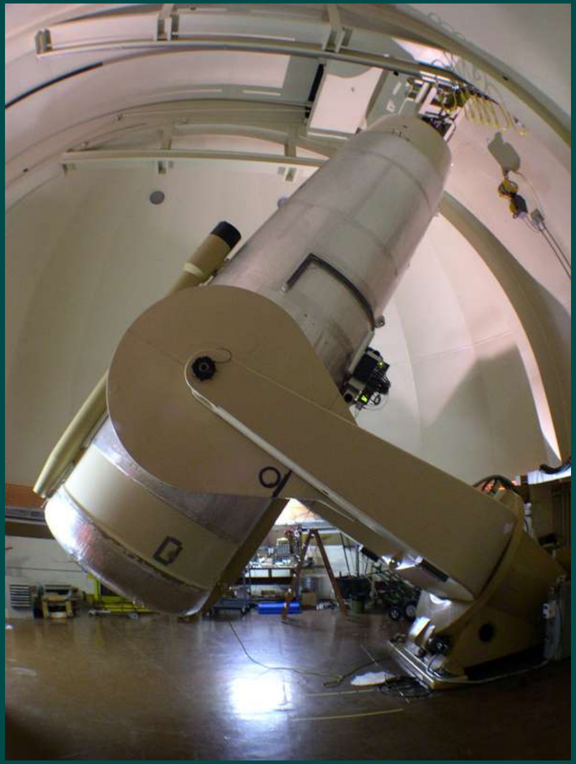
Feeding the Pipeline: The SNfactory Supernova Search



Richard Scalzo

NERSC User Group Meeting

October 4, 2005

Outline

Background

- Interest in supernova science
- Observational challenges in finding supernovae

The SNfactory search pipeline

- Description of hardware
- Past searches and challenges in development
- Present and future

Why supernovae are interesting

Two types of SNe: "type Ia" and "core-collapse".

SNe Ia

- Model: Thermonuclear explosion of degenerate star(s) (carbon-oxygen white dwarf + binary companion star)
- Characteristic Si II absorption line in spectrum
- Brightness very uniform → good distance indicators

Other SNe (types Ib, Ic, II)

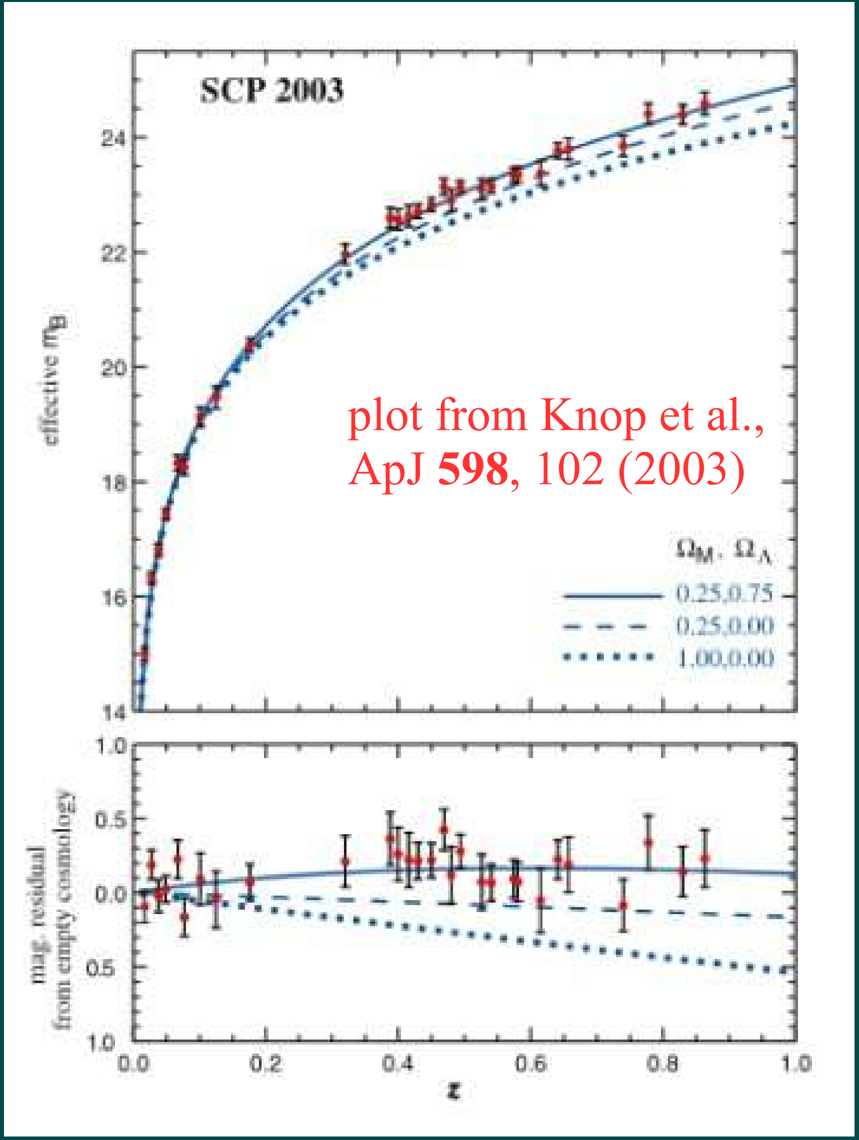
- Gravitational collapse of core of young, massive star
- Spectra are diverse; evidence of stellar envelope (H, He)
- Rate tracks star formation rate; may be GRB progenitors

SNe Ia as cosmological "std. candles"

Discovery of "dark energy" and accelerated expansion depended on measurements of brightness of SN Ia.

Want to measure: density (Ω_Λ), equation of state (w) and its rate of change (w').

Currently very few nearby Hubble-flow SNe Ia used in cosmological fits. Will also need to control systematics of brightness calibration.



SNe Ia as cosmological "std. candles"

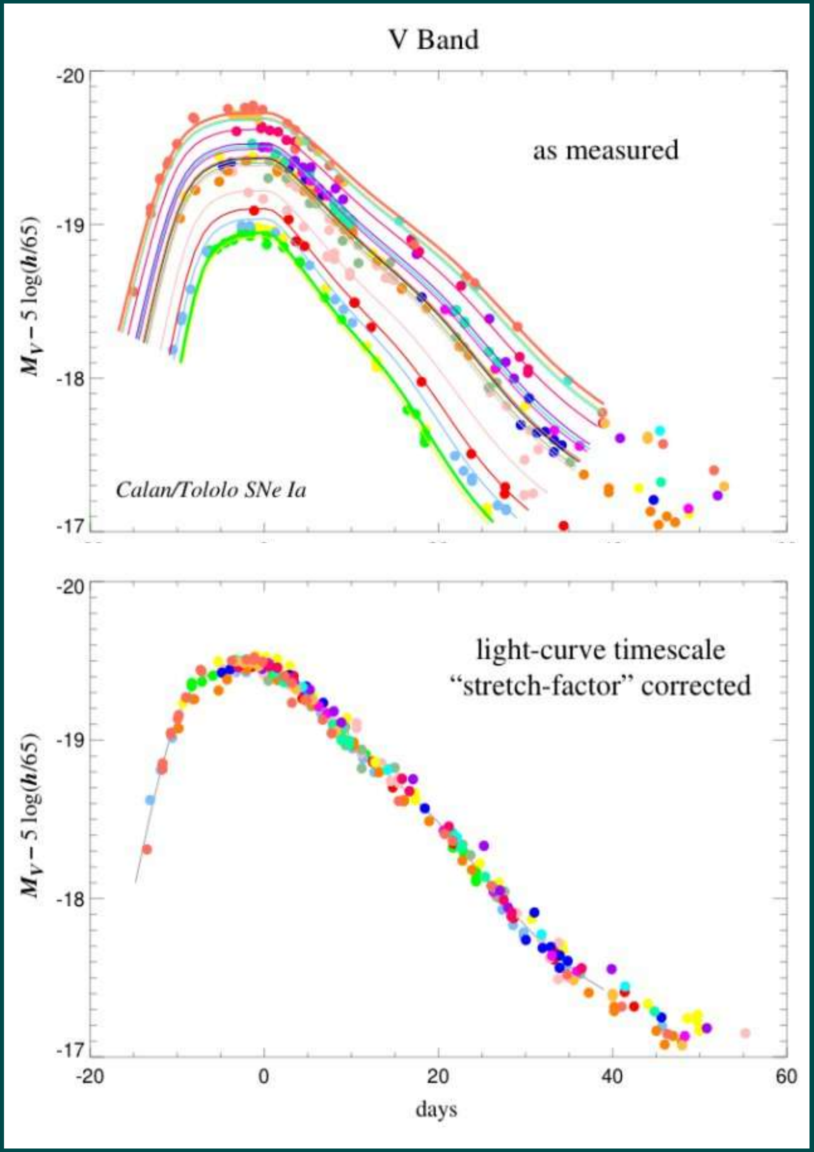
Different SNe Ia have same maximum brightness to 40%. B-band light curves obey a 1-parameter scaling relation, reducing scatter to 15%.

Are there other parameters?

What's the physical origin of remaining small variation?

Details of Ia progenitors?

How best to correct for dust extinction, evolution with redshift, other systematics?

