The *Kepler Mission:* Good Planets are Hard to Find and Vice Versa

Jon M. Jenkins
SETI Institute/NASA Ames Research Center

Wednesday May 23, 2012

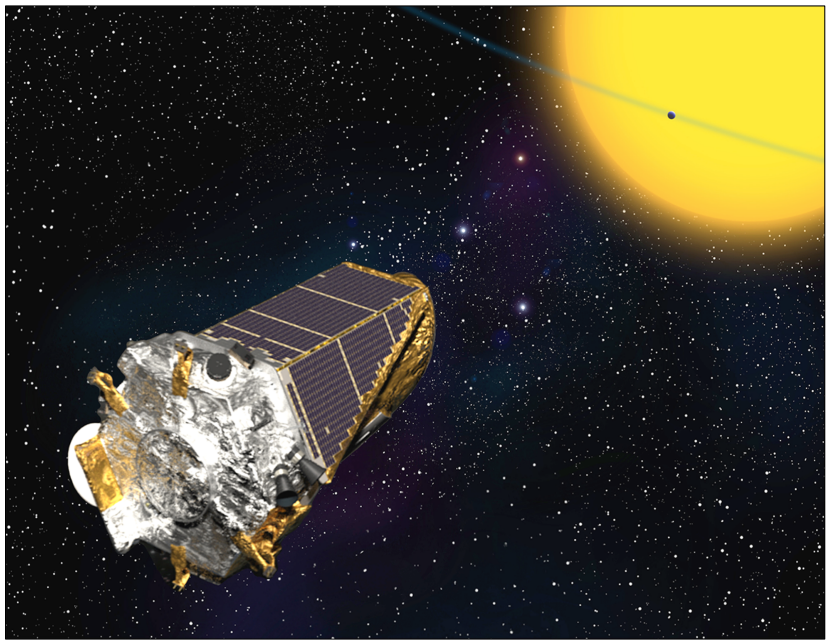
16th Annual CASIS
Signal and Imaging Sciences Workshop

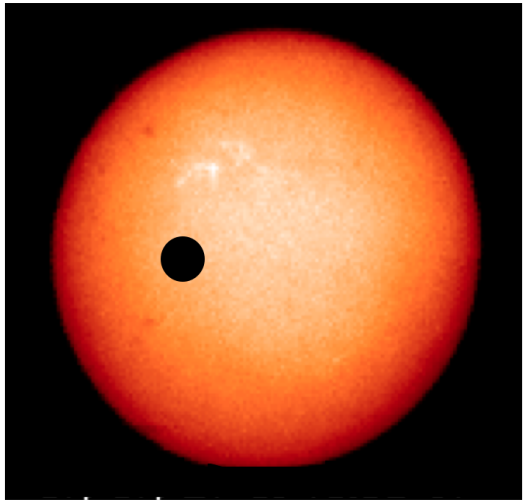
Ames Research Center
Ball
JPL
LASP
LHS
SAO
SETI Institute
STScI

Overview

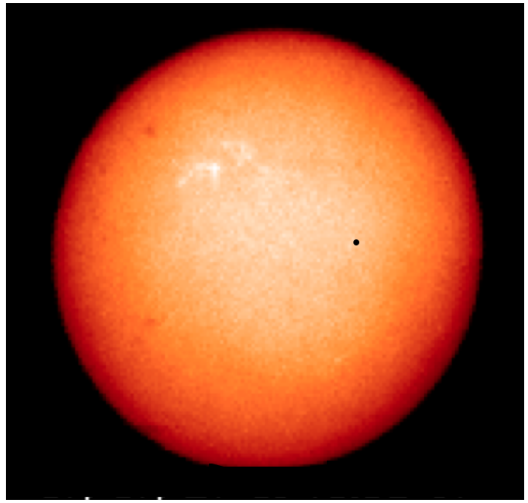
Kepler
A Search for Earth-size Planets

- Key Challenges
- The Kepler Mission
- Key Science Results
- Correcting Systematic Errors
- Searching for Transits
- Summary





Jupiter:
1% area of the Sun (1/100)



Earth or Venus
0.01% area of the Sun (1/10,000)

Critical Questions:

Kepler
A Search for Earth-size Planets

- Are terrestrial planets common or rare?
- What are their sizes & orbital distances?
- How often are they in the habitable zone?
- What is their dependence on stellar properties?

Bottom Line: Are We Alone?

Key Challenges

Key Technical Challenges

Kepler
A Search for Earth-size Planets

1. Terrestrial planets are really small ~100 ppm deep signatures

⇒ 20 ppm photometric precision

2. Transits occur on timescales of one hour to ~ ½ day

⇒ Sample at ½ hour cadence

3. Transiting planets are rare (geometric probability of alignment is ~0.5% for Earth analogs)

⇒ Observe ~100,000 stars

4. We don't know when a transit will occur

⇒ Stare at one FOV for 3.5-10 years and don't blink

The *Kepler Mission*

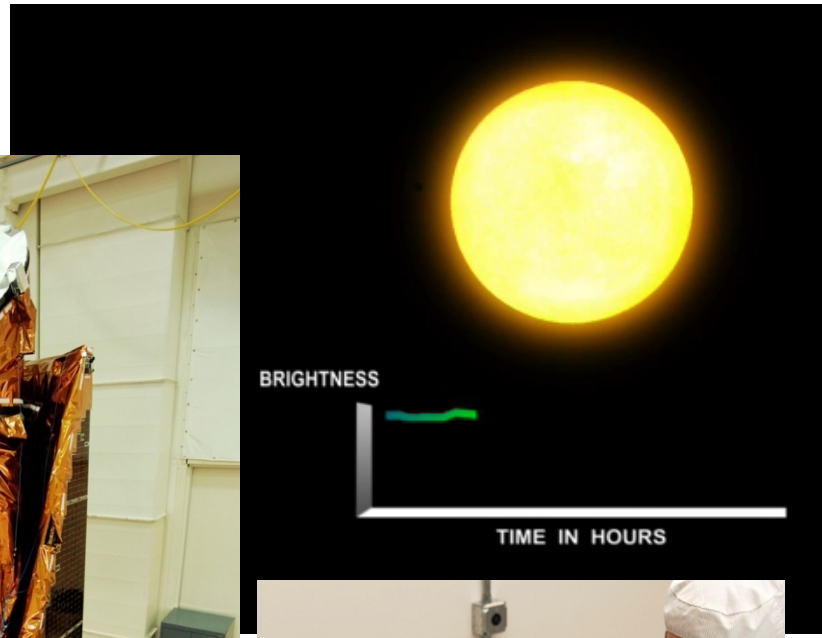
The Solution: The Kepler Mission

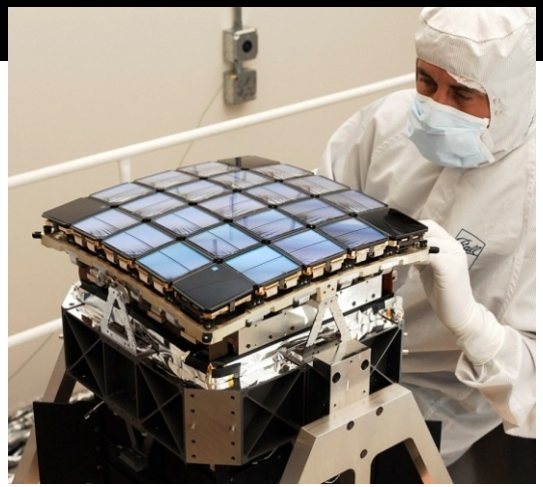
Kepler
A Search for Earth-size Planets

KEPLER: A Wide Field-of-View Photometer that monitors ≥100,000 Stars for 3.5 yrs with precision to find Earth-size planets in the Habitable Zone

Transit Detection using:

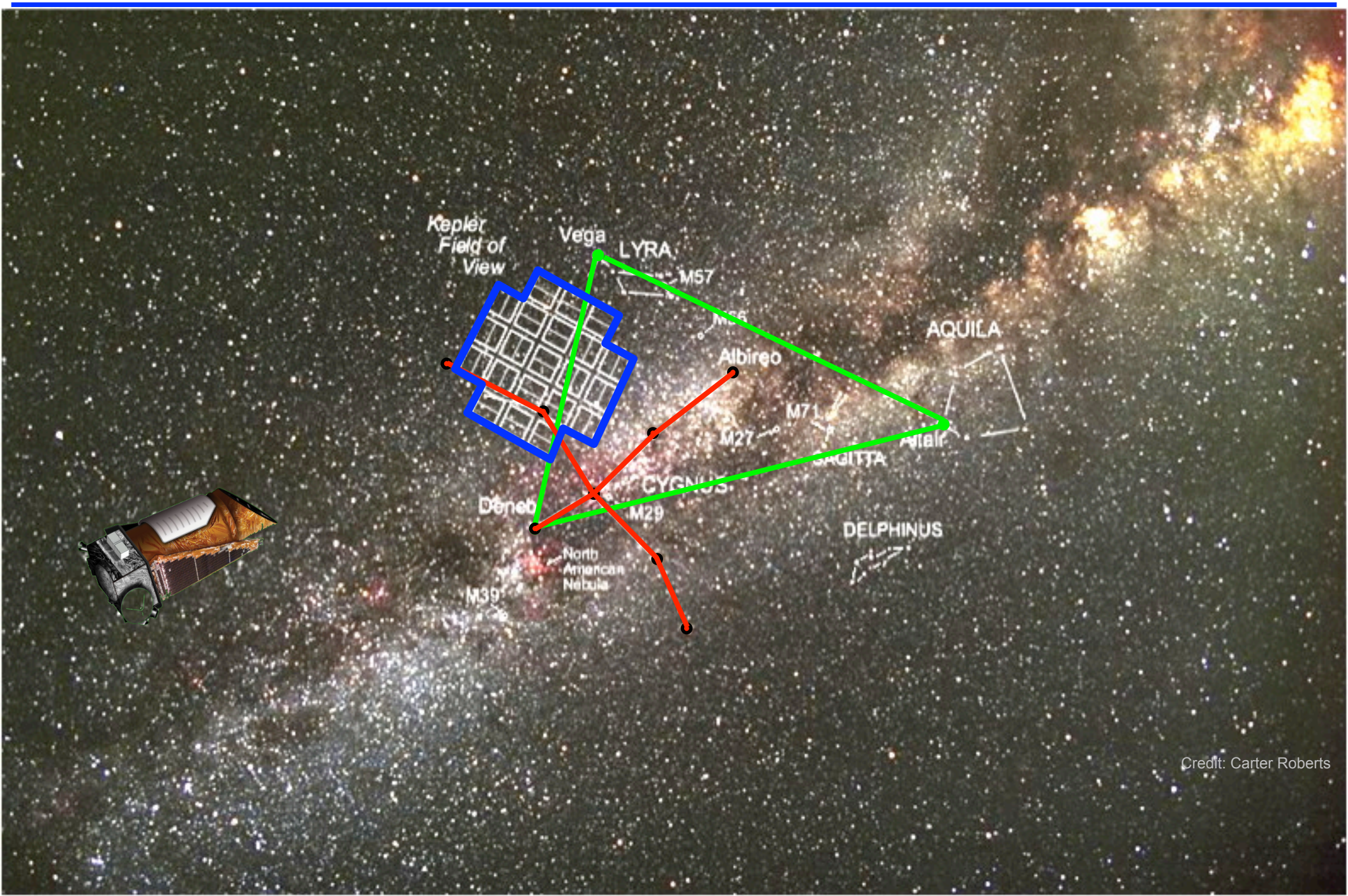
- 0.95 meter aperture
- Wide FOV: 100 sq deg
- 42 CCDs
- >= 3.5 years
- Fixed pointing
- Heliocentric orbit
- 170k targets: 30 min
- 512 targets: 1 min





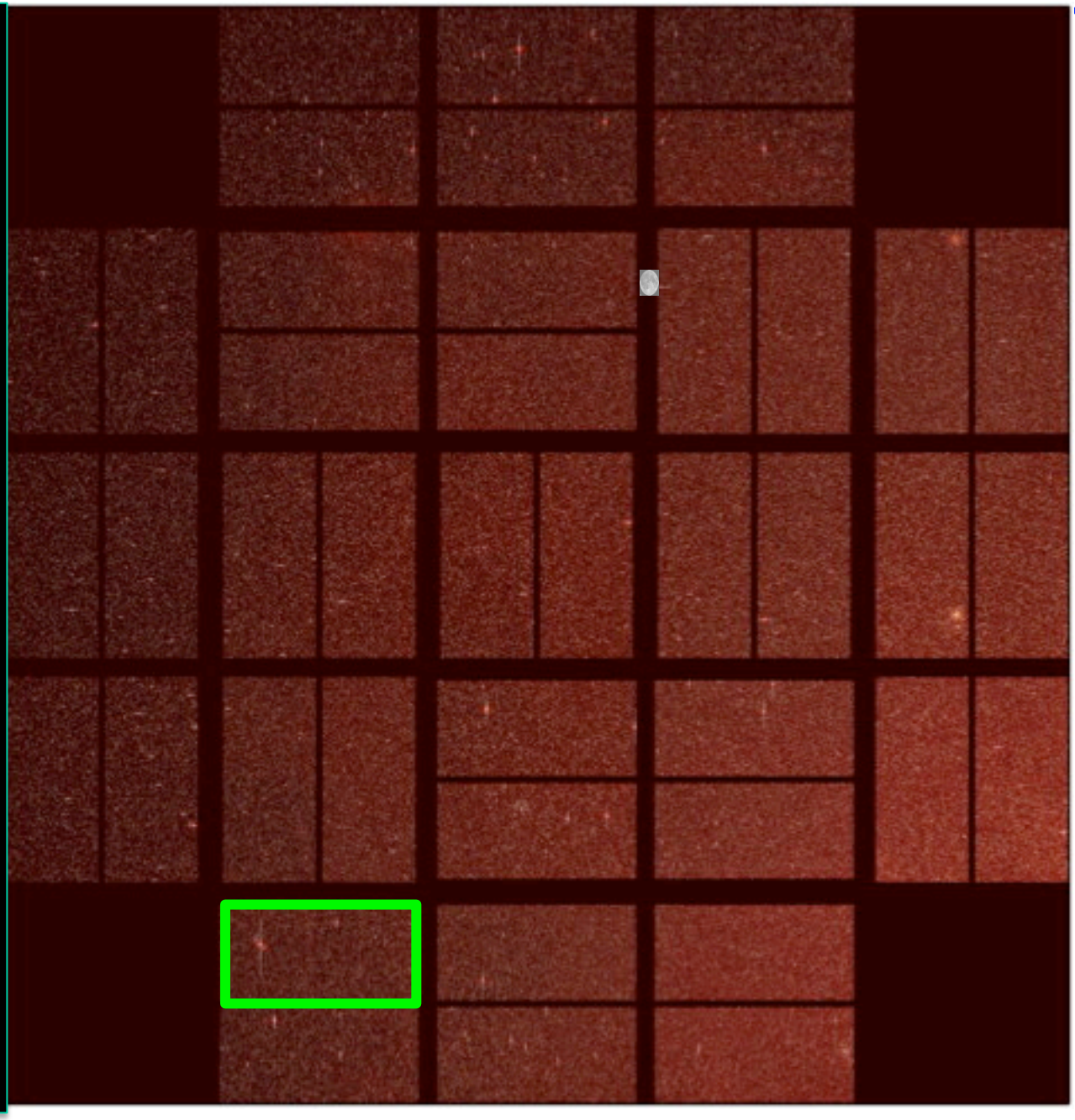
Kepler Field Of View

A Search for Earth-size Planets



First Light Image

Kepler
A Search for Earth-size Planets



Launched March 7 2009

SETI Institute
Jon M. Jenkins

Kepler: Big Data, Big Challenges

Kepler
A Search for Earth-size Planets

Big Data:

- >150,000 target stars
- 6×10^6 pixels collected and stored per ½ hour
- ~40 GB downlinked each month
- $>40 \times 10^9$ points in the time series over 3.5 years

Big Processing Challenges:

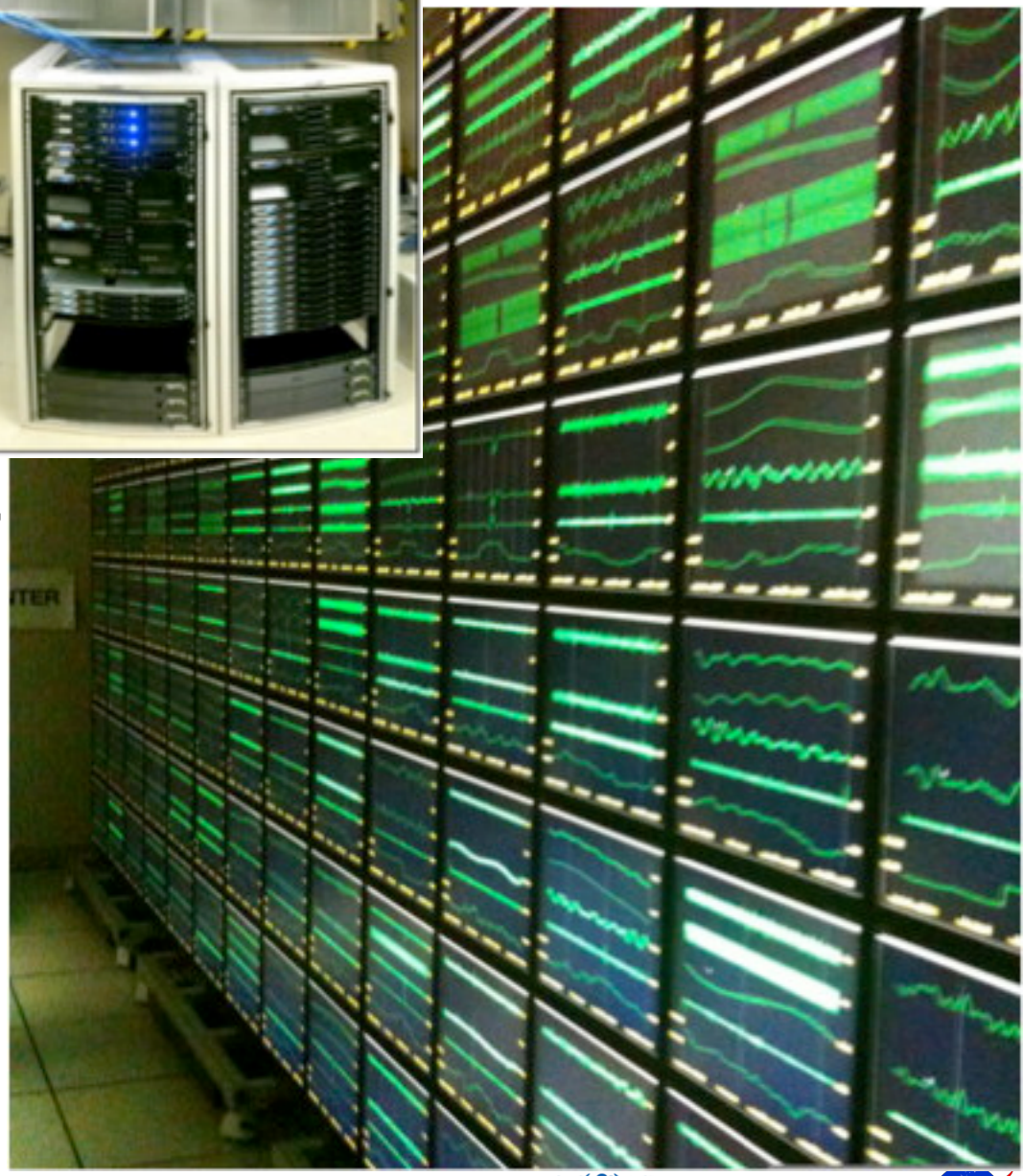
- Instrumental effects are large compared to signal of interest
- Observation noise is non-white and non-stationary
- $\sim 80 \times 10^9$ tests per star for planetary signatures for 2 years of data
- $\sim 1400 \times 10^9$ tests per star for planetary signatures for 8 years of data
- Computations grow as $[O(N^2)]$

