Managing More Bits Than Barrels: Learning From Leveraging Subsurface Data → Big Data

One Vision. One Journey. One Team.

Peter Breunig, Chevron

Former seismic processing geophysicist who has listened to some talks about "big data" and has one partially completed proof of concept in his group.

Sept 20, 2011

Agenda

Upstream subsurface workflow and goals.

Seismic method primer – 3 slides

Business value, past and present.

Seismic method – workflow and similarities to big data stuff.

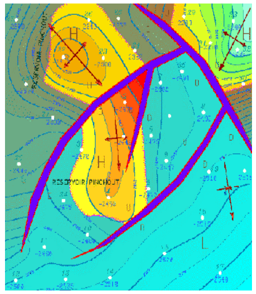
Future exponential data growth areas

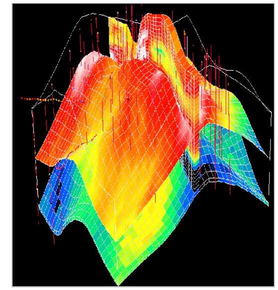
Takeaways

- “Big data” will provide lots of opportunity in the upstream
  - Seismic methods have incorporated some of these “technologies for a long time”
  - Geologic/Geophysical interpretation methods have been merging info between data types to “raise alarms” and present info for decisions for years.
- New technologies are continuing to arrive that break down approximations, and barriers to new technologies.
  - Big data should pay attention to where the new opportunities are.
- We are trying something in this space – Hadoop.

