Big Mirrors, Bayesian Evangelists and the Public:
How Advances in Mirror Technology, detectors, databases, machine learning and crowd-sourcing is driving Astronomy into the future

Michael Way (NASA/Goddard Institute for Space Studies)
Outline

• **Mirrors**: Single dish to segments
• **Detectors**: paving with plates or silicon?
• **Data**: From ascii files to databases
• **Crowd-sourcing**: AAVSO ➔ Zooniverse
• **Machine Learning**
• **Examples**
Mirrors: single to segments

- Lippersheey/Galileo refractor: 1.5cm? 1608/1609
- Newton/Hooke reflector: 3.3cm/18cm: 1668/1674
- Herschel: 1.26m: 1789
- Leviathan/Hooker: 1.83m/2.54m: 1845/1917
- Hale/BTA: 5.08m/6m: 1948/1976
- Keck/GTC: 10m/10.4m: 1993/2009
- TMT/E-ELT: 30m/42m: 2020?
Mirror Technology

Telescope Aperture Size: through the centuries

Telescope Aperture Size: through the centuries
Telescopes/Culture

Oct. 3, 1947: Birth of Palomar’s ‘Giant Eye’

By Tony Long  October 3, 2011 | 6:30 am  Categories: 20th century, Astronomy, Engineering

1947: After 13 years of grinding and polishing, the Palomar Observatory mirror is completed at Caltech.

It was, at the time, the largest telescope mirror ever made in the United States, measuring 200 inches in diameter.

Following its completion, the disk was mounted in Palomar’s Hale Telescope and first used in January 1949 to take pictures of the Milky Way. Edwin Hubble was the first astronomer to make images using the new scope.

The Hale Telescope

The telescope of which the 200-inch mirror in the basin was closed in 1949 to the senior astronomer, George E. Shapley. Funds were raised by the Rockefeller Foundation and St. Paul’s, Palos Verdes, California, referred to as the observatory site. Our dedication ceremony was held on May 4, 1950.

The Director of Hale was Dr. R. A. Lynds, who died ten years before completion of the project.

Galileo’s primitive telescope was measured only 3/4-inch in diameter (smaller than the 200-inch mirror). Grind the disk, 3/4-inch in diameter is the size of the disk, the larger it is, the better. Experimental results indicate that the Hale Telescope is getting up where 6 million times distant than the naked eye can see.

Using photographic plates, astronomers can see stars and planets that are 8,000,000 miles away and stars so far away as to have permanent positions.

The 200-Inch Disk

The giant mirror disk in the Cassegrain focus system from the Cassegrain end. Water was poured into the 200-inch mirror at a temperature of 100°F.

The water was used to cool the mirror at the front of the glass. The water was purged and then purged again. The water was purged with the 200-inch mirror. Water is purged at the front of the glass. This process is repeated until the temperature of the water is 100°F.

Experimental tests of the 200-inch mirror were done on the last day before the tests were completed. The mirror was covered with a blanket and held at 100°F for 24 hours. The water was purged and then purged again. Water is purged at the front of the glass. This process is repeated until the temperature of the water is 100°F.

The mirror was then cooled with water and the temperature of the water was then reduced to 100°F. The water was then purged and then purged again. Water is purged at the front of the glass. This process is repeated until the temperature of the water is 100°F.

The mirror was then cooled with water and the temperature of the water was then reduced to 100°F. The water was then purged and then purged again. Water is purged at the front of the glass. This process is repeated until the temperature of the water is 100°F.

The mirror was then cooled with water and the temperature of the water was then reduced to 100°F. The water was then purged and then purged again. Water is purged at the front of the glass. This process is repeated until the temperature of the water is 100°F.

The mirror was then cooled with water and the temperature of the water was then reduced to 100°F. The water was then purged and then purged again. Water is purged at the front of the glass. This process is repeated until the temperature of the water is 100°F.

The mirror was then cooled with water and the temperature of the water was then reduced to 100°F. The water was then purged and then purged again. Water is purged at the front of the glass. This process is repeated until the temperature of the water is 100°F.
Future: Larger in Space too?