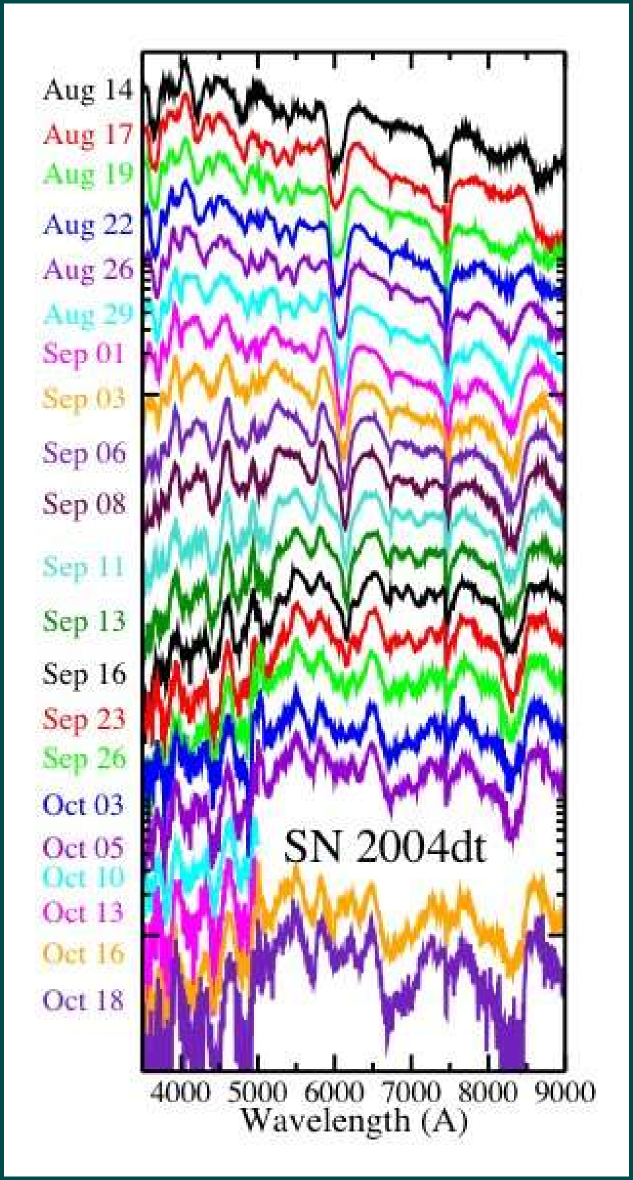


The Nearby Supernova Factory

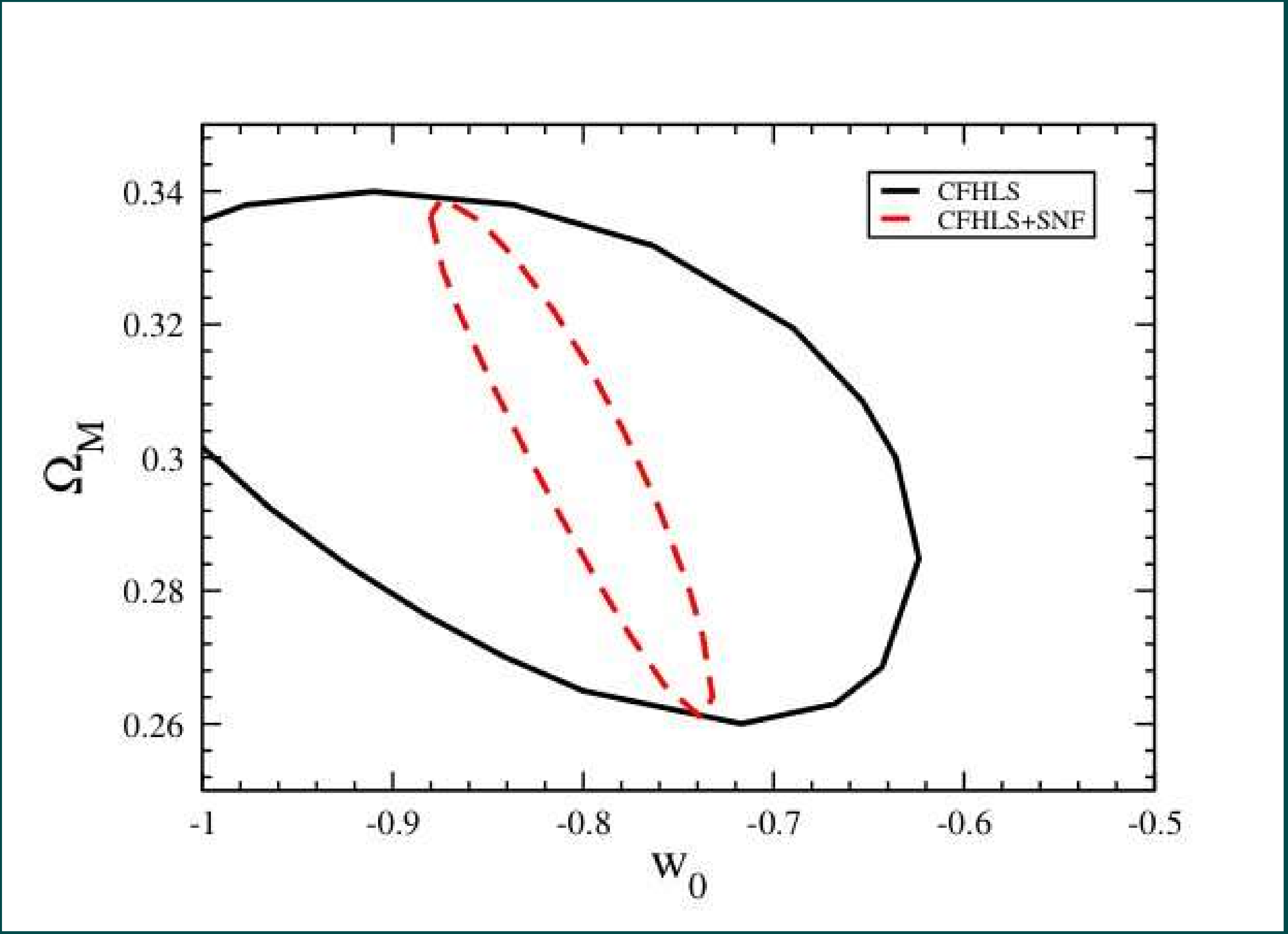
. . . will produce a large, high-quality data set of ~300 low-redshift SNe Ia over the next few years.

- More low-z SNe in Hubble plot, tighter constraints on cosmology
- Improved brightness calibration of SNe Ia for future experiments
- Strong constraints on SN Ia progenitors & explosion models

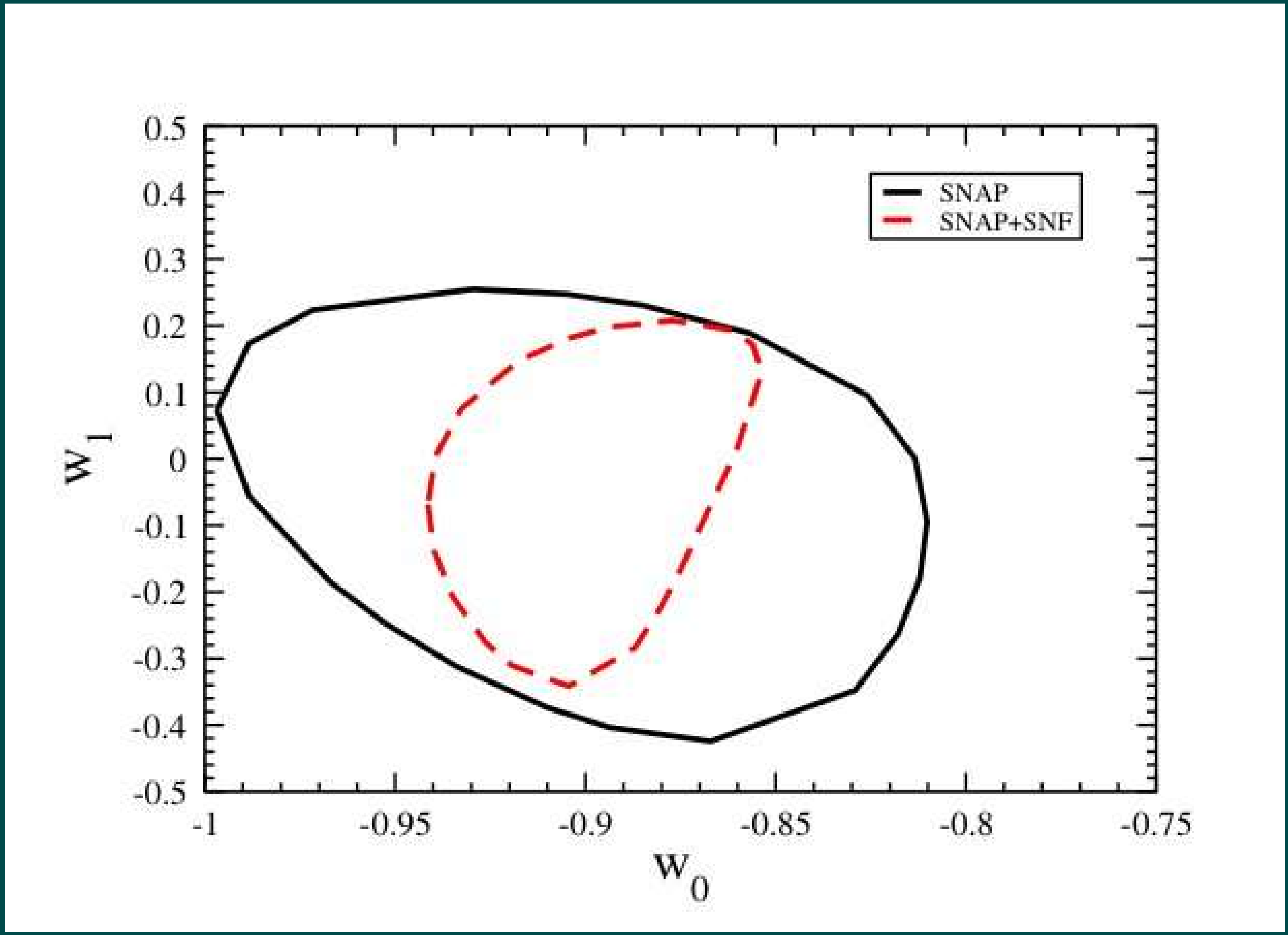
Shown at right: spectral time series (real data!). In final data product the spectra will be flux-calibrated.