Hardware Architecture

Kepler
A Search for Earth-size Planets

64 hosts, 712 CPUs, 3.7 TB of RAM,

148 TB of raw disk storage

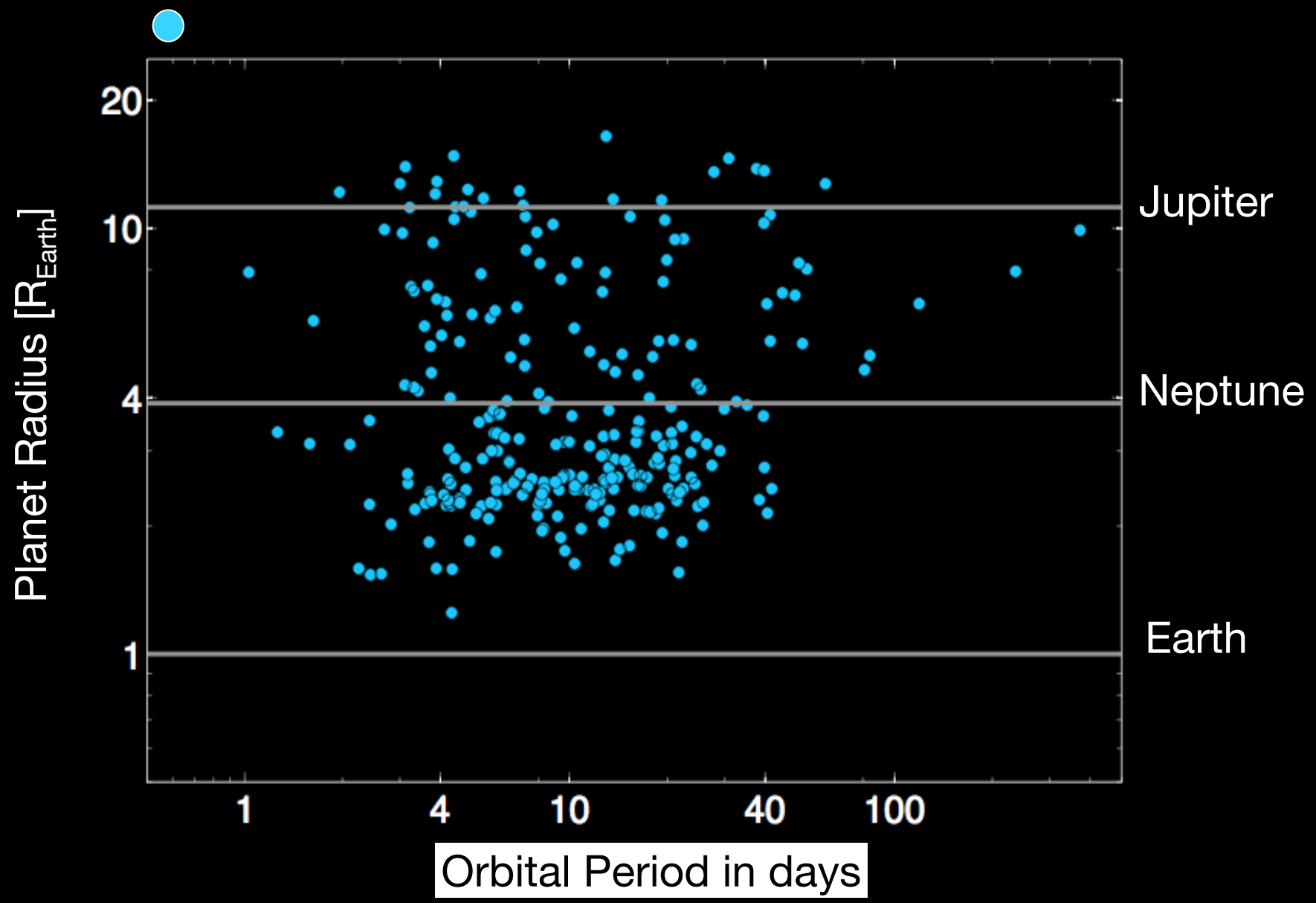


Key Science Results

Kepler's First Discoveries

Kepler
A Search for Earth-size Planets

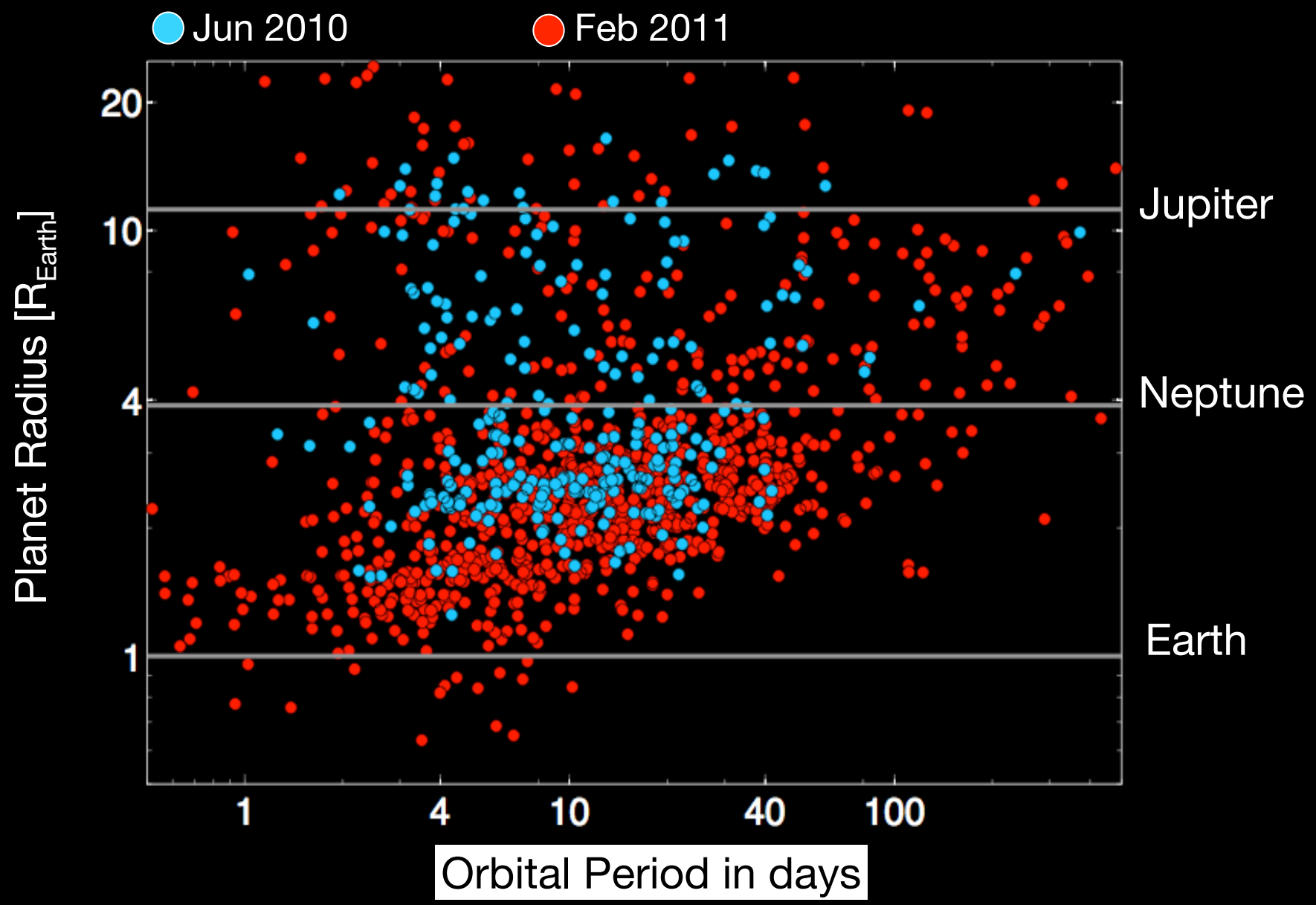
First month yields many planets



A Year of Data Makes a Big Difference

Kepler
A Search for Earth-size Planets

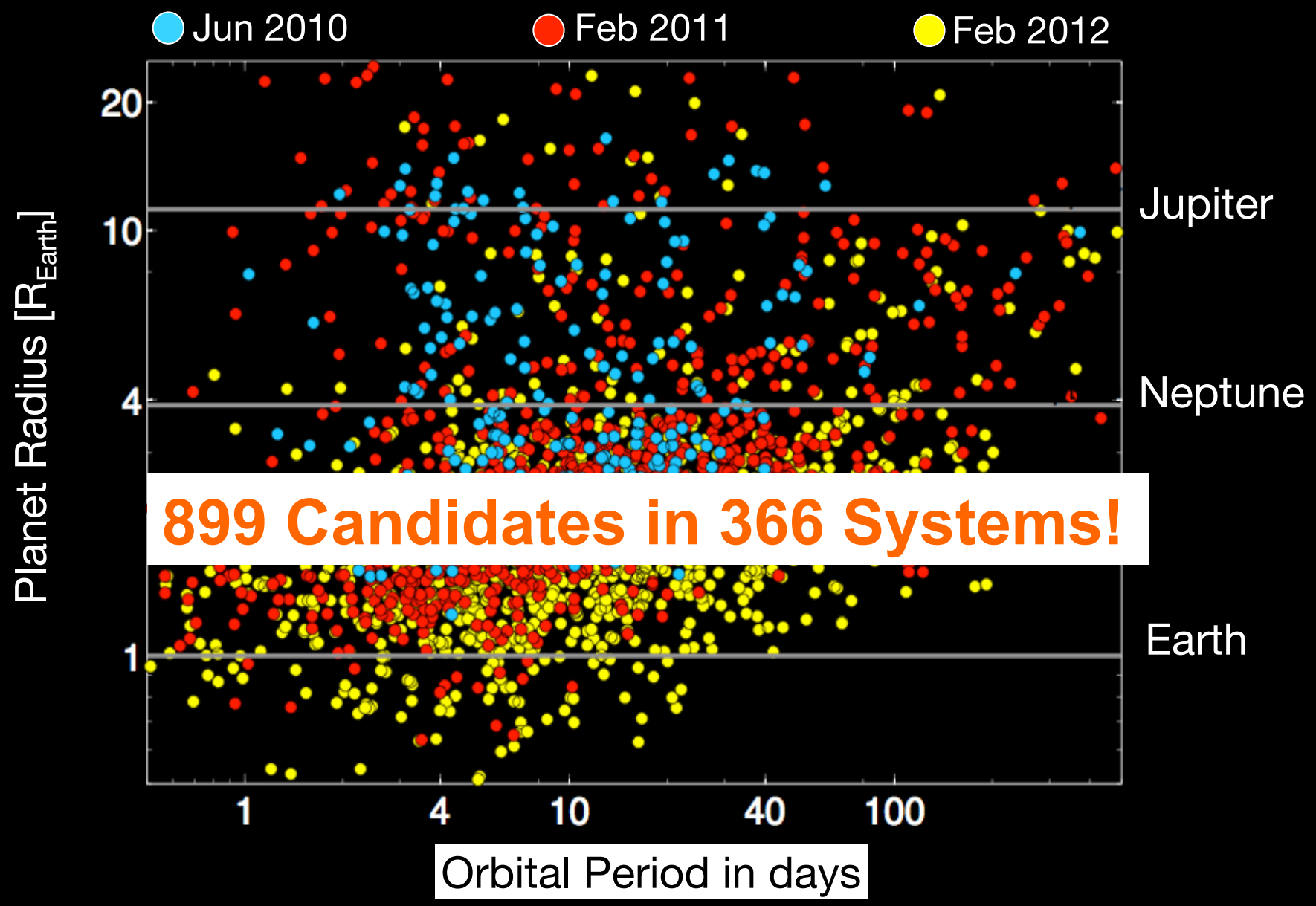
Full year yields smaller planets and longer periods



Over 2300 Candidates with 16 Months of Data

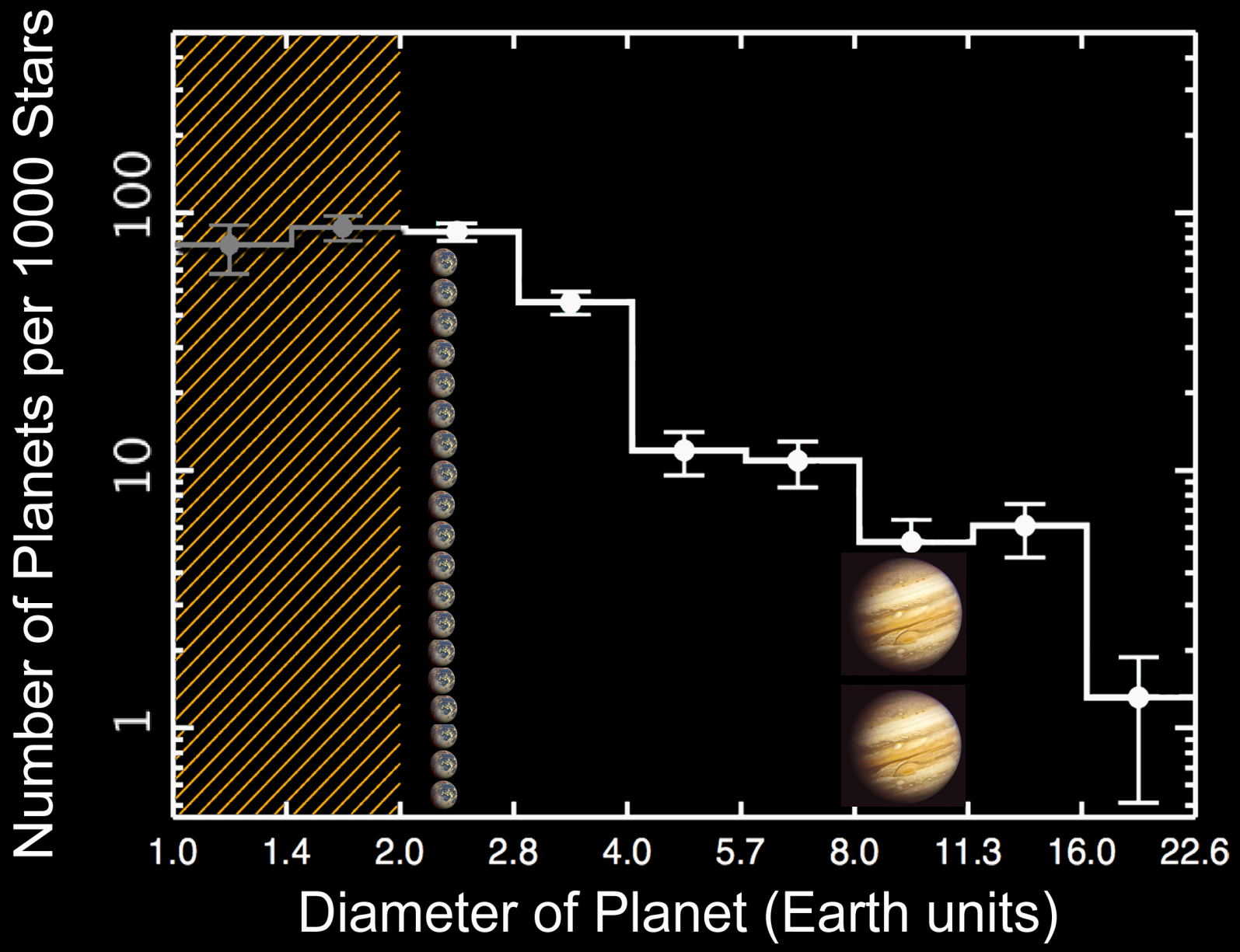
Kepler
A Search for Earth-size Planets

Improved pipeline reveals many previously unseen planets



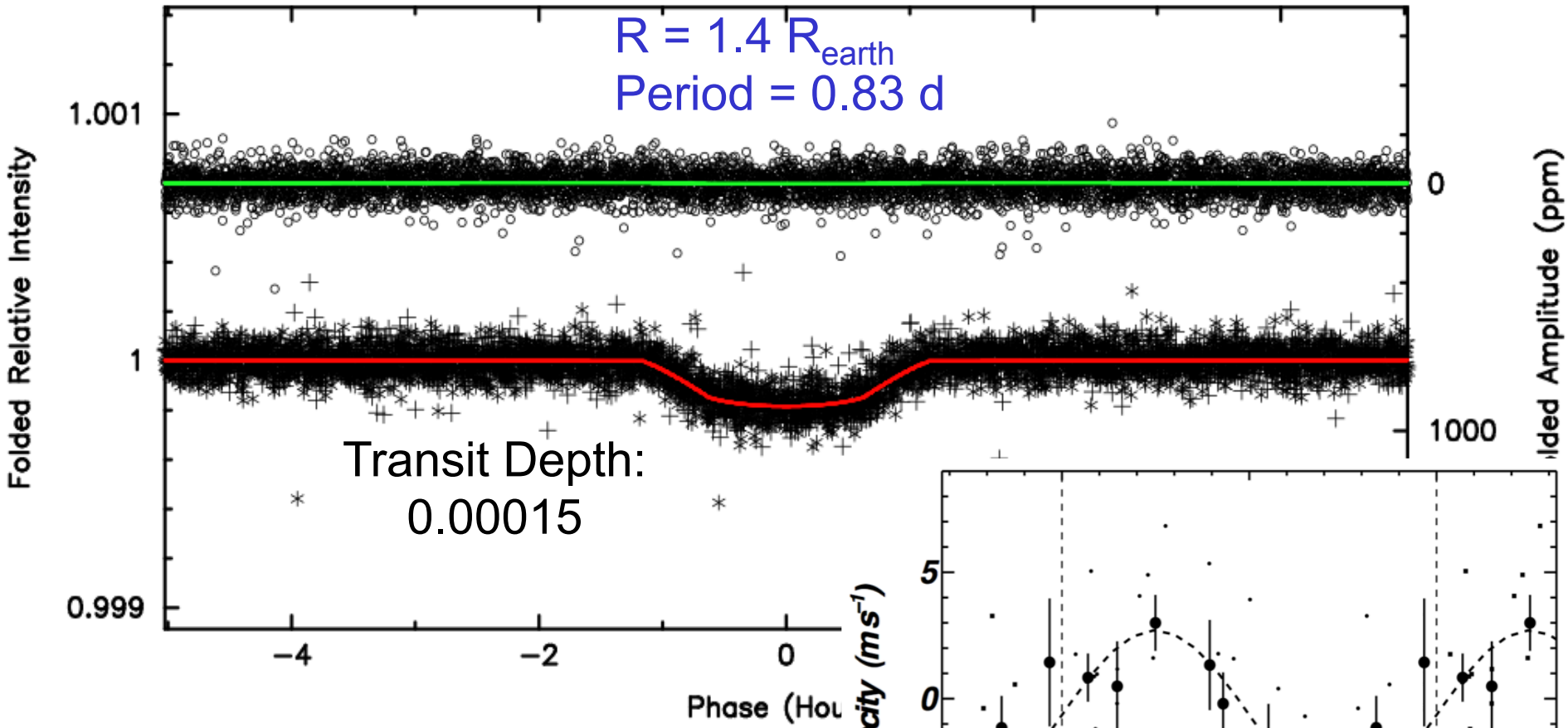
899 Candidates in 366 Systems!

