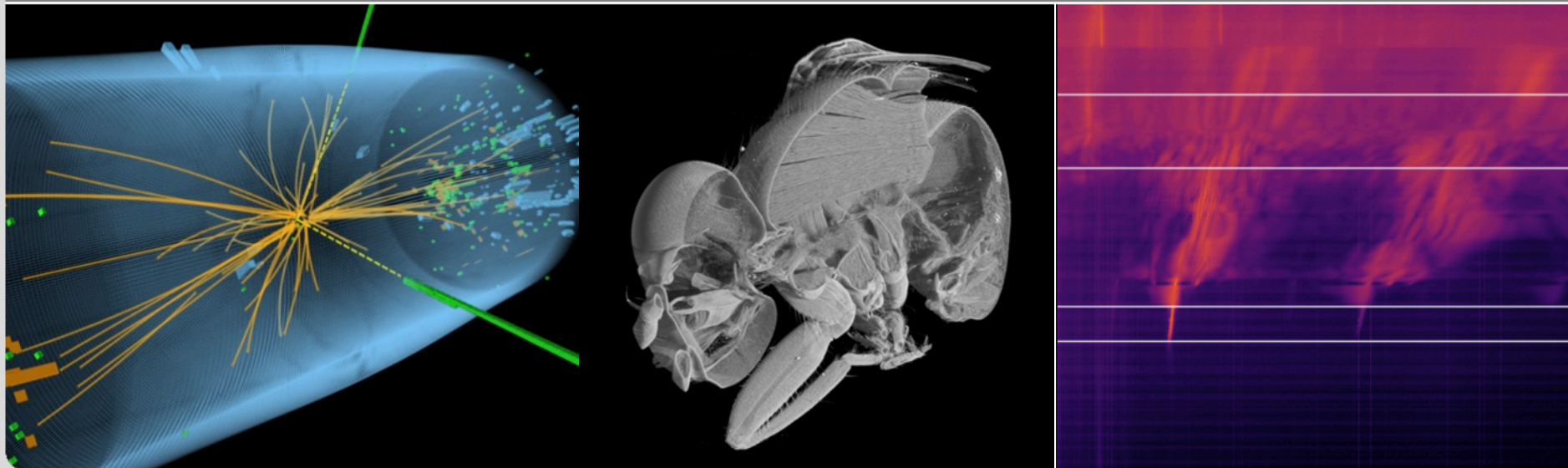


*ARD topical workshop on “ps - fs Electron and Photon Beams”*

# Online processing with GPUs in data acquisition systems for improved data quality and control

Andreas Kopmann

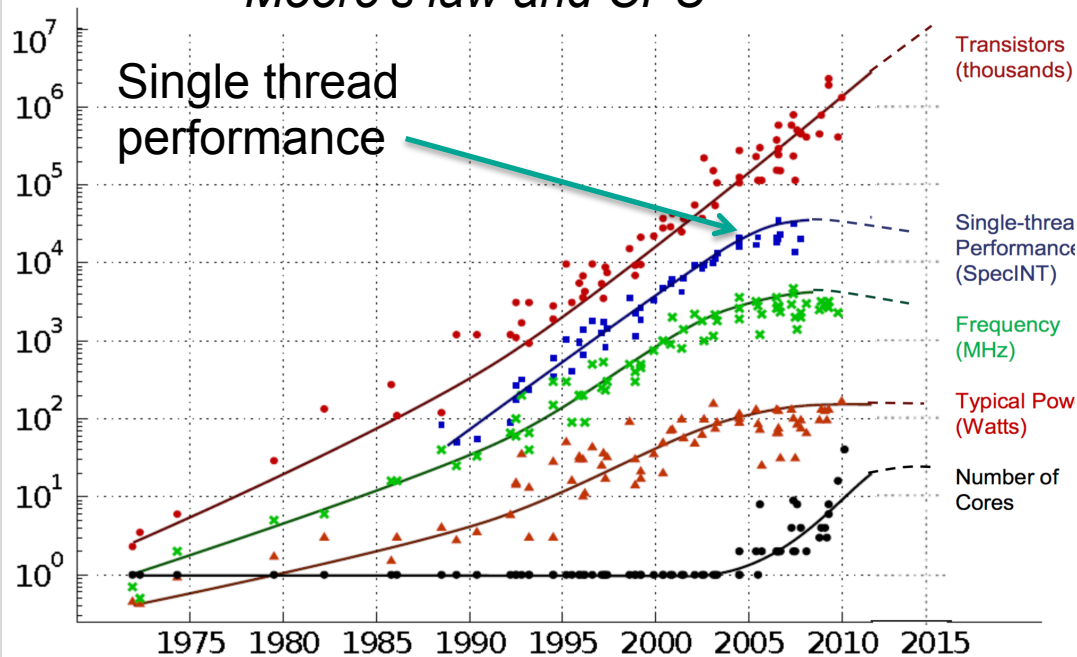
Institute for Data Processing and Electronics



# Scientific computing with GPUs

- Single core age ended -> *parallel programming is required!*
- GPUs are fast, cheap and scalable (up to 4 in a PC)
- TeraFLOP applications: e.g. tomography (2TFLOP/GB)