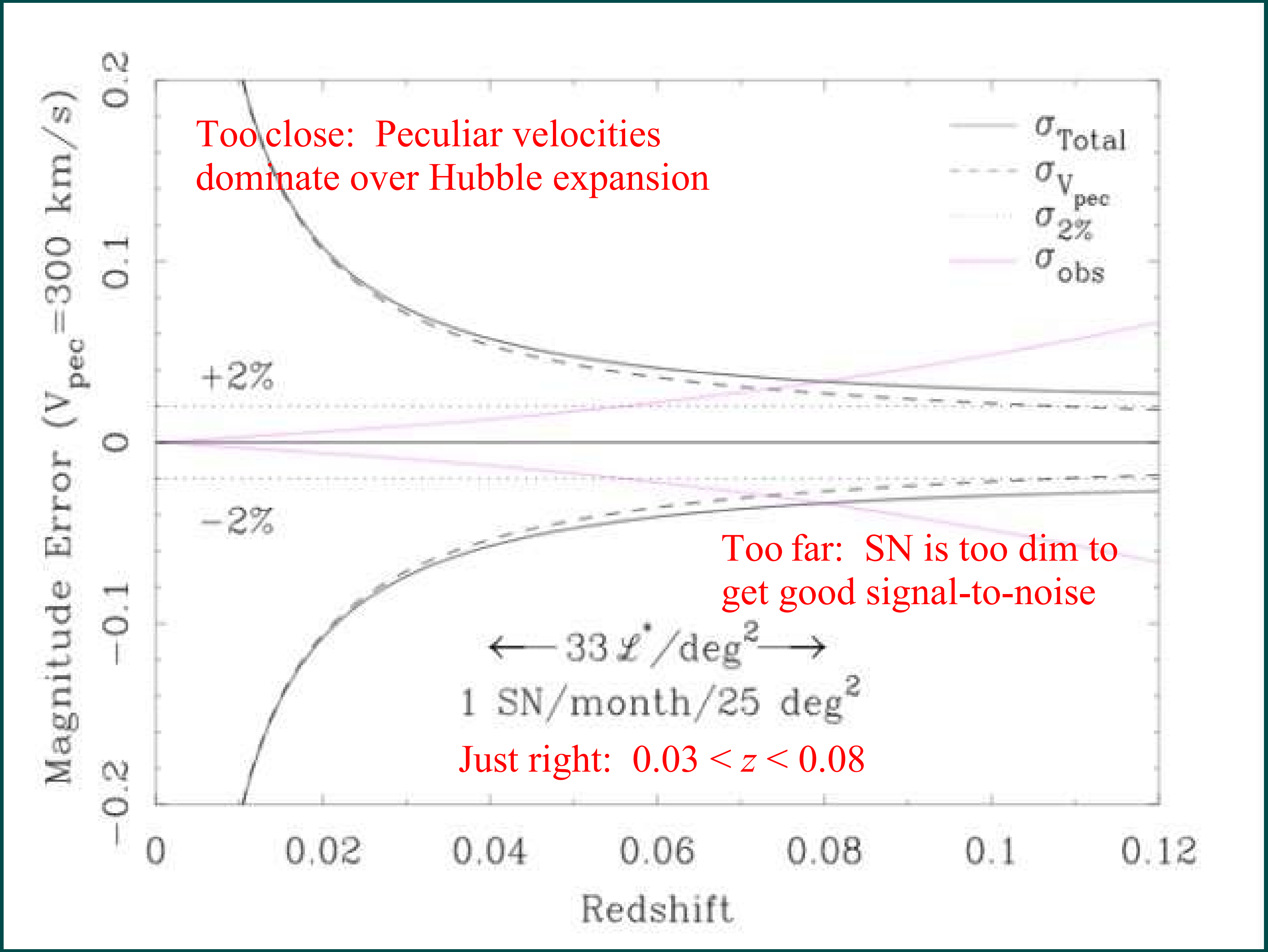
Statistical impact of SNfactory



Statistical impact of SNfactory



What's a "nearby" supernova?



Why are SNe so hard to find?

- *They're rare!* Only about 0.001-0.01/yr/galaxy.
- *They're transient!* Width of light curve peak is ~ 10 days in rest frame of SN ($z << 1$).
- *We need to catch them early!* For SNfactory science goals, we should catch the SNe *before* they reach their maximum brightness, to study the whole light curve.
- *Large background:* asteroids, variable stars, AGN, . . .

High-redshift SN searches are usually done by eye on very deep images, and followed up with HST.

Low redshift → small volume → need a *lot* more area. Therefore the search *must* be automated.

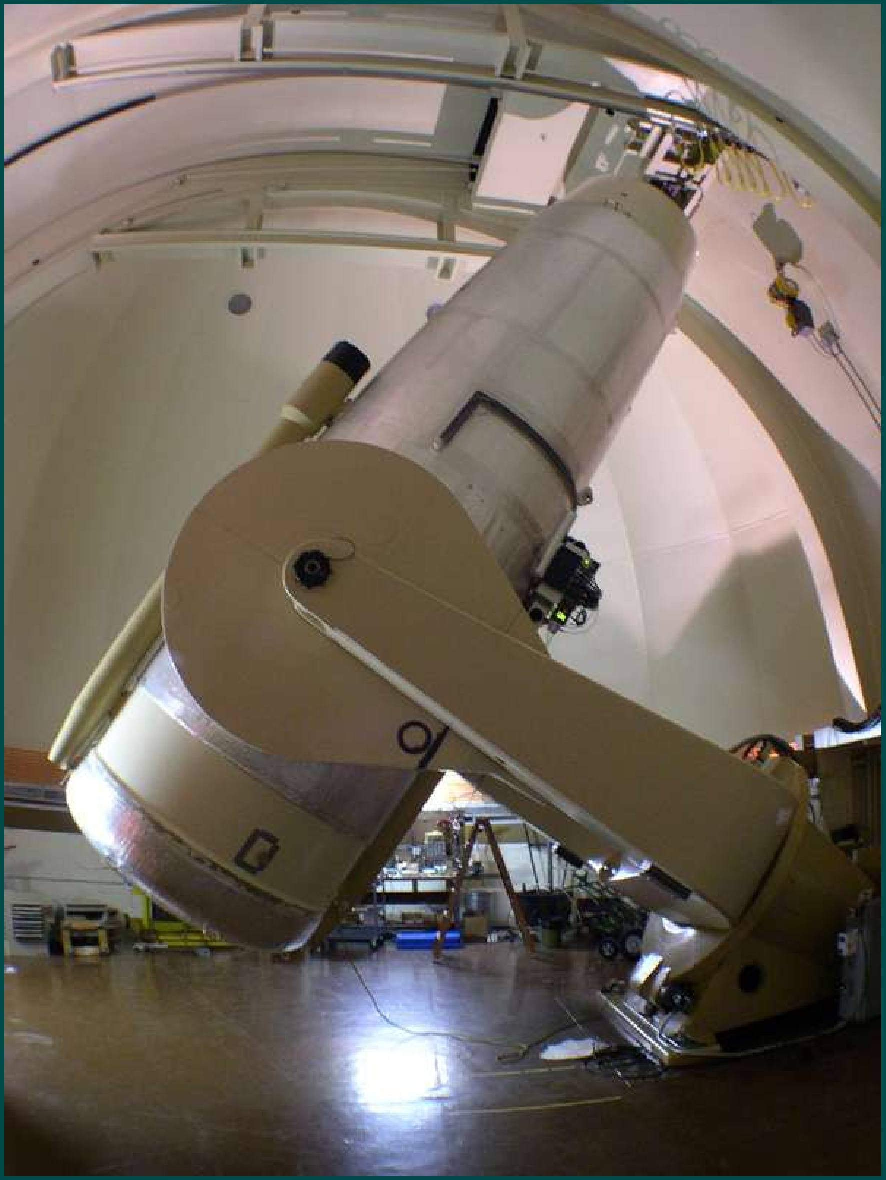
The SNfactory supernova search

Uses a small ground-based telescope on Mt. Palomar: the Samuel Oschin 1.2-m.