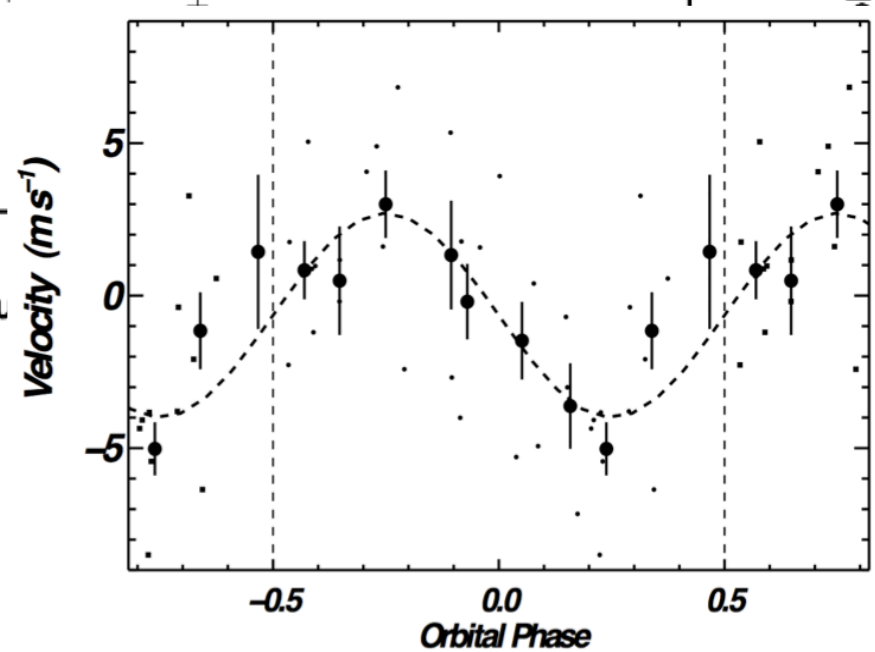
Distribution of Planet Diameters



Kepler-10b: Kepler's First Rocky Planet

Kepler
A Search for Earth-size Planets



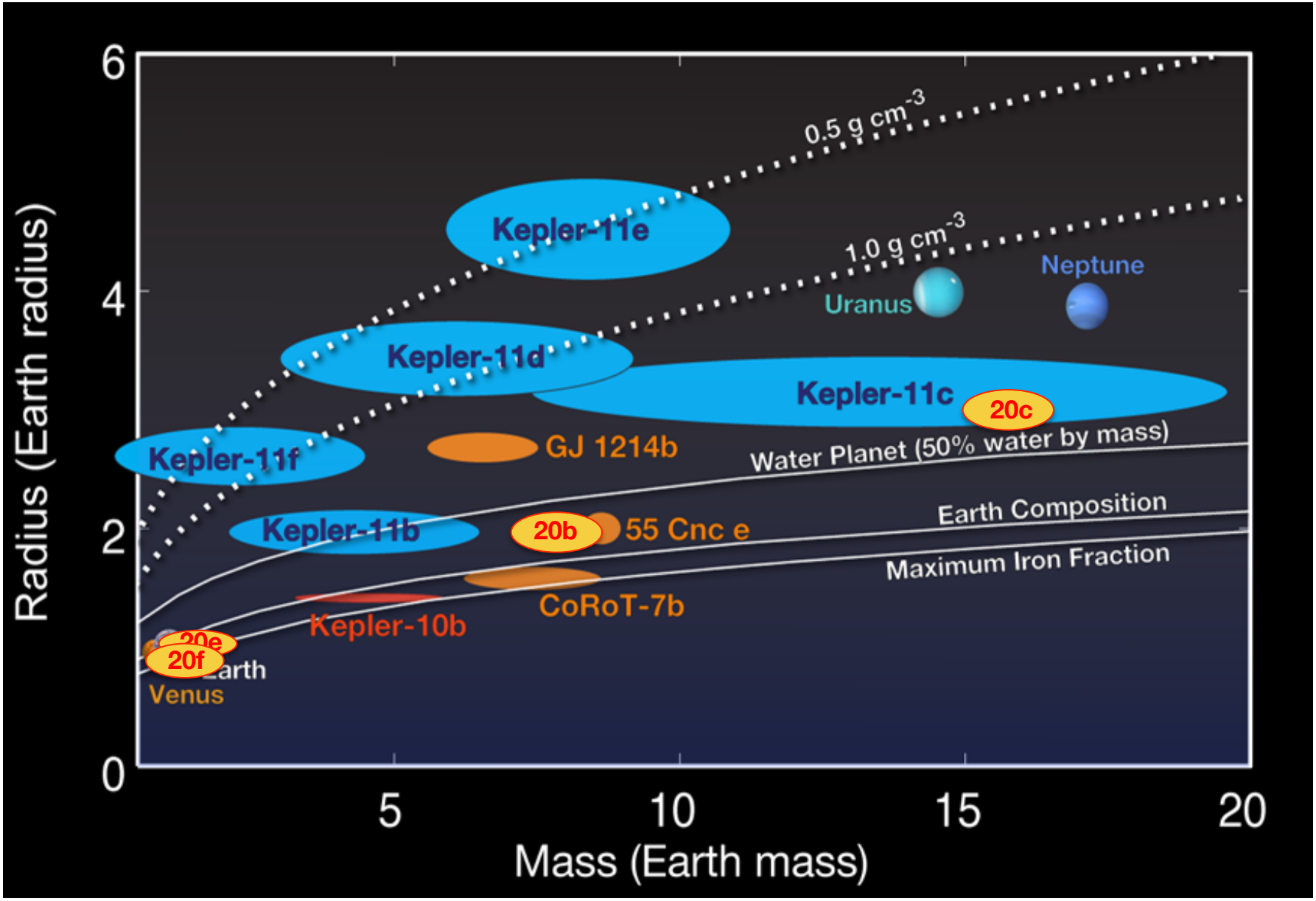


- 4.6 M_{earth}, 1.4 R_{earth}
- Density = 9 g/cm^3

Jon M. Jenkins

Mass-Radius Relationship

Kepler
A Search for Earth-size Planets

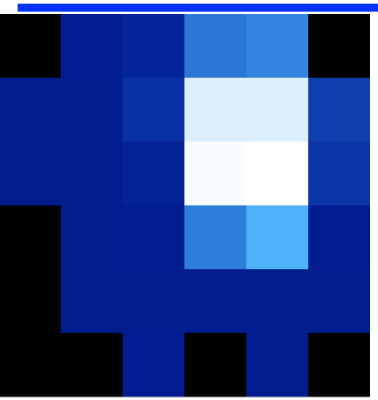


Jon M. Jenkins
SETI Institute

Processing *Kepler* Data

The *Kepler* Science Pipeline: From Pixels To Planets

Kepler
A Search for Earth-size Planets



Raw Data

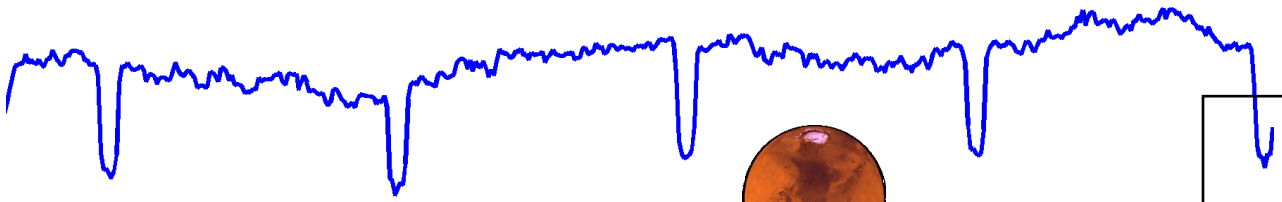
CAL
Pixel Level Calibrations

Calibrated Pixels

PA
Photometric Analysis

Sums Pixels Together/ Measures Star Locations

Raw Light Curves/ Centroids

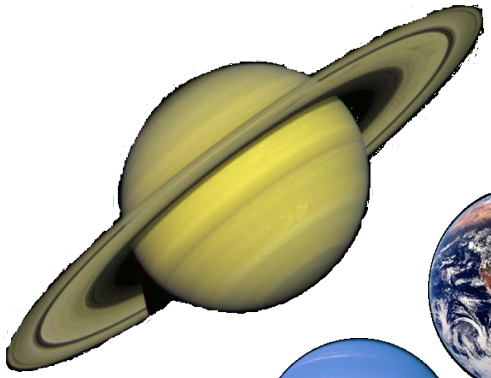


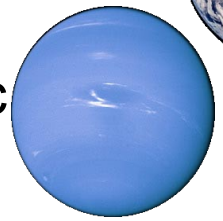
Corrected Light Curves

PDC
Presearch Data Conditioning

Removes Systematic Errors

TPS
Transiting Planet Search





DV
Data Validation

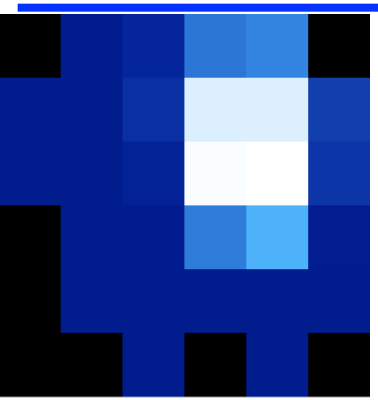
TCEs:
Threshold Crossing Events

Diagnostic Metrics

SETI Institute
Jon M. Jenkins

Step 1: Calibrating Pixels

Kepler
A Search for Earth-size Planets



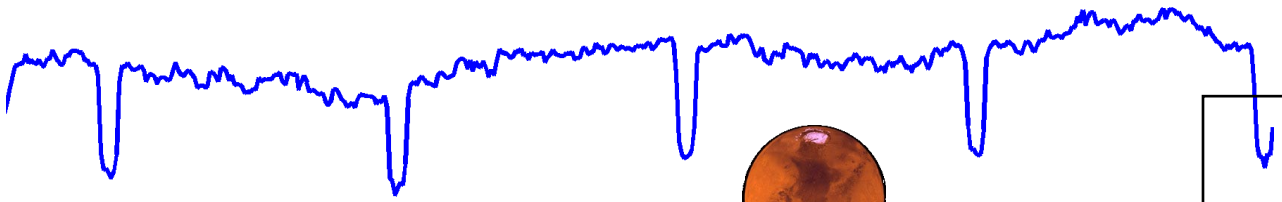
Raw Data

CAL
Pixel Level Calibrations

Calibrated Pixels

PA
Photometric Analysis

Sums Pixels Together/ Measures Star Locations



Raw Light Curves/ Centroids

Corrected Light Curves

PDC
Presearch Data Conditioning

Removes Systematic Errors

TPS
Transiting Planet Search

TCEs:
Threshold Crossing Events

DV
Data Validation

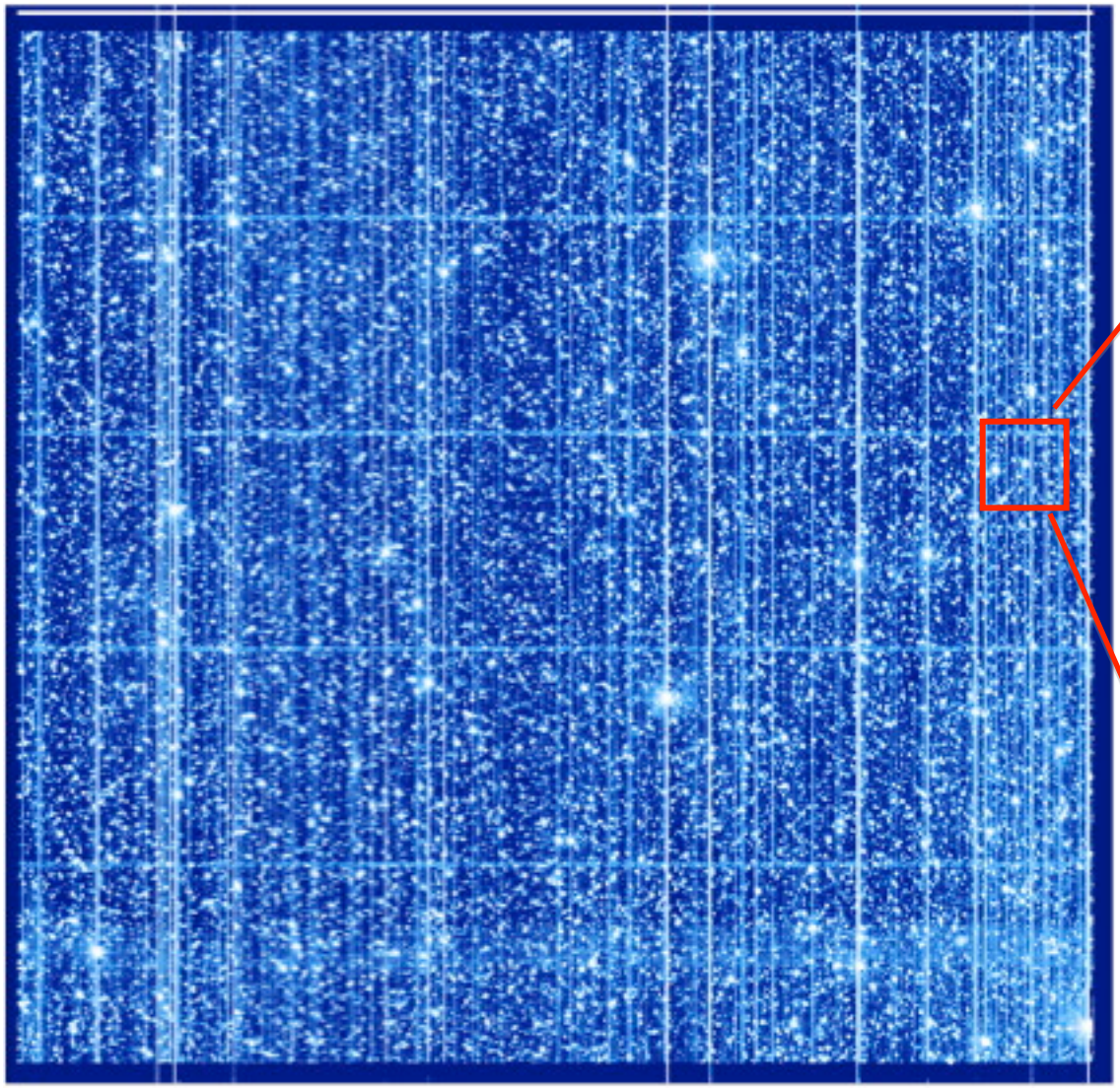
Diagnostic Metrics

Jon M. Jenkins
SETI Institute

Pixel-Level Data From Kepler

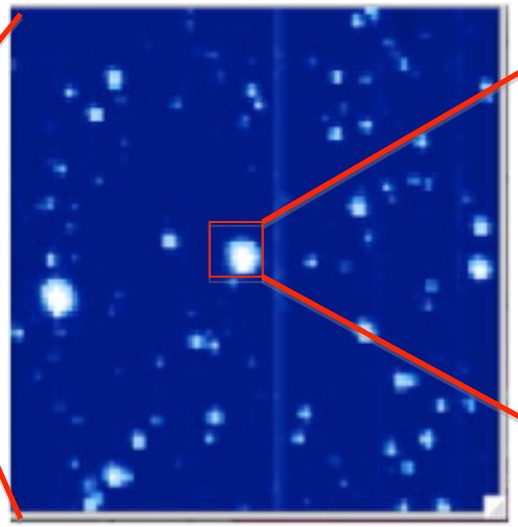
Kepler
A Search for Earth-size Planets

Module 17 Output 2



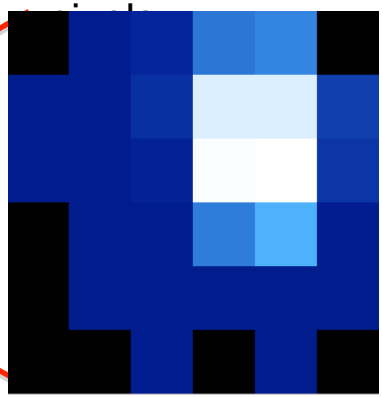
1.13 (h) x1.22 (w) degrees

Zoomed image near planet HAT-P-7b



0.09x0.09 degrees
80x80 pixels
6400 pixels total

HAT-P-7b



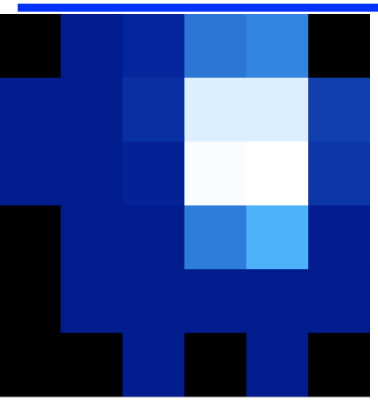
6.6x6.6 millidegrees
28 pixels collected
Black = no data

