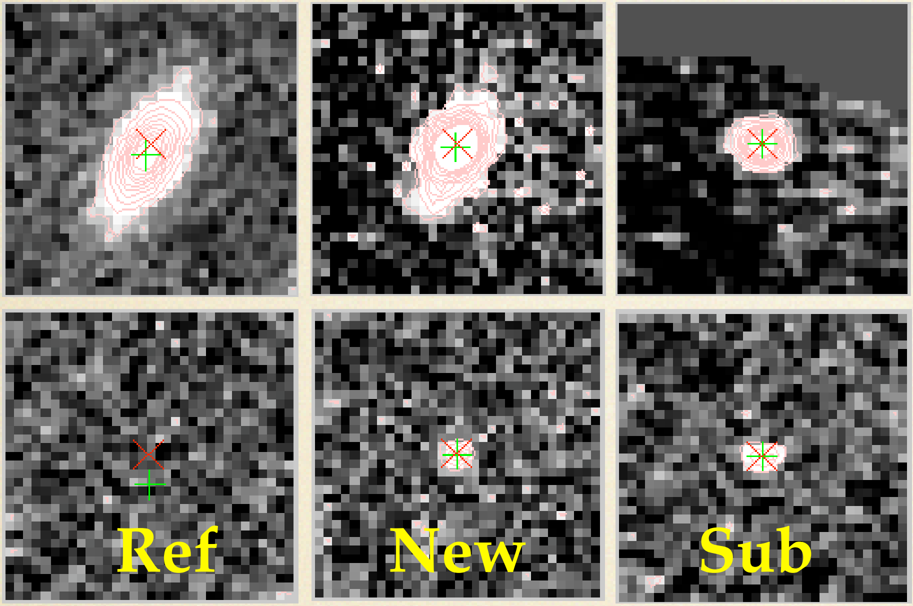
Supernova Factory Search

Stephen Bailey, LBNL
SN/Transient Search Workshop
KITP, 27 April 2007

- Overview of SNfactory Search
- Changes for 2007
- Better object selection methods
- Advice for future projects