The Big Picture of subsurface work

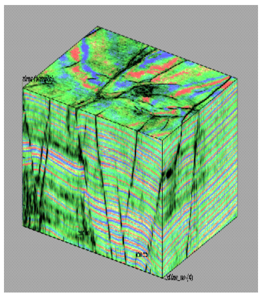
Mapping



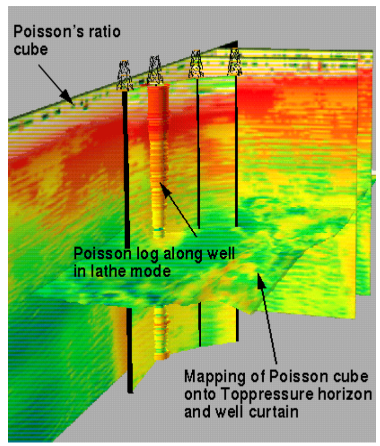


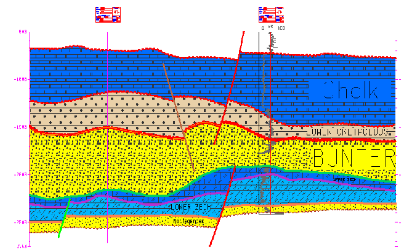
Reservoir characterization

Seismic Interpretation



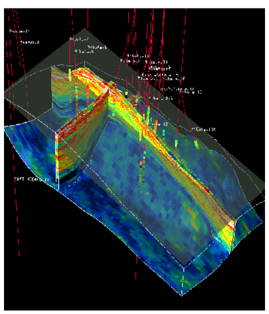
Common Earth Model

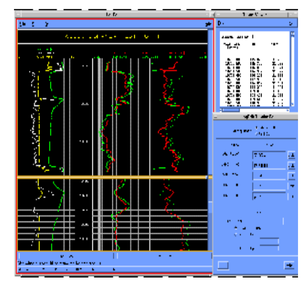




Cross-Sections

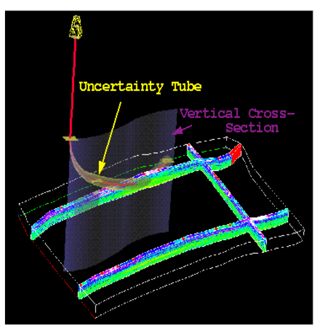
Stratigraphic Modeling

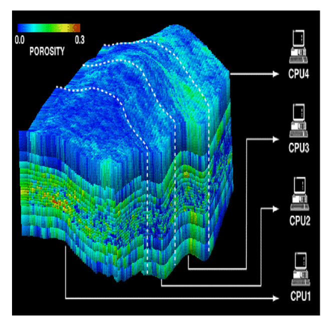




Petrophysics

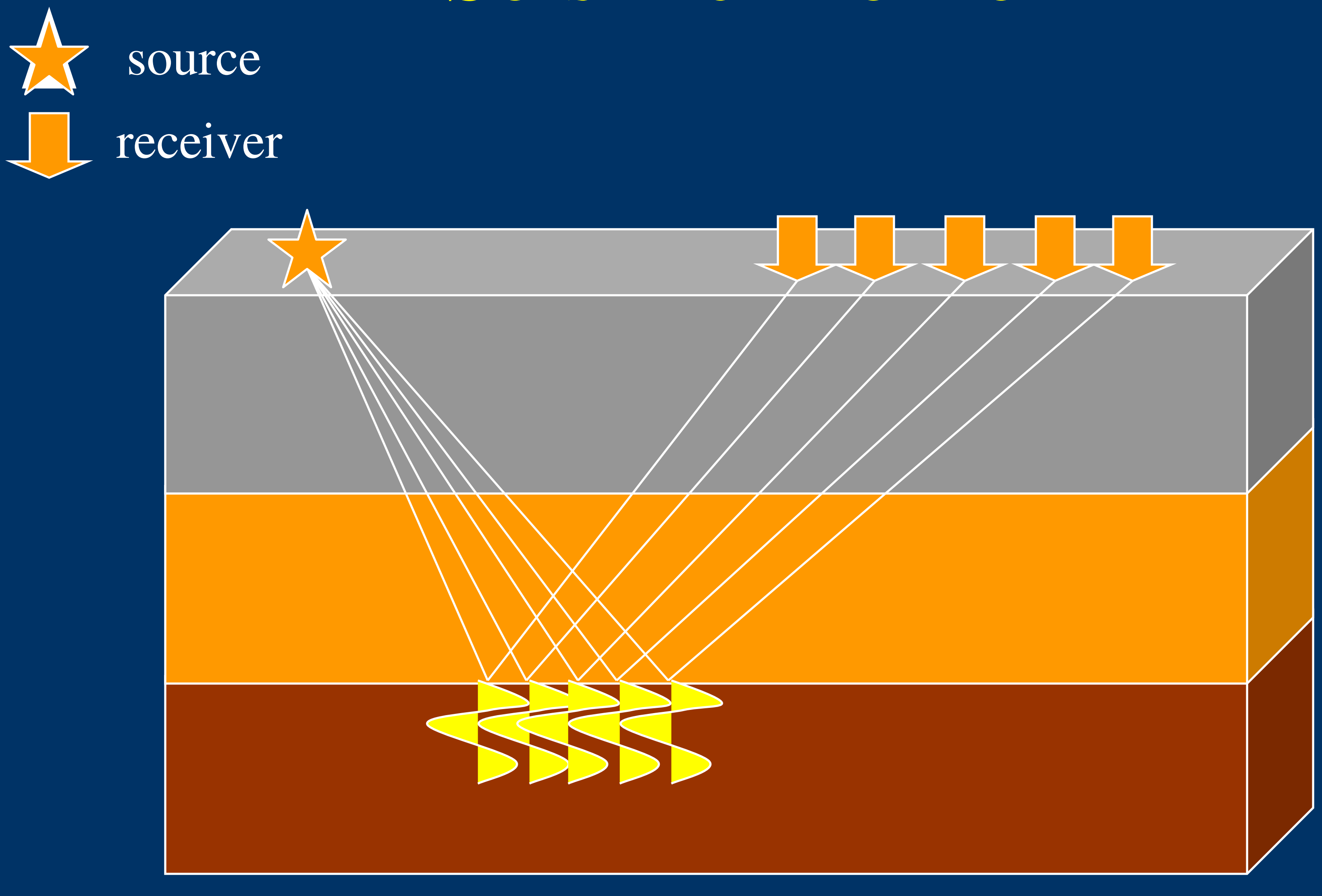
Well planning & Drilling Simulation



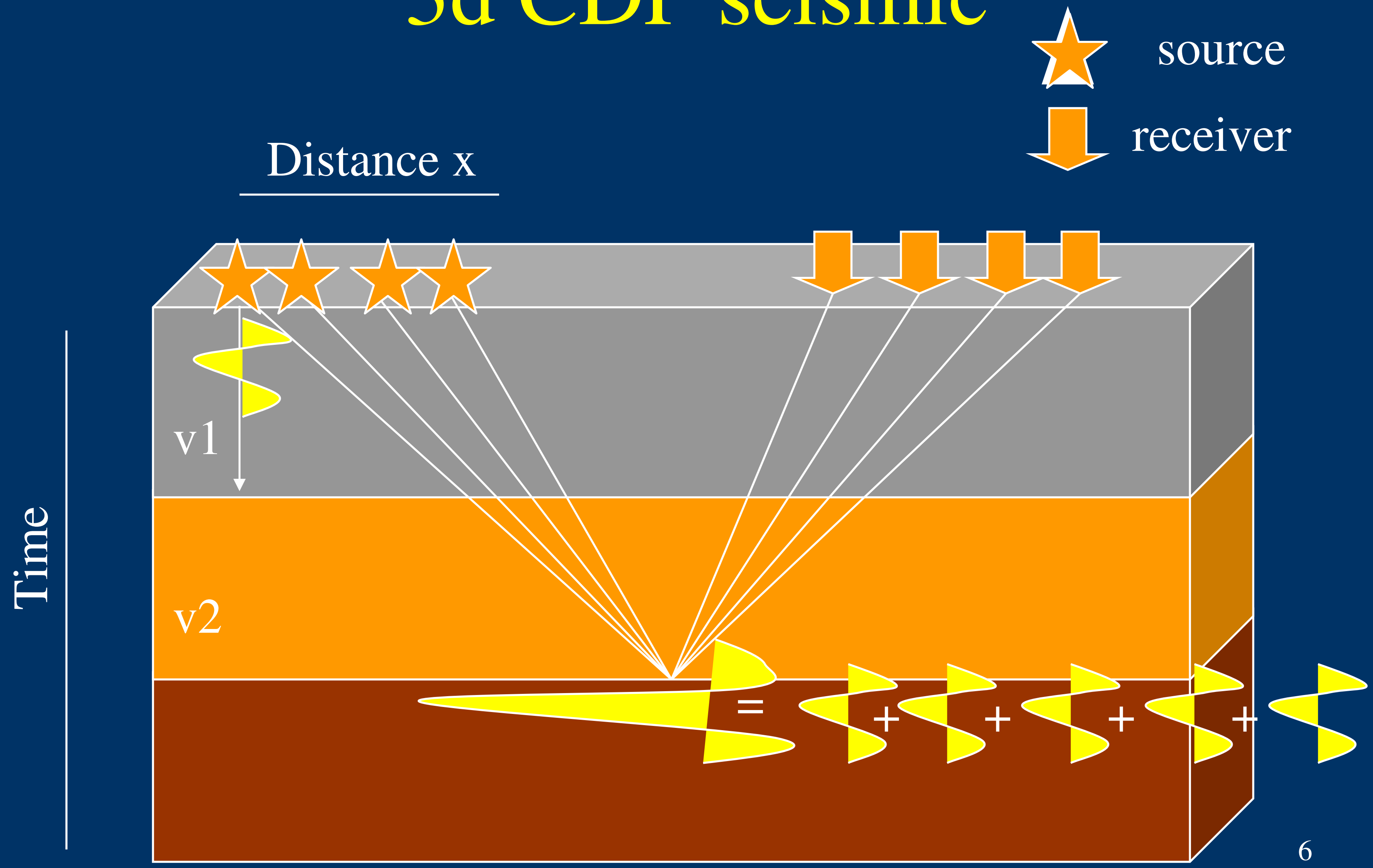


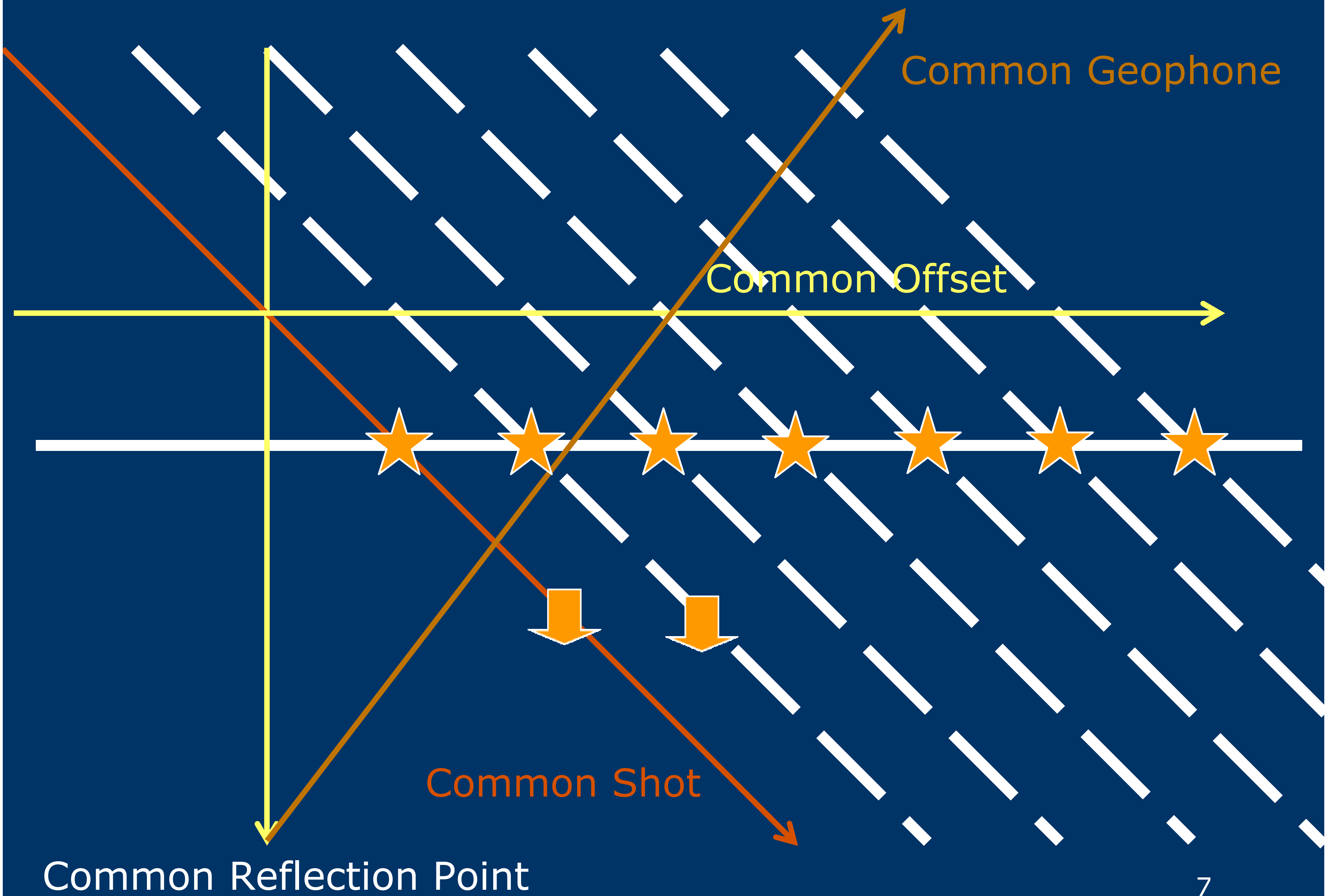
Reservoir Simulation

Seismic Profile



3d CDP seismic




Common Geophone
Common Offset
Common Shot