Scaled to show faint detail

SETI Institute
Jon M. Jenkins

Step 2: Photometric Analysis

Kepler
A Search for Earth-size Planets



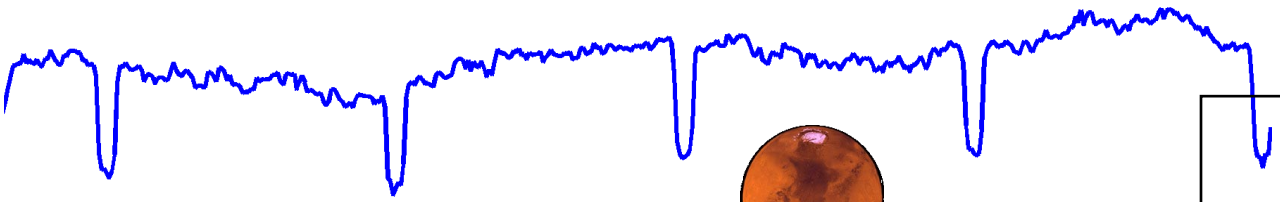
Raw Data

CAL
Pixel Level Calibrations

Calibrated Pixels

PA
Photometric Analysis

Sums Pixels Together/ Measures Star Locations



Raw Light Curves/ Centroids

Corrected Light Curves

PDC
Presearch Data Conditioning

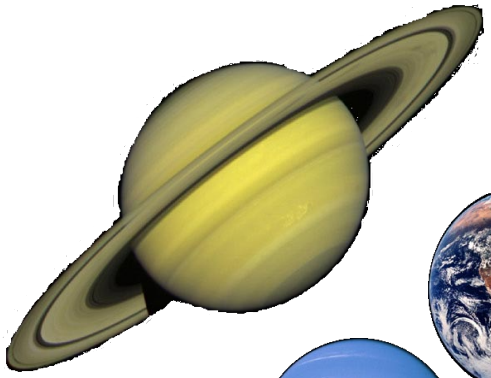
Removes Systematic Errors

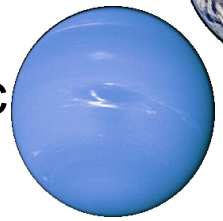
TPS
Transiting Planet Search

TCEs:
Threshold Crossing Events

DV
Data Validation

Diagnostic Metrics

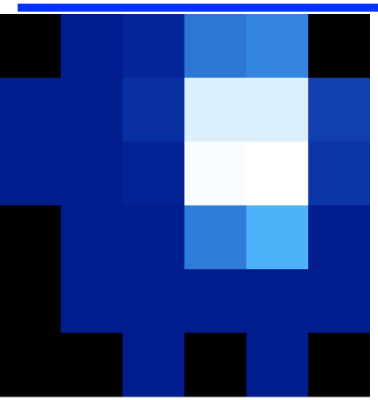




SETI Institute
Jon M. Jenkins

Step 3: Removing Systematic Errors

Kepler
A Search for Earth-size Planets



Raw Data

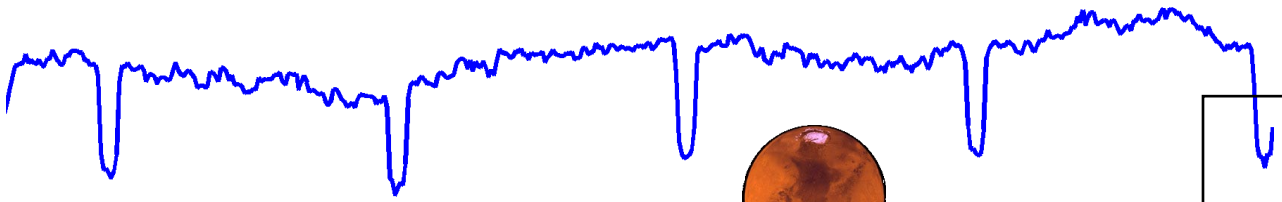
CAL
Pixel Level Calibrations

Calibrated Pixels

PA
Photometric Analysis

Sums Pixels Together/ Measures Star Locations

Raw Light Curves/ Centroids



Corrected Light Curves

PDC
Presearch Data Conditioning

Removes Systematic Errors

TPS
Transiting Planet Search

TCEs:
Threshold Crossing Events

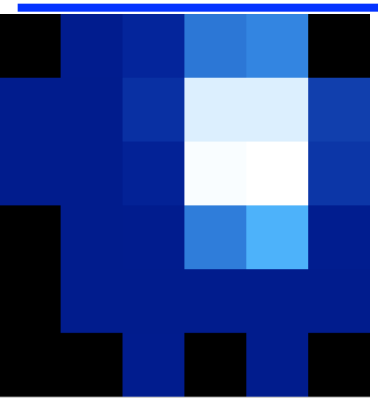
DV
Data Validation

Diagnostic Metrics

Jon M. Jenkins

Step 4: Searching for Planets

Kepler
A Search for Earth-size Planets



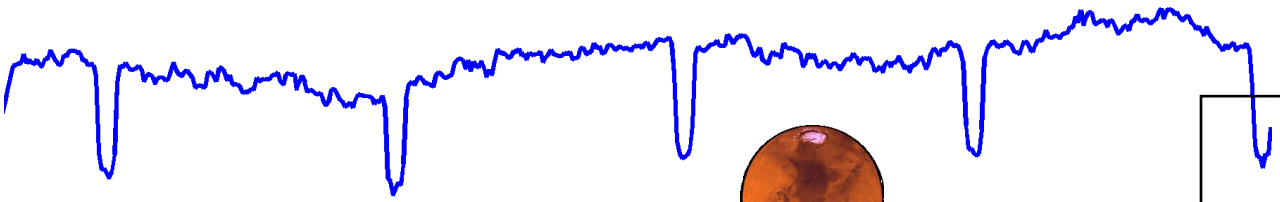
Raw Data

CAL
Pixel Level Calibrations

Calibrated Pixels

PA
Photometric Analysis

Sums Pixels Together/ Measures Star Locations



Raw Light Curves/ Centroids

Corrected Light Curves

PDC
Presearch Data Conditioning

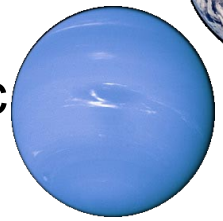
Removes Systematic Errors

TPS
Transiting Planet Search

TCEs:
Threshold Crossing Events

DV
Data Validation

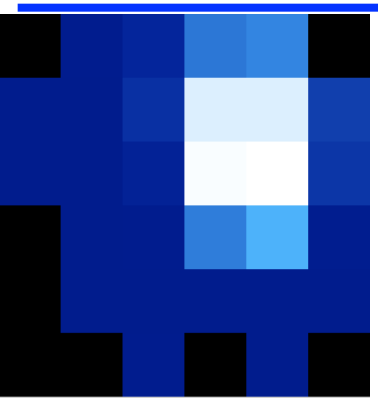
Diagnostic Metrics



SETI Institute
Jon M. Jenkins

Step 5: Validating Potential Planets

Kepler
A Search for Earth-size Planets



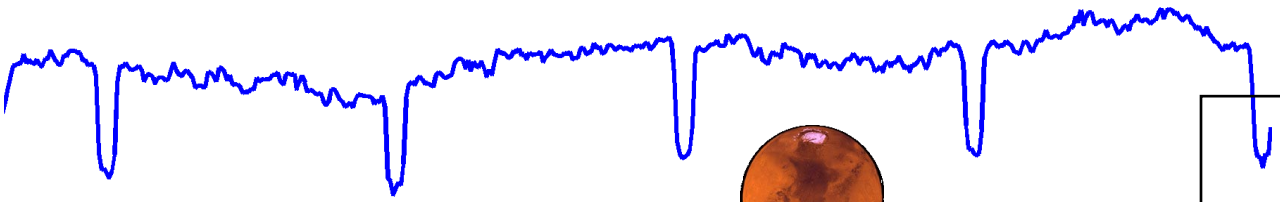
Raw Data

CAL
Pixel Level Calibrations

Calibrated Pixels

PA
Photometric Analysis

Sums Pixels Together/ Measures Star Locations



Raw Light Curves/ Centroids

Corrected Light Curves

PDC
Presearch Data Conditioning

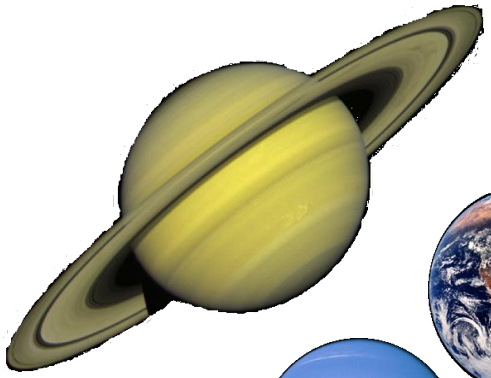
Removes Systematic Errors

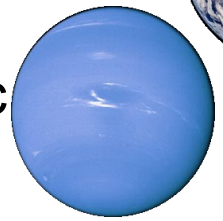
TPS
Transiting Planet Search

TCEs:
Threshold Crossing Events

DV
Data Validation

Diagnostic Metrics



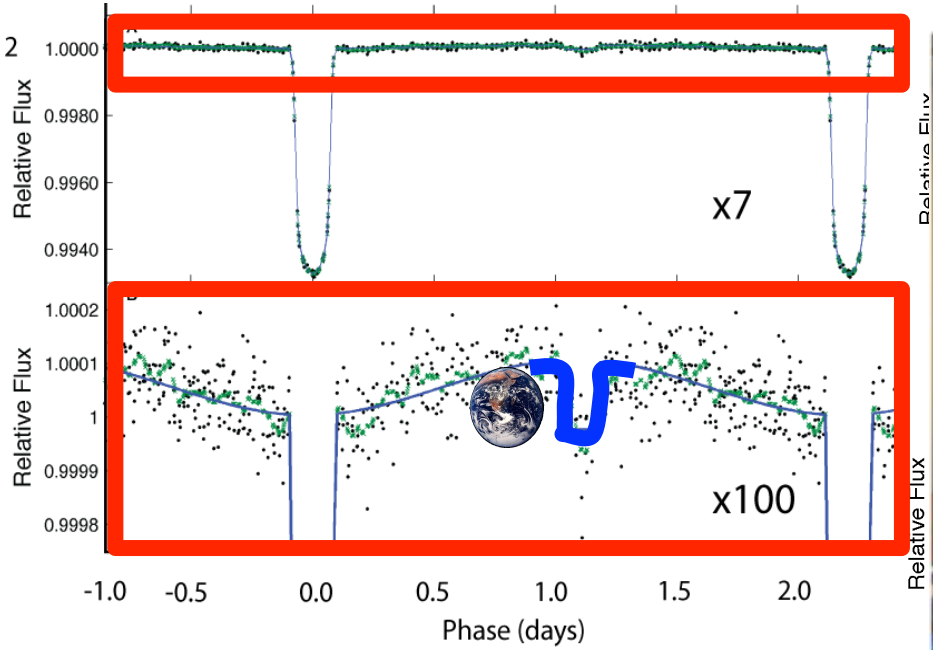


SETI Institute
Jon M. Jenkins

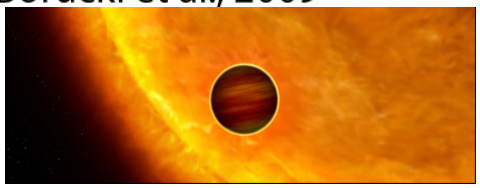
What Does a Planet Sound Like?

Kepler
A Search for Earth-size Planets

HAT-P-7B



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



Another Star

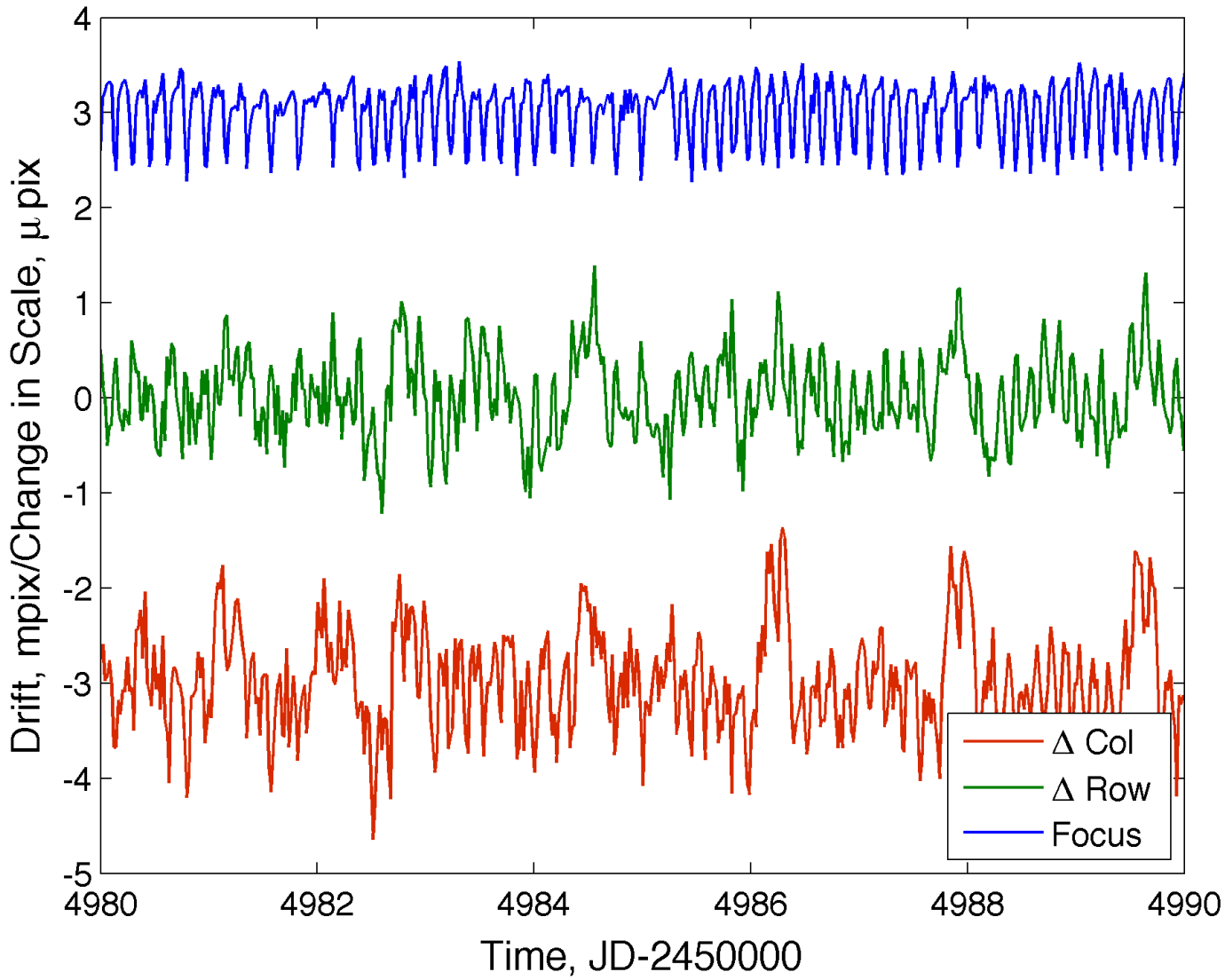
Jon M. Jenkins
SETI Institute

Correcting Systematic Errors

Short Timescale Instrumental Errors

Kepler
A Search for Earth-size Planets

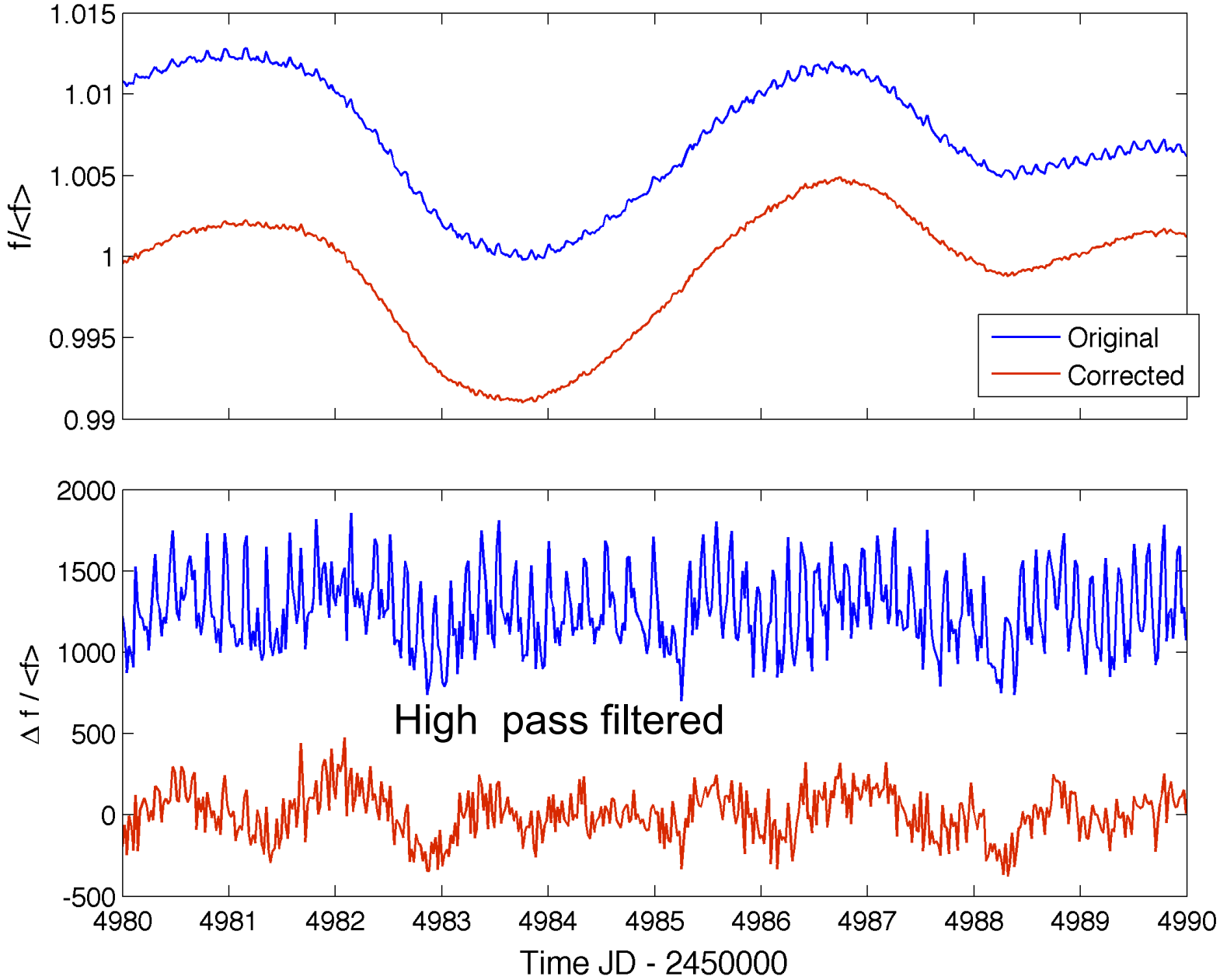
Signature of a heater cycling on the reaction wheels 3/4