Moore's law and CPU



CPU vs GPU

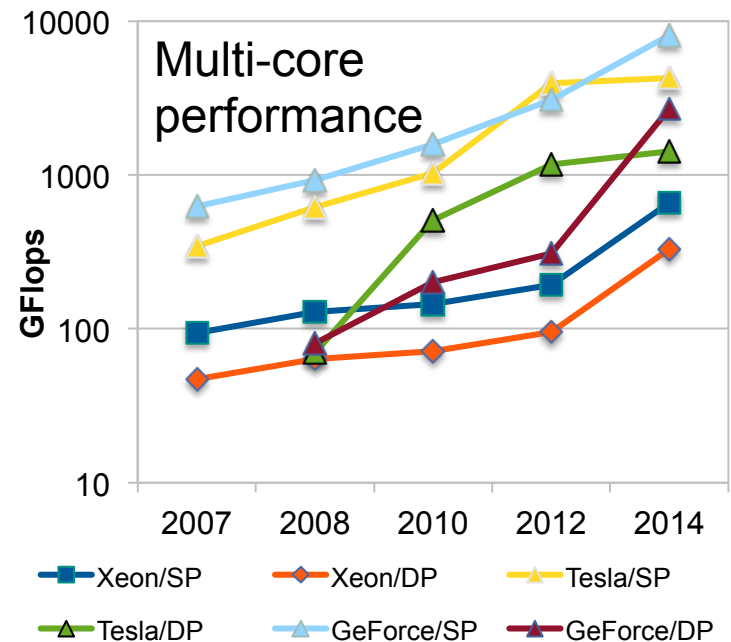
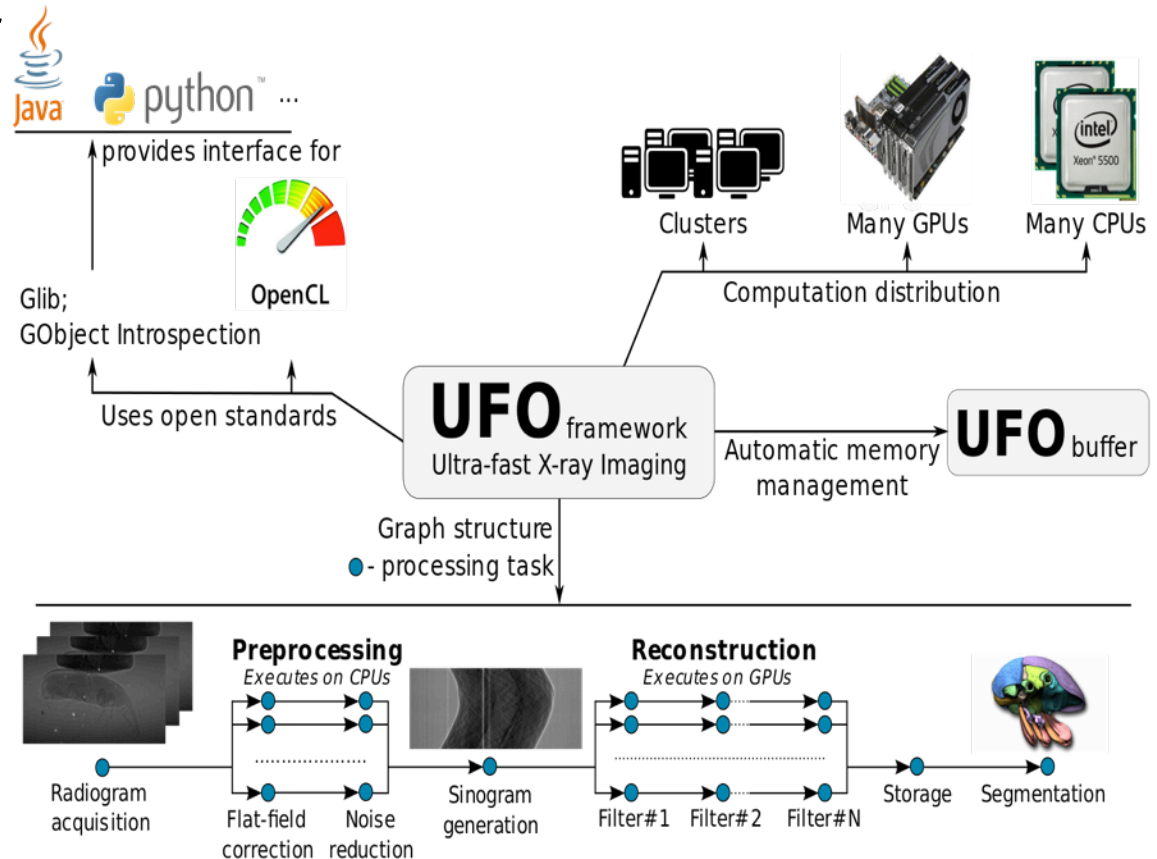


Figure: Chuck Moore, AMD Technology Group CTO

# Ultra-fast X-ray imaging – “UFO Framework”

## Goals:

- Pipelined processing of **data streams**
- (Re-)use of **optimized algorithms**
- Automatic scheduling
  - **CPUs, GPUs, ...**
- **Integration in**
  - Control system
  - Analysis tools
- **Easy to use for**
  - Users, Admins, Developers



Vogelgesang M et al, Proc HPC-CESS (2012) 824-829  
 Vogelgesang M et al, Proc ICALEPS (2013)

<http://ufo.kit.edu>

# Custom DAQ Electronics

## UFO DAQ Platform

Embedded FPGA  
for online data  
analysis  
⇒ open design  
⇒ library of  
standard IP cores

Modular sensor interface  
⇒ rapid development for  
new sensors

High-throughput link  
(PCIe, <15GB/s)  
⇒ streaming to  
GPU for processing

### Applications:

- Phase contrast tomography P05/07
- KAPTURE electron beam characterization, ATP  
*see M Caselle*

Enabling HPC technologies  
for DAQ systems



# But why online processing?

- Improve data quality
  - Optimize experimental setup
- Reduce data size
  - Apply pre-analysis steps, when possible
- Design new experiments
  - Automatic control



Minerva, Max Planck Institute for Gravitational Physics



# Data-Intense Applications in Photon Science and Machine Physics

## ■ Facilities

- PETRA3
- XFEL
- Acc. Test Platform ATP: ANKA, FLUTE, ..