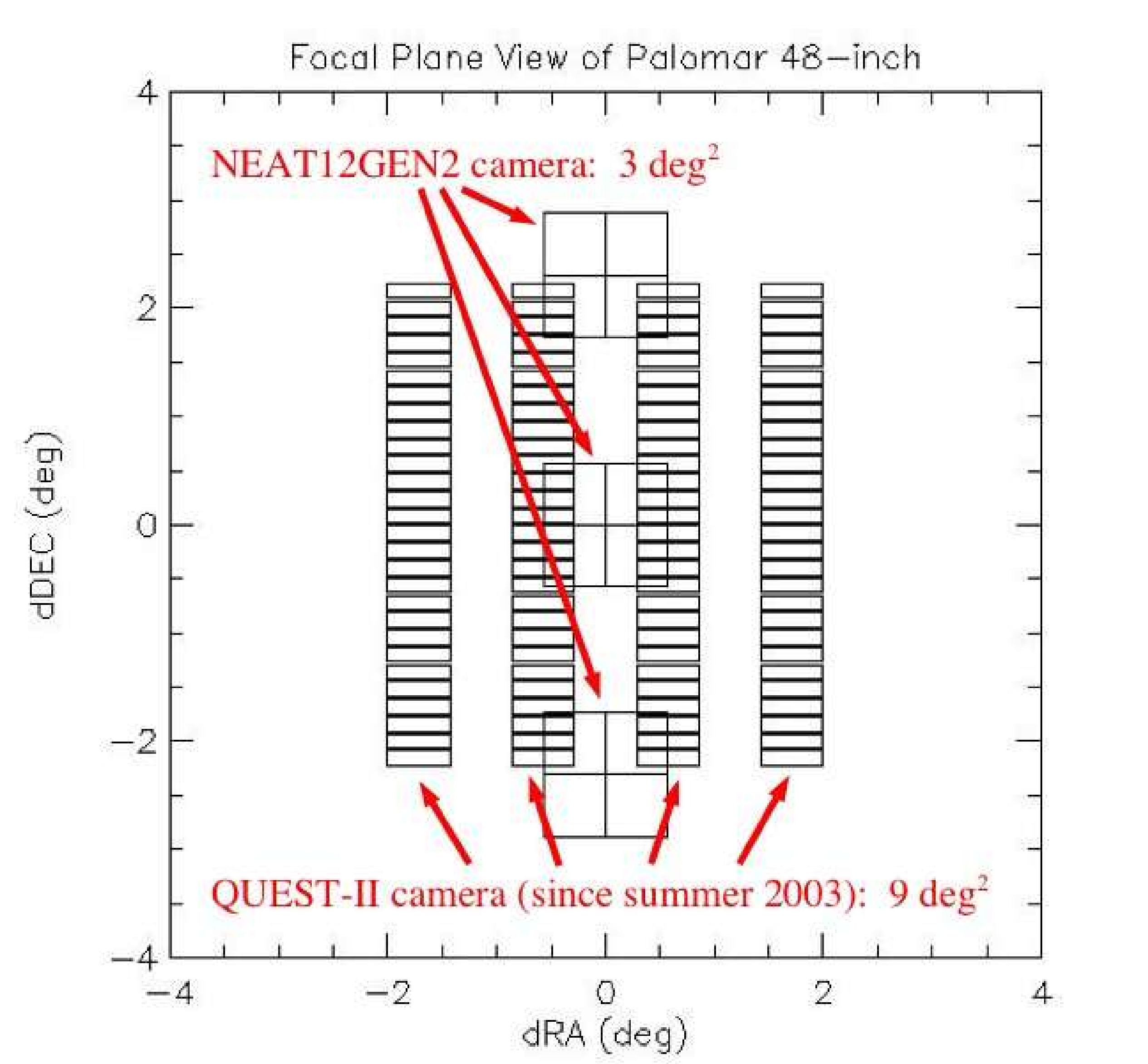
Time on the telescope is split between two groups, NEAT and QUEST, who give us access to their data for the supernova search.

Up to July 2003: NEAT12GEN2 camera

Present: QUEST-II camera built at Yale.



Focal plane of the 1.2m



Where the search data comes from

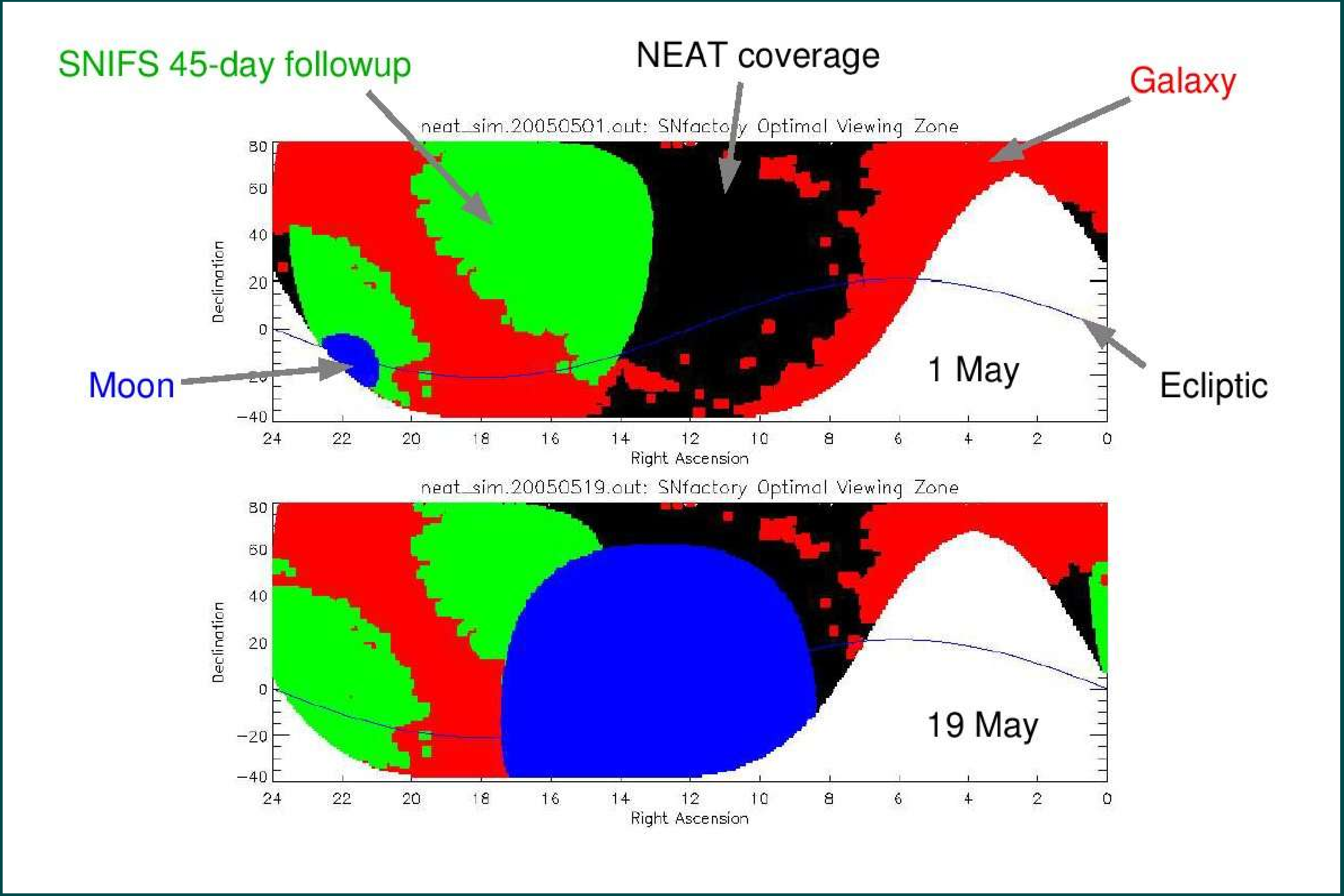
Near Earth Asteroid Tracker (NEAT): ~60% time.

- Searching for asteroids and minor planets
- They use the QUEST-II camera in point-and-track mode with a red (RG-610) filter.
- 30000 images / 500 deg^2 / 85 GB of data per night

QUasar Equatorial Survey Team (QUEST): ~40% time.