***Kepler* is sensitive to its thermal environment**

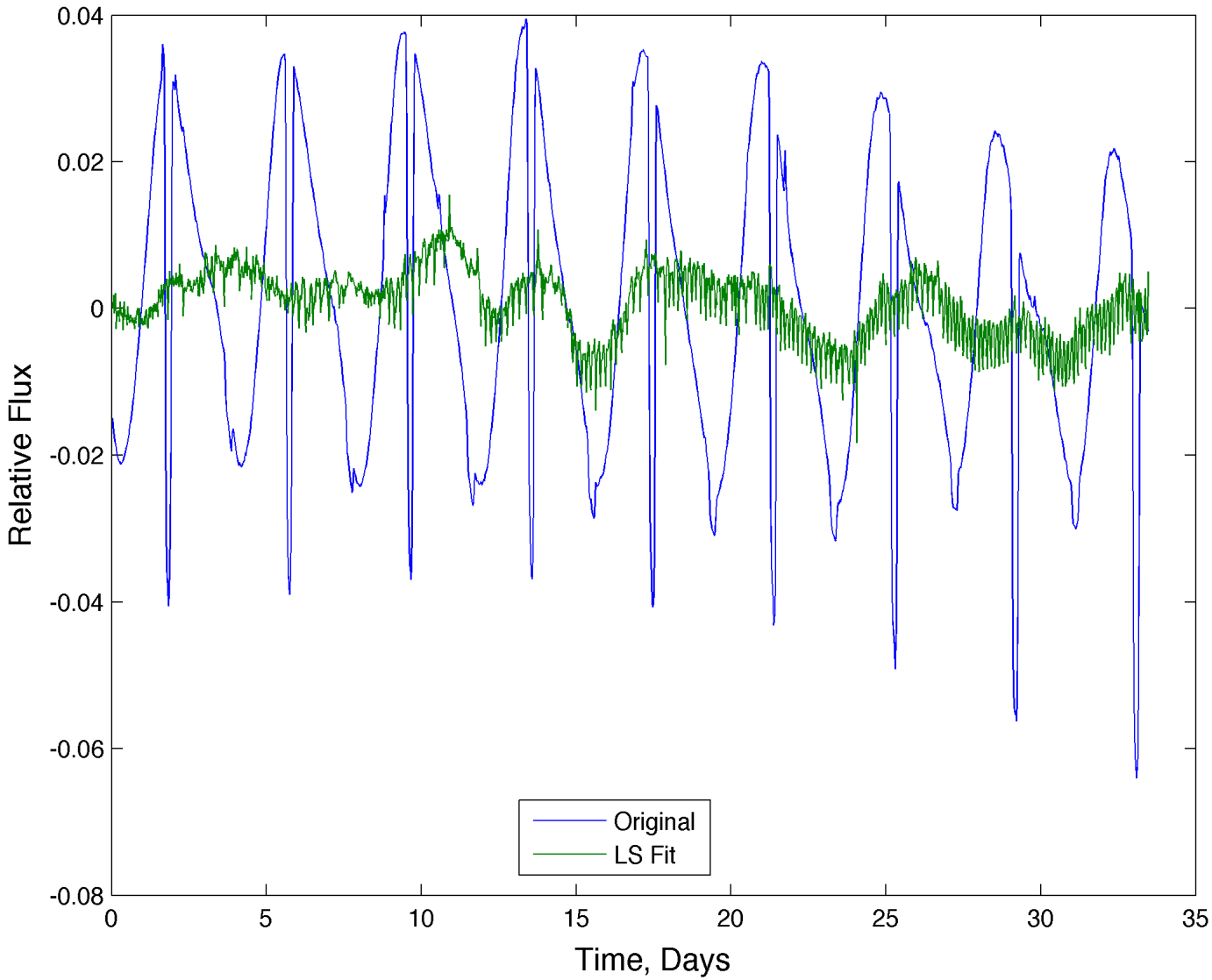
Jon M. Jenkins

Correcting Instrumental Effects



Jon M. Jenkins

PDC Often Over-Fits Variable Stars



A star fitted with systematics (green) extracted by SVD from 1400 “quiet” stars from Q1 on channel 2.1.

Jon M. Jenkins

PDC Is Fundamentally Flawed

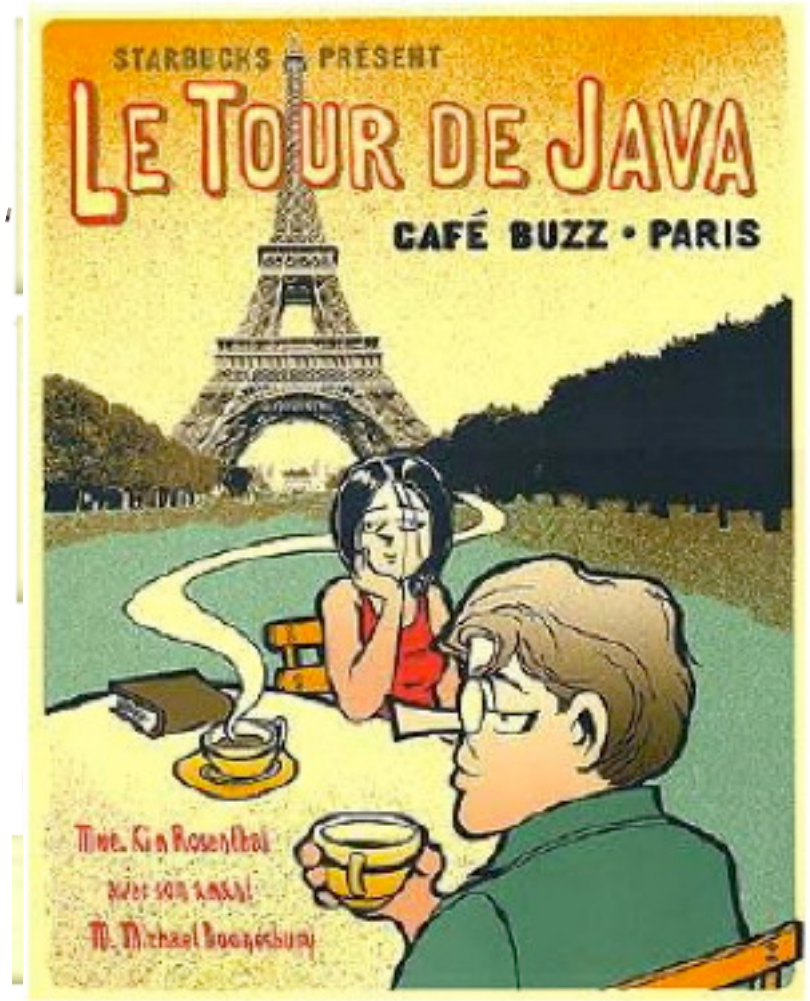
Kepler
A Search for Earth-size Planets

PDC co-trends against instrumental signatures using least squares (LS) approach

LS attempts to explain ***all*** of a given time series, not just the part the model can explain well

There is no way a simple LS fit can “put on the brakes”

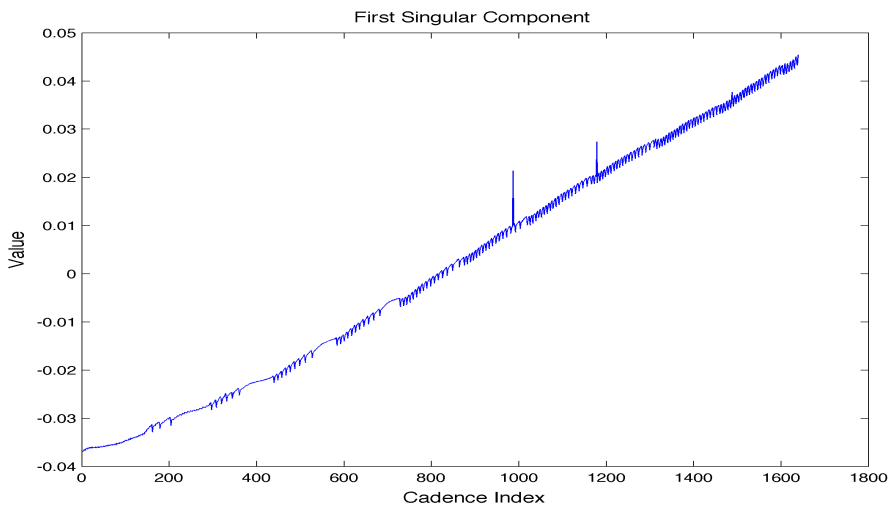
PDC often trades bulk RMS for increased noise at short time scales

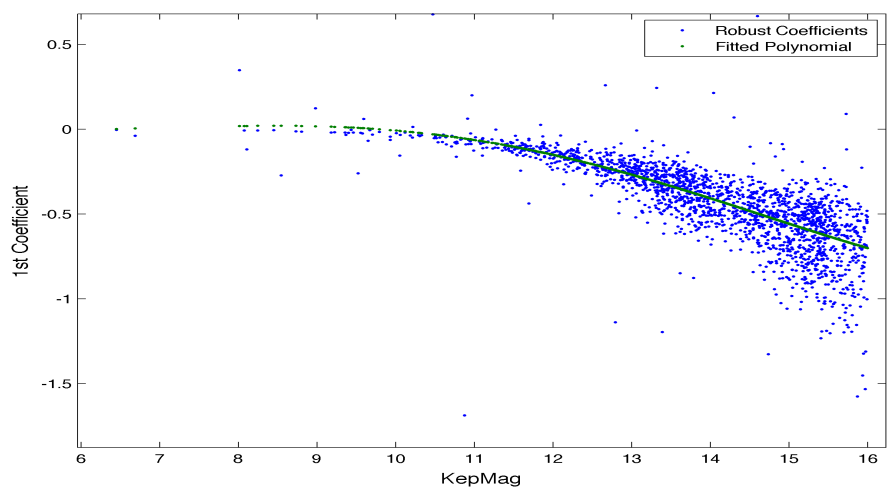


A Bayesian Solution

- Examine behavior of ensemble of stars responding to systematics
- Formulate prior probability distributions for model coefficients
- Maximize Posterior Distribution: $\hat{\theta} = \arg\max_{\theta} \left\{ \log\left[p(x|c)\right] + \log\left[p(c)\right] \right\}$

Maximum Likelihood Prior PDF

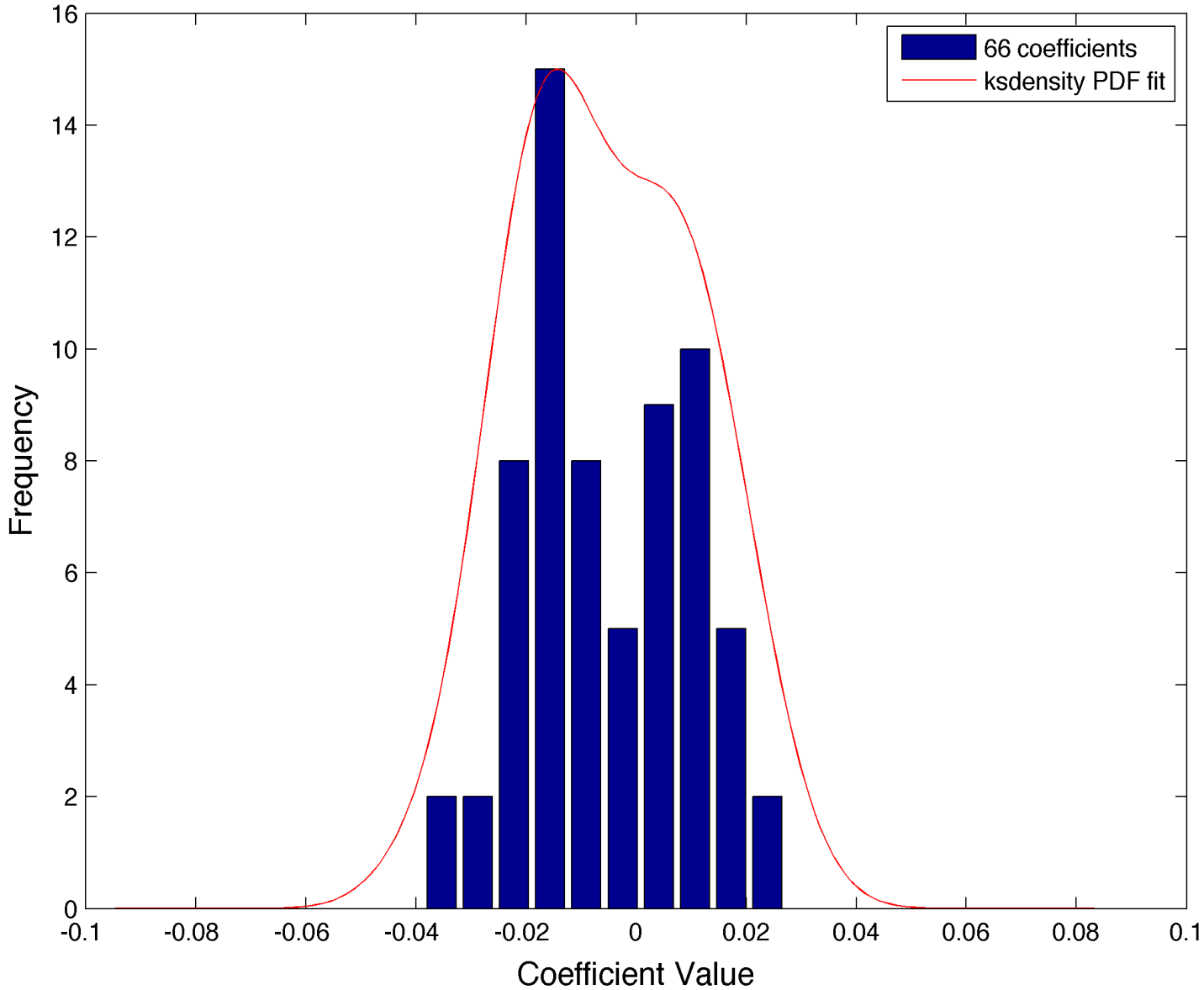




"A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule."