## ■ Data-intense methods

- Tomography
- Scattering
- Electron Beam monitoring
- ...

## ■ Detectors

- Commercial cameras
- X-ray detectors
- Fast digitizers



High-speed cameras  
(e.g. pco.dymax)



Large-area X-ray  
detectors  
(e.g. Pilatus 6M)



Custom detectors  
(e.g. KAPTURE)

# Detector Rates in Photon Science and Machine Physics

Device	Rate MB/s	Rate TB/8h	Ch.	ADC bit	Rate Hz	Beamline
“CL Base”	255	7				e.g. pco.dymax
“CL Full”	800	22	4M	16	100	e.g. pco.edge
Pilatus 6M	223	6	6M	20	10-12,5	P11
UFO Camera	1125	31	20M	12	30	P05/07
KAPTURE	3815	105	4	12+4	500M	ANKA
KAPTURE2	15259	419	4-8	12+4	1G	
KALYPSO2	4883	134	512	14	5MHz	

- CL = Camera Link
- All devices without compression

# “Real-life” Data Rates (2015-16)

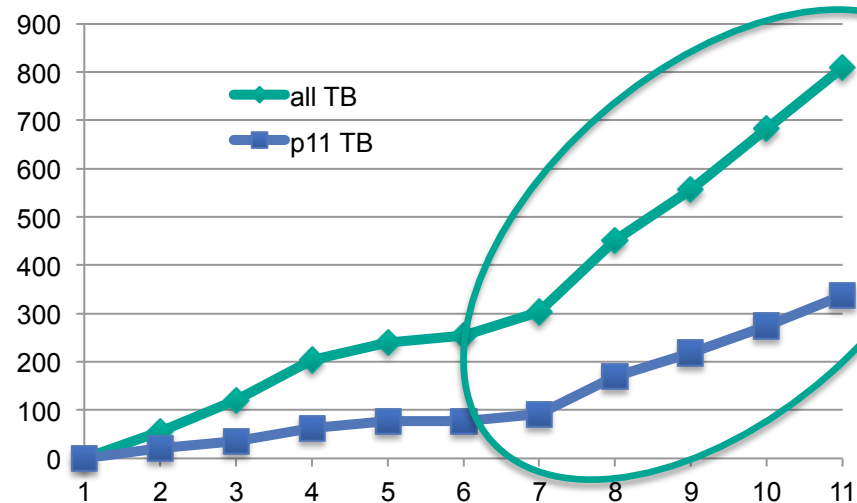
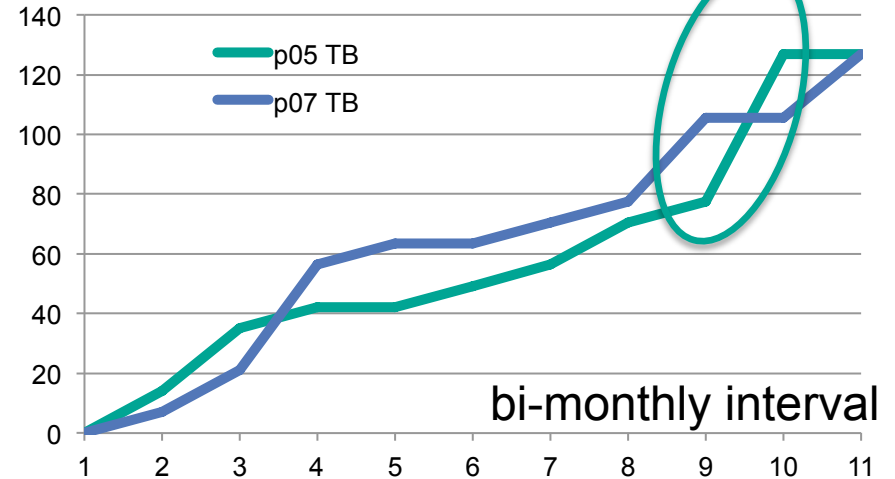
## PETRA3

- Tomography (P05/P07)
  - Mean 75 TB/a
  - Max 300 TB/a (eq. 75h op)
  
- Imaging / Diffraction (P11)
  - Mean 200 TB/a
  - Max 370 TB/a (eq. 500h op)

## ANKA

- Tomography (TopoTomo)
  - Mean 40 TB/a (eq. 42h op.)
- KAPTURE
  - Mean 10 TB/a (eq. 1h op.)