MOIRE: **Membrane Optical** Imager for Real-Time Exploitation

*Artist's Concept*
Detectors (Optical)

Eyes, Plates, Photometers and CCDs…
Eye: QE 1—10%  
D=0.7cm, f=2.2cm, F≈3.2
Upgrades are available
Detectors/QE: Paving the focal plane

• Niépce (1826) Camera Obscura
• Glass Plates (Wet) QE 1%
  – 1839 Daguerre: Moon images
  – 1840 Draper
    • 13cm reflecting telescope
    • Moon (20 min)
  – 1845: Fizeau & Foucault: Sun Images
  – 1850: Bond & Whipple: first star photo - Vega
  – 1872 Miller & Huggins: First spectrograph – Sirius

Earliest extant Daguerre image?: 1851 by John Adam Whipple
Detectors/QE: Paving the focal plane

• Glass Plates (Dry) QE 1—3%
  – 1876 Huggins (spectrograph)
  – 1880 Henry Draper – first photo of Great Neb. in Orion
  – 1883 Common : First objects fainter than seen by eye?
    • 91cm reflecting/60 min exposure

Orion Nebula by Common  Early Common Observatory
Detectors/QE: Paving the focal plane

- Large Scale Surveys become feasible with photography
  - 1887 Astrographic Catalogue and Carte du Ciel
    - Aperture ~33 cm, scale: 60 arcsec/mm
    - Field of view: $2^\circ \times 2^\circ$ (Moon diameter ~0.5°)
  - 1948 Oschin Schmidt (Palomar): 1.22m, FoV = $4^\circ \times 4^\circ$ degrees
    - 14” x 14” glass plates (41”/mm) – paving the focal plane with glass
  - 1970 DuPont 2.54m ($2.1^\circ \times 2.1^\circ$)
Digital(?) Detectors

- 1892 – 1980s: Photoelectric photometers
  - Think about it as a single element CCD
  - Digital output? More likely paper tape…
  - X-rays: V2 rockets used in 1949!

19 stage linear photomultiplier tube developed at the Paris Obs ➔
Detectors/QE: Paving the focal plane

• Charged Coupled Devices (CCDs) QE: 10—90%
  – Boyle and Smith 2009 Nobel Prize for 1970 CCD development at Bell Labs (the 7th from this lab)
Until CCDs were big enough we continued to fill the focal plane with photographic glass plates AND Fibre Spectrographs (1980):

THE ASTROPHYSICAL JOURNAL, 242:L69–L72, 1980 December 1
© 1980. The American Astronomical Society. All rights reserved. Printed in U.S.A.

MULTIPLE OBJECT SPECTROSCOPY: THE MEDUSA SPECTROGRAPH

John M. Hill, J. R. P. Angel, John S. Scott, and Delvin Lindley
Steward Observatory, University of Arizona

and

Paul Hintzen
NASA Goddard Space Flight Center
Received 1980 August 4; accepted 1980 August 28

ABSTRACT

We have built and tested an instrument to obtain simultaneous spectra of many objects in the field of view of the Steward 90 inch (2.29 m) telescope. Short lengths of fused silica fiber 300 μm in diameter are used to bring the light from galaxy images at the Cassegrain focus into a line along the spectrograph slit. From a single exposure of the cluster Abell 1904, which has a redshift of ~20,000 km s^{-1}, we have determined the redshifts of 26 individual galaxies, each with a precision of ~100 km s^{-1}. The present device, while already giving a sixfold reduction in the mean telescope time per galaxy, has significant light losses because it is not ideally matched to the telescope. An instrument being designed for the prime focus will transmit light from each object as efficiently as a conventional spectrograph.

Subject headings: galaxies; redshifts — instruments
Circa 1990:
Dupont 2.54 m (2° FoV)
Las Campanas, Chile
Sloan Digital Sky Survey Telescope

Today:

Subaru Prime Focus Spectrograph

30 2048x2048 CCDs
600 (1000) fiber spectrograph
2400 fiber positioner, Prime Focus, 1.5 deg FoV
Data:
From ascii files to databases
Revolutions in Data Handling

• Traditionally
  – nearly all data acquisition & data reduction was handled by individual Astronomy researchers

• Today
  – most astronomical data is collected in surveys or by service observations

Part of this has been made possible by the digital and internet revolutions
CCDs (and some photomultipliers) gave us a fully electronic record: Thankfully portable storage was catching up too!

