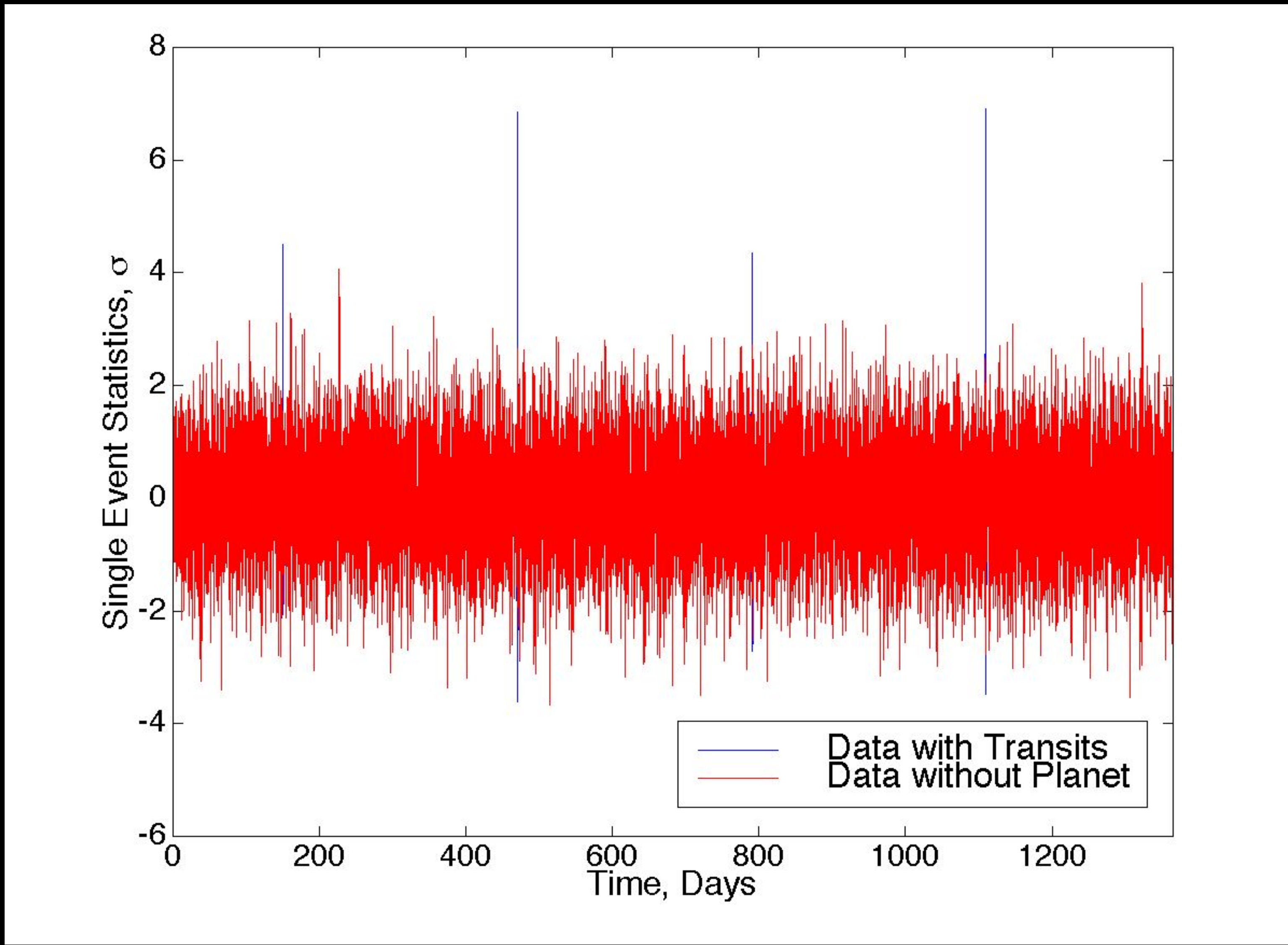


Calibrated FFI

Light Curves Also Need 'Calibration'