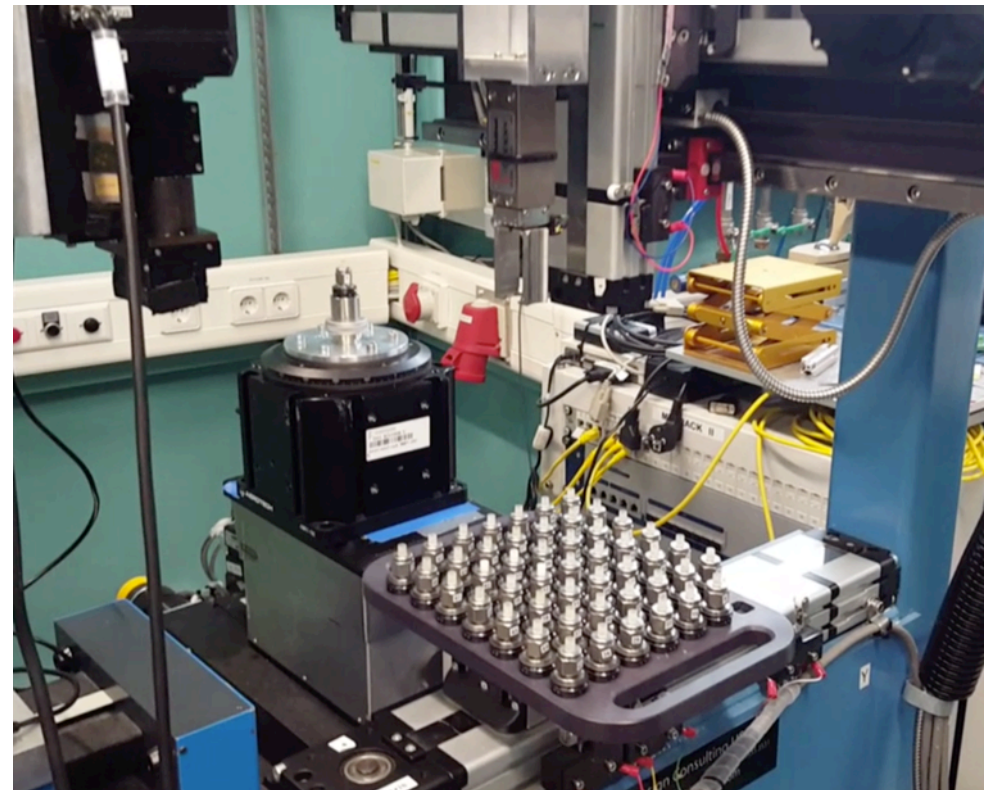
extracted from previous chart





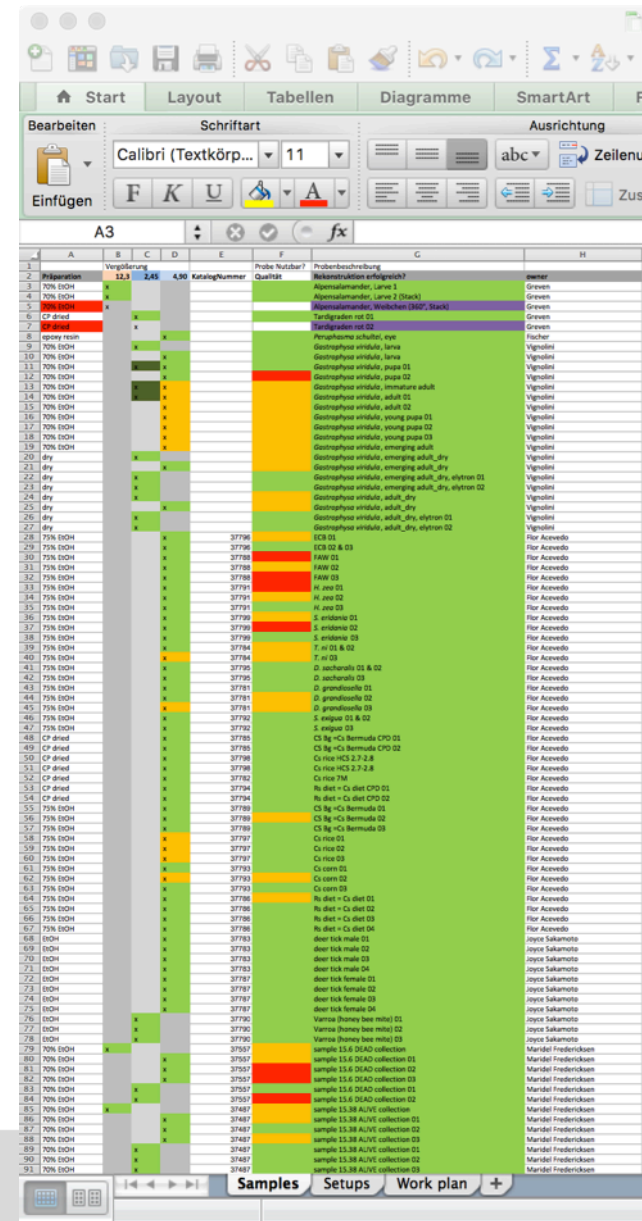
# Beamline Automation

- Robots for sample changers
- Sample identification (e.g. by QR code)
- Camera revolver, lenses
- Advanced control systems



# Tomography Services for Morphology

- ANKA domain expert collects samples
  - Check of sample quality
- Measurement at ANKA
  - Beamtime 6 days (e.g. Nov 2015)
  - Up to 300 samples from 14 scientists
  - Methods: Radiology, Tomography, Cine-Tomography
  - Samples: in alcohol, dried, amber, ...
  - More than 90% of the samples are usefull
  - Interested scientists might join
- Datasets are automatically transferred to processing storage
  - Analysis by users