- 9 track tapes [1970] (800-6250 bpi) = 170MB
- TK50/DLT [1984/89 DEC] = 94MB/2.6GB → 110GB
- 8mm Exabyte = 112m (2.5/5GB), 54m (1.2/2.4GB)
- 4mm DAT = 90m (2GB), 120m (4GB with DAT2)
- CD = 700MB (100 year lifespan?)
- DVD = 4.7–7GB, BlueRay = 128GB
- Today? It all sits on disk, or in the cloud? …
Tape rescues big data

Sep 26th 2013, 15:07 by Economist.com

WHEN physicists throw the “on” switch on the Large Hadron Collider (LHC), between three and six gigabytes of data spew out of it every second. That is, admittedly, an extreme example. But the flow of data from smaller sources than CERN, the European particle-research organisation outside Geneva that runs the LHC, is also growing inexorably. At the moment it is doubling every two years. These data need to be stored. And that need for mass storage is reviving a technology which, only a few years ago, seemed destined for the scrapheap: magnetic tape.

LTO-6: 2.5TB, 160MB/s
Distributing Digital Surveys

- Digitized Sky Survey 1980s: **102 CDROMs**
- Two Micron All Sky Survey (early 1990s):
  - first large scale fully digital survey (two 1.3m) and catalog.
  - 471 million objects detected
  - Released on **5 double-sided DVDs (43GB)**
  - Full fidelity images ~10TB
1970-80s

Late 1980s/early 1990s

1990s

2000s
“The Next Generation”
Sloan Digital Sky Survey (2000—Present)

– 120 Megapixel camera, 1.5x1.5 degrees
– 600 (1000) multi-object spectrograph
– ~1 Billion Objects detected, over 1 million spectra
2005 and beyond

SDSS Query / CasJobs

'qso1' Details

<table>
<thead>
<tr>
<th>JobID</th>
<th>TaskName</th>
<th>Context</th>
<th>Queue</th>
<th>Submitted</th>
<th>Started</th>
<th>Finished</th>
<th>Status</th>
</tr>
</thead>
</table>

Executed on: DR7
Rows: 76454

Message: Query Complete

Query:

```sql
Select p.ObjID, p.ra, p.dec, 
s.z, s.zErr, s.zConf, s.zStatus into mydb.qsol from SpecOBJAll s, PhotoObjAll p 
WHERE s.specobjid=p.specobjid 
and s.zWarning=0 
and (SpecClass=dbo.fSpecClass('QSO') or SpecClass=dbo.fSpecClass('HIZ_QSO')) 
and ((flags & 0x8) = 0) and ((flags & 0x2) = 0) and ((flags & 0x400000) = 0)
```
Unexpected Collaborators?

• SDSS dB was built in collaboration with Microsoft
  – Jim Gray + Alex Szalay and others…
  – Interesting large problem, open source data model
  – Still possible to download O(100s GB) data sets
• SciDB: built for next generation data sets (LSST)
  – http://www.scidb.org
  – Not possible to download PB sized data and use it
    • 1PB over 10Gb/s line is 10 days, 1PB = $200 in 2020?
    • I/O not keeping up with other Moore type laws (arXiv:1108.5124v1)
  – R-interface for expert users
Digital Surveys of Today/Tomorrow…

– PanStarrs (2011—2020?)
  • 64x64 array (600x600 CCD) 1.4 Gigapixels
  • ~3TB/night

  • 74 CCDs
  • 1TB/night raw data

– Large Synoptic Survey Telescope (2020—2030)
  • 1PB/night raw data
  • Database size: ~10PB
  • 60PB of images
Crowdsourcing:
From AAVSO to Zooniverse
Crowdsourcing is older than you think?

- AAVSO = American Assoc. of Variable Star Observers
  - Amateur Astronomers contributing observations of variable stars since 1911!

- 1999: SETI @ Home and copycats
  - Called “Volunteer Computing”
  - Distributed computing, not crowdsourcing
Modern Citizen Science

• 2000: Clickworkers (NASA/Ames)
  – Identifying & Classifying ages of Martian craters from Viking Orbiter images
    • Kanefsky, Barlow and Gulick.