Actual PDFs Are Non-Gaussian

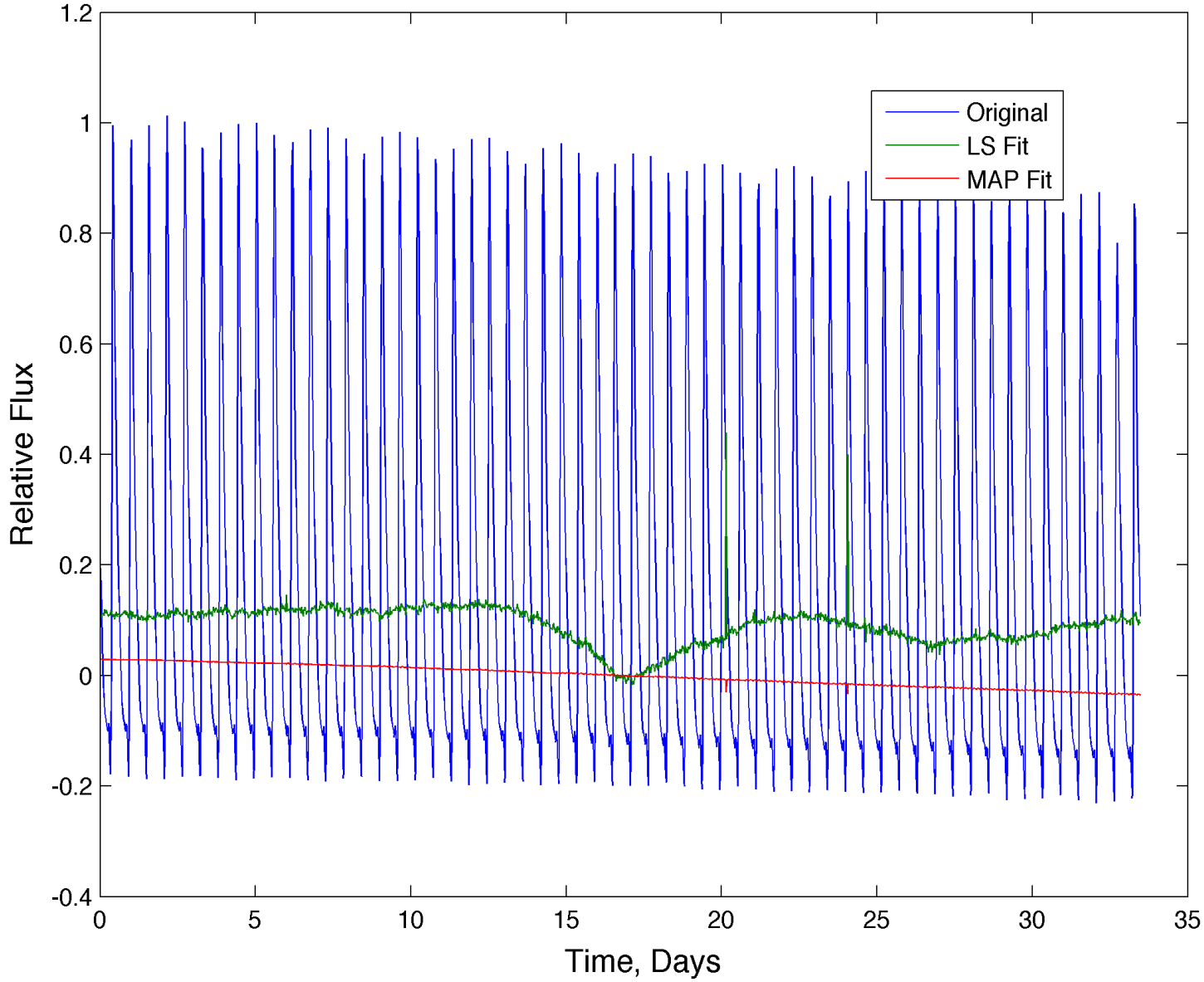
A Search for Earth-size Planets



Jon M. Jenkins

Example 1

An RR Lyrae star that is treated poorly by robust LS, but is well-served by MAP

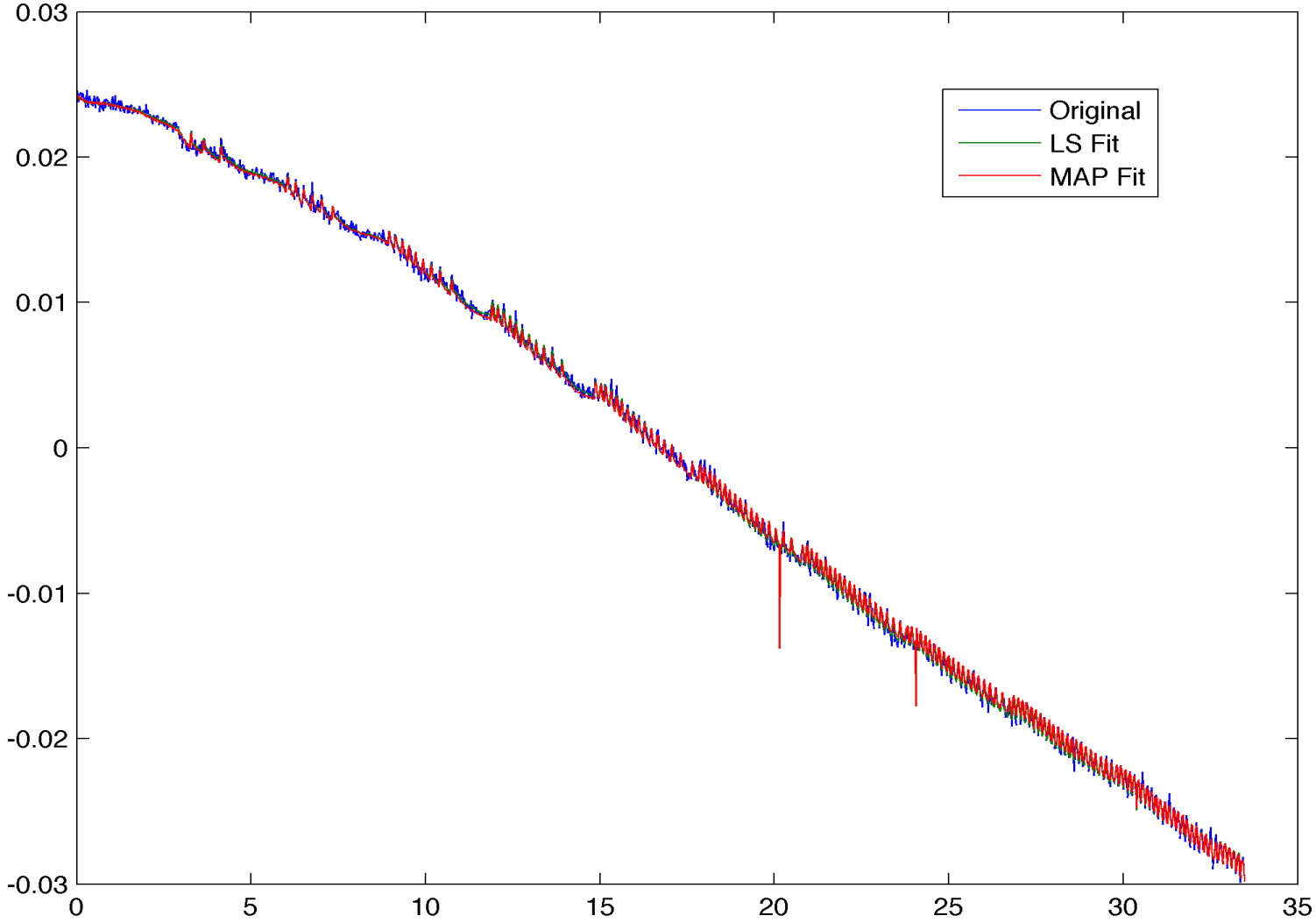


Jon M. Jenkins

Example 2

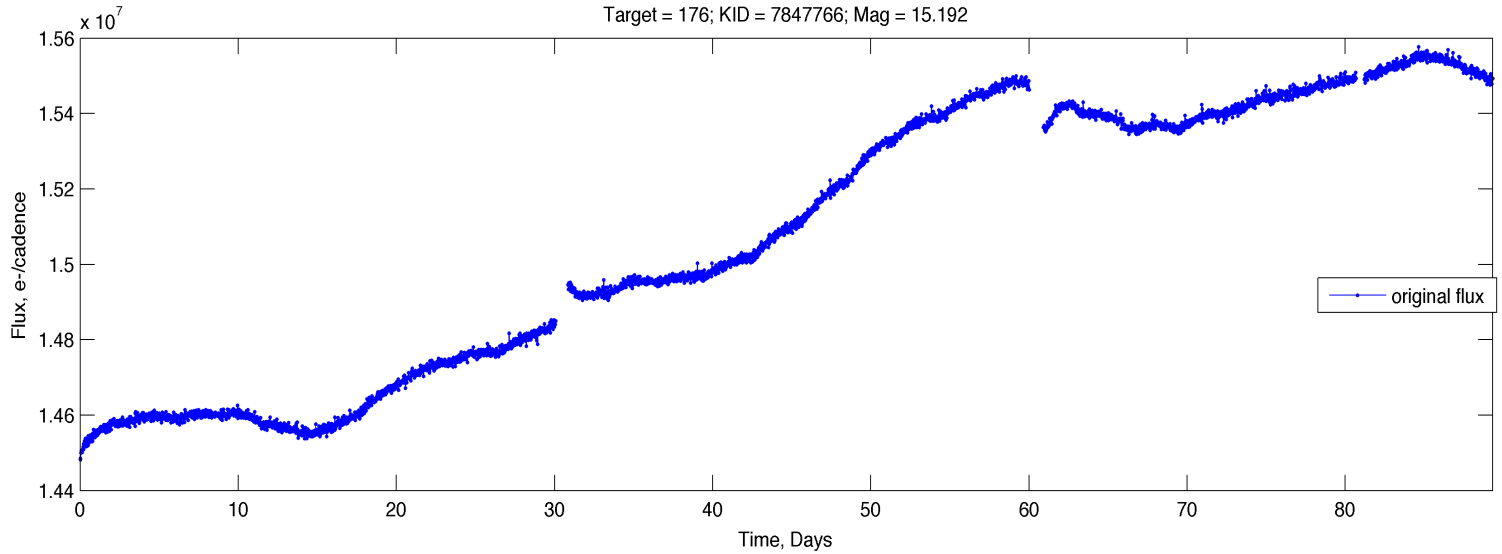
A "quiet" star dominated by systematics.

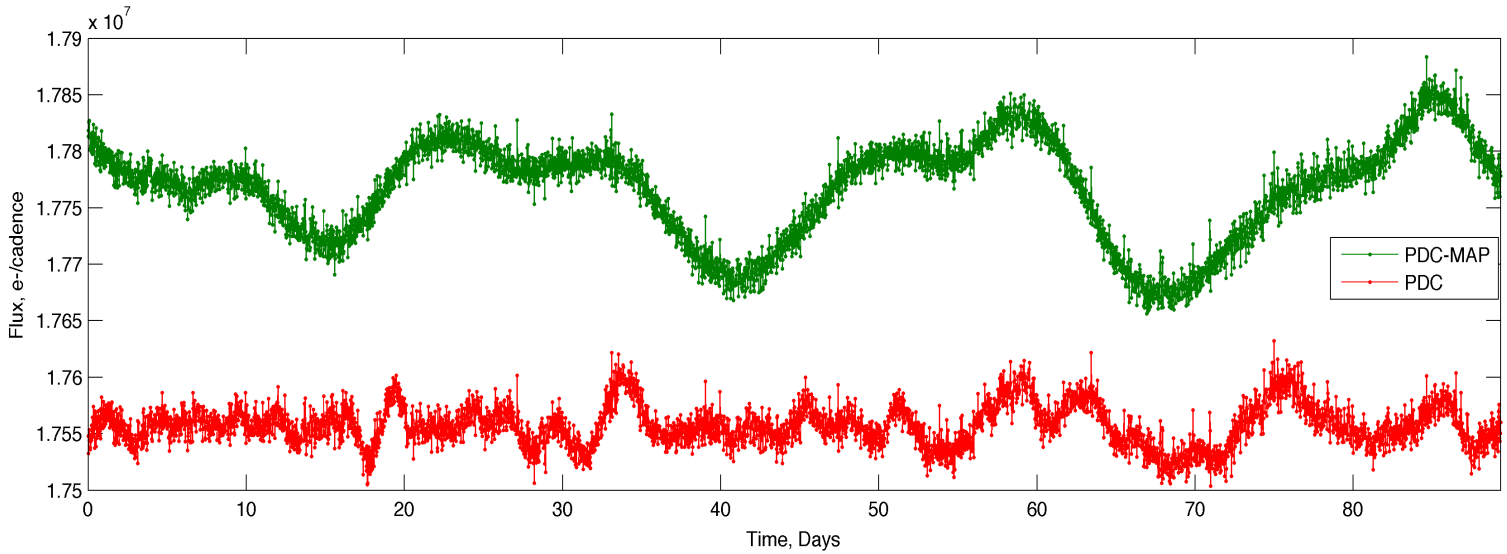
The LS and the MAP fits are comparable.



PDC-MAP Example

A Search for Earth-size Planets





Jon M. Jenkins

Searching for Transit Signatures

The Search Problem

A Search for Earth-size Planets

Detection Statistics For Colored Noise

w is (colored) Gaussian noise with autocorrelation matrix *R*
x is the data
s is the signal of interest

Decide *s* is present if

$$T = \frac{x^T R^{-1} s}{\sqrt{s^T R^{-1} s}} = \frac{(Hx)^T (Hs)}{\sqrt{(Hs)^T (Hs)}} = \frac{\tilde{x}^T \tilde{s}}{\sqrt{\tilde{s}^T \tilde{s}}} > \gamma$$

How do we determine R?

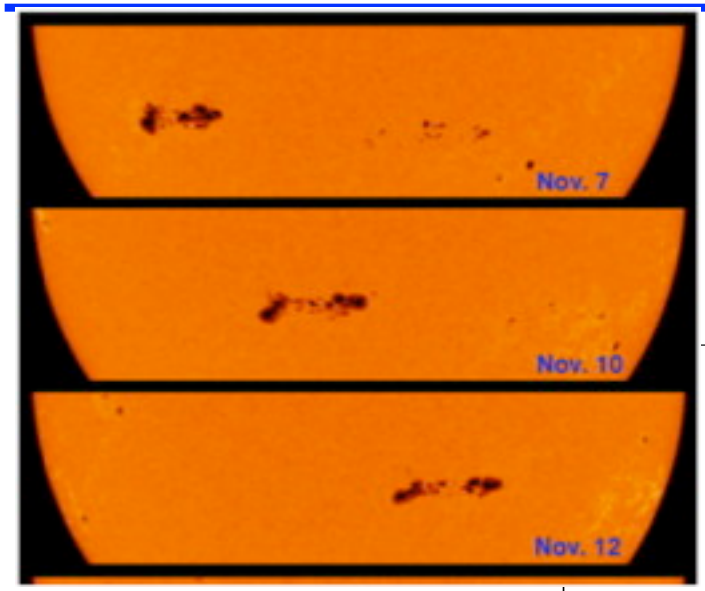
Looks like a simple matched filter!

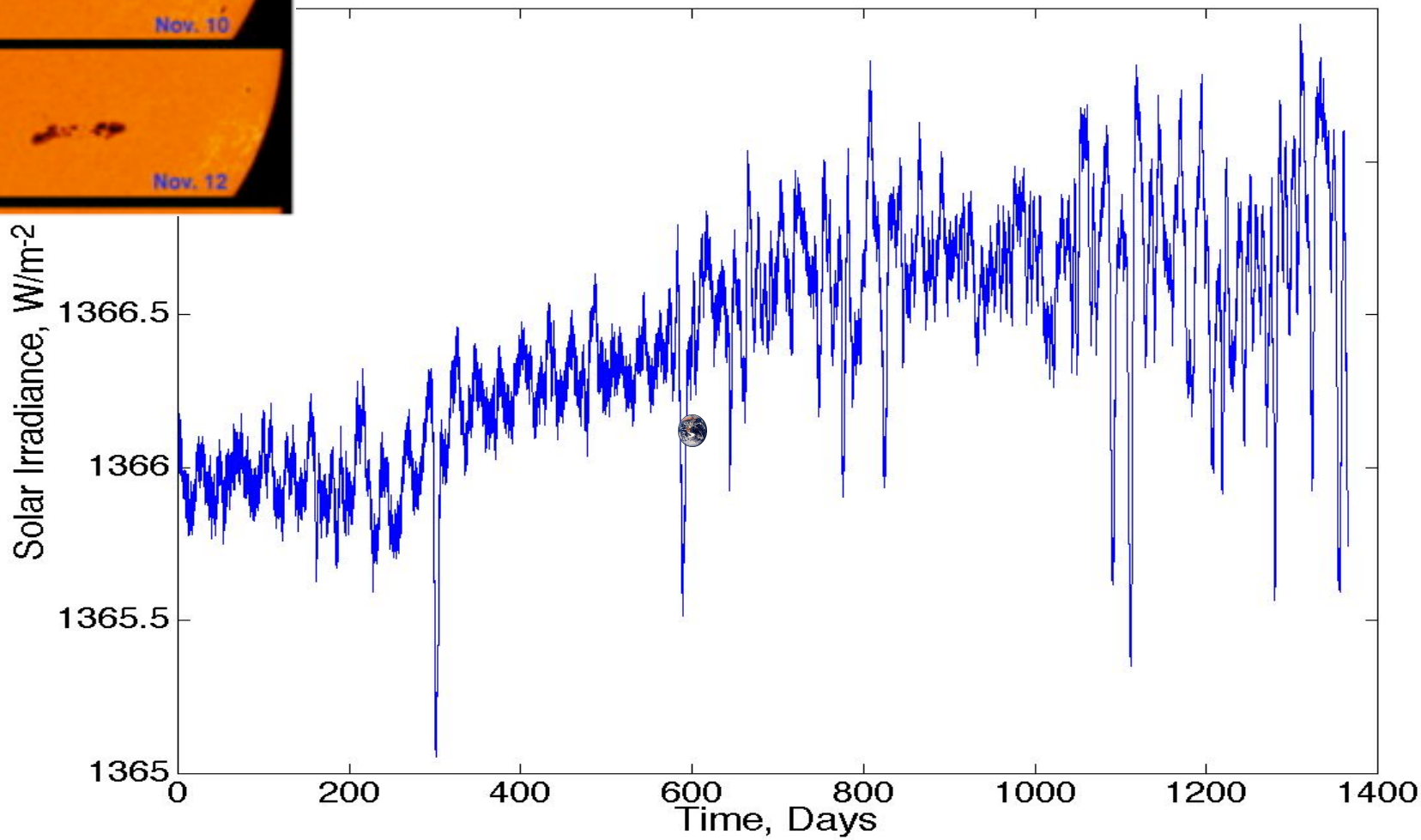
If the noise is stationary, we can work in the frequency domain:

$$T = \int \frac{X(f)S^*(f)}{P(f)} df \Bigg/ \sqrt{\int \frac{S(f)S^*(f)}{P(f)} df}$$

Solar Variability

A Search for Earth-size Planets





Jon M. Jenkins

PSDs for Solar-Like Variability

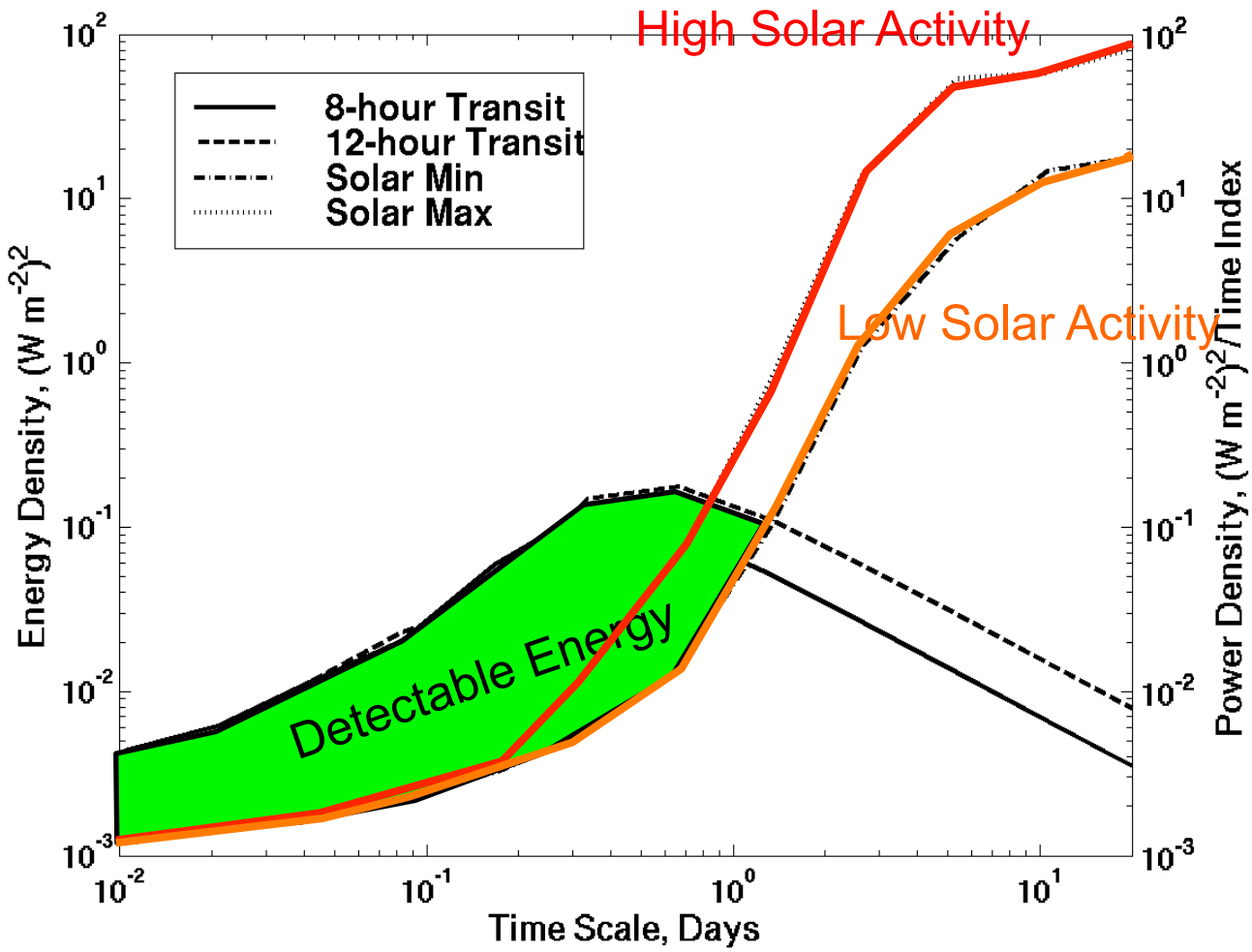
Kepler
A Search for Earth-size Planets

Is stellar variability stationary?

No!