	A	B	C	D	E	F	G	H
	Preparation	1.1	2.45	4.90	Katalognummer	Probe Nummer?	Probenbeschreibung	
2	70% DIH2O	x					Alpenalmander, Larve 1	Green
3	70% DIH2O	x					Alpenalmander, Larve 2 (Dunk)	Green
4	70% DIH2O	x					Alpenalmander, Weibchen (3007, tauch)	Green
5	70% DIH2O	x					Tardigraden tot 01	Green
6	CP dried	x					Tardigraden tot 02	Green
7	CP dried	x						Green
8	epoxy resin	x					Penicillium schultzi, eye	fischer
9	70% DIH2O	x					Gastrophysa viridula, larva	Vignolini
10	70% DIH2O	x					Gastrophysa viridula, larva	Vignolini
11	70% DIH2O	x					Gastrophysa viridula, pupa 01	Vignolini
12	70% DIH2O	x					Gastrophysa viridula, pupa 02	Vignolini
13	70% DIH2O	x					Gastrophysa viridula, immature adult	Vignolini
14	70% DIH2O	x					Gastrophysa viridula, adult 01	Vignolini
15	70% DIH2O	x					Gastrophysa viridula, adult 02	Vignolini
16	70% DIH2O	x					Gastrophysa viridula, young pupa 01	Vignolini
17	70% DIH2O	x					Gastrophysa viridula, young pupa 02	Vignolini
18	70% DIH2O	x					Gastrophysa viridula, emerging adult	Vignolini
19	70% DIH2O	x					Gastrophysa viridula, emerging adult_dry	Vignolini
20	dry	x					Gastrophysa viridula, emerging adult_dry	Vignolini
21	dry	x					Gastrophysa viridula, adult_dry	Vignolini
22	dry	x					Gastrophysa viridula, emerging adult_dry, elytrium 01	Vignolini
23	dry	x					Gastrophysa viridula, emerging adult_dry, elytrium 02	Vignolini
24	dry	x					Gastrophysa viridula, adult_dry	Vignolini
25	dry	x					Gastrophysa viridula, adult_dry, elytrium 01	Vignolini
26	dry	x					Gastrophysa viridula, adult_dry, elytrium 02	Vignolini
27	dry	x						Vignolini
28	75% DIH2O	x			37786		ECB 01	Ror Accardo
29	75% DIH2O	x			37786		ECB 02 & 03	Ror Accardo
30	75% DIH2O	x			37788		FAW 01	Ror Accardo
31	75% DIH2O	x			37788		FAW 02	Ror Accardo
32	75% DIH2O	x			37788		FAW 03	Ror Accardo
33	75% DIH2O	x			37791		R. v. jesi 01	Ror Accardo
34	75% DIH2O	x			37791		R. v. jesi 02	Ror Accardo
35	75% DIH2O	x			37791		R. v. jesi 03	Ror Accardo
36	75% DIH2O	x			37799		S. erianota 01	Ror Accardo
37	75% DIH2O	x			37799		S. erianota 02	Ror Accardo
38	75% DIH2O	x			37799		S. erianota 03	Ror Accardo
39	75% DIH2O	x			37784		T. vi. 01 & 02	Ror Accardo
40	75% DIH2O	x			37784		T. vi. 03	Ror Accardo
41	75% DIH2O	x			37795		D. saeharae 01 & 02	Ror Accardo
42	75% DIH2O	x			37795		D. saeharae 03	Ror Accardo
43	75% DIH2O	x			37781		D. grandiosella 01	Ror Accardo
44	75% DIH2O	x			37781		D. grandiosella 02	Ror Accardo
45	75% DIH2O	x			37781		D. grandiosella 03	Ror Accardo
46	75% DIH2O	x			37782		S. vespa 01 & 02	Ror Accardo
47	75% DIH2O	x			37782		S. vespa 03	Ror Accardo
48	CP dried	x			37785		CS Bg + CI Bermuda CPO 01	Ror Accardo
49	CP dried	x			37785		CS Bg + CI Bermuda CPO 02	Ror Accardo
50	CP dried	x			37786		CS vicia HCS 2.1 2.8	Ror Accardo
51	CP dried	x			37788		CS vicia HCS 2.2 2.8	Ror Accardo
52	CP dried	x			37782		CS vicia 7M	Ror Accardo
53	CP dried	x			37794		Ru diet + Cs diet CPO 01	Ror Accardo
54	CP dried	x			37794		Ru diet + Cs diet CPO 02	Ror Accardo
55	75% DIH2O	x			37789		CS Bg + CI Bermuda 01	Ror Accardo
56	75% DIH2O	x			37789		CS Bg + CI Bermuda 02	Ror Accardo
57	75% DIH2O	x			37789		CS Bg + CI Bermuda 03	Ror Accardo
58	75% DIH2O	x			37787		CS vicia 01	Ror Accardo
59	75% DIH2O	x			37787		CS vicia 02	Ror Accardo
60	75% DIH2O	x			37787		CS vicia 03	Ror Accardo
61	75% DIH2O	x			37783		CS corn 01	Ror Accardo
62	75% DIH2O	x			37783		CS corn 02	Ror Accardo
63	75% DIH2O	x			37793		CS corn 03	Ror Accardo
64	75% DIH2O	x			37786		Ru diet + Cs diet 01	Ror Accardo
65	75% DIH2O	x			37786		Ru diet + Cs diet 02	Ror Accardo
66	75% DIH2O	x			37786		Ru diet + Cs diet 03	Ror Accardo
67	75% DIH2O	x			37788		Ru diet + Cs diet 04	Ror Accardo
68	DIH2O	x			37783		deer tick male 01	Joyce Salamotto
69	DIH2O	x			37783		deer tick male 02	Joyce Salamotto
70	DIH2O	x			37783		deer tick male 03	Joyce Salamotto
71	DIH2O	x			37783		deer tick male 04	Joyce Salamotto
72	DIH2O	x			37787		deer tick female 01	Joyce Salamotto
73	DIH2O	x			37787		deer tick female 02	Joyce Salamotto
74	DIH2O	x			37787		deer tick female 03	Joyce Salamotto
75	DIH2O	x			37787		deer tick female 04	Joyce Salamotto
76	DIH2O	x			37790		Varnia (honey bee mite) 01	Joyce Salamotto
77	DIH2O	x			37790		Varnia (honey bee mite) 02	Joyce Salamotto
78	DIH2O	x			37790		Varnia (honey bee mite) 03	Joyce Salamotto
79	70% DIH2O	x			37557		sample 13.6 DIAD collection	Mandel Fredericksen
80	70% DIH2O	x			37557		sample 13.6 DIAD collection 01	Mandel Fredericksen
81	70% DIH2O	x			37557		sample 13.6 DIAD collection 02	Mandel Fredericksen
82	70% DIH2O	x			37557		sample 13.6 DIAD collection 03	Mandel Fredericksen
83	70% DIH2O	x			37557		sample 13.6 DIAD collection 04	Mandel Fredericksen
84	70% DIH2O	x			37557		sample 13.6 DIAD collection 05	Mandel Fredericksen
85	70% DIH2O	x			37557		sample 13.8 ALIVE collection	Mandel Fredericksen
86	70% DIH2O	x			37487		sample 13.8 ALIVE collection 01	Mandel Fredericksen
87	70% DIH2O	x			37487		sample 13.8 ALIVE collection 02	Mandel Fredericksen
88	70% DIH2O	x			37487		sample 13.8 ALIVE collection 03	Mandel Fredericksen
89	70% DIH2O	x			37487		sample 13.8 ALIVE collection 04	Mandel Fredericksen
90	70% DIH2O	x			37487		sample 13.8 ALIVE collection 05	Mandel Fredericksen
91	70% DIH2O	x			37487		sample 13.8 ALIVE collection 06	Mandel Fredericksen