• 2006: Stardust@Home
  – Search aerogel images for tiny dust impacts gathered from tail of Comet Wild
• 2007-Present: GalaxyZoo
  – Classifying Galaxies in the Sloan Digital Sky Survey
  – Largest by eye “professional catalog” ~1400 objects
  – GZ2: 16 million classifications of 304,122 objects
  – Use Machine Learning to train automated classifiers…

• Zooniverse does a lot more:
  – Planet Hunting
  – Transcribe old Weather Logs
  – Ancient lives (reading old papyri)

• SETI Live (Feb 29, 2012) March 10th Economist ✰
Bayesian Evangelism and Machine Learning
Rev. Bayes 1702-61

1795 (1809): Gauss
1805: Legendre
1808: Adrian

Least Squares

Shannon (1916-2001)

Machine Learning
Machine Learning

Given new large complex multivariate data, Machine Learning & Data Mining are becoming more commonly used.

- 2600+ plus papers with Bayes in title or abstract
- Large numbers of citations for most popular papers
- Increase in number year by year…
Bayes Evangelism

Review of Bayesian methods in cosmology: Trotta (2008), arxiv: 0803.4089

The rise of Bayesian methods in astrophysics

Number of papers normalized to 1980

"Bayesian" papers

All papers in astrophysics

Year
An iterative technique for the rectification of observed distributions

L. B. Lucy*

Departments of Physics and Astronomy, The University of Pittsburgh, Pittsburgh, Pennsylvania 15213
(Received 15 January 1974; revised 26 March 1974)

DOI: 10.1051/0004-6361:20031117
© ESO 2003

A synthetic view on structure and evolution of the Milky Way

A. C. Robin¹, C. Reylé¹, S. Derrière², and S. Picaud¹

Bayesian-Based Iterative Method of Image Restoration*

William Hadley Richardson

Visibility Laboratory, University of California, San Diego, San Diego, California 92152
(Received 15 September 1970)
---# of Papers with these in title/abstract (2013/12)---

• 1579 “Neural Networks” (well known?) 12288 citations
• 1276 “Data Mining”: 7490 citations
• 429 “Machine Learning”: 2684 citations
• 125 “Self Organizing Maps” (obscure?) 778 citations

---Some conferences related to Stats/ML in Astronomy---

• Statistical Challenges in Modern Astronomy
  – 5 conferences+books 1991—2011, 7 summer schools
• Astrostatistics/Astroinfomatics (https://asaip.psu.edu/)
• SciCoder workshops (http://www.scicoder.org)
• CESS (http://www.giss.nasa.gov/meetings/cess2011)
BOOKS

Advances in Machine Learning and Data Mining for Astronomy

Edited by Michael J. Way, Jeffrey D. Scargle, Kamal M. Ali, and Ashok N. Srivastava

5 volumes! 2003-2013
Example from Today?
SDSS+ Crowd-Sourcing + Machine Learning

“Galaxy Zoo: reproducing morphologies via machine learning”, Banerji et al. 2010
1. GalaxyZoo morphologies from volunteers ($T_{eye}$)
2. Primary & Secondary Isophotal Parameters ($T_{SDSS}$)

Neural network

Inputs=$T_{sdss}$

Outputs=$T_{eye}$
Primary & Secondary Isophotal Parameters ($T_{SDSS}$)

Table 1. First set of input parameters based on colours and profile fitting.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dered_g-dered_r</td>
<td>$(g - r)$ colour</td>
</tr>
<tr>
<td>dered_r-dered_i</td>
<td>$(r - i)$ colour</td>
</tr>
<tr>
<td>deVAB_i</td>
<td>de Vaucouleurs fit axial ratio</td>
</tr>
<tr>
<td>expAB_i</td>
<td>Exponential fit axial ratio</td>
</tr>
<tr>
<td>lnLexp_i</td>
<td>Exponential disc fit log likelihood</td>
</tr>
<tr>
<td>lnLdeV_i</td>
<td>de Vaucouleurs fit log likelihood</td>
</tr>
<tr>
<td>lnLstar_i</td>
<td>Star log likelihood</td>
</tr>
</tbody>
</table>

Table 2. Second set of input parameters based on adaptive moments.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>petroR90_i/petroR50_i</td>
<td>Concentration</td>
</tr>
<tr>
<td>mRrCc_i</td>
<td>Adaptive (+) shape measure</td>
</tr>
<tr>
<td>aE_i</td>
<td>Adaptive ellipticity</td>
</tr>
<tr>
<td>mCr4_i</td>
<td>Adaptive fourth moment</td>
</tr>
<tr>
<td>texture_i</td>
<td>Texture parameter</td>
</tr>
</tbody>
</table>

Table 5. Summary of results for the entire sample when using input parameters specified in Tables 1 and 2.