SNfactory Search Goals

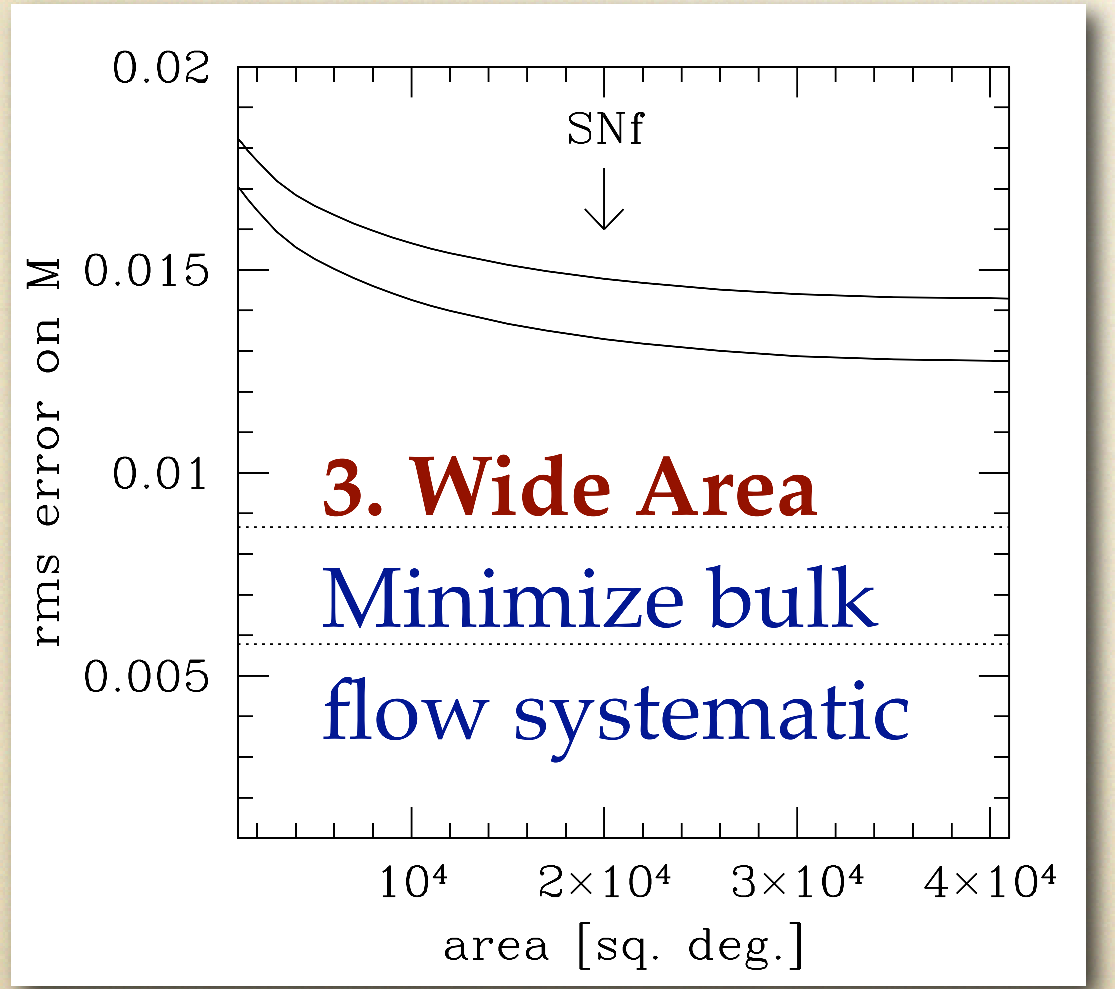


1. Not targeted at known galaxies

Match sample selection of high-*z* surveys

2. Redshift 0.03 to 0.08

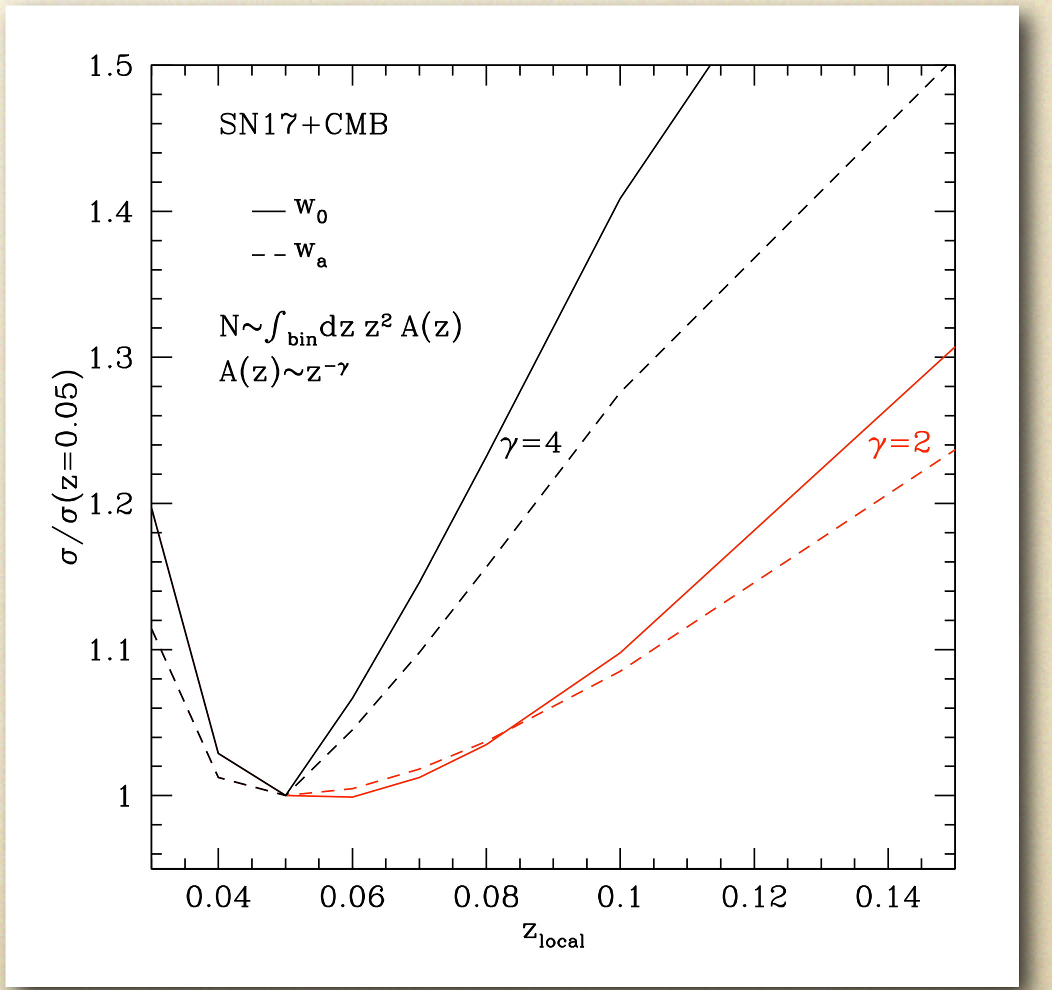
maximize power of sample



3. Wide Area

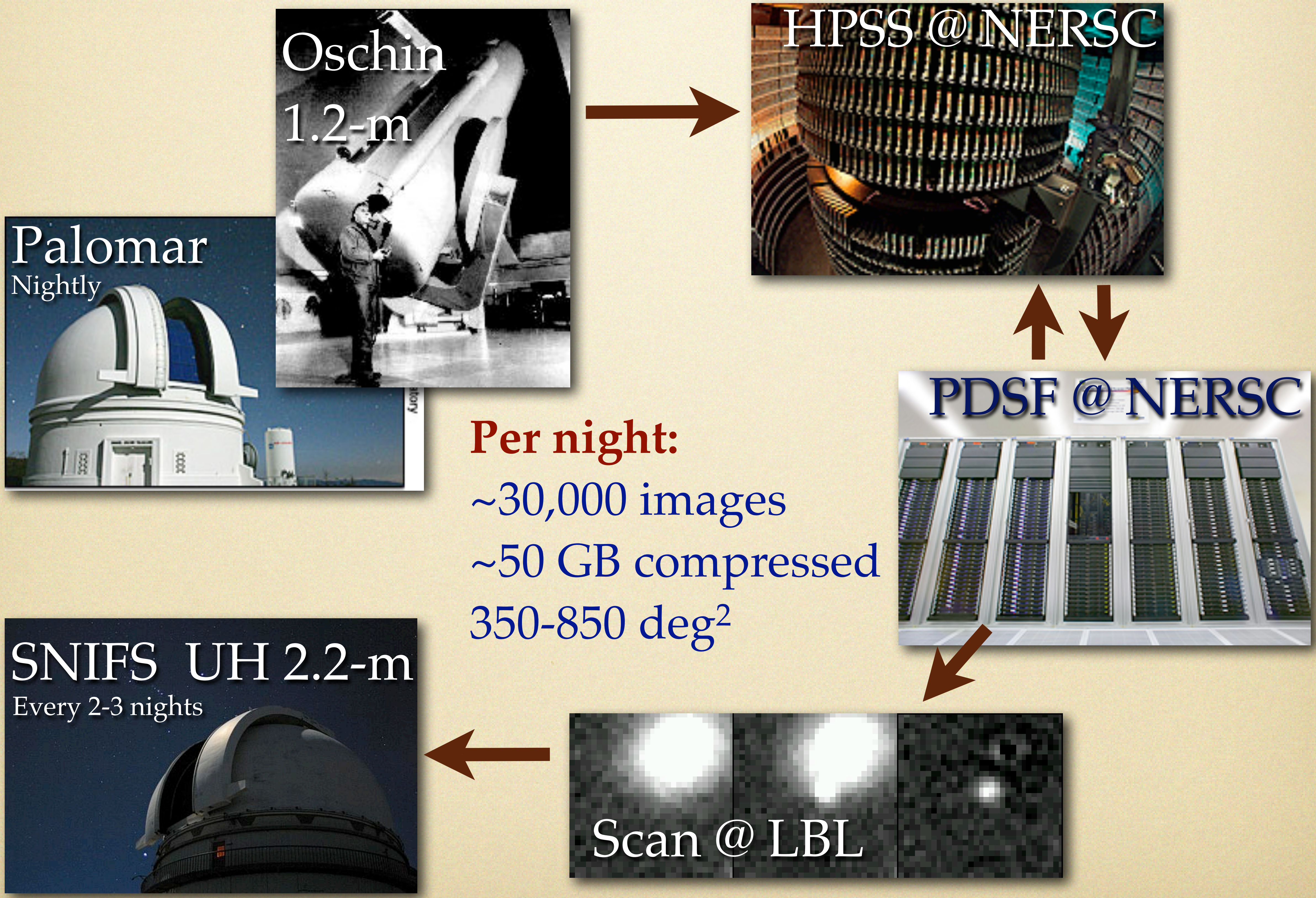
Minimize bulk flow systematic

Hui & Greene 2006



Linder 2006

Search Data Flow



Oschin 1.2-m

HPSS @ NERSC

Palomar
Nightly

PDSF @ NERSC

Per night:

~30,000 images

~50 GB compressed

350-850 deg^2

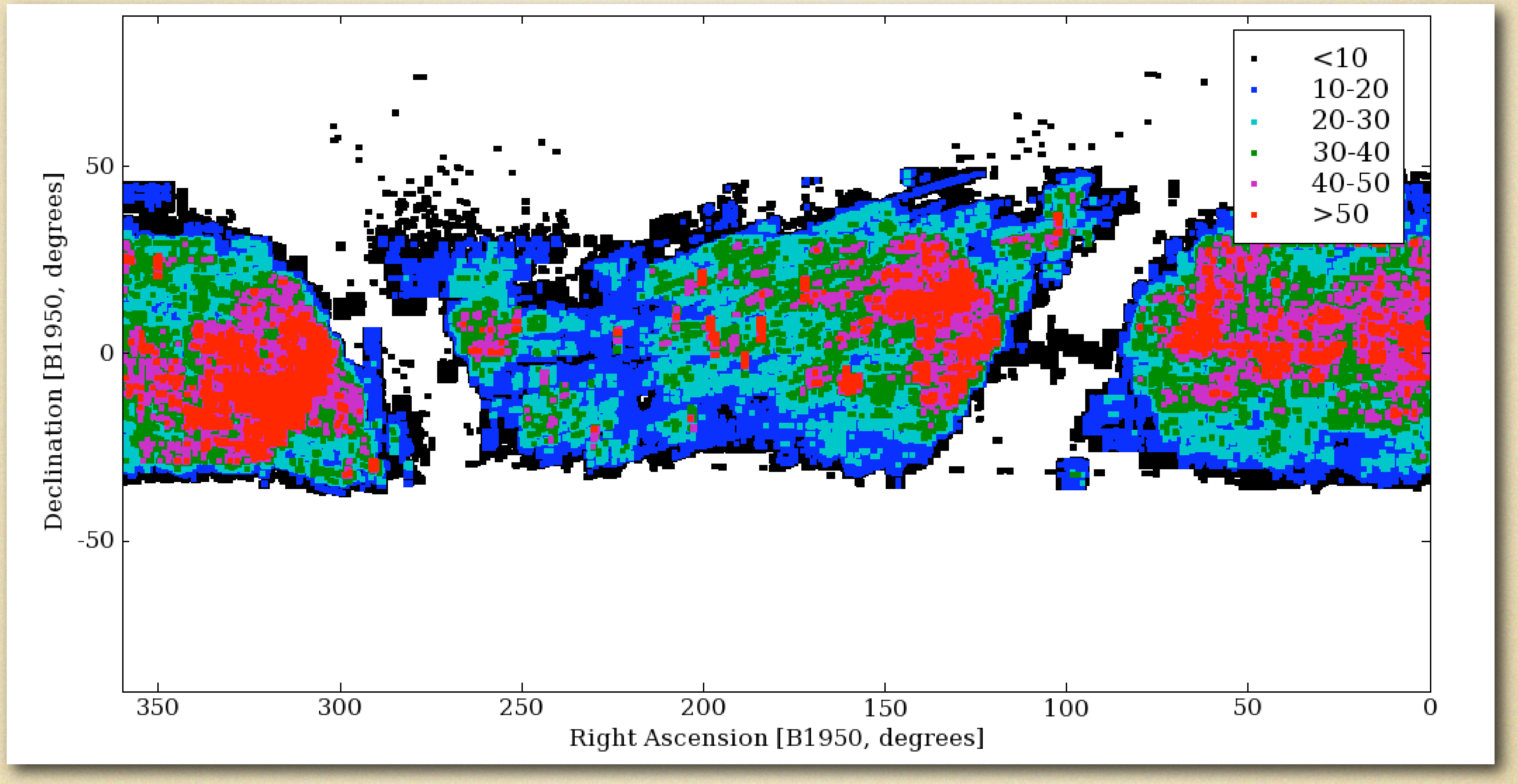
SNIFS UH 2.2-m
Every 2-3 nights

Scan @ LBL

Basic Info (2005 - 2006)

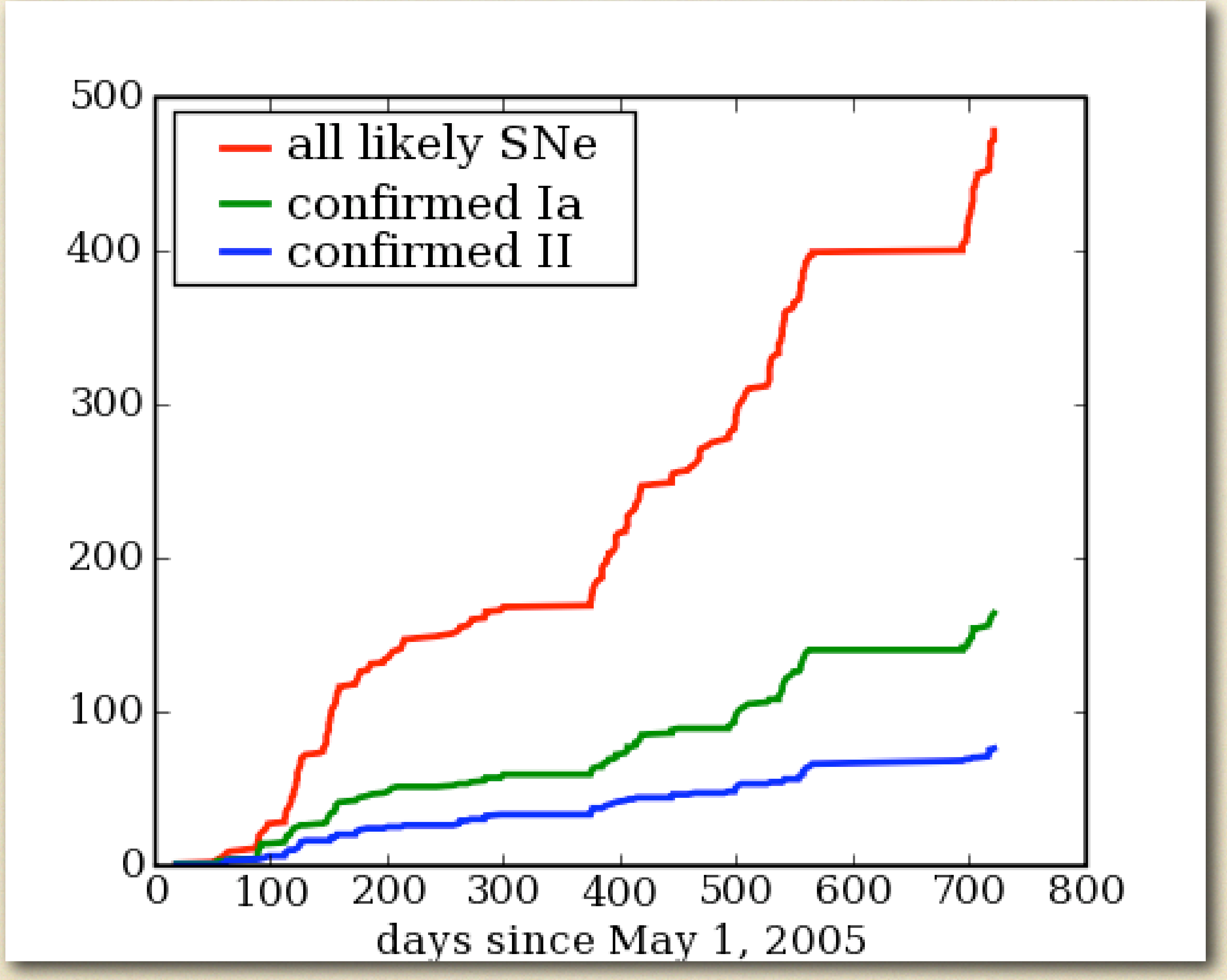
- Palomar Oschin 1.2-m telescope
  - 2″- 3″ typical seeing
  - 112 CCDs, 2400 x 598 pixels, 0.878″/pixel
- NEAT: RG610 filter (long pass redward of 610 nm)
  - 3 x 60 second exposures spread over ~1 hour
- QUEST: *(UB)RI* or *ri(z)* in driftscan
- Typical subtraction depth 19.5 to 20.5