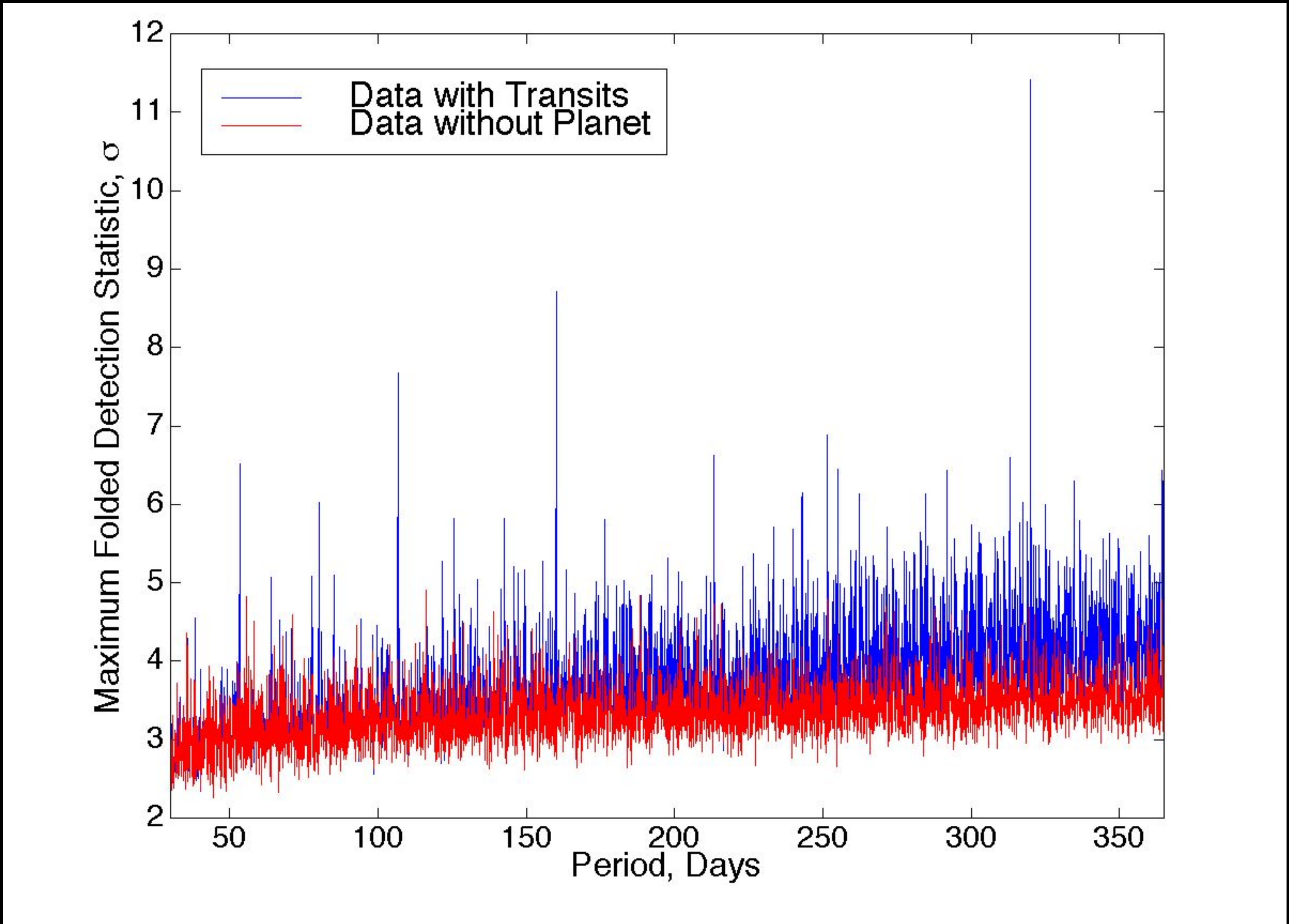


Single Transit Statistics



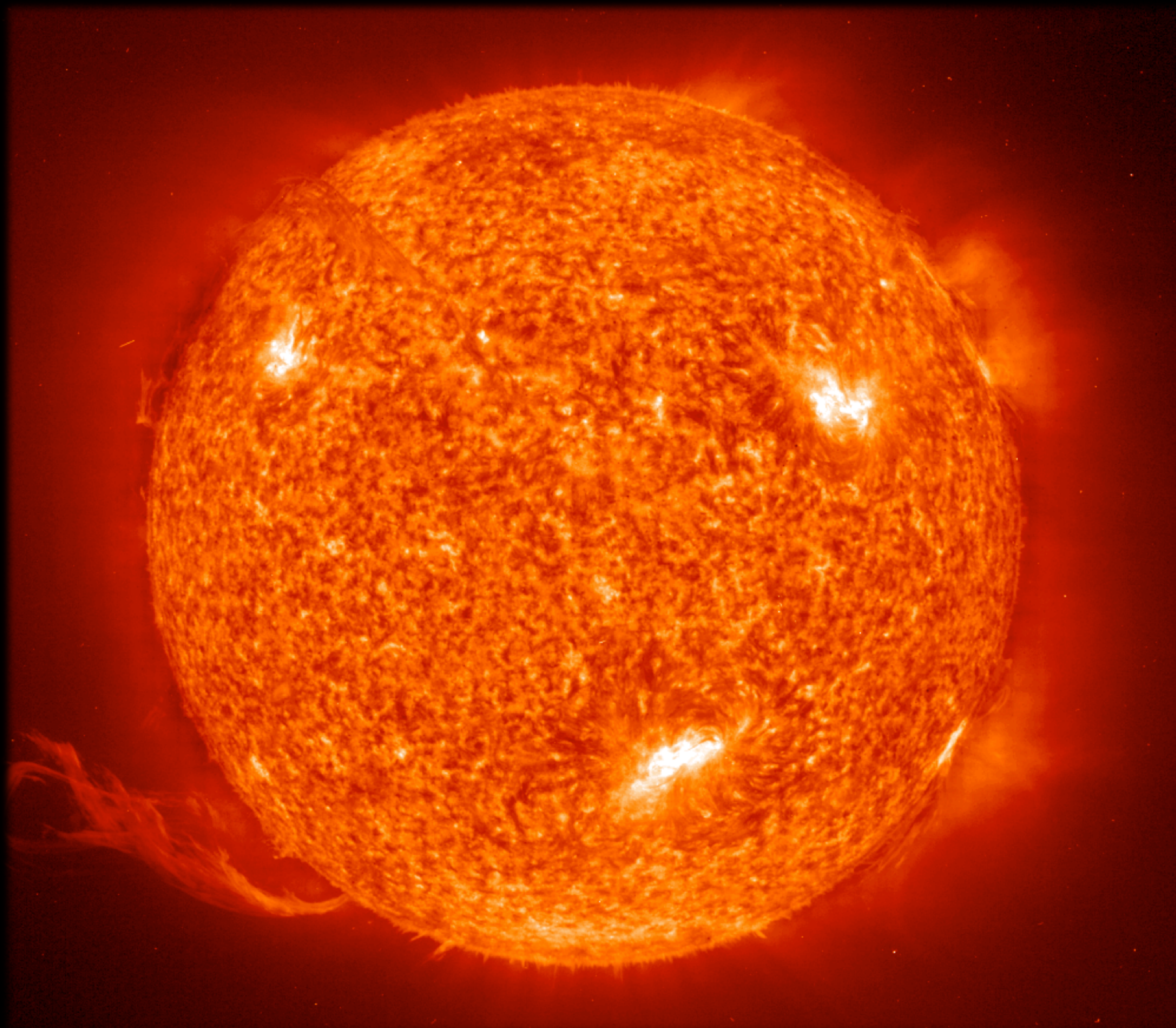
Jon Jenkins

Folded Transit Statistics



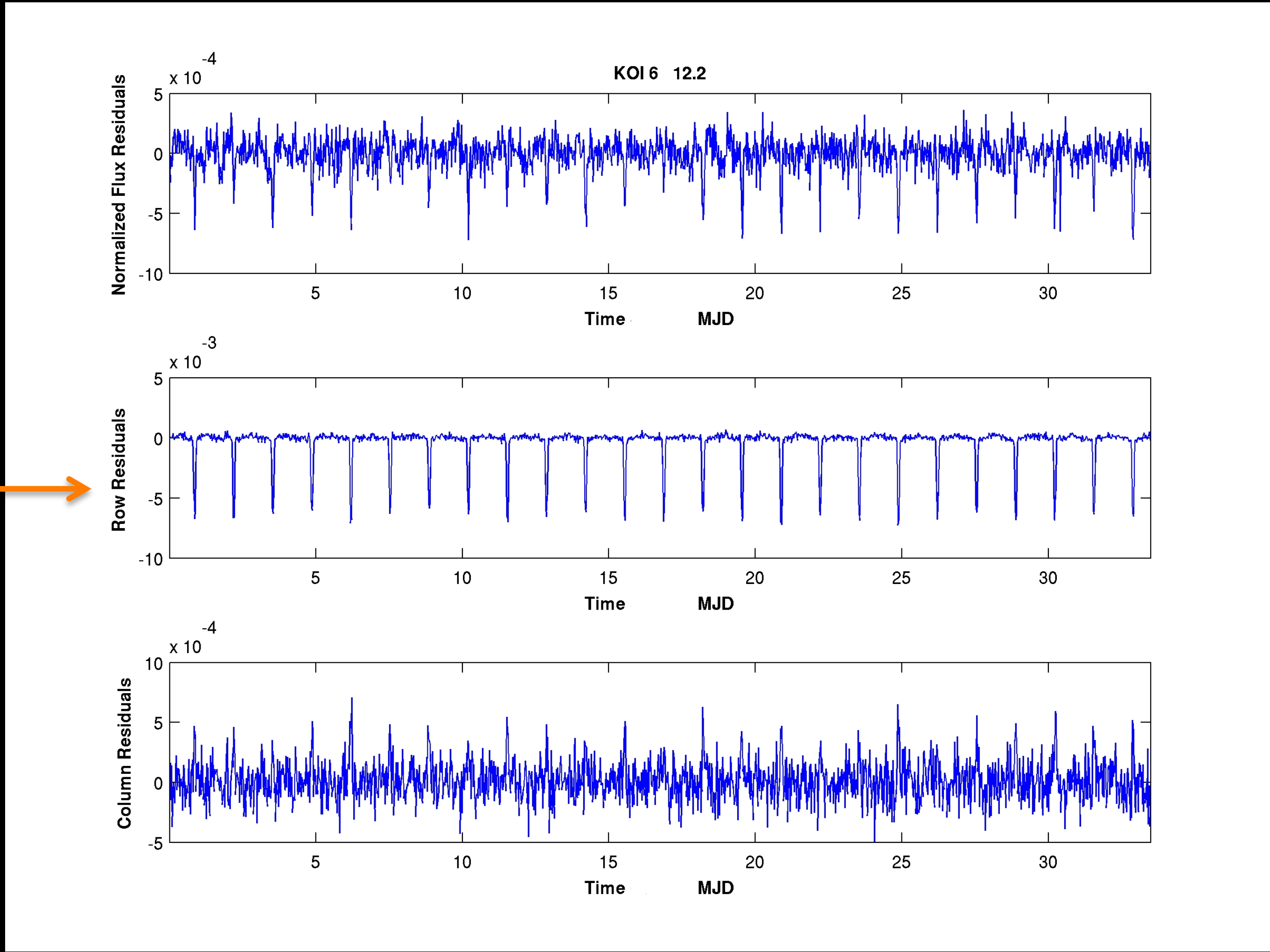
Jon Jenkins

Limb Darkening



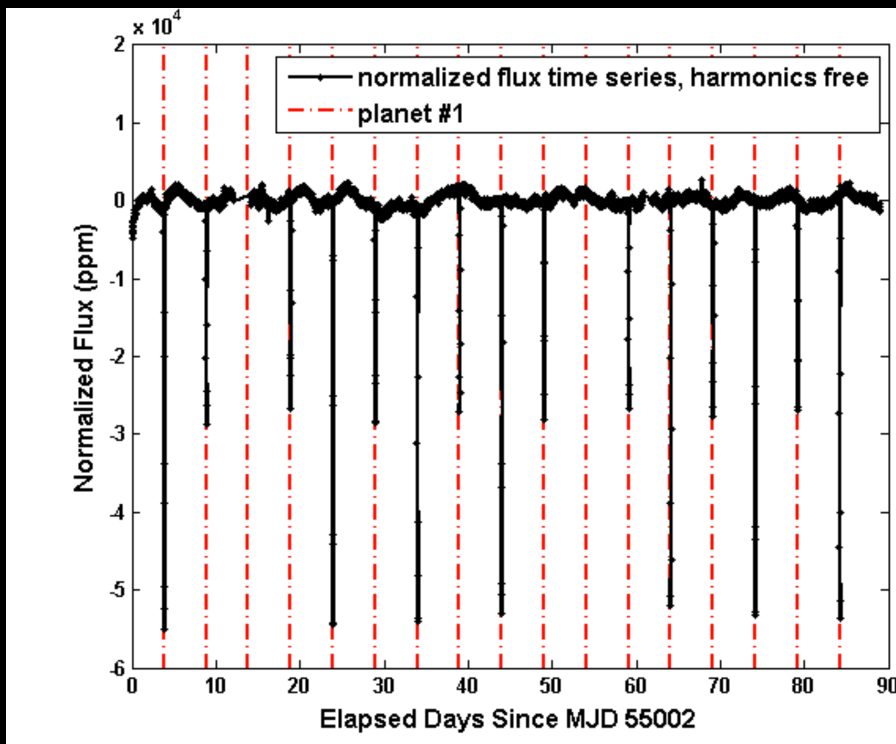