- Searching mainly for quasars; but also core-collapse SNe
- They use the camera in drift-scan mode with a different filter (e.g. *R, I, B, U*) on each of the four "fingers".
- 16000 images / 300 deg^2 / 45 GB of data per night

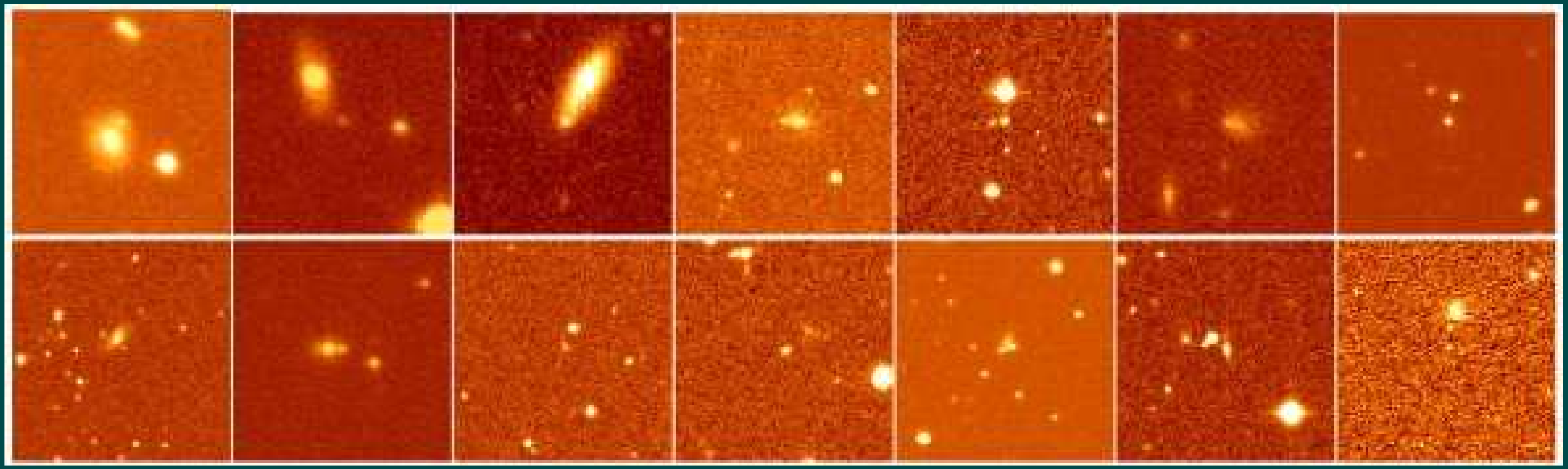
Where we look in the sky



Prototype search (up to July 2003)

Developed and run on PDSF. (W. M. Wood-Vasey, Ph.D. thesis 2004)

- Used the original NEAT12GEN2 camera ("3-banger"): 3 square degrees, 3 CCDs (1 image = 60' × 60')
- 750 images / 250 deg^2 / 24 GB per night
- Very successful search – found 80 SNe in the first year of running. Proof of principle for search algorithms.



Scaling up from the prototype

- Factor of 3 more area (3 $deg^2 \rightarrow$ 9 deg^2)
- Factor of 37 more images (3 CCDs → 112 CCDs)
- New images are long and skinny; bad edge effects when trying to automatically identify star fields.
- Problems discovered when trying to scale up:
  - Code base for software originally intended for *manual, collaborative* searches – not a pipeline!
  - Heavy dependence on NFS, database, tape storage made search slow and prone to erratic crashes
- Challenge: Search 80 GB of data in less than 24 hours with 100% efficiency.

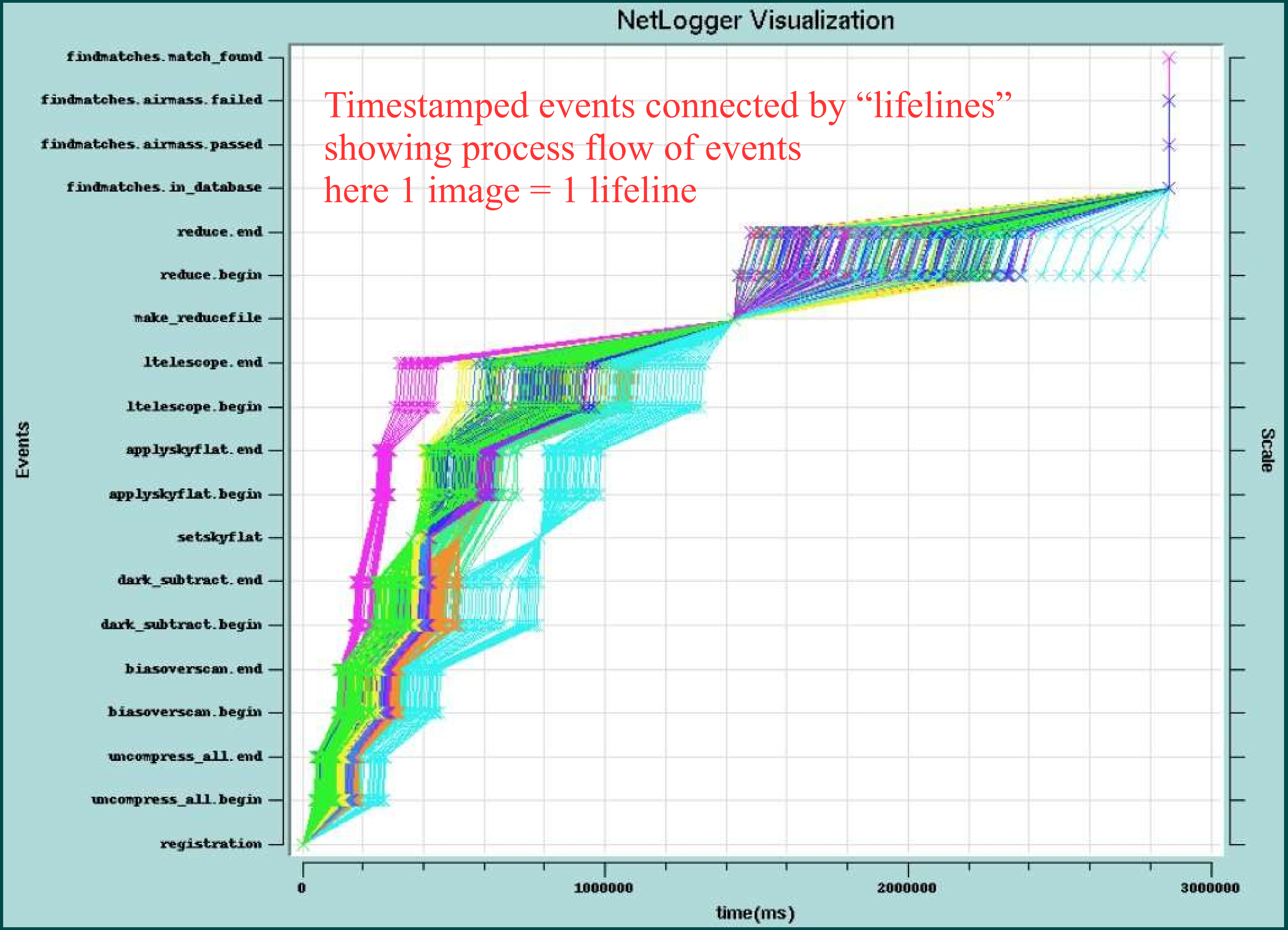
NetLogger to the rescue!

Methodology & toolkit for monitoring status of distributed applications. (B. Tierney & D. Gunter, LBNL Tech Report LBNL-51276)

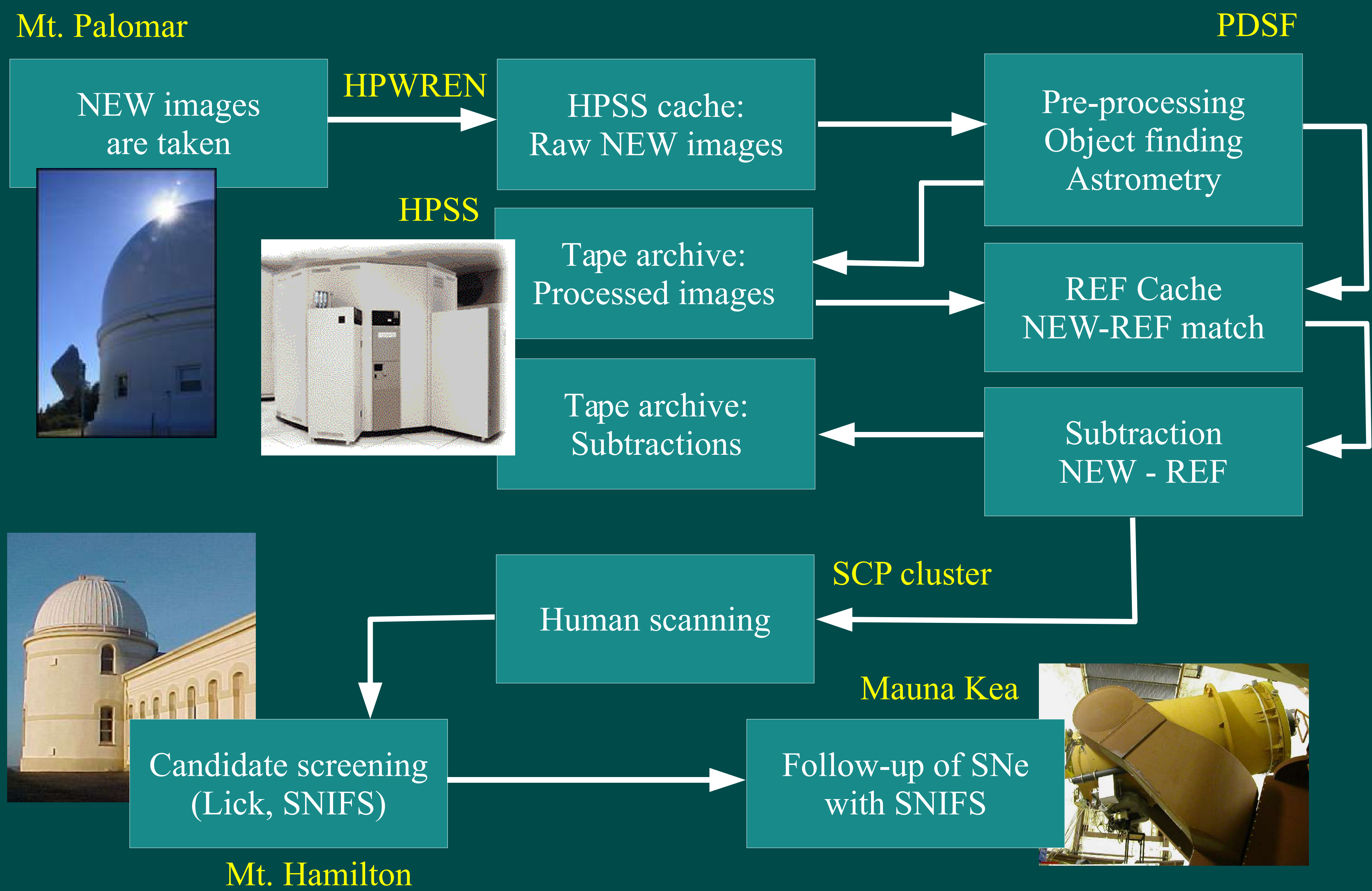
- Timestamped messages written out asynchronously at critical system events, collected by daemon
- Visualization tool shows:
  - Did a process crash?
  - If so: At what time, in what stage, on which node?
  - How long did each stage take to run?
- Use information to tune and debug pipeline.