Common Reflection Point

1980's 2D seismic work flow and relation to big data activities

1. Record – tape (then disk, then ?) ➔ managing large data ➔ storage
2. *Reformat for internal system ➔ sorting
3. *Apply field geometry ➔ ?
4. Recover the amplitudes in the waves ➔ ?
5. Deconvolution ➔ ?
6. *Apply field statics – account for variations in the near surface elevations and velocity. ➔ ?
7. *Determine velocities ➔ opportunity to automate velocity analysis?
8. *Residual statics ➔ opportunity to further automate this?
9. *CDP stacking – improve signal to noise ➔ sorting, S/N enhancement?
10. Filtering (appropriate hertz) ➔ ?
11. *Seismic migration ➔ sorting, summing, possibly other operations.
12. *Signal enhancement exercises
    - Frequency and Amplitude analysis
    - Edge detection
    - Coherency enhancements
    - Deterministic deconvolution
    - Multiple elimination
    - Palinspastic Restoration
    - Extraction of info from seismic.
        - Principle component analysis,
        - Auto picking
        - Geo-body recognition
13. Interpret the data ➔ see above.
    - *Visualization – Seismic and earth model environments visualized

* - sorting of some type – diracc dial

HPC Value
Chevron Cray 1985-1989

- The Cray cost roughly $10mm over 3 years.
- $10,000/day.
- Feed the beast was the mantra.

HPC Value
Deepwater Drill Ship 2007

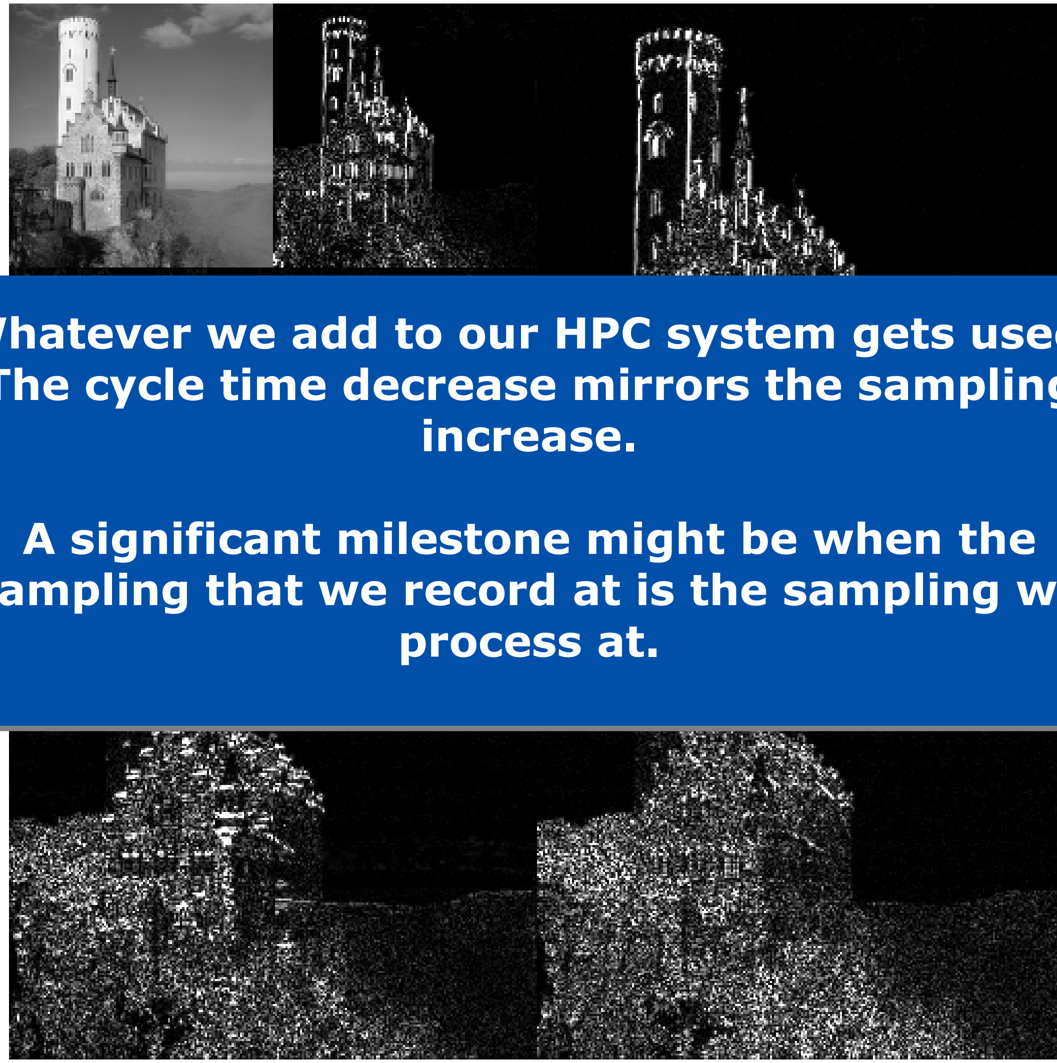
- Drill Ship costs roughly $1b over 3 years.
- $1,000,000/day.
- Which beast are we feeding?