We must work in a joint time-frequency domain

Wavelets are a natural choice



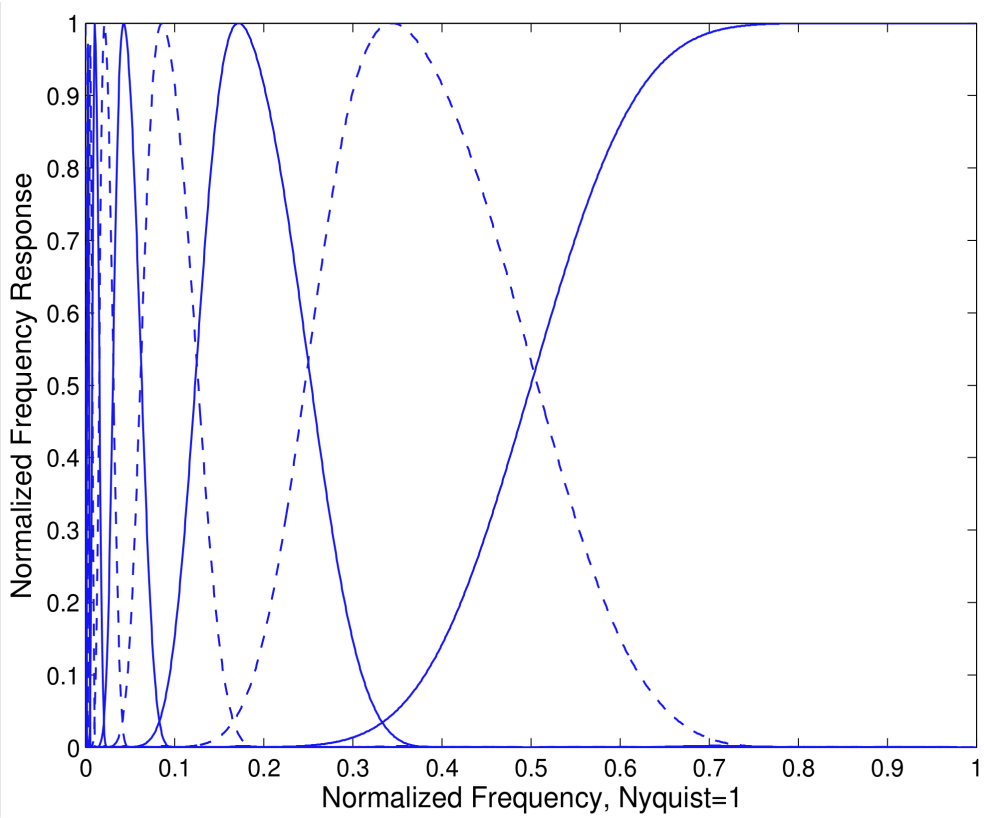
A Wavelet-Based Approach

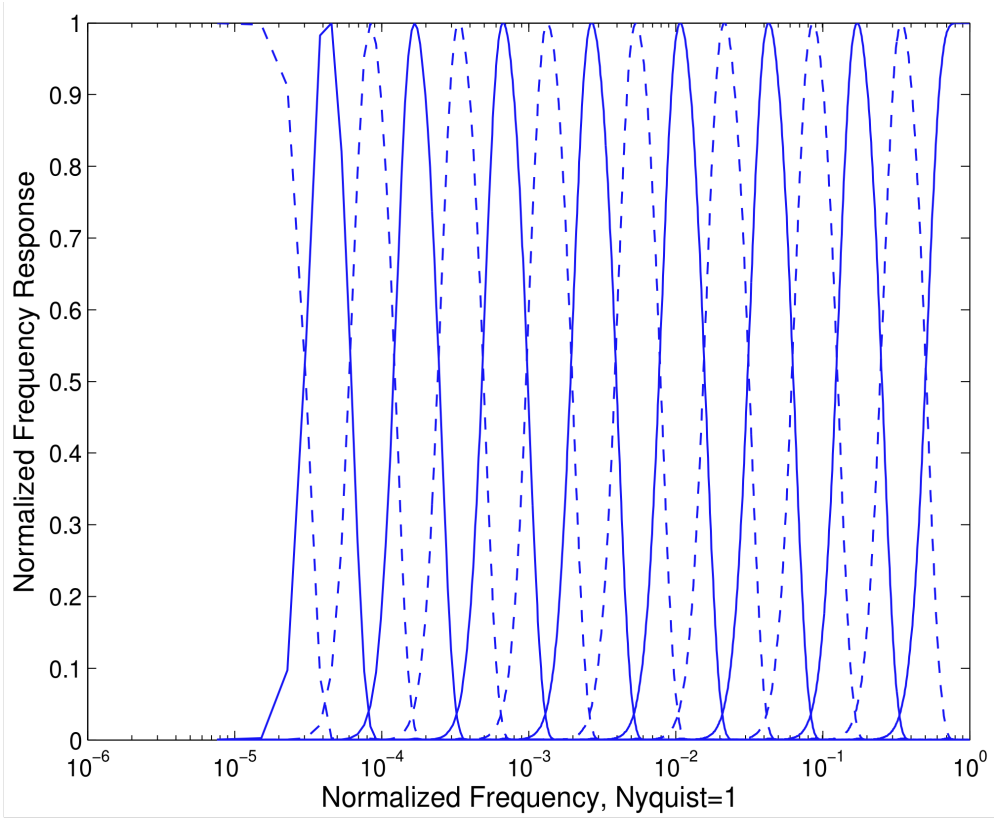
Kepler
A Search for Earth-size Planets

Filter-Bank Implementation of an Overcomplete Wavelet Transform

The time series *x(n)* is partitioned (filtered) into complementary channels

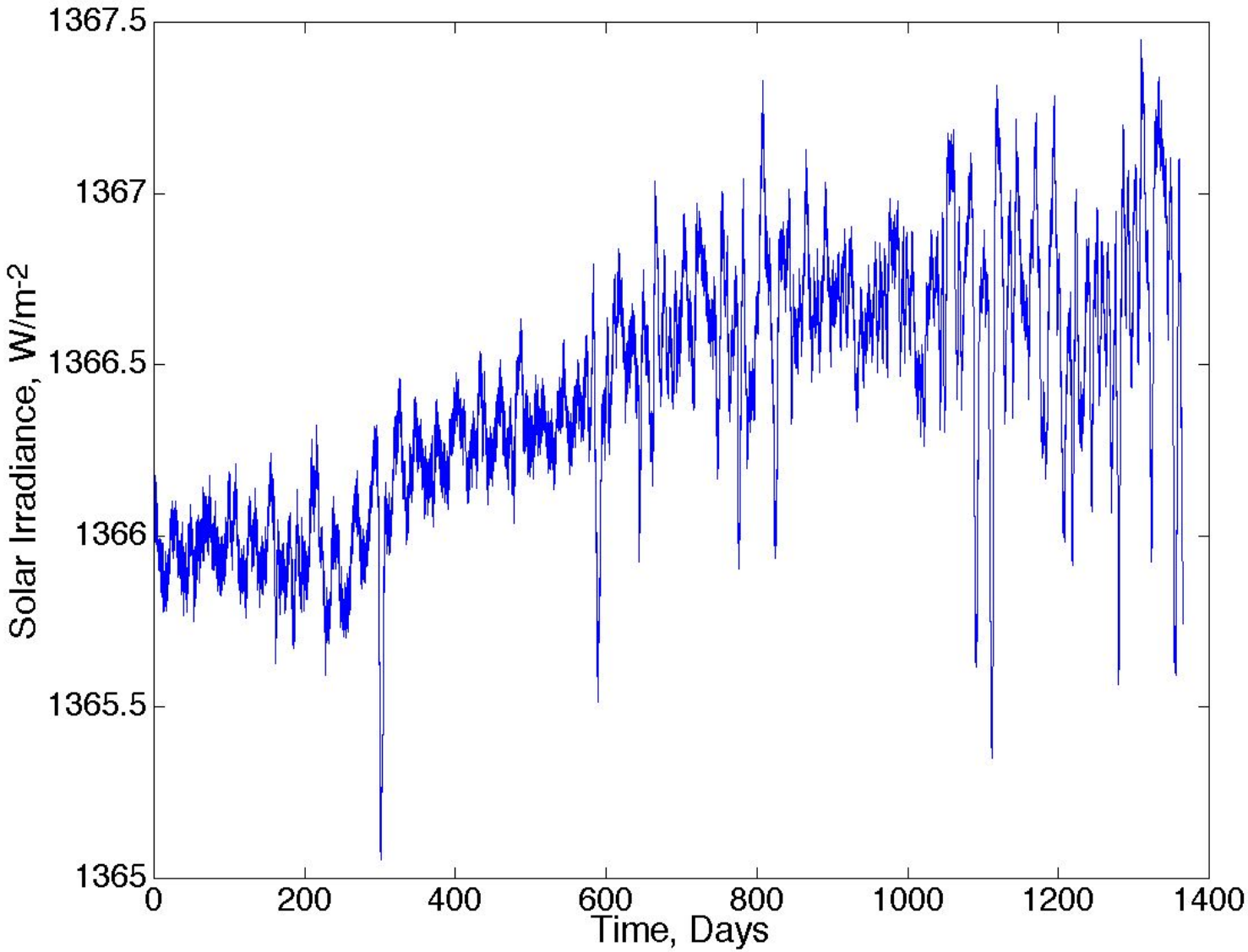
$$W_X(i,n) = \{h_1(n) * x(n), h_2(n) * x(n), \ldots, h_M(n) * x(n)\}$$
$$= \{x_1(n), x_2(n), \ldots, x_m(n)\}$$





Kepler-like Noise + Transits

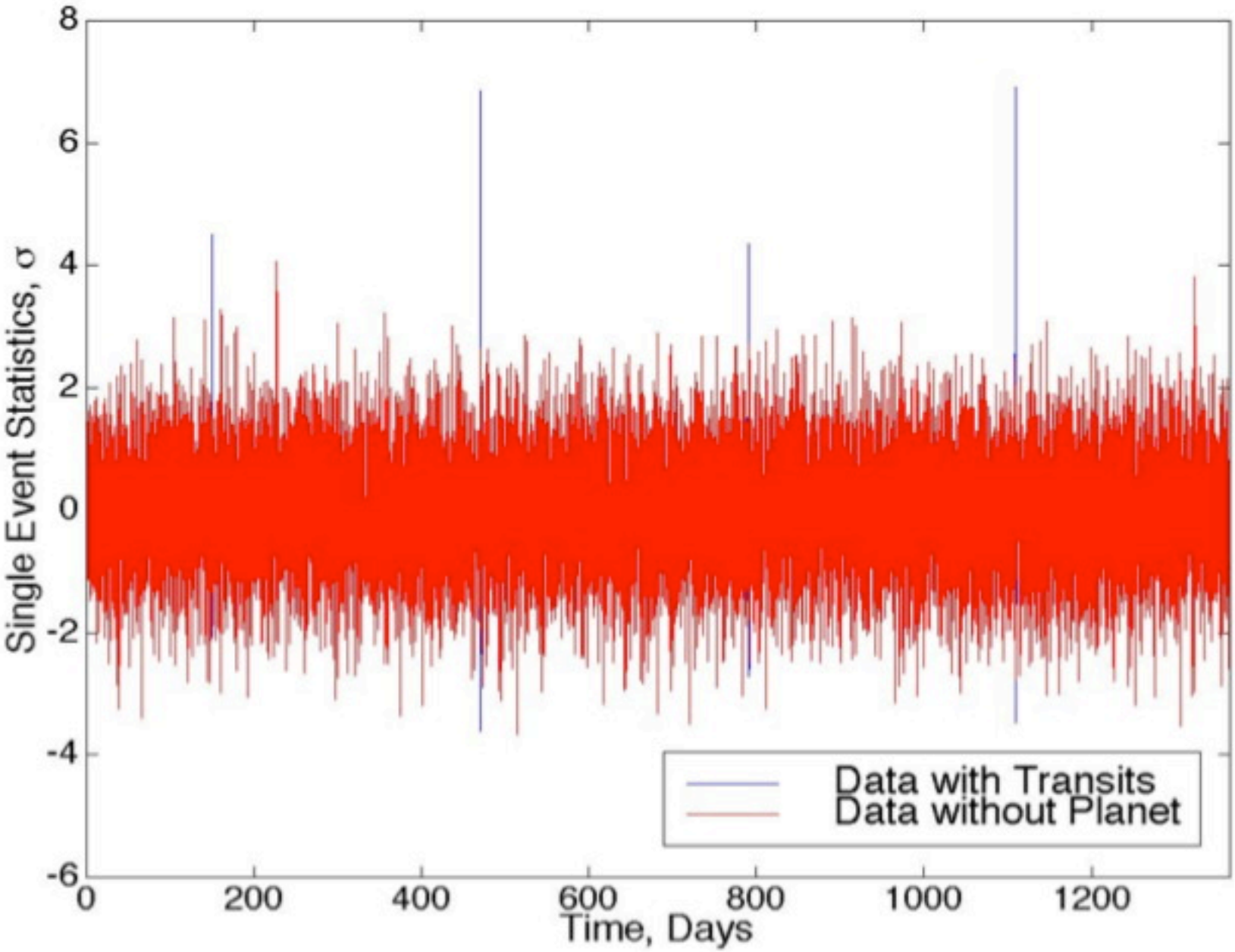
A Search for Earth-size Planets



Jon M. Jenkins

Single Transit Statistics

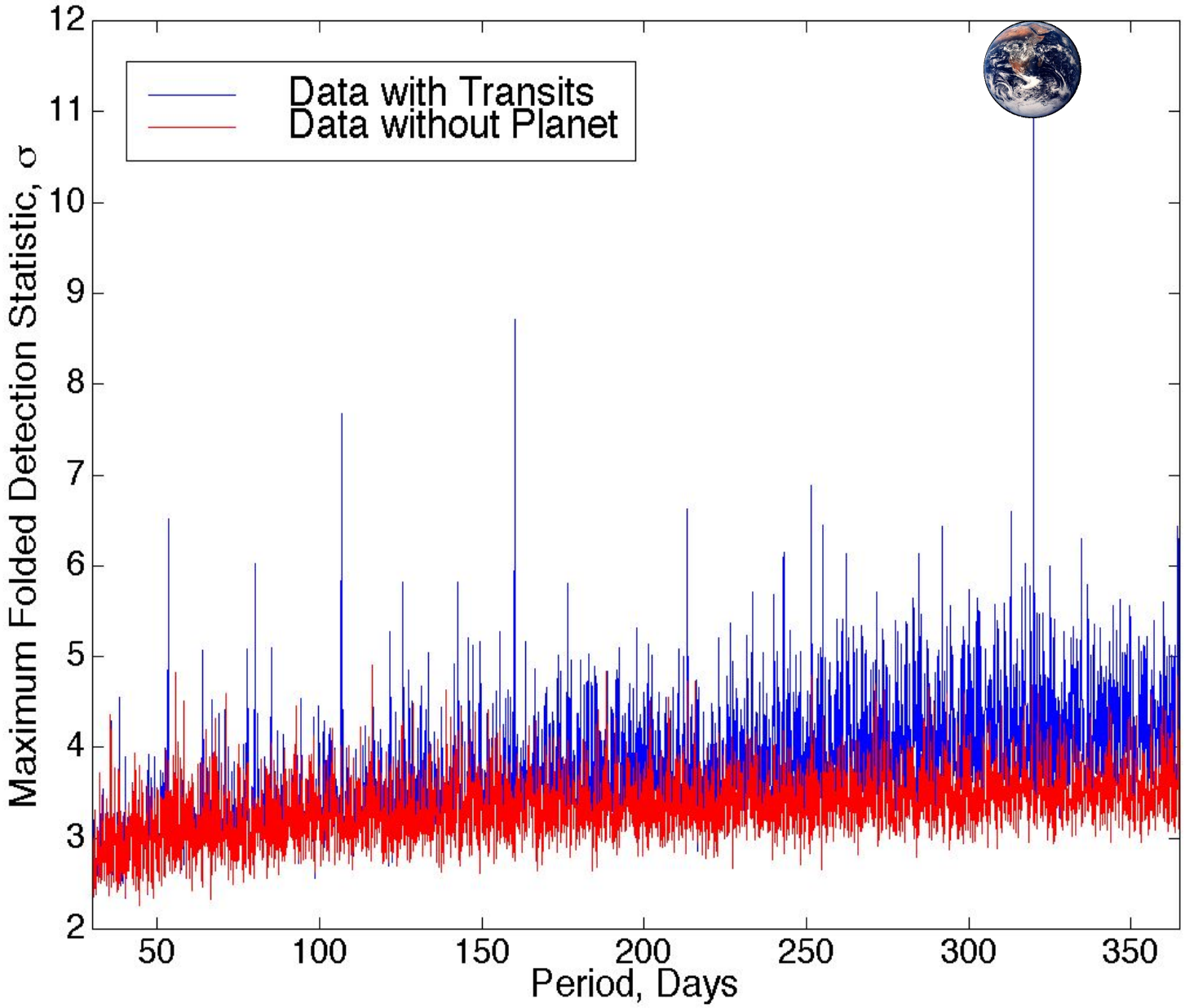
Kepler
A Search for Earth-size Planets



Jon M. Jenkins

Folded Transit Statistics

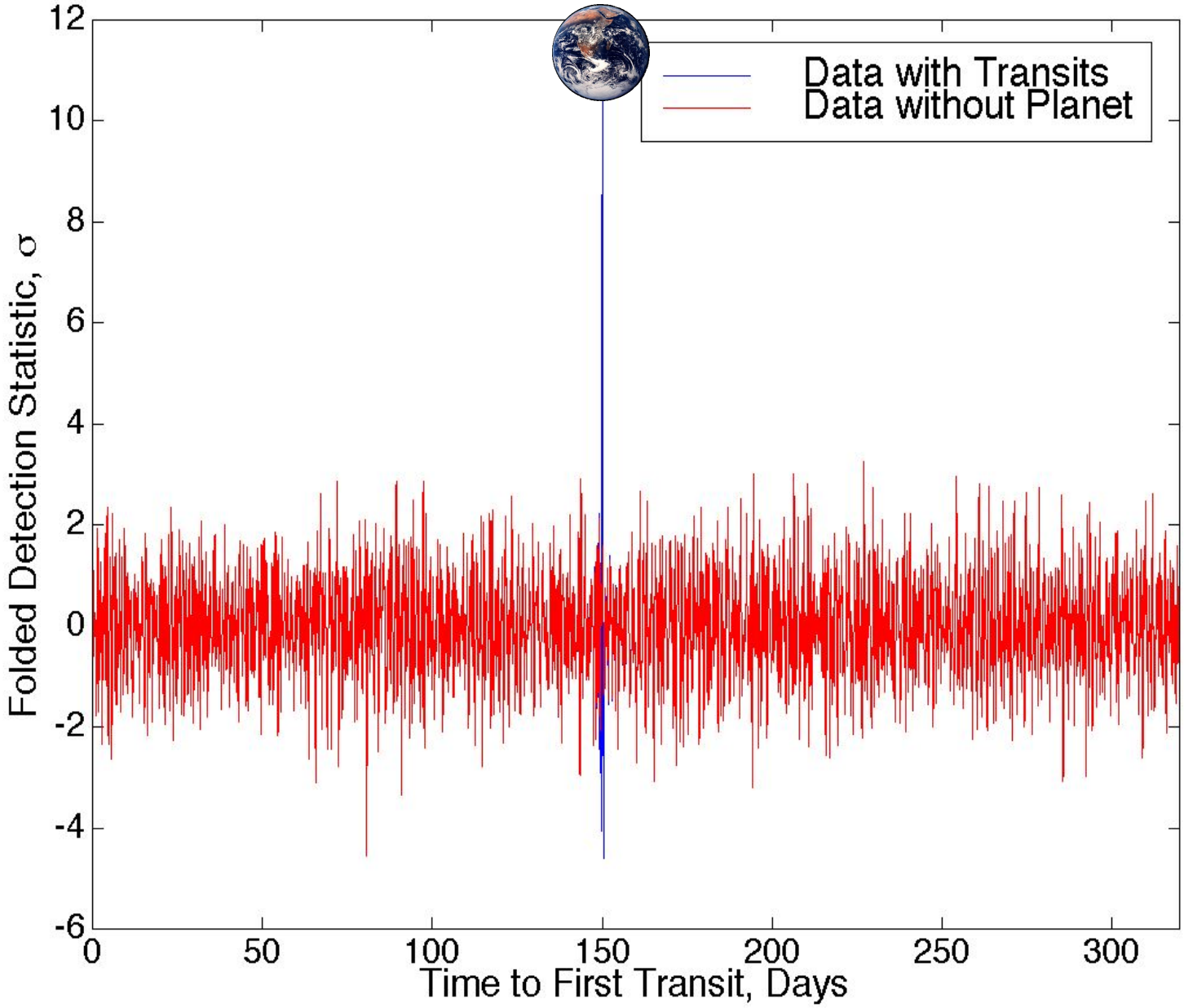
Kepler
A Search for Earth-size Planets



Jon M. Jenkins

Time to First Transit (Epoch)

Kepler
A Search for Earth-size Planets



Jon M. Jenkins

Summary

Kepler
A Search for Earth-size Planets

- *Kepler* has initiated an explosion of exoplanet research, effectively tripling the number of known exoplanets
- There are an astonishing number of multiple planet systems and these are remarkably coplanar – like the solar system
- *Kepler's* noise regime requires new algorithms to properly handle the huge dynamic range in stellar variability and the complex instrumental signatures
- Adaptive matched filters are required to account for non-stationary, non-white noise
- **We're just getting started: hold on to your hats!**