We liked Netlogger and helped developers to test it.

Example NetLogger status screen



Outline of the pipeline



Downloading from Mt. Palomar

High Performance Wireless Research and Education Network (HPWREN) at UCSD

(supported by NSF grants 0087344, 0426879)

45 Mbps off the mountain. NEAT and QUEST use about 10% of that to transfer files to HPSS as images are read out, approximately in real time.

Pre-processing

The next morning, raw images are pre-processed on PDSF.

- *Fetch images* from HPSS using `htar`.
- *Uncompress* images (`gunzip`) onto local scratch.
- *Subtract bias & dark current:* remove instrumental background (zero-point offset for pixel values).
- *Build a "sky flat":* measure pixel-to-pixel gain variations across the camera. The night sky is assumed to be a surface of uniform brightness.
- *Flat-field:* divide out pixel-to-pixel gain variations.
- *Register image* in database and copy to shared disk.

Reference cache

To find a new transient object, you need to compare a recent picture of the sky with one taken before the object appeared.

We keep a library of reference images *(REFs)* on HPSS: every image the SNfactory has ever received (40+ TB).

HPSS metadata queries are expensive, so retrieving REFs one by one is very slow.

A local cache (1.5 TB) is kept on PDSF disk vault. We cache only REFs corresponding to sky well above horizon.

The cache is updated daily.

Object finding and astrometry

Next we want to know what objects are visible in the image and establish an accurate (< 1 arcsec) coordinate system.

- *Surface image* (`edgesurface`): remove spatial variations in night sky background.
- *Find objects* (`isofind`): trace contours of constant surface brightness, find moments of light distribution.
- Identify and mask out *saturated stars* and *CCD defects*
- Retrieve list of USNO catalog stars overlapping w/image
- *Astrometric matching* (`quickmatch`): find constellations of catalog stars in the image and use their relative positions to calculate a coordinate transformation.

Association of NEWs and REFs

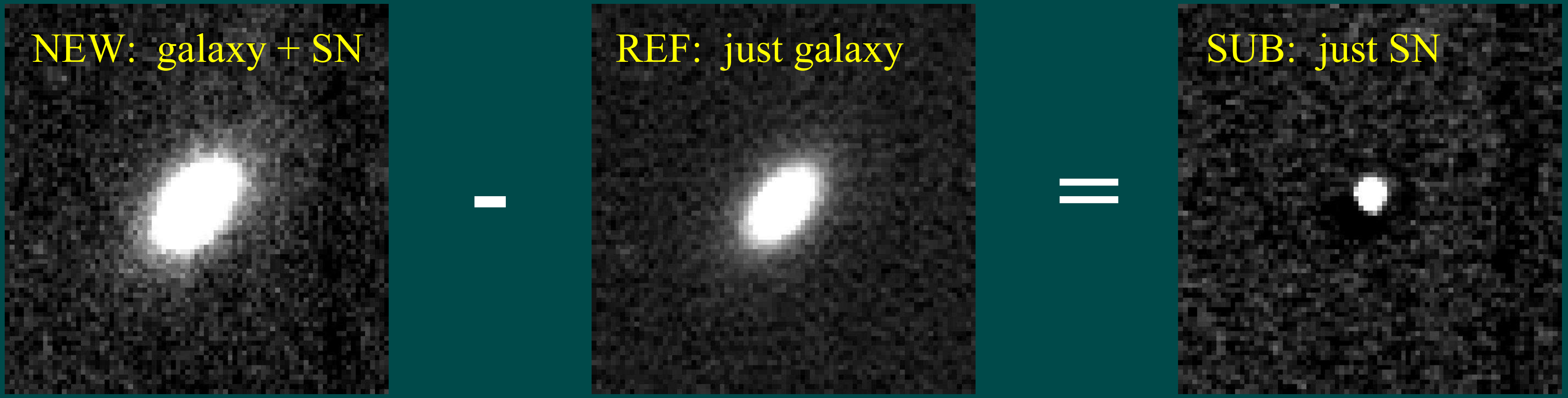
Intermediate step: Find which REF images correspond to which NEW images.

- Use the image corners to see which images line up
- Use only images which have had astrometry calculated
- Require at least 2 NEW images (for quality control)
- Require at least 3 REF images (for area, depth)

Danger we found: Often telescope pointing error prevents long, skinny QUEST-II images from lining up with each other. For now, use only NEAT12GEN2 images as REFs.

Finding SNe by image subtraction

One convenient way to compare NEW and REF is to subtract the REF from the NEW:



For this to work, we need to normalize the depth and the "seeing" (PSF) of the REF and NEW images to each other. This is not as easy as we would like to make it look!

Our subtraction process

Many steps (here we use two NEW images):

- Quality vetting on all images (NEW1, NEW2 and REFs)
- Transform all REFs to NEW1 coordinate system
- Reject asteroids by their motion NEW1 → NEW2
- Co-add NEW1 and NEW2 to form a single NEW
- "Union co-add" of REFs to form a single large REF
- Normalize NEW and REF stars to same mean brightness
- Convolve NEW and REF images with Gaussians to blur PSFs to worst common denominator
- Actually perform the subtraction.

“Autoscanning”

After subtracting, find leftover objects in subtracted images and calculate numerical “scores” for them. They might be:

- Catalog objects (stars or galaxies) which saturated in one of the images: Mask these out.
- Catalog objects for which the PSF or object flux was not well-matched (these are misshapen or look like donuts)
- Cosmic rays: Rare on ground, look like single bad pixels
- Detector artifacts (“ghosts”, “LEDs”, stuck shutter etc.)
- Non-SN variable objects (asteroids, CVs, AGN, . . .)

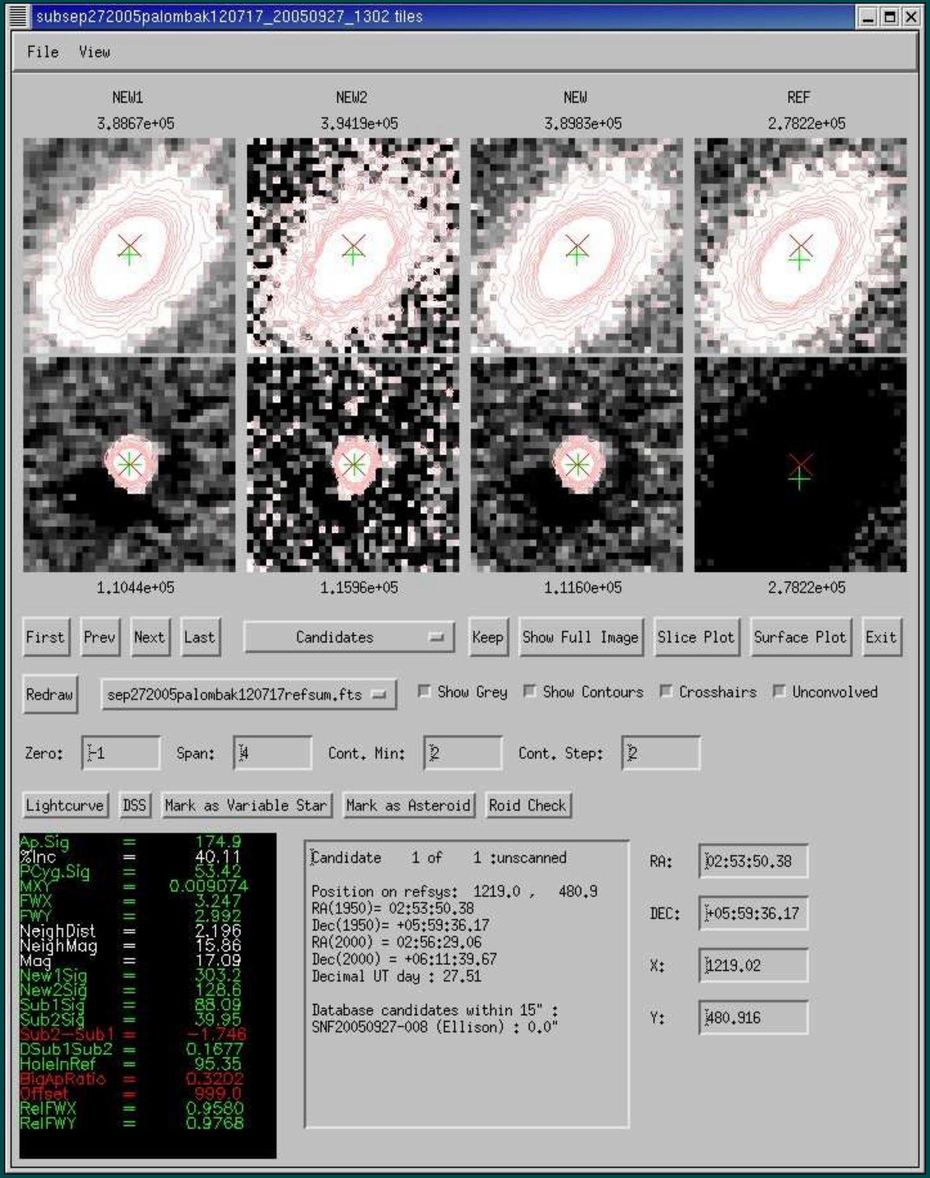
At present, we still need humans to tell these from real SNe.