What becomes critical to the digital technology part of the energy business.

- ✓ Technology application is critical to adding value.
- ✓ Remote operations will be critical in deepwater.
- ✓ Remote operations may be critical in shelf and land environments.
- ✓ Big service companies providing all the innovation is not going to happen (margins are too small).
- ✓ Improved **resolution** within the reservoir is critical because:
  - ✓ deepwater wells cost a lot,
  - ✓ fully exploiting existing assets is essential.
- ✓ Integration opportunities become critical.
- ✓ Innovation will come from the "fringes of technology", improving equations, **reducing approximations** and refinement of measurement.
- ✓ Workflow efforts will be critical to define business value.

Sampling and Resolution

Original Resolution, recorded resolution



- Whatever we add to our HPC system gets used. The cycle time decrease mirrors the sampling increase.

- A significant milestone might be when the sampling that we record at is the sampling we process at.

Utilizing the Hadoop Framework in the seismic processing space at Chevron

- Leveraging IBM Big Insights technology, which includes the Hadoop component stack
- Deployed small cluster in Chevron ITC/TMA lab consisting of 12 Linux nodes
- Targeted POC focusing on 2 to 3 applied functional aspects of the seismic imaging processing workflow
  - Data is SEGY format, fairly large, one shot contains +- 400K traces, = roughly 2.5GB. POC dataset is 50TB, of which, we are using a small percentage due to the small size of the POC lab cluster
  - POC functionality:
    1. Sorting of five dimensional traces
    2. One-dimensional filtering/transformation
    3. Two-dimensional filtering/transformation
- Challenges:
  - Merging the seismic computing world with the Hadoop framework. Approach for POC is to leverage the seismic unix modules, an open source seismic processing library.
  - Access to, and time with seismic SMEs at Chevron
- Potential Hadoop Pros:
  - Highly fault tolerant, compute moves to the data, large extra-Chevron community of expertise
- Potential Hadoop Cons:
  - Data isolation requirements (shared nothing architecture), may not be as fast as MPI

Present Day Methods

Differential/Wave Imaging Methods

3D Reverse-Time Migration (Time extrapolation, $$)

3D Wavefield Migration (Depth extrapolation, $)

Integral/Ray Imaging Methods

3D Kirchhoff / Gaussian Beam

3D Acoustic/PseudoAnisotropic Wavefield Modeling

2D Full Wavefield Inversion (proof of concept)

Some not so Future Methods

- 3D Elastic Anisotropic Modeling
- 3D Elastic Anisotropic Reverse Time Migration & Imaging with Multiples
- 3D Full Wavefield (constrained) Inversion of Primaries & Multiples
- Iterative Wavefield Modeling for Stochastic Inversion
- 60's → Digital, 70's → Wave equation migration (post stack), 80's → Dip Moveout, 90's → Pre stack depth migration, 00's → Anisotropy → Oz Yilmaz ~ 1999.
- 10's → Acquisition, 20's → Full modeling?