Centroid Timeseries

Row centroid shows large shifts correlated with flux

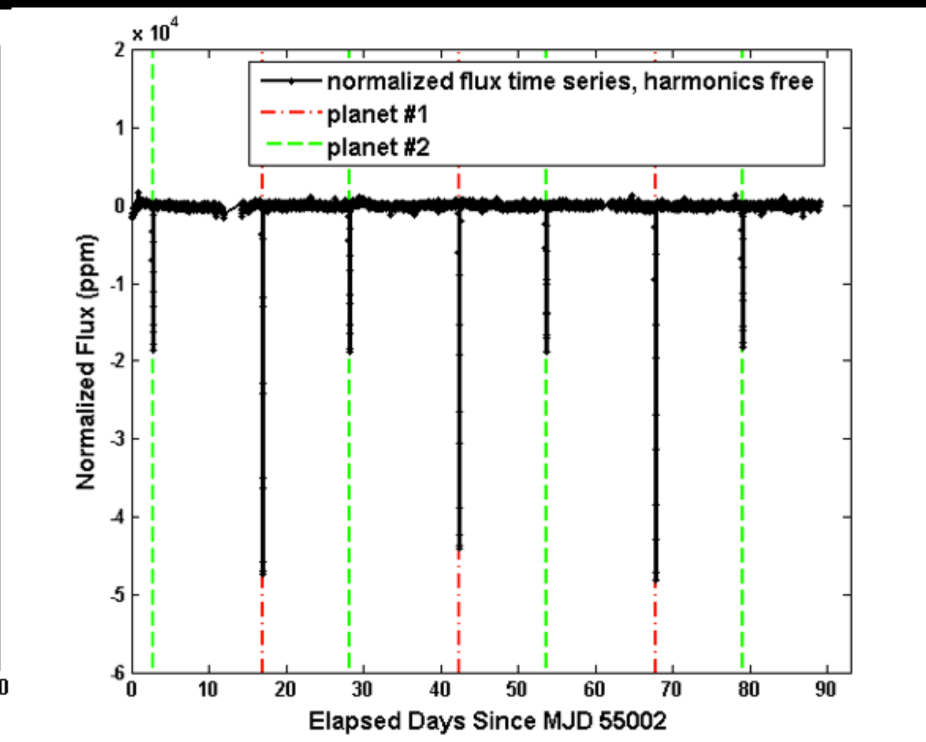


Eclipsing Binary Discrimination Test Example

Flight data, 1 TCE

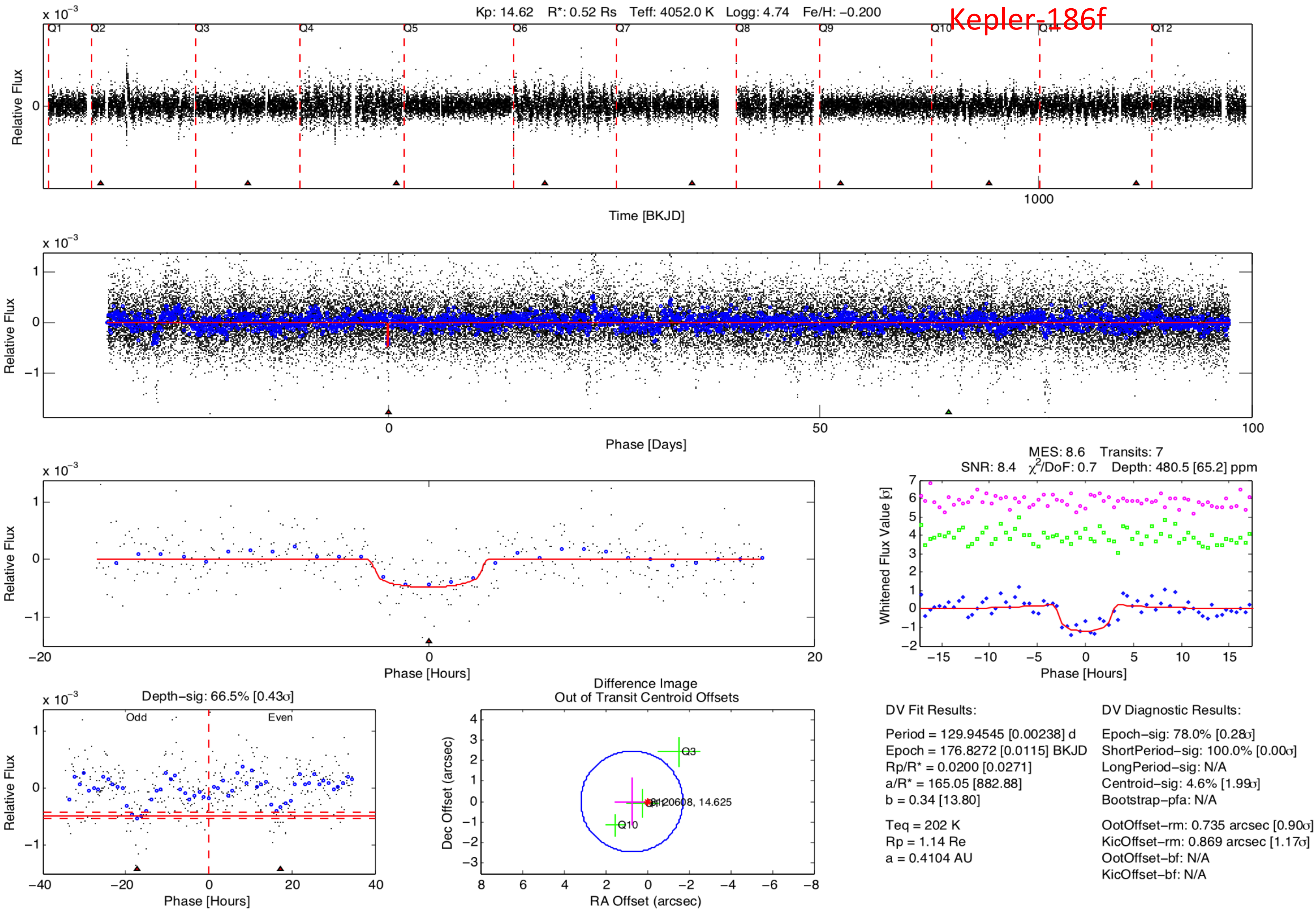


Flight data, 2 TCEs



- Odd depths ≠ even depths
- Unlikely a planet

- $P_1 = P_2$
- Unlikely a planet



Software Revision: svn+ssh://murzim/repo/soc/tags/release/8.3.1@49247 –– Date Generated: 30−Oct−2012 16:27:15 Z

30-Oct-2012

SOC Clusters

The Pleiades Supercomputer

For More Information

- http://kepler.nasa.gov
- http://www.nasa.gov/kepler/
- http://planetquest.jpl.nasa.gov/kepler/
- http://exoplanetarchive.ipac.caltech.edu
- https://archive.stsci.edu/kepler/
- http://exoplanets.org
- http://en.wikipedia.org/wiki/Exoplanet