Managing Kepler's Data Volume

Kepler
A Search for Earth-size Planets

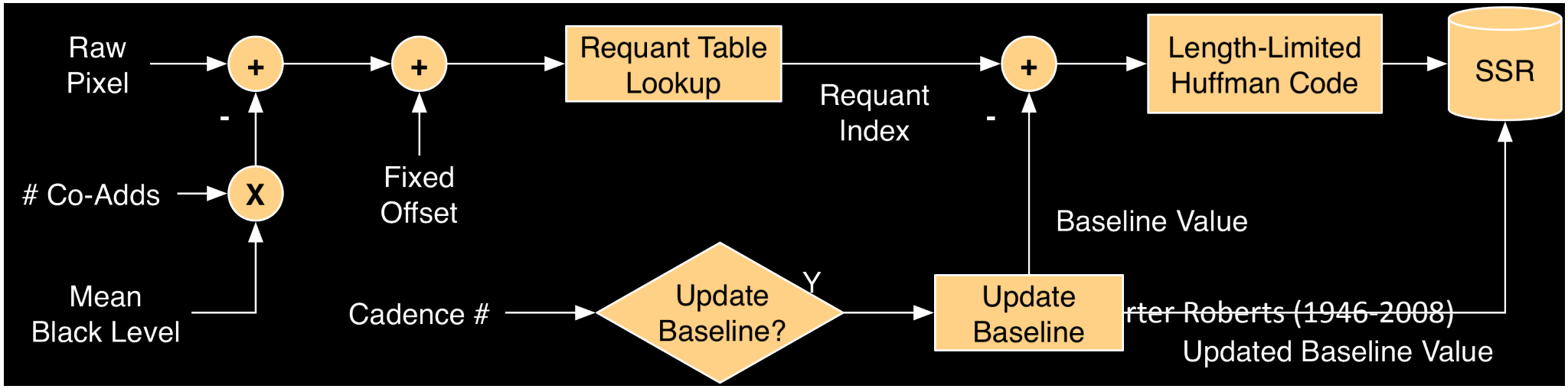
Key Requirements:

Collect and store image data for up to 170,000 stars for up to 66+ days

Downlink 31 days of data in less than 24 hours

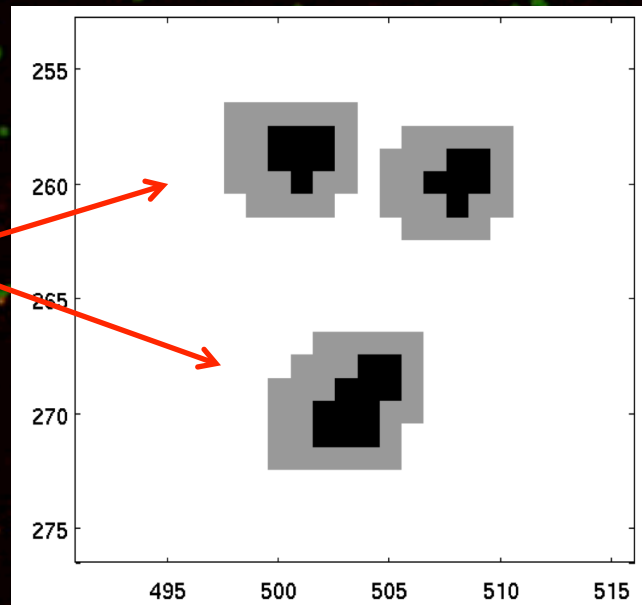
Solution:

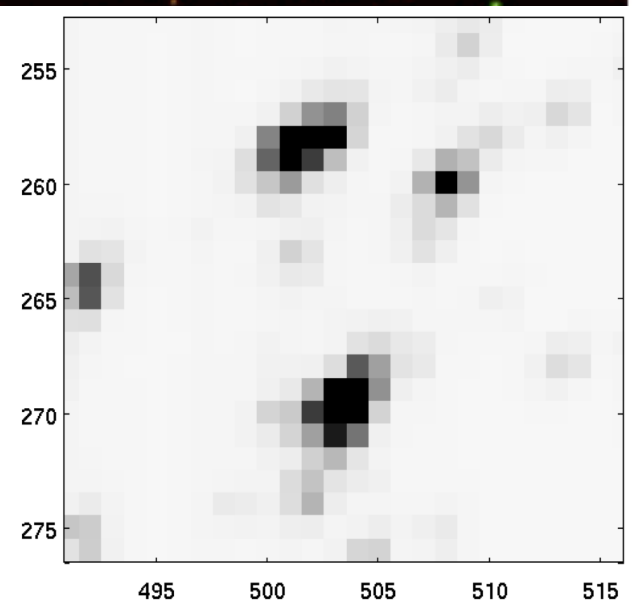
- Collect ***only*** pixels of interest
- Compand data to manage quantization noise as a term in noise budget
- Detrend the data to centralize the measurements to be transmitted
- Entropically encode the data before CCSDS packetization (must be robust)



Step 1: Selecting Pixels of Interest

Pixels of interest specified for downlink





Generate synthetic images of each star and its background scene

Perform SNR calculation to identify pixels of interest for photometry

Add a 1-pixel buffer ring

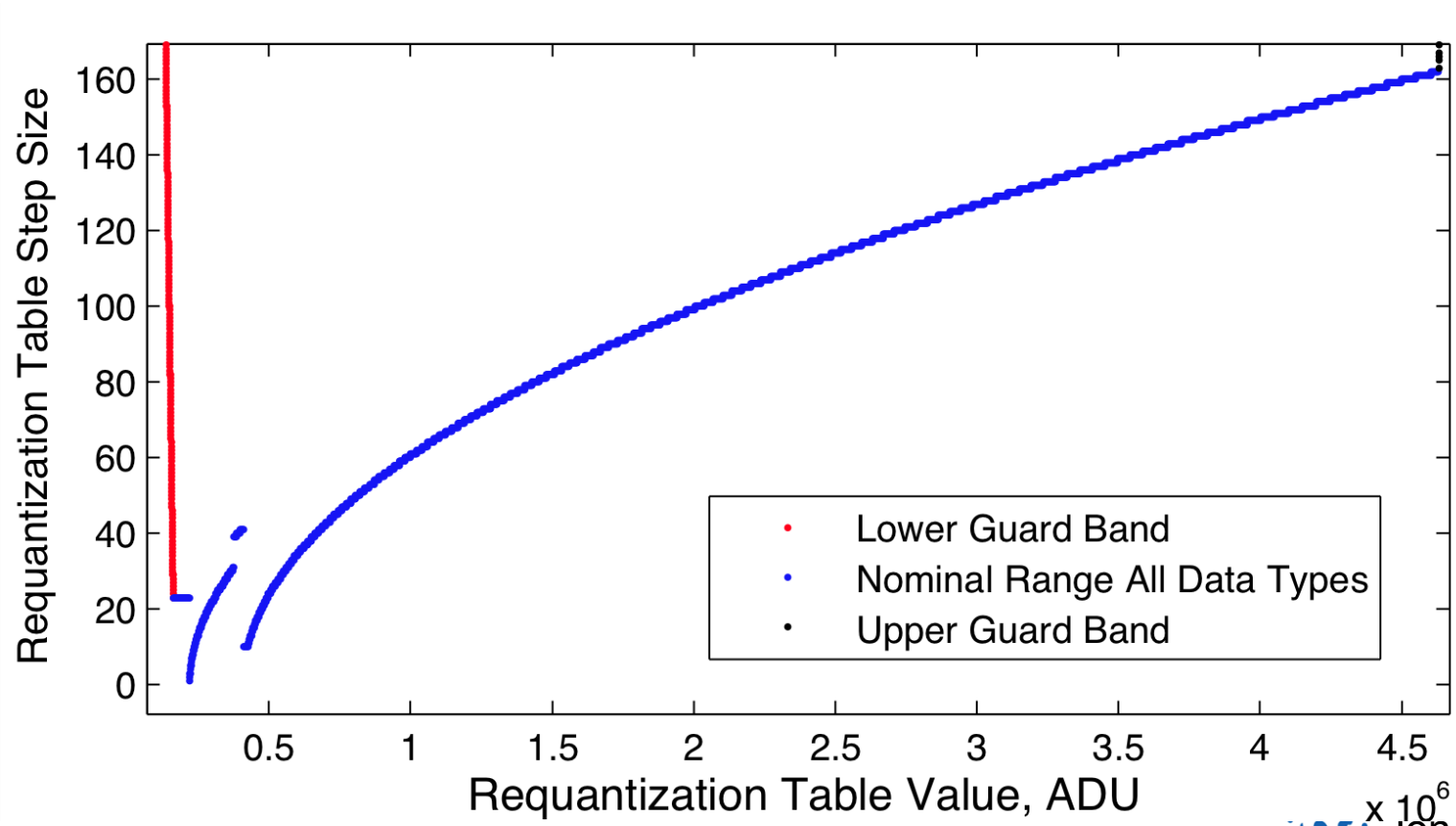
Image by Carter Roberts (1946-2008)

Choose smallest of 1024 masks that will cover POI

Only 4% excess pixels collected

Step 2: Requantization

- Manage quantization noise in noise budget – ¼ of intrinsic measurement noise
- Must account for dispersion of read noise, gain, and bias across 84 channels!
- Raw pixel measurements are represented with 23 bits
- Requantization compresses the data to 16 bits



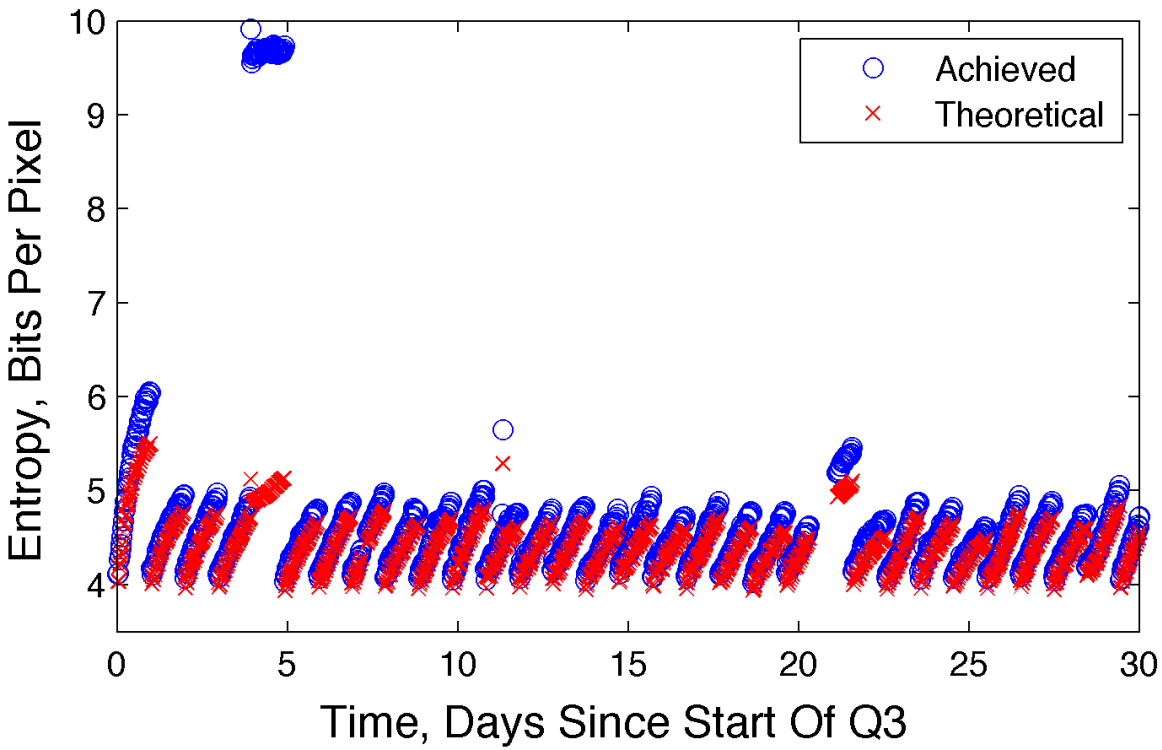
Step 3: Huffman Coding

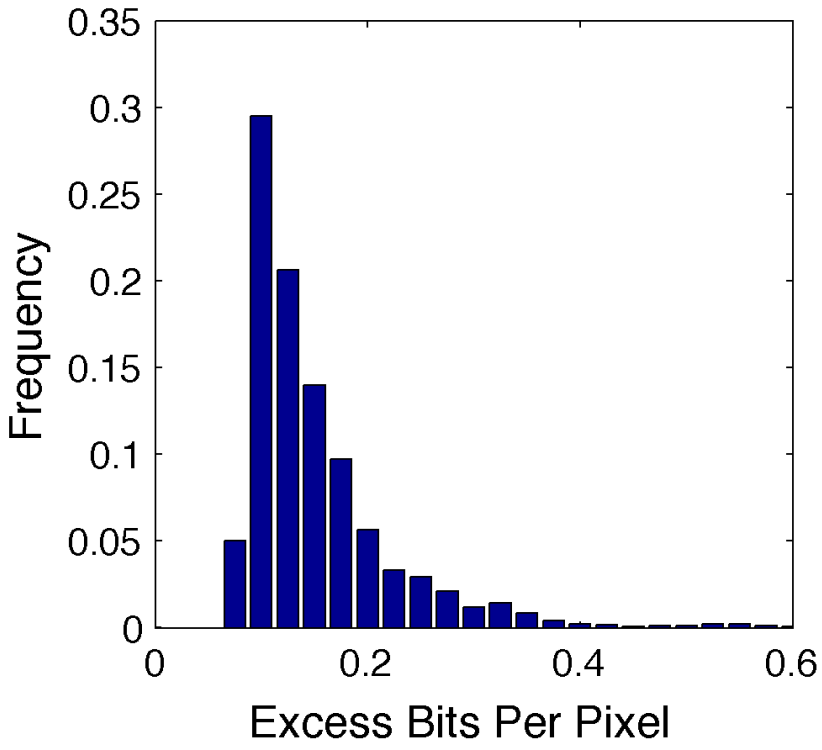
Kepler
A Search for Earth-size Planets

Subtract a baseline measurement (changes on a daily basis)

Entropically encode residuals with a Length-Limited Huffman Code

Compresses the data to 4.4 bits/symbol





Watch the Transit of Venus on June 5th!

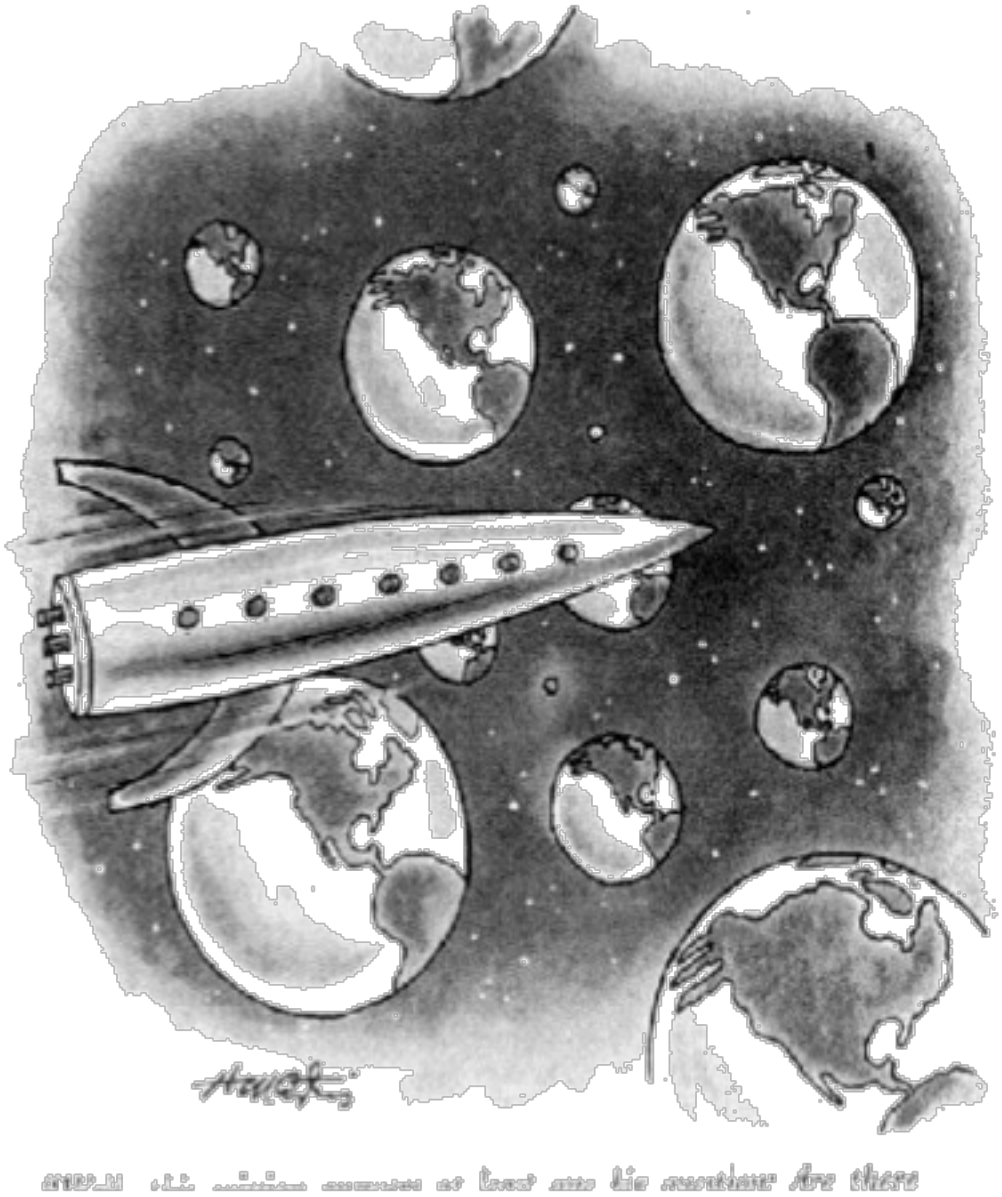
Kepler
A Search for Earth-size Planets

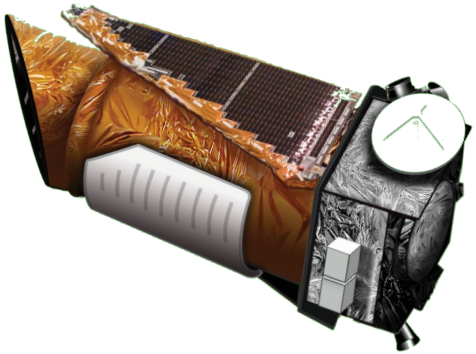
Image Credit: Jan Herold/Wikipedia (CC BY-SA 3.0)

Jon M. Jenkins
SETI Institute

The Future!?!

Kepler
A Search for Earth-size Planets





"Well, this mission answers at least one big question: Are there other planets like ours in the universe?"

Drawing by H. Martin; © 1991
The New Yorker Magazine, Inc.