Coverage

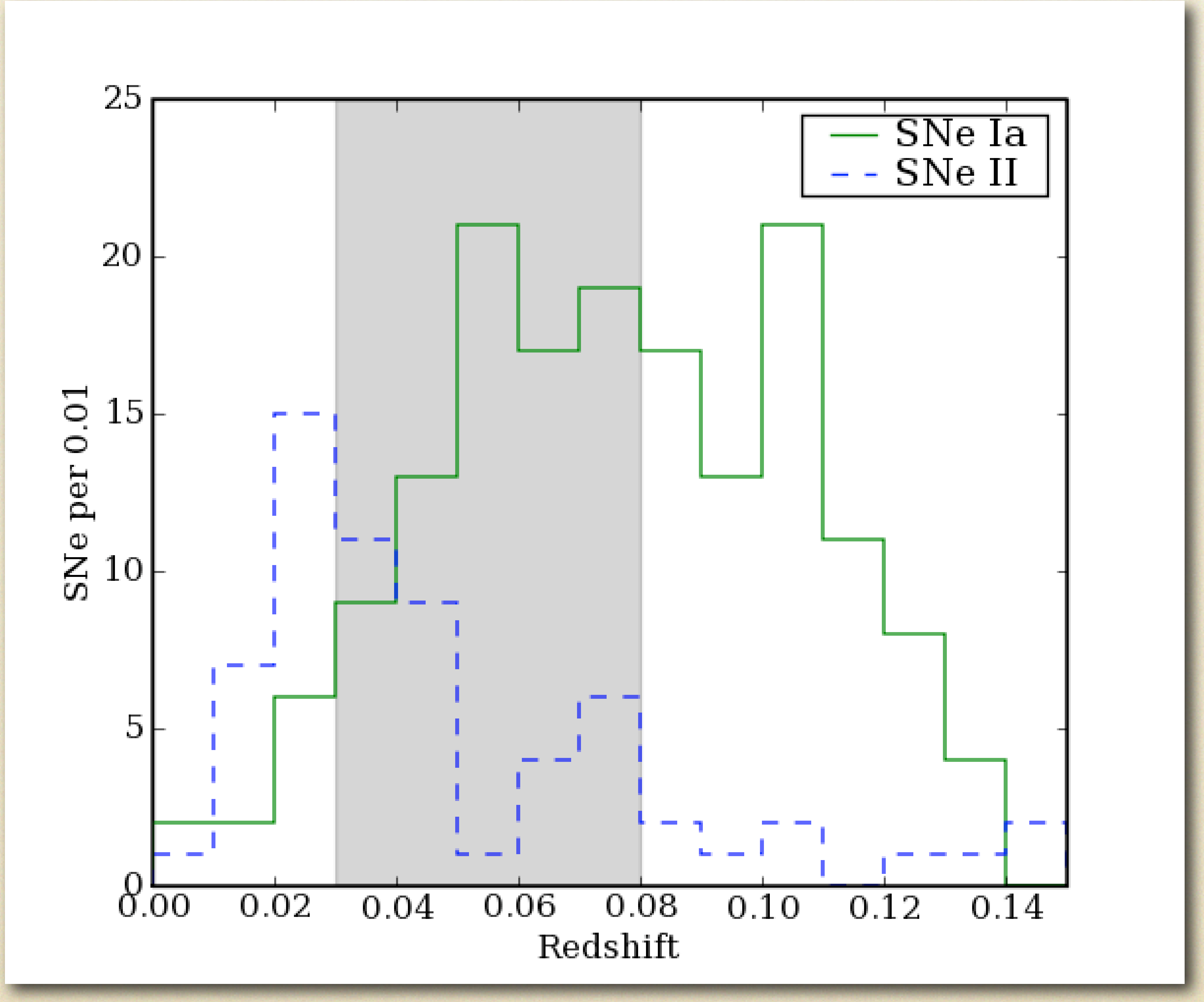


- 350 - 850 deg^2 per night
- 2500 - 3000 deg^2 monitored per month (more on cadence later)
- ~20,000 deg^2 over course of survey