From Pixels to Planets

Part I: Science

[SLIDE 1: From Pixels to Planets]

The search for exoplanets

- **[SLIDE 2: The Known Planets in 1994]**
- The first exoplanet about a main-sequence star confirmed in 1995
  - Before then we knew of only one planetary system
  - Theories of planet formation based on N = 1
- Various approaches
  - Direct imaging
    - **[SLIDE 3: Direct Imaging: Beta Pictoris b]**
    - Remove light from star
    - For nearest stars, big planets, far from their stars
  - Radial velocity
    - **[SLIDE 4: Planetary System Center of Mass]**
    - **[SLIDE 5: Radial Velocity]**
    - **[SLIDE 6: Doppler Measurements]**
    - Doppler effect
    - Favors large planets and small orbital radii
  - Astrometry
    - For confirmation
    - Doesn't work for discovery
  - Transit
    - **[SLIDE 7: Earth-size Planets: Detection Method]**
    - Favors short orbital periods
    - 0.5% of earth-size planets in orbits of 1 AU around sun-like stars
  - Gravitational microlensing
    - **[SLIDE 8: Gravitational Microlensing]**

The Kepler Mission

- Mission goals
  - How common are earth-size planets in the habitable zone of their stars?
    - **[SLIDE 9: The Habitable Zone]**
    - Habitable zone if surface temperature permits water in the liquid state
      - Surface temperature allows there to be liquid water on the surface
      - Depends on star temperature/size and planet's distance from star
    - Size matters
      - Too small and it can't hold an atmosphere

Too large and it's gaseous

A space telescope

[SLIDE 10: Bill Borucki]

Borucki, W.J., and A.L. Summers.

The photometric method of detecting other planetary systems.

Icarus 58, 121Ã134, 1984

In December 2001, a proposal was funded on the 5th try

Space vs. ground telescopes: an order of magnitude more precise

[SLIDE 11: HAT-P-7b: Ground vs. Space]

Atmosphere

Clouds

Dust, pollution, light pollution

Rotation of the earth

Day-night cycle

The Earth gets in the way

Uses the method of transits

[SLIDE 12: Kepler's Field of View]

Stares continuously at the same 115-square-degree patch near the Milky Way

Need to detect 10-100 ppm dimming to detect earch-sized planets

A flea crawling across a car headlight miles away

Guest Observer program

[SLIDE 13: Kepler Observations of Supernova KSN 2011b]

Telescope available to make observations for other research programs

Recent results of observing the entire life cycle of a supernova

So far Kepler found 4607 planet candidates, 1021 of which are confirmed

[SLIDE 14: Exoplanet Discoveries]

[SLIDE 15: Sizes of Kepler Candidates]

Mostly between the sizes of Earth and Neptune

[SLIDE 16: NASA Kepler's Hall of Fame]

More than 100 candidates in the habitable zone

[SLIDE 17: 5253 Planet Transits in a Single Week]

Kepler discovered more than half of the confirmed exoplanets to date

Very few Kepler planet candidates are rejected

The satellite and what it does

[SLIDE 18: The Spacecraft]

[SLIDE 19: The Focal Plane]

[SLIDE 20: Full Frame Image]

A space telescope with 42 CCDs on the focal plane

[SLIDE 21: Earth-trailing Orbit]

Orbits the sun trailing the Earth to

Assure thermal stability

Avoid occlusion
Avoid gravitational effect
[SLIDE 22: Launch]
Launched March 6, 2009
Begain science operations May 12, 2009
Designed to last 8 years
Funded for 3.5 years
Extended in November, 2012
Long enough to observe 3 transits with an orbital period of one earth year
[SLIDE 23: Reaction Wheel]
First reaction wheel failed July, 2012 and the second May, 2013
Can no longer point to its region with sufficient precision
Roll controlled with thrusters
Refinements in the software mean more discoveries with the same data
[SLIDE 24: K2 Captures Neptune and Moons]
Repurposed as K2 in June, 2014
A series of 80-day campaigns, each staring at a region in the ecliptic plane
Entirely guest observers, though some are at Ames
Onboard fuel expected to last until December, 2017

Scientific legacy of the Kepler mission

Resoundingly vindicates the transit method
Thought to be an average of one planet per star
Planets in the habitable zone are common
Hot Jupiters
A gas giant can have very small orbital radius
Less than 1%
Planet orbits don't have to be near-circular
Planet orbits don't have to be perpendicular to the star's axis of rotation
Planetary systems tend to be coplanar
There are planets between Earth and Neptune in size
Most commonly observed
Absent from our Solar System
Characterize stellar variability
Made asteroseismology studies of 1000 targets

Future missions

TESS: Transiting Exoplanet Survey Satellite
NASA and MIT
Scheduled to launch in 2017
To search for transits, as Kepler did
Four telescopes