<table>
<thead>
<tr>
<th></th>
<th>Early type (per cent)</th>
<th>Galaxy Zoo Spiral (per cent)</th>
<th>Point source/artefact (per cent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Early type</td>
<td>92</td>
<td>0.07</td>
</tr>
<tr>
<td>N</td>
<td>Spiral</td>
<td>0.1</td>
<td>92</td>
</tr>
<tr>
<td>N</td>
<td>Point source/artefact</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Cultural Changes

How are these things changing the way we work?

• Avoid applying for grants and telescope time and requisite travel funding

• Avoid specialized instrument knowledge, data acquisition, data reduction, backup/storage concerns
  – Positive and Negative aspects

• More time for thinking up good questions?!
  – Submit a query (10 min)
  – Download the data
  – Write your paper
My world... then

• PhD: Spent 2 years in Chile collecting spectroscopic redshifts in (Abell) clusters of galaxies
  • Incredibly specialized instrument knowledge
    ▪ Detector was photon counting, but also imaging like CCD
    ▪ Fiber spectrograph: instrument calibration, sky subtraction, etc.
  • Heavy use of time/labor
    ▪ Plates drilled in Pasadena, shipped to Chile weeks before
    ▪ A 3 night run required showing up 3-4 days before to prepare
      • Marking up plates, checking out instrument, getting to observatory
    ▪ very long nights: plugging plates beforehand, collecting calibration frames (sky flats) until well after twilight
  • 1—2 years “reducing” the data for thesis
My world… today

“Galaxy Zoo Morphology & Photometric Redshifts in the Sloan Digital Sky Survey”
1) SDSS Data Release 7 (Oct 2008)
2) GalaxyZoo Data Release 1 (Feb 2011)
   – Morphologies for SDSS galaxies
   – Banerji et al. 2010 isophotal parameters of use
3) Gaussian Process Regression (Foster et al. 2009)
4) Cross-match catalog built in 5 min (March 2011)
5) Paper written in 2 weeks
   – Received March 25, 2011 : Accepted April 21, 2011
Differences in data compression?

Thesis:
40 DAT (~100GB) + 4 years →

Last Project:
30TB + 1000 years? →
I’m not an Astronomy Chauvinist!

These changes/challenges are not limited to Astronomy

http://www.sciencemag.org/site/special/data
October 2013 in Wired & Quanta

* 2013/10/02: “A Digital Copy of the Universe”

* 2013/10/03: “Big Data Is Too Big for Scientists to Handle Alone”
http://www.wired.com/wiredscience/2013/10/big-data-science

* 2013/10/09: “Scientific Data Has Become So Complex, We Have to Invent New Math to Deal With It”
http://www.wired.com/wiredscience/2013/10/topology-data-sets

* 2013/10/11: “Biology’s Big Problem: There’s Too Much Data to Handle”
http://www.wired.com/wiredscience/2013/10/big-data-biology

http://www.wired.com/wiredscience/2013/10/computers-big-data
Not limited to Natural Science either:

The Humanities

We know **data** is also allowing new collaborations within the humanities & even between science and the humanities:

- History
- Sociology
- Anthropology, Archaeology -- Carbon Dating
- Geology
- Genetics
Lessons for the Business World!?  

Schumpeter

Titans of innovation

What can business learn from Big Science?

Apr 27th 2013 | From the print edition

As a technical feat, ATLAS takes some beating. It is the world’s biggest microscope, used by physicists at CERN, a large laboratory near Geneva, to probe the fundamental building blocks of matter. Its barrel-shaped body, 45 metres long, 25 metres tall and weighing as much as the Eiffel tower, was assembled in a cavern 100 metres beneath the Swiss countryside from 10m parts, nearly twice as many as in a jumbo jet. It generates more data each day than Twitter does.
Even Foreign Affairs

The Rise of Big Data
How It's Changing the Way We Think About the World

By Kenneth Neil Cukier and Viktor Mayer-Schoenberger

http://www.foreignaffairs.com/articles/139104/
kenneth-neil-cukier-and-viktor-mayer-schoenberger/the-rise-of-big-data
Nor are these changes going to solve all of our problems…

Conclusions

1. Science is becoming more data intensive

2. We are exploring new collaborations to deal with the data deluge forced upon us by technological advances

3. This also leads to rethinking our methodologies:
   - Probability theory, statistics, machine learning, data mining, etc…
   - And perhaps our productivity?