# Quantitative Studies of Fly Puparia

- Digitalization of 30 million-year-old phosphatized insects
- Little-known fossil type!

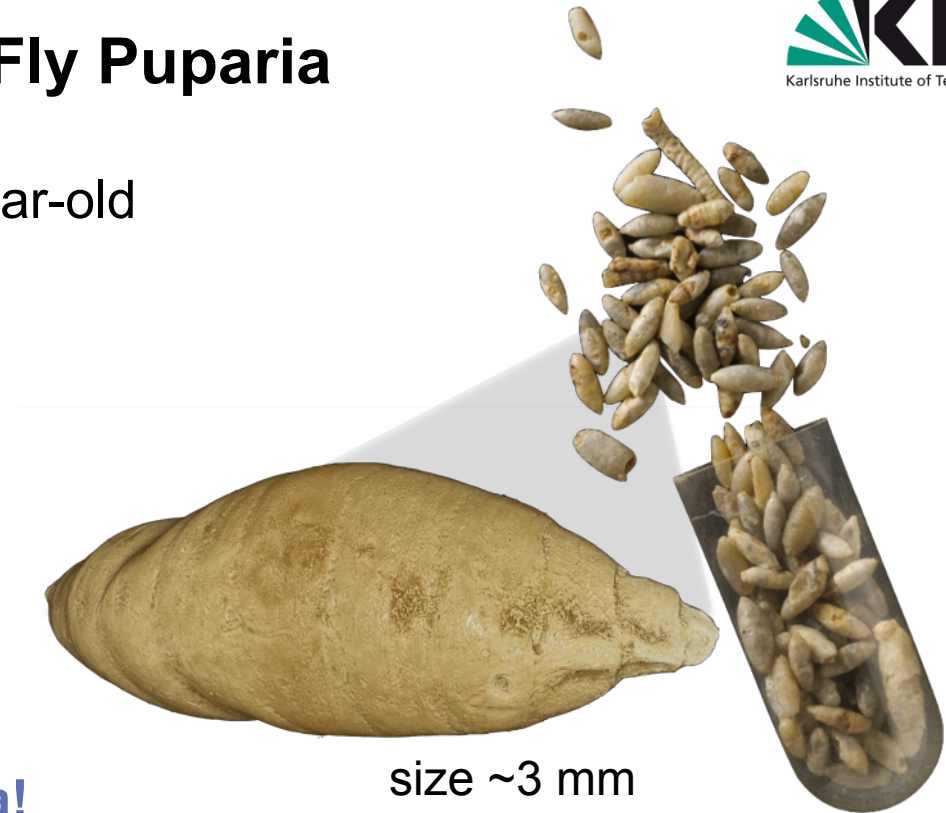
## Scanning:

- February 2016: 29 puparia
- August 2016: 1379 puparia

**Result: 1408 scanned puparia!**

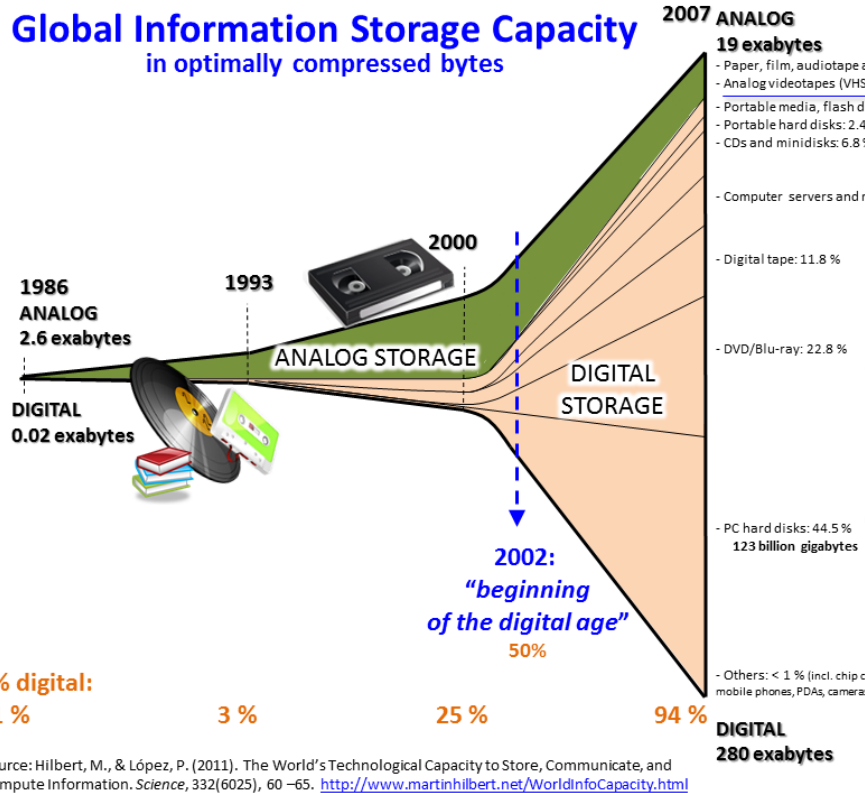
**Datasets: ~60 TB**

**Automation enables  
quantitative studies**



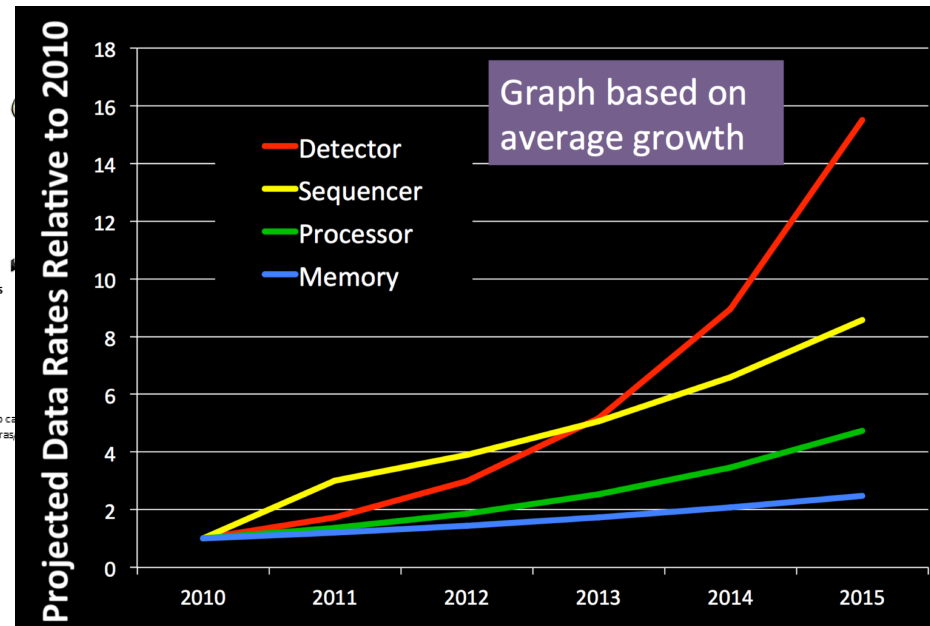
# Digital transformation and large datasets

- “Big datasets are so large or complex that traditional data processing applications are inadequate” (Wikipedia)



Nersc: “Data growth is outpacing Computer growth”

Source: Yelick K, LBNL/Nersc



According to Moore’s law: factor 2 every 2a



# Data Challenges:

## 1. Reduce Data as Early as Possible

- Adopt pre-processing, reconstruction to detector rates
- Avoid user interaction, define workflows
- Make use of trigger to filter relevant data
- Assess quality of data
  
- Heterogeneous FGPA-GPU computing is very powerful
- But tasks are often domain specific
  
- Solution:  
Modularity and frameworks with core functionality

## 2. Adopt Analysis for Large Datasets

- Huge amount of data
  - User often do not have the necessary resources (computers, memory, software) available in order to process the datasets at their home institutions
  
- Data analysis is still very time-consuming
  - Segmentation of tomographic datasets is usually done manually
  
- Solutions:
  - Establish analysis centers at the facilities
  - Involve more scientists by collaborative analysis and open data
  - Re-evaluate known algorithms with recent computing performance

Verbundforschung:

astor --> nova



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386



Zentrum für Material- und Küstenforschung



Friedrich-Schiller-Universität Jena



ERNST MORITZ ARNDT  
UNIVERSITÄT GREIFSWALD



Wissen  
lockt.  
Seit 1456

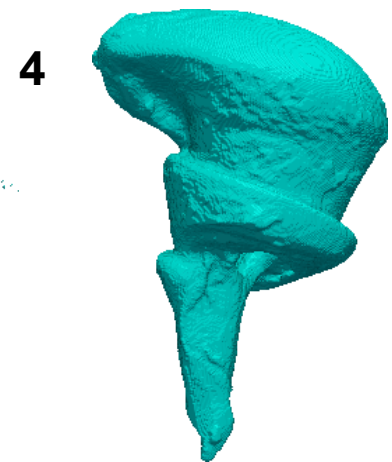
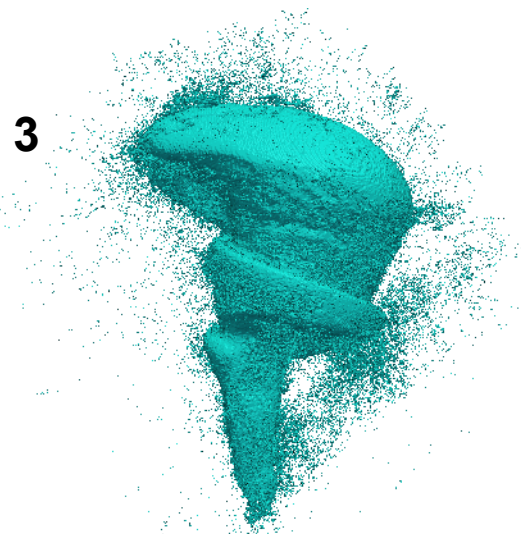
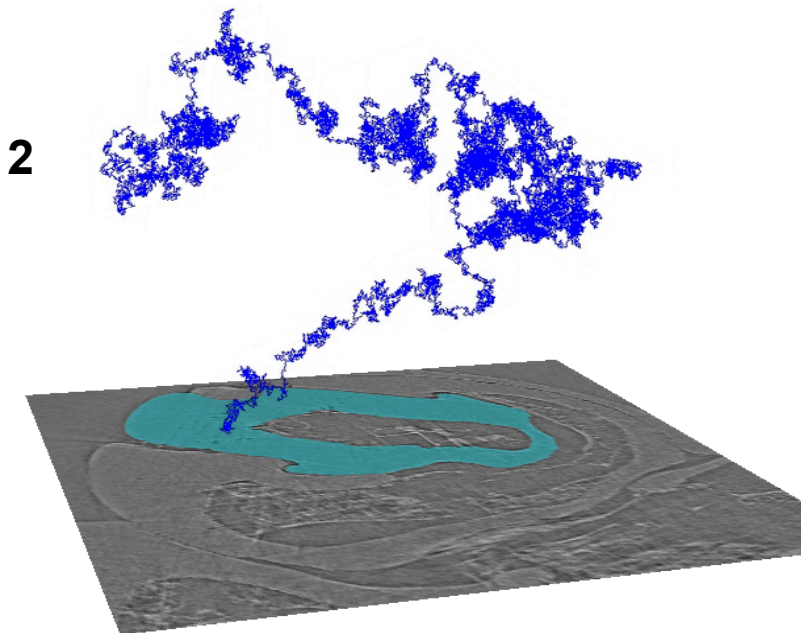
EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