Discoveries

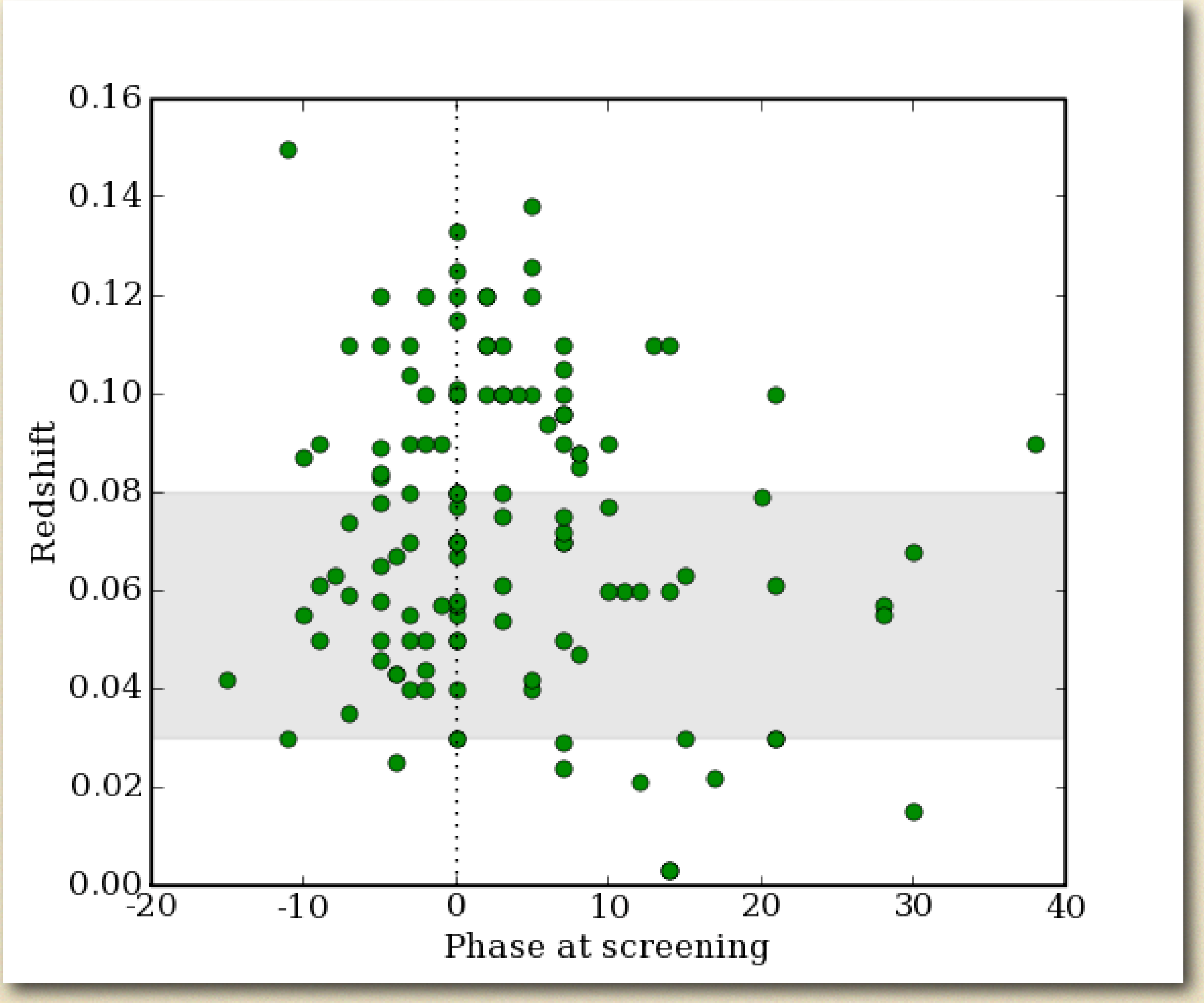


Redshift Distribution



(2005 - 2007; 80% of type II sample)

Redshift vs. Phase

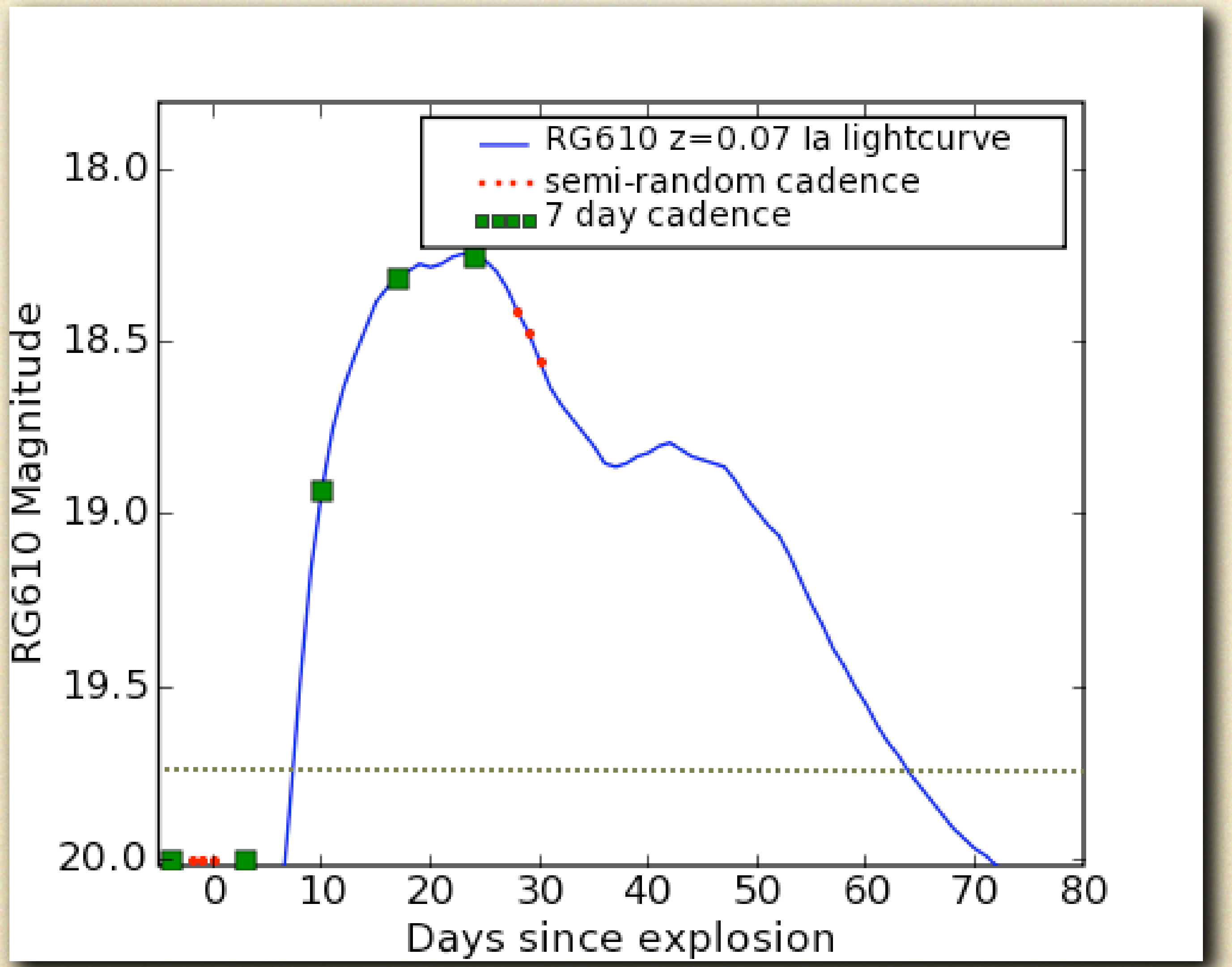


2005, 2006 SNe Ia

Discovery vs. Followup

- SNfactory is Discovery + Followup
- But our Followup facilities ≠ Discovery facilities
- Advantages:
  - Search optimized for searching
    (area, cadence, filter, exposure time, etc.)
  - Followup optimized for followup
- Disadvantages:
  - Must find things early; search data less useful for lightcurve
  - Coordination of Discovery → Followup resources