Seismic numerical requirements for 3DFD

	Low ($)	Medium ($$)	High ($$$)
Earth Complexity	Const Den	Acoustic (shear=0)	Elastic (shear involved)
Ops / spacetime pt	50	100	200
Shot Vol X, Y, Z (km)	10 10 5	20 20 8	30 30 12
Max Frequency*	15	30	60
Geo resolution (m)	30	15	8
Vel Accuracy (rel.err)	0.010	0.003	0.001
Simulation Time (sec)	6	9	12
Total #Shots	500	5000	50000
Total Runtime Mem (Gb)	0.4	20	1000
$\log_{10}$(Total Float Pt Ops)	16	19	22
1000-cpu-days	0.01-1	10-100	10000
Cost **	$.02K-$2K	$20K-$200K	$20000K

(* Total cost goes as 4th power (xyzt) of max frequency → 10% extra bandwidth costs 40% more)
(** based on $20K 8cpu Opteron box depreciated over 3 years: $20000 / (8cpu*24*365*3 hr) ~ $0.10/cpu-hr)
Message passing vs shared memory,
Hardware improvements drive algorithmic improvements

Modeling Lessons Being Learned

Value in modeling lies in its fidelity to the observed data
Custom coding (earth & seismic) usually *necessary*
Accurate guides (seismic samples, type logs) are *crucial*
Imperative to *iterate* on the model (2D ac/el → 3D anis)
Provide enough lead *time*
Each $ucce$$ive $tep ~order of magnitude cost increase

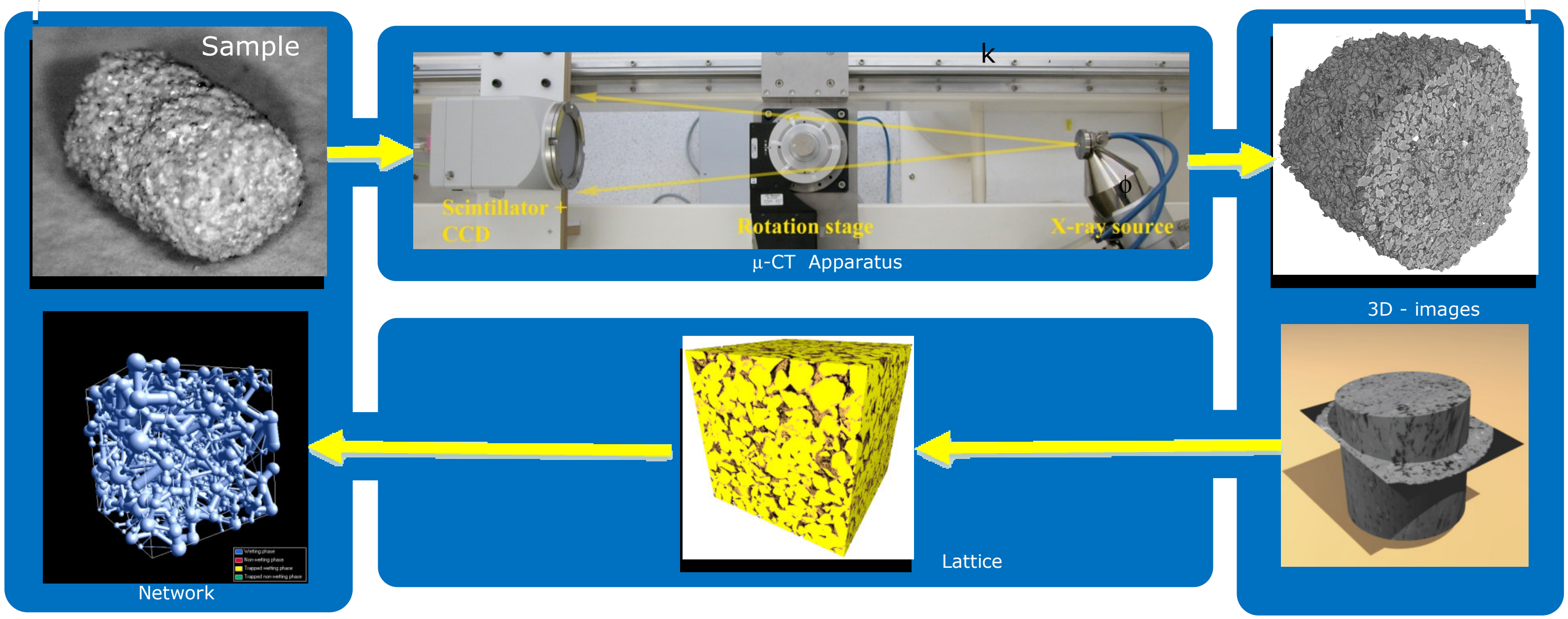
	VOI modeling →	Acquisition →	Drilling →	Development
log($Cost):	5	8	9	10

So:
Don't skimp upfront on the overall modeling

Digital Rocks -- Emerging Technology

Advances in high resolution imaging and computation enable rapid calculation of laboratory scale data