Human scanning

Images are now copied from PDSF to cluster of SCP machines at LBL.

A simple GUI interface allows a trained scanner to examine each candidate:

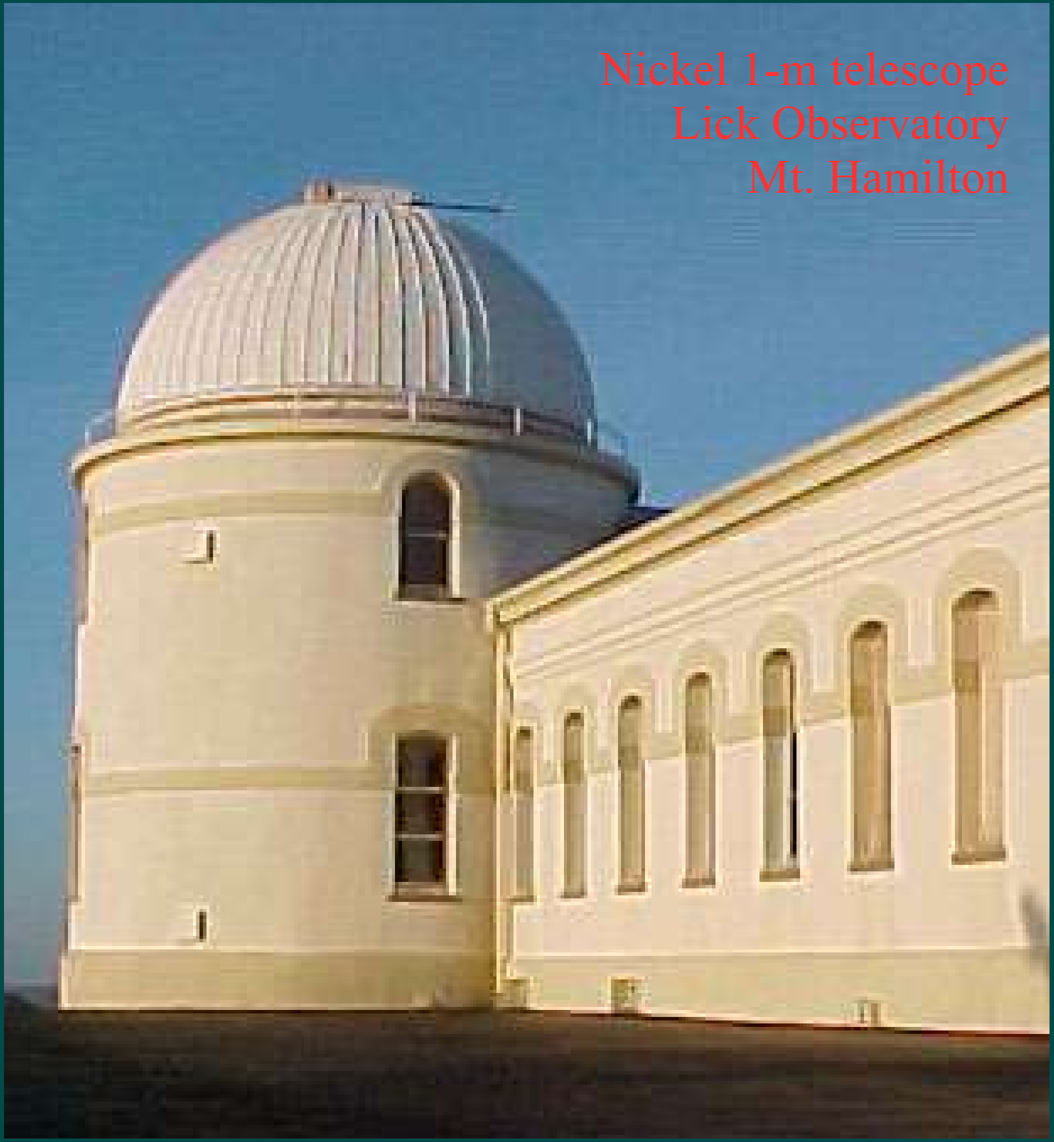
- Clean, round PSF?
- Visible host galaxy?
- Not on center of host?
- Not in 'roid database?
- etc.



Screening of promising candidates

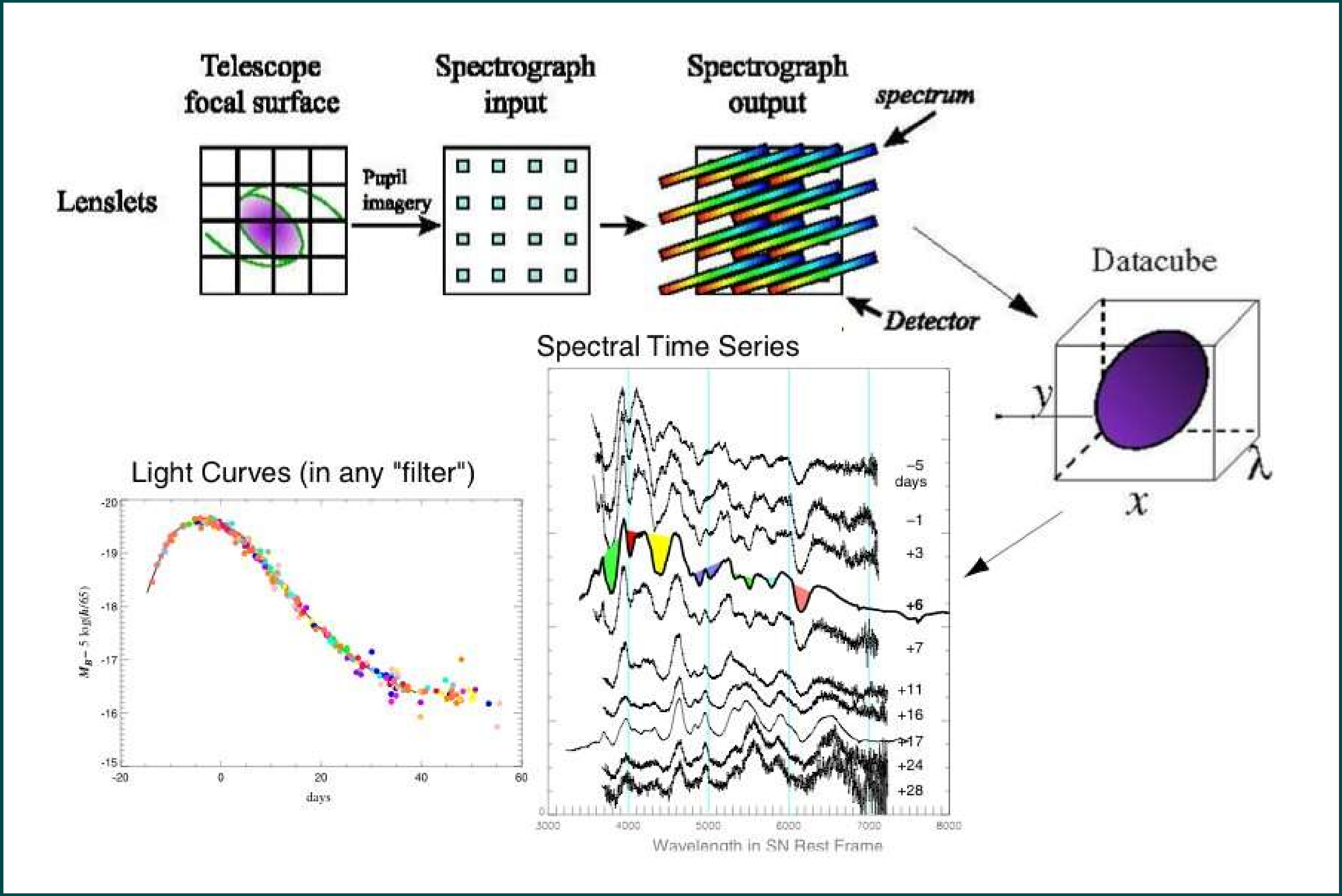
Favorite candidates are observed using another telescope. Check whether the color and rate of increase of brightness are typical of healthy young SNe Ia.