Cadence (sub)Optimization



- Historic NEAT pointings were less than optimal for SN search
- We’re working to improve cadence and pointings for SNe

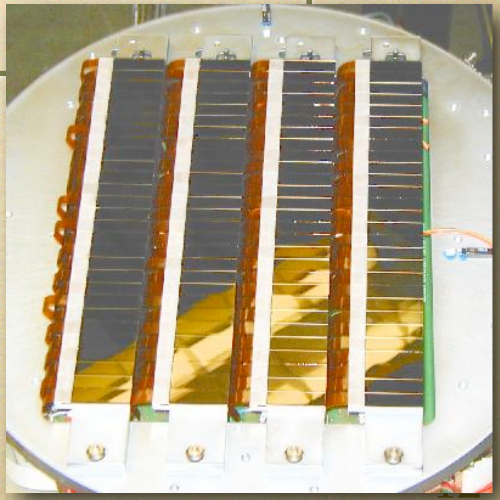
2007: Rolling Trigger Search

- Now optimize QUEST time for ~45% SN program interleaved with TNO program
  - Previously NEAT and QUEST picked their own pointings, driven by NEO and TNO programs
- Point-and-track RG610 only, April - November
  - SN: 5-7 day cadence, 2 x 60 sec exposures, away from moon
  - TNO: single 1 day cadence, 2 x 240 sec exp., near opposition
- Optimization is a work in progress
- Benefits: earlier discoveries, higher purity, better distribution in time, away from moon

Many Timescales

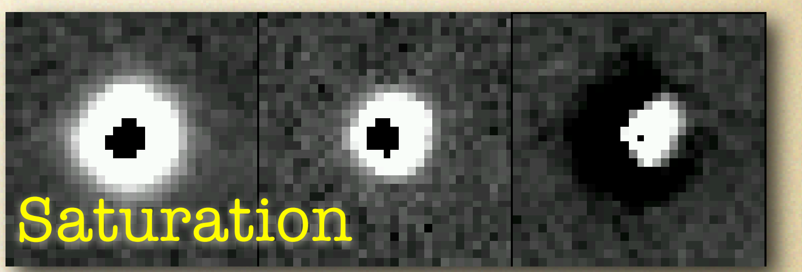
Timescale	Comments
100 - 300 sec	Interleaved TNO pairs†
1-2 hour	Nightly pairs/triplets
1-2 nights	TNO survey
3-4 nights	Interleaved SN pairs†
5-7 nights	SN survey
up to 7 years	Historical data

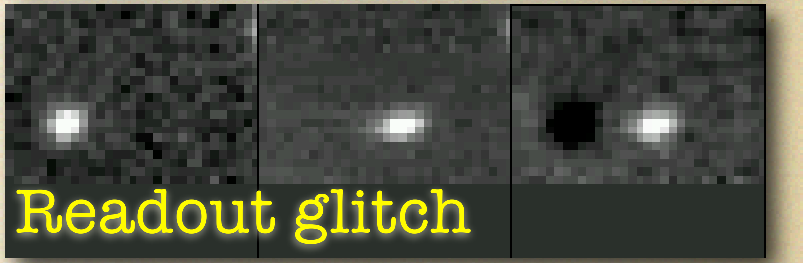
† 5% of survey; from overlap in exposures interleaving columns of camera

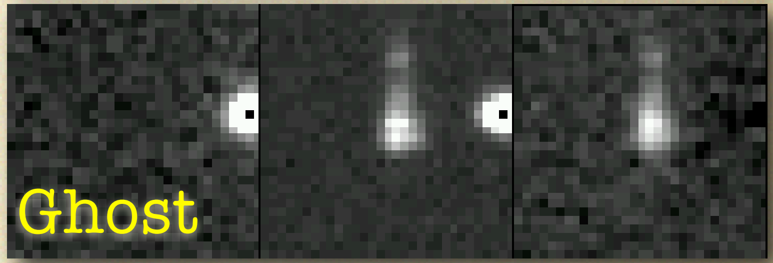


The Problem of False Positives

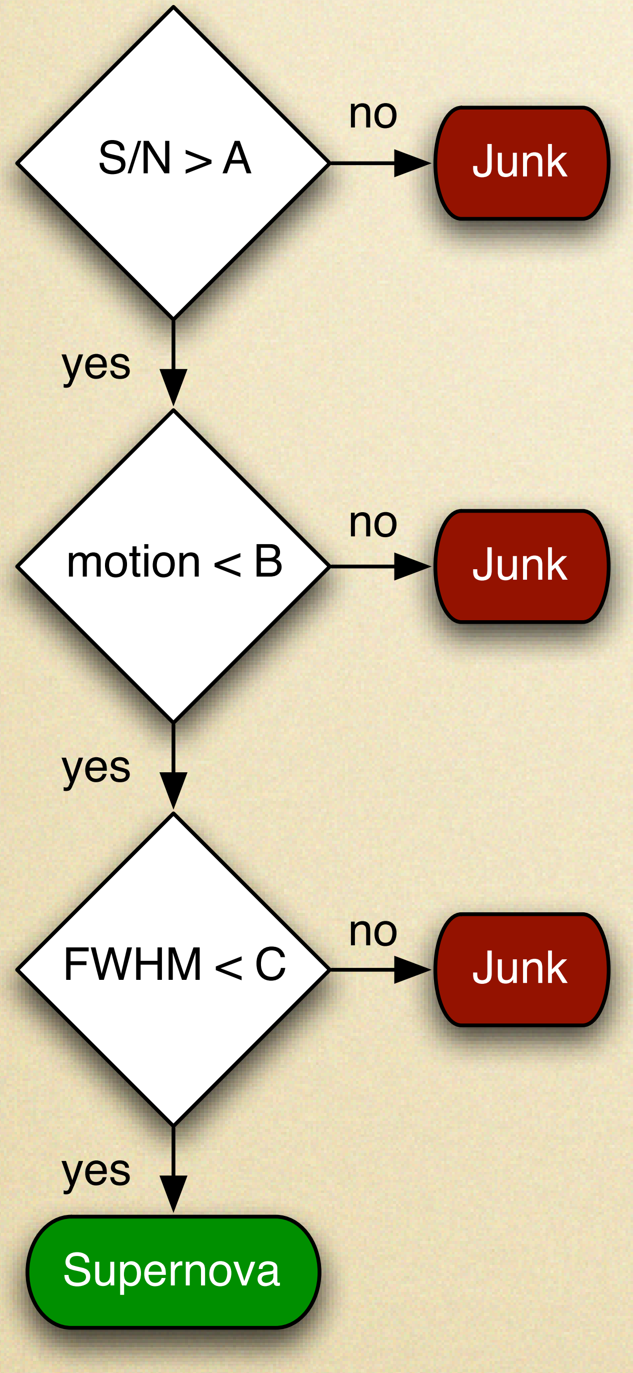
- Extreme Case: SDSS
  - 2005: 190,020 scans for 129 SNe Ia
  - 2006: 2 epochs before scanning
    - 14,441 scans for 193 SNe Ia
- SNfactory pre-Aug 2006:
  - ~1000 scans per day, save ~10 objects
  - < 60% efficiency for SNe
- We needed a better method of distinguishing good candidates from junk