Quantitative 3-d pore scale measurements



Calculations on lattice or network

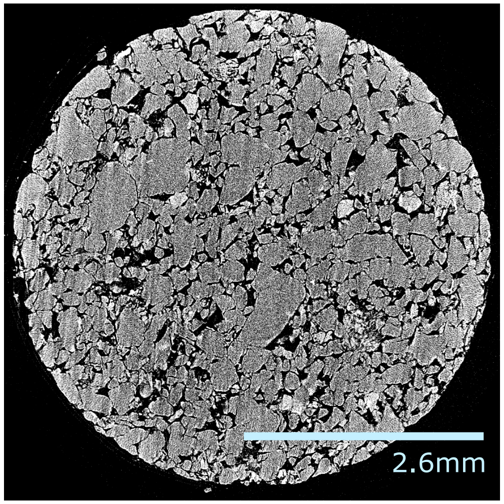
Porosity *Permeability* *Formation Factor*

Capillary Pressure *Relative permeability*

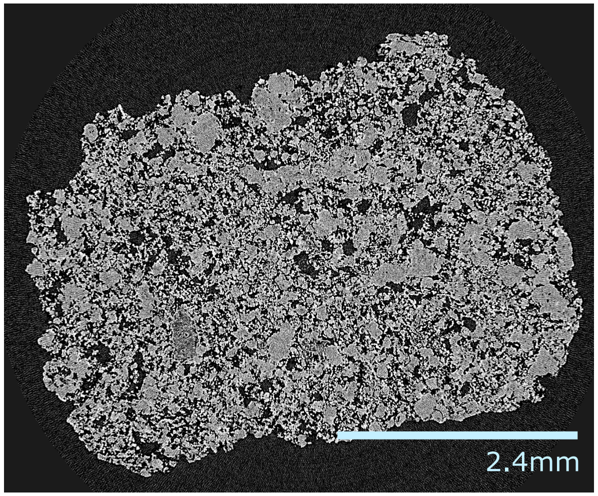
Resistivity Index *Elastic* *NMR*

Pore scale Imaging (.01 – 10 microns) Can generate large data sets

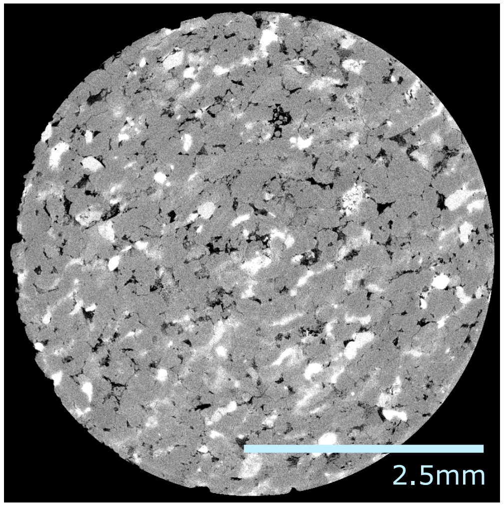
- Image of 1 cm^3 volume can exceed 100 Gb
- Imaging rock in a typical 1000 meter core exceeds 1 Exabyte