For *very* promising candidates, we'll go ahead and take a spectrum.

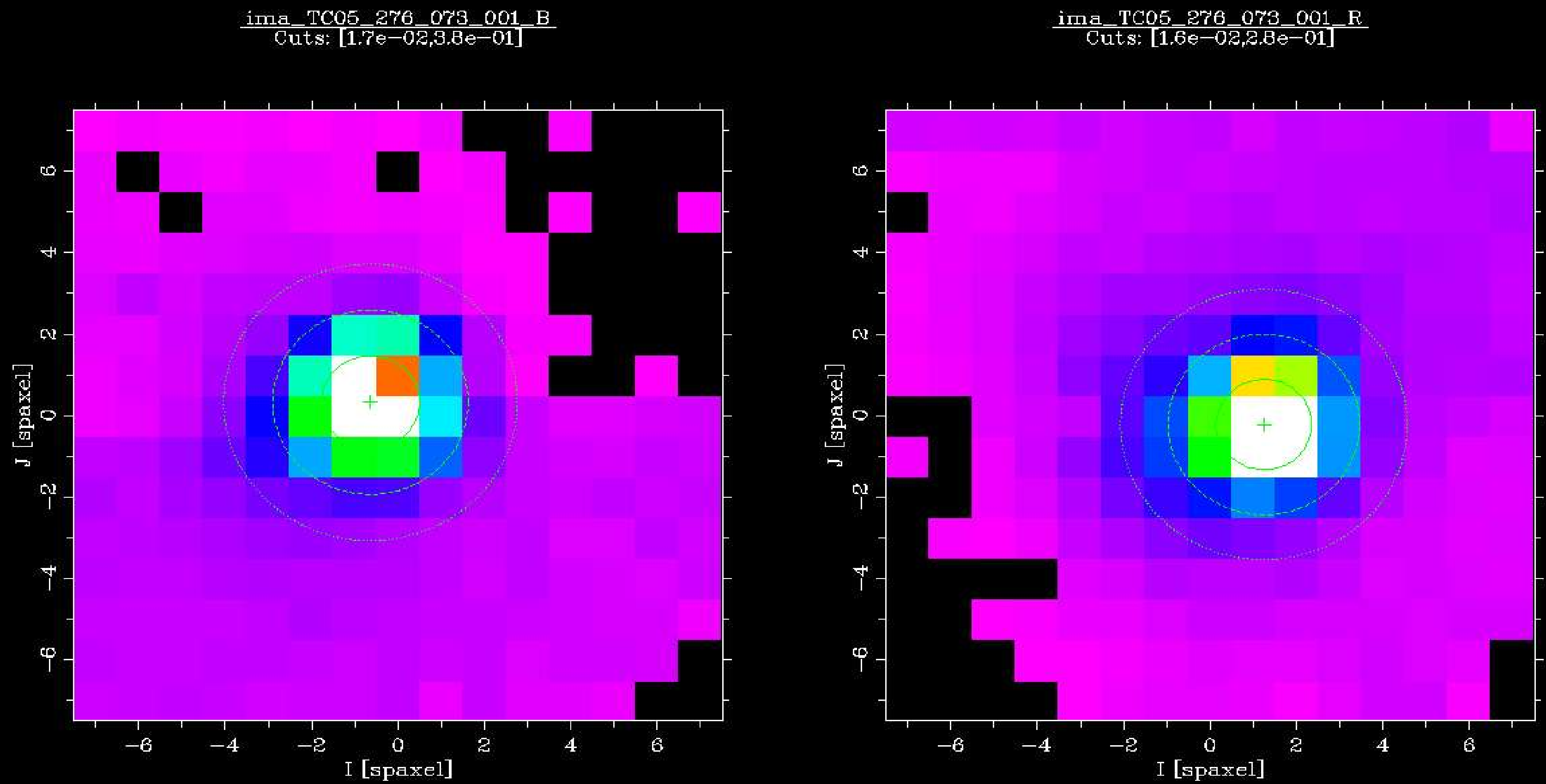


Follow-up with SNIFS

How SNIFS works



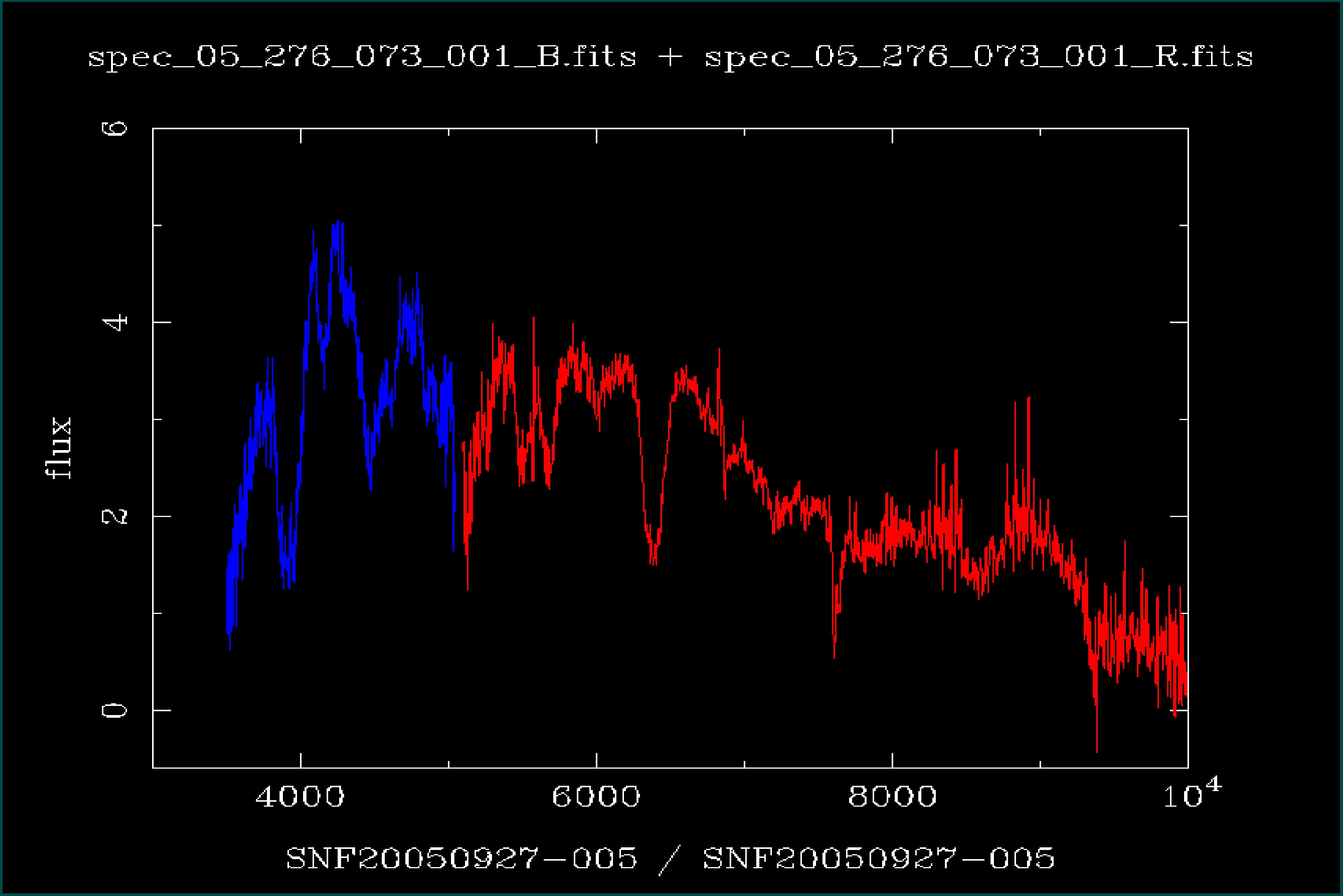
SNIFS reconstructed images



Blue channel (3500-5500 A)

Red channel (5500-10000 A)

SNIFS reconstructed SN Ia spectrum



NERSC resources used by the search

- HPSS: Raw / processed / subtracted images are stored at the rate of about 150 GB/day (during a two-week run).
- PDSF Batch: About 6000 CPU-hours used per lunation.
- PDSF Disk: REF cache uses 1.5 TB over 7 disk vaults; comparable amount needed for processed NEW images. Will migrate vaults soon to GPFS clustered file system.
- NetLogger: Now installed on PDSF for general use.
- Last lunation produced ~150 scanned candidates and some 44 SNe, of which 13 were confirmed as type Ia.

Future work

We've shown that such a search can be highly successful! But there is still room for improvement:

- Better tuning of autoscanning cuts to retain more *young* (pre-max) SNe Ia while removing roids, junk etc.
- Most of our SNe come from NEAT point-and-track data. We've just started searching with QUEST drift-scan.
- Use multi-color data in drift scan to screen candidates. (Also use of McDonald telescope at UT for screening.)
- Software improvements: more automation, reliability, user-friendliness. Reduce manpower needed for search.