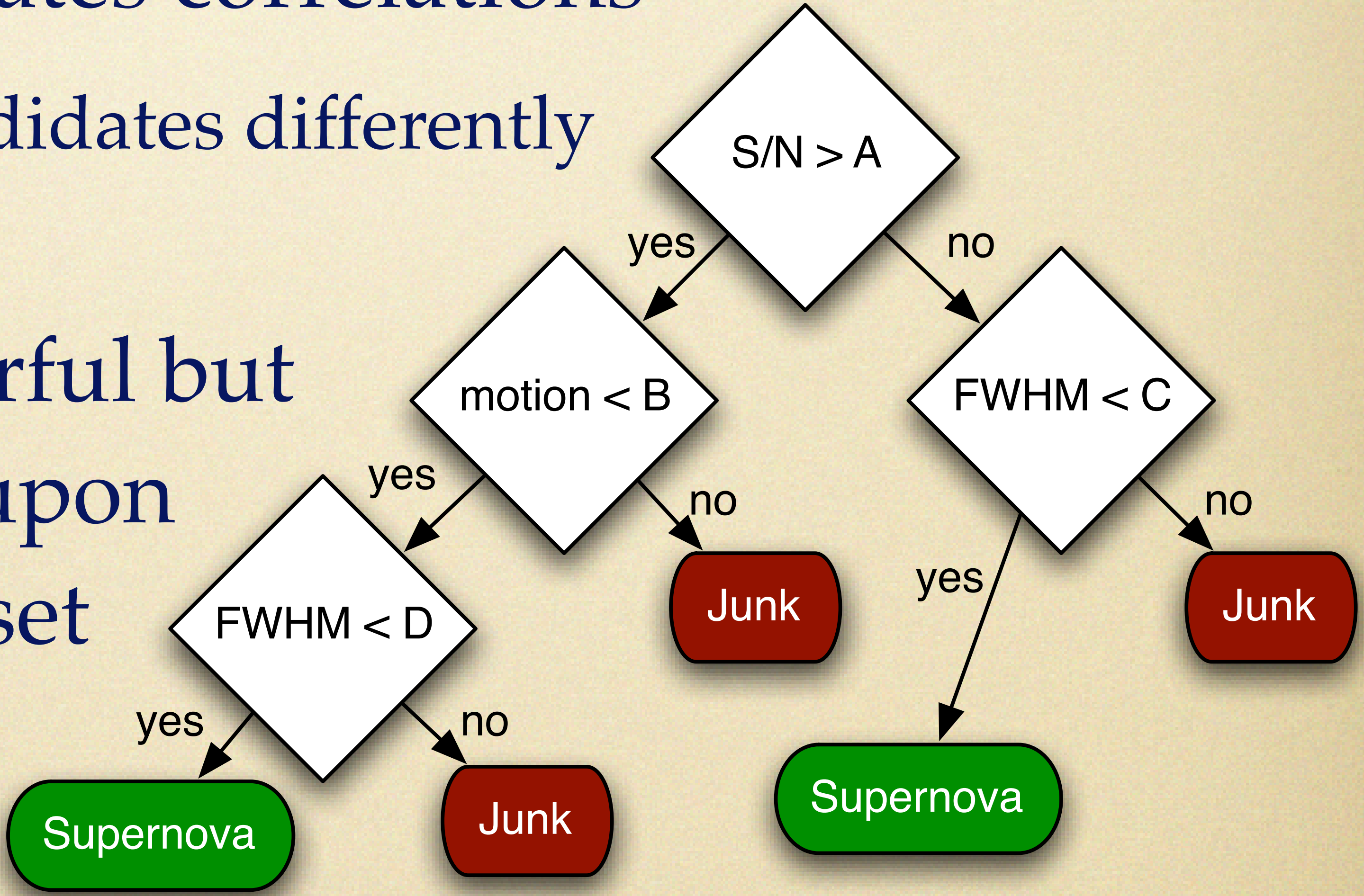
Threshold Cuts



- Traditional, easy, straight-forward
- Doesn't handle correlations
- No subtlety – barely failing one cut is same as badly failing all of them
- Outliers are problematic
  - every cut must be very efficient
  - $(1-\epsilon)^N << 1$

Decision Trees

- Make a tree of cut decisions
  - Automated procedure picks branch points
- Naturally accommodates correlations
  - e.g. treat high S/N candidates differently than low S/N ones
- Single trees are powerful but unstably dependent upon specific training dataset