19% 300md
8bit 4000^3 @ 1.5μm - 50Gb

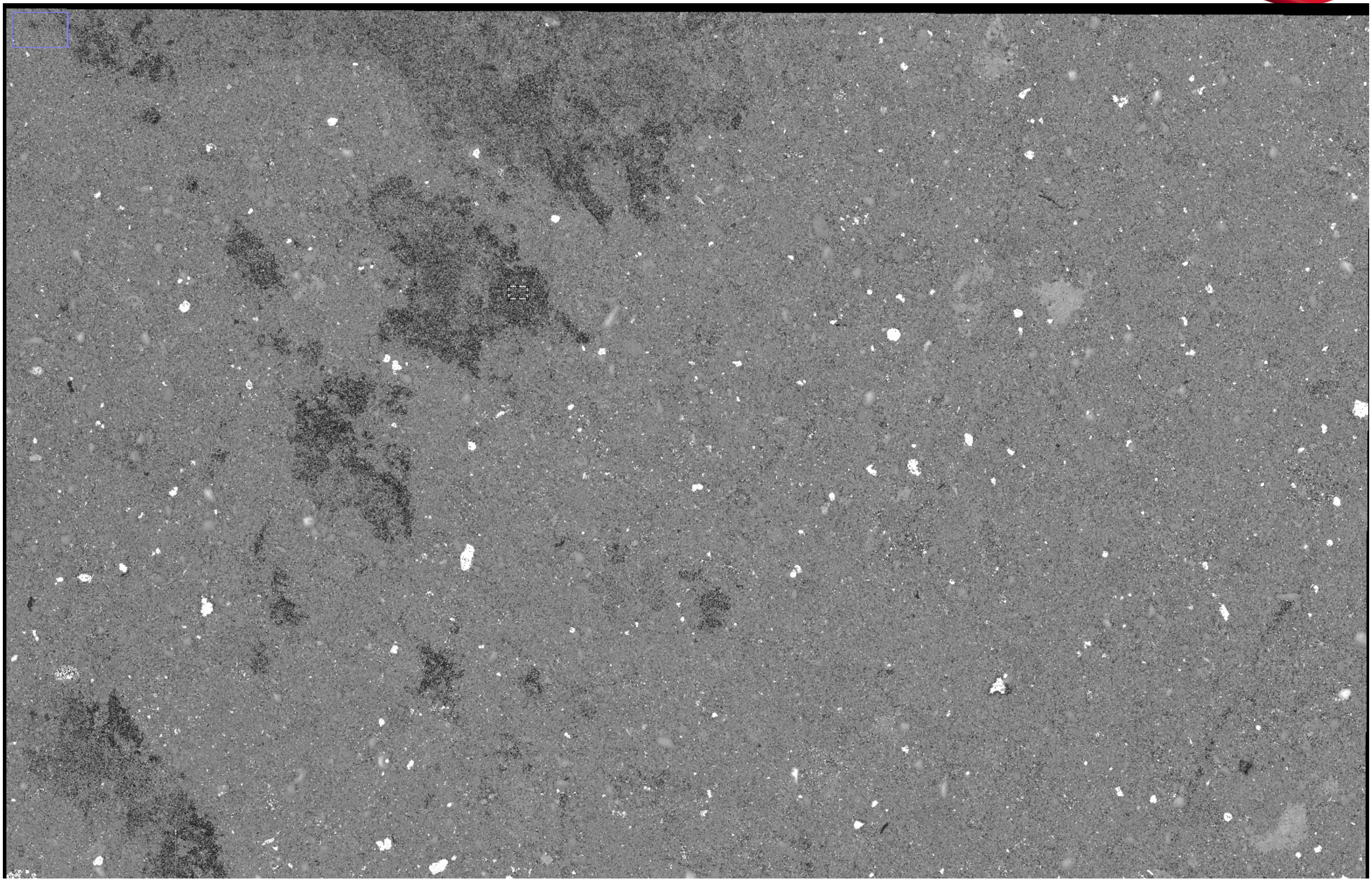


38% 700md
16bit 4000^3 @ 1.8μm - 100Gb



9% 0.01md
16bit 4000^3 @ 2.7μm - 100Gb

Example of very high resolution images with ability to zoom into regions and make calculations

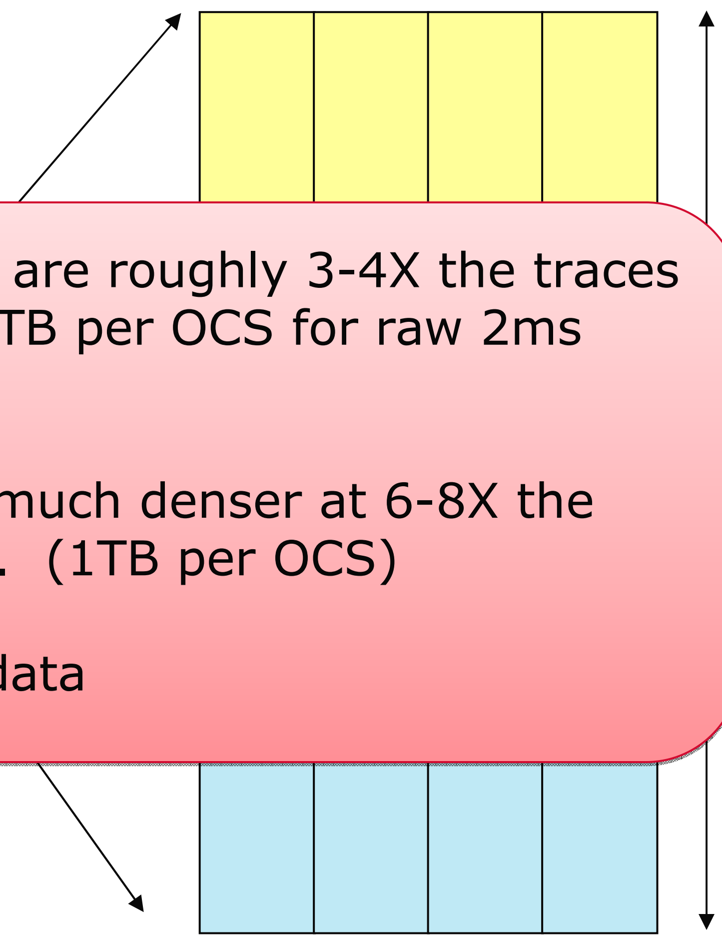


150Gb - 300,000^2 pixels - 0.03µm/pixel - zooming region about 50µm

New Acquisition - Wats

WATS – How It's Done:

- *Multiple Passes*
- *Multiple Source Vessels*



Exploration WAZ programs are roughly 3-4X the traces per sq/km vs. NATS (0.50 TB per OCS for raw 2ms traces).

Development WAZ can be much denser at 6-8X the traces per sq/km vs. NATS. (1TB per OCS)

So with success → 8x the data

Multiple passes provide an effective streamer width not otherwise obtainable. Each pass provides one receiver "tile".

Multiple sources provide an effective streamer length not otherwise obtainable.