# 3D Segmentation Algorithms

## ■ Algorithm – *requires performance of GPUs*

1. Preprocess: Label the segments in some well chosen slices.
2. Start a great number of weighted random walks at each labeled pixel.
3. The number of hits by random walks which were started in the same segment leads to the probability that a voxel belongs to this segment.
4. Post processing using Active Contour Method in 3D.



P Lösel, V Heuveline, SPIE 2016, Paper 9784-92  
 U Heidelberg



casual interpolation





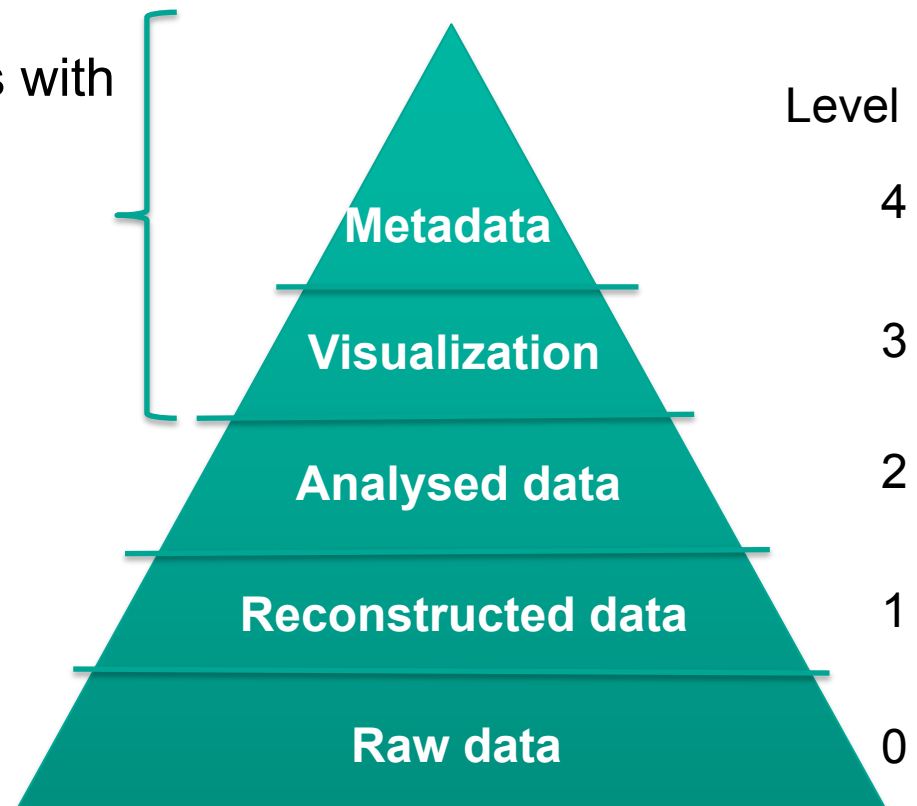
# BIOMEDISA interpolation

### 3. Hierarchical Organization of Datasets

- Long-term archive have long access times
- Organize large dataset in levels with decreasing size is needed
- Only metadata is available in data catalogs

#### Solution:

- Extend classical metadata by visualization data
- Develop 3D-visualization for data catalogs



# Web Data Catalog with 3D Visualization

## volume visualization

volume name: /sara current texture res: 4096 px  
512 1024 2048 4096  
zoomed out



**zooming**

x-axis

z-axis

y-axis

zoom in

zoom out

**layer**

show advanced controls

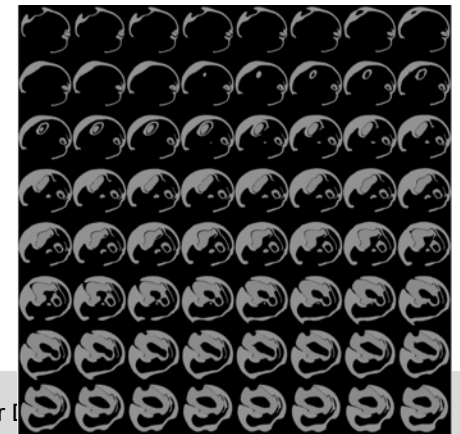
interactive 3D visualization  
from slice maps

- Various levels of resolution
- Slice maps
- Features: cuts, thresholds, illumination models , ...

gray value selection

105 255

low



# Conclusion

- Data rates will reach 1PB/a soon
- Exascale data center will come
  
- **Still datasets need to be analyzed**
  - Online data assessment
  - Improve Analysis methods for large datasets
- **Data management matters**
  - Scalable data formats with extended metadata + scalable visualization
  - Dependencies between datasets  
e.g. higher resolution data should replace older

*Solutions are often domain specific and will not appear automatically by generic “Big Data” research*