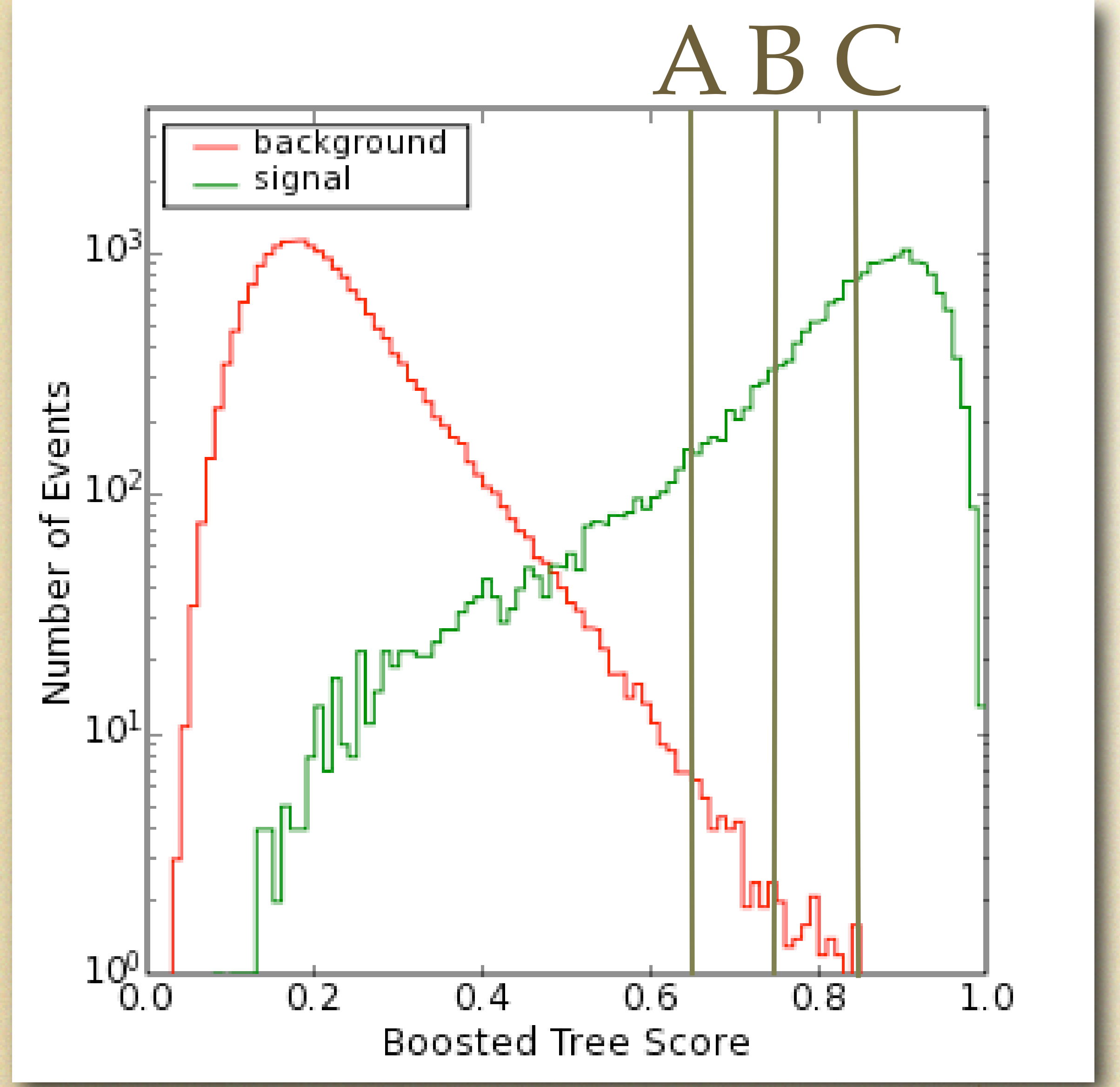


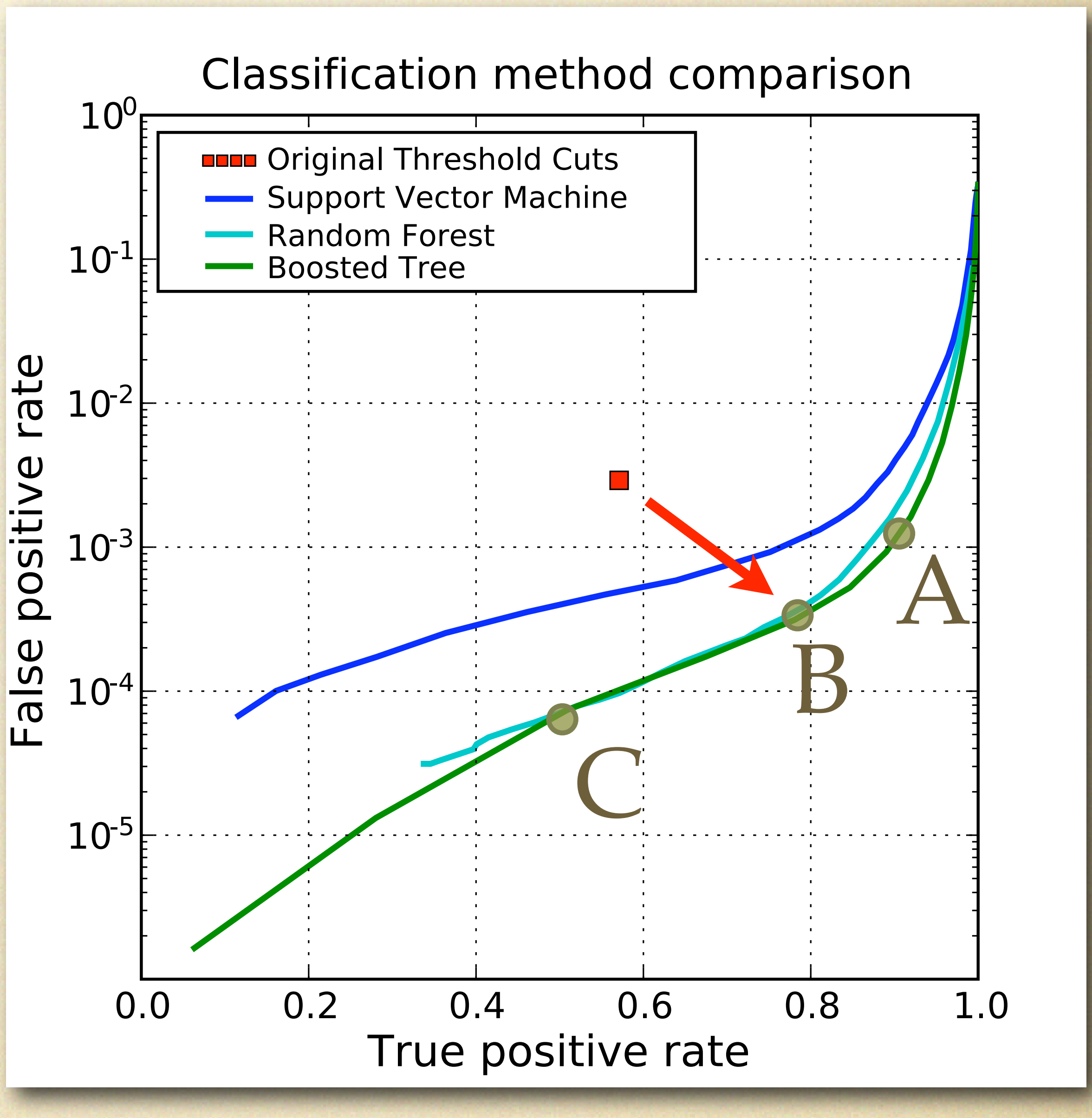
Boosted Decision Trees

- Train a tree using events of known type (sig / bkg)
- Increase the weight of misclassified events
- Train another tree with the altered weights
- Repeat many times

- This makes a series of progressively better trees, focusing on events which are difficult to classify
- Final answer is weighted average of individual trees

Separating Junk from SNe





~36,200 scans for 30 Ia (May-Jul) → ~4,300 scans for 51 Ia (Sep-Nov)

Similar to SDSS 2006 scan rate but with single night scan trigger

Advice for the Future

- Scaling issues matter
  - Buying *N* times as many computers won't automatically let you process *N* times as much data
  - Do mock data challenges before you ever get real data
  - Inherit ideas and algorithms, but not code
- Realistic simulations are important
  - Optimize tradeoffs between NEO, TNO, and SN programs
  - Understand effects of search output vs. followup resources
- Use something better than cuts
  - Boosted Decision Trees work great for us