To scan the entire sky, not just a 115-square-degree field
To concentrate on targets closer and brighter than Kepler's
PLATO: Planetary Transits and Oscillations of Stars
ESA
Scheduled to launch in 2024

Part II: Data Analysis

What happens on the spacecraft

Save aperture pixels for about about 170,000 pre-selected target stars
Because the probability of observing a transit is 0.5%
When a photon hits a pixel, its electron count increments
Light collected in 6-second "observations" and then shifted out over 0.5 sec
270 observations are summed over 29.4-minute "cadences" for dim stars
Precise enough to capture transits of 1-24 hr
1-minute short cadences of bright stars for asteroseismology
Requantized, compressed science data are downlinked once per month
The telescope rolled every 3 months to keep solar panels facing the sun

A peek behind the scenes at the software

What it takes to find planets in pixels
Several categories of software
On-board the satellite
Ground station for satellite communication and control
Importing data into the database and file store
Simulations for algorithm development and testing
Science pipeline
Vetting

Why a pipeline?

Can't see transits by eye at the pixel level
Data reduction pipelines are standard in astronomy
Modules for discrete phases of data analysis
Different expertise required for each module
Each module can require days to weeks for processing all inputs
Different modules can be run on different platforms
The results of each pipeline module have to be saved anyway

Interaction of Java and MATLAB

[SLIDE 25: Baseline Architecture]
One million lines of code, half Java and half MATLAB
A Java GUI console launches a pipeline

- The Java pipeline infrastructure creates tasks and deploys them to workers
- Each module has a Java process
  - Marshals the configuration parameters
  - Fetches module inputs from persistent storage
- The Java process then invokes the MATLAB portion of the module
- The MATLAB program places its outputs in the file system
- The Java process that launched resumes and collects and persists the outputs
- The Java Pipeline Infrastructure also maintains an audit trail
  - What version of the software was run
  - On what version of the data
  - By which operator
  - On which platform
  - With which configuration

Pre-Pipeline

- Target selection: which stars are worth observing
  - Broad selection in size, brightness, age, and temperature
  - Separated from other target stars
  - Minimize crowding
  - Distributed across the field of observation
  - Have known properties (Kepler Input Catalog)
- **[SLIDE 26: Target and Aperture Definition]**
- Aperture definition: which pixels are needed to observe each target star
  - A menu of 1,024 apertures
  - Padded for safety

The Pipeline modules

[SLIDE 27: Kepler Science Processing Pipeline]

(1) Pixel Calibration

- **[SLIDE 28: CCD Readout]**
- **[SLIDE 29: CAL: Pixel-Level Calibrations]**
- Remove instrument-induced artifacts
- Example: Smear caused by not having a shutter
  - Each pixel in a column receives light from all the stars that affect that column
  - Appears as a vertical line
  - Measure smear with a shielded pixel
  - Subtract the measured smear from the signal
- All pixels are treated equally

(2) Photometric Analysis

- Start with calibrated pixel light curves for up to 180,000 apertures
  - Produce raw flux light curves for each target

  Produce target photocenters (centroids) for each cadence
Identify and remove cosmic rays from target and background flux level
  About 3 hits per pixel per day
  Detect outliers and replace with median value
Clean up the background stars
  From background pixels, generate a polynomial for each cadence
  For each target pixel, subtract its value of the polynomial
Sum up the values for all pixels in the aperture
  The units are electrons per cadence
Calculate the centroid for each cadence
  Flux weighted sum or Pixel Response Function
  Called Simple Aperture Photometry
Put time stamps on the cadences
  Orbit of the spacecraft distorts transit times and orbital periods
    4 minutes because of 60-degree angle
  Adjust to when light would reach the solar system center of mass

(3) Pre-Search Data Conditioning

[SLIDE 30: Light curves also need 'calibration']

Start with raw flux light curves for each target for each quarter
Produce a noise-reduced "conditioned" light curve for later stages`
Identify and correct random flux discontinuities
  Discontinuities caused by cosmic rays
    1% abrupt drop in flux
    Partial exponential recovery
    Permanent damage to CCD pixels
Remove systematic effects
  Account for anomalies
    Monthly earth pointing for downlink
    Commanded attitude adjustment
    Quarterly roll
    Occasional safe mode
  Signatures correlated with ancillary engineering data
    E.g., temperature changes alter focus
  Astrophysical events identified and replaced with interpolated values
    Large transits (e.g., brown dwarfs)
    Flares
    Eclipses
    Microlensing events
Remove the effect of crowding
  Dilutes the transit and underestimates planet radius
  The crowding metric is the proportion of flux due to the target

The amount of crowding is determined along with the aperture

Condition for Transiting Planet Search

Remove outliers

Fill data gaps with interpolated values

This is the last of the "front end"

(4) Transiting Planet Search

This is the first of the "back end"

The first time that all quarters are examined at once

Stitch together the quarters

Every quarter, the module/output for a target changes

Minimize edge effects

Variable stars

Remove low-frequency sinusoidal components

Remove low-order polynomial effects

The mean of the relative flux time series becomes zero

A negative relative flux suggests a transit

Maximize uniformity of transit depths

[SLIDE 31: Single Transit Statistics]

Search transit durations and transit periods for evidence of transits

Statistic called Combined Differential Photometric Precision

For a single event

[SLIDE 32: Folded Transit Statistics]

Then fold the light curve

Yields multiple-event statistic

If the statistic passes 7.1-sigma cutoff, it is a Threshold Crossing Event

(5) Data Validation

Works on Threshold Crossing Events generated by TPS

Reduces false alarms

Fit a mathematical model to the light curve

Seven Parameters needed to describe a star-plus-planet system

1. Star radius (known)
2. Eccentricity (insufficient time resolution; assumed circular)
3. Epoch (time of the first transit)
4. Orbital period
5. Impact parameter
6. Ratio of planet radius to star radius
7. Ratio of semi-major axis to star radius

Transit depth and transit duration derived from the model

[SLIDE 33: Limb Darkening]

Takes "limb darkening" into account

Report if model-fitting fails
Determine if it is happening to a background star
[SLIDE 34: Centroid Timeseries]
Centroid tests
The centroid moves at the same time as the TCE
The centroid of the difference image moves
[SLIDE 35: Eclipsing Binary Discrimination Test Example]
Matches the signature of an eclipsing binary system
Odd and even transits analyzed separately
Reject if statistically significantly different transit depths
Reject if statistically significantly different epochs
Works with one TCE or two TCEs
If it survives
Subtract model from light curve and repeat TPS on the residual
Repeat until no more TCEs are found or a limit is reached
Generate a report for each target that has TCEs
[SLIDE 36: DV Report Page]
Target data from Kepler Input Catalog
Each planet candidate
Measurements and statistics generated from search
Inferred characteristics
Planet radius
Orbital radius from Kepler's 3rd law
Surface temperature

Export

Outputs archived at multiple stages of processing
Raw and calibrated pixel light curves
Target flux light curves at varying degrees of processing
Folded light curves
Centroid time series
Allows others to analyze the same data

Vetting

TCE Review Team (TCERT)
Determines whether a TCE is a Planet Candidate
Automated vetting
Robo-Vetter manually captures vetting expertise
Auto-Vetter uses machine learning to capture human expertise

Confirmation that a Kepler Planetary Candidate is actually a planet

Use converging sources of evidence outside of Kepler

    - Radial velocity
    - Ground-based imaging (e.g., Keck in Hawaii)
    - Space telescopes (Hubble, James Webb)
- Follow-On Program

Use of computing resources

- The processing is batch, not real time
- A group of 4 operators launches the pipelines
- The unit of work depends on the module
    - Logical constraints
        - Per pixel (CAL) vs. per target (everything else)
        - Multiple quarters (only TSP & DV) vs. single quarter (everything else)
    - Unit of work depends on other considerations
        - CAL: by module/output, for any number of quarters, binned by cadence
        - PA + PDC: by module/output, for any number of quarters
        - TPS + DV: by sky group, for all quarters, binned by target
    - An autonomic system might determine unit of work dynamically
        - Logical requirements
        - Optimize use of resources
        - Optimize quality of results

[SLIDE 37: SOC Clusters]

- The Science Operations Center has 6 dedicated clusters
    - Initially each had 16 workers of 8 cores each, for 144 cores per cluster
        - Clusters are not evenly sized in workers, memory, disk
    - One RDBMS server, one Time Series server, the rest for pipeline tasks
    - They are configured differently for different uses
        - Flight segment
            - Revise the target and aperture definitions quarterly
            - Plan the quarterly roll
            - Idle most of the time
        - Quarterly archive
            - Create data for export about one month in three
            - When idle, used for TPS and DV or reprocessing with newer SW version
        - Preliminary processing of monthly downlink
            - Reduced cluster size
            - Tune parameters
            - Photometer performance assessment
            - Idle one week of four
        - Test clusters
            - Integration runs of one module at a time (except TPS & DV)
            - Full pipeline for validation and verification
            - Use a snapshot of a production cluster

- Can't fully exploit parallelism
  - Cluster memory and disk needs limit the number of simultaneous threads

[SLIDE 38: The Pleiades Supercomputer]

When processing needs overwhelms the clusters, use the Pleiades supercomputer

- Why use the NAS
  - A run may take weeks to complete on the clusters
    - 21 GB or raw pixel data per month for 4 years
    - LC CAL takes 35-45 days Q0-Q17 in serial on cluster, 11 days on NAS
  - The clusters may be needed for jobs that can't wait
- The most demanding modules were ported to run on Pleiades
  - First TPS and DV
  - Later CAL
- Difficulties could negate the value of using the NAS
  - First have to transfer all the data that the pipeline needs
  - Sometimes the data get corrupted
  - Sometimes tasks wait on queue for their preconditions to be satisified
  - Sometimes tasks exceed their processing quotas and are terminated

We welcome ideas for introducing autonomicity to our science processing

[SLIDE 39: For More Information]