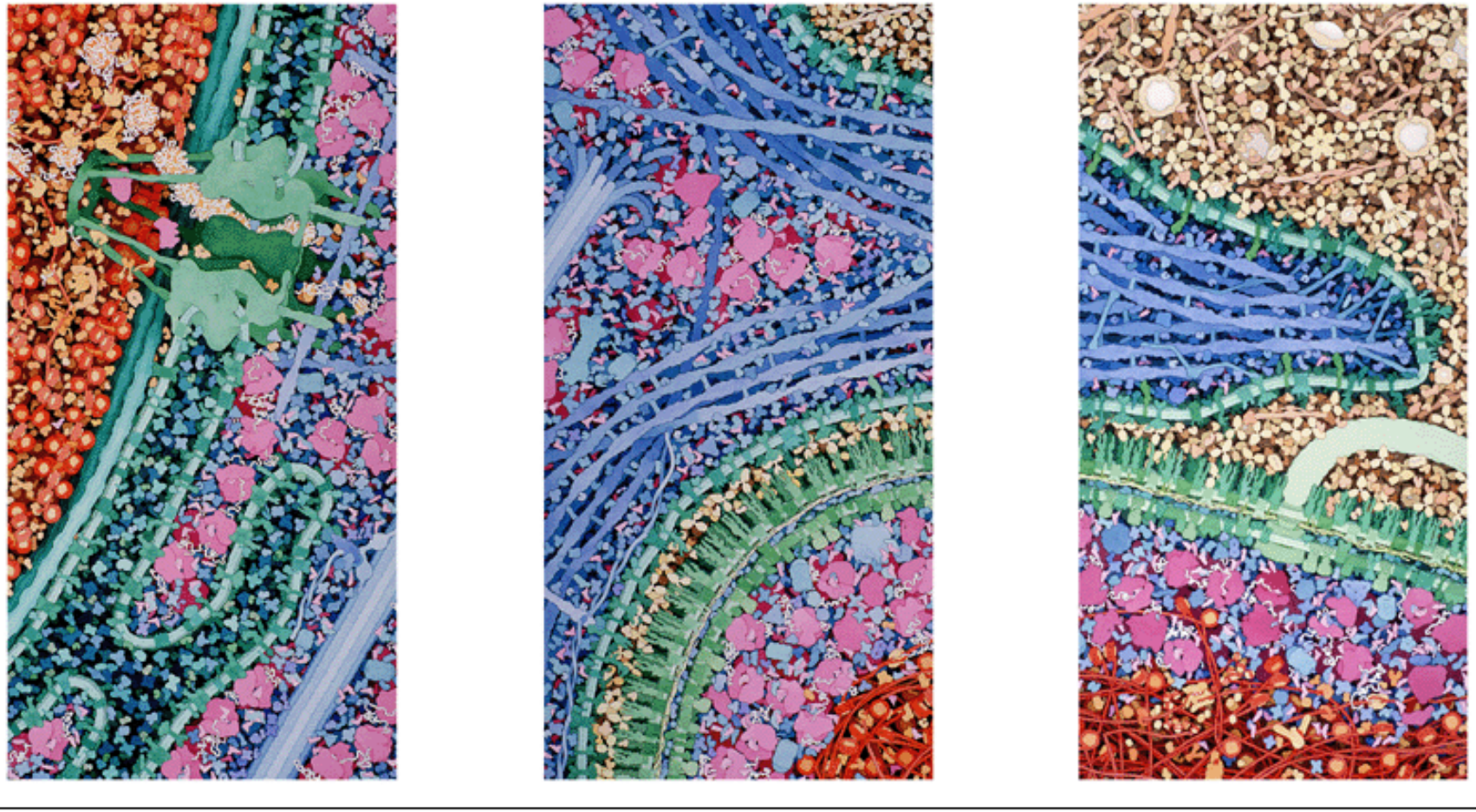
Computational problems in 3DEM

C.O.S. Sorzano

Biocomputing Unit, National Center of Biotechnology (CSIC), Spain

INSTRUCT associated Image Processing Centre for Microscopy

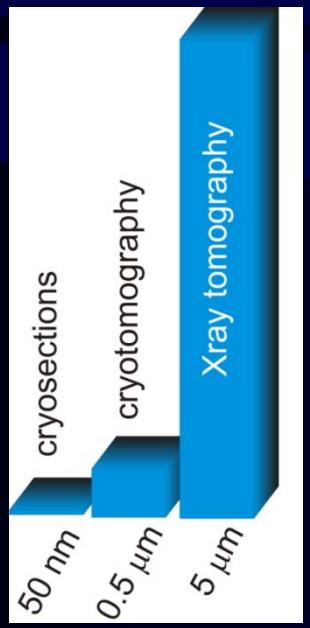
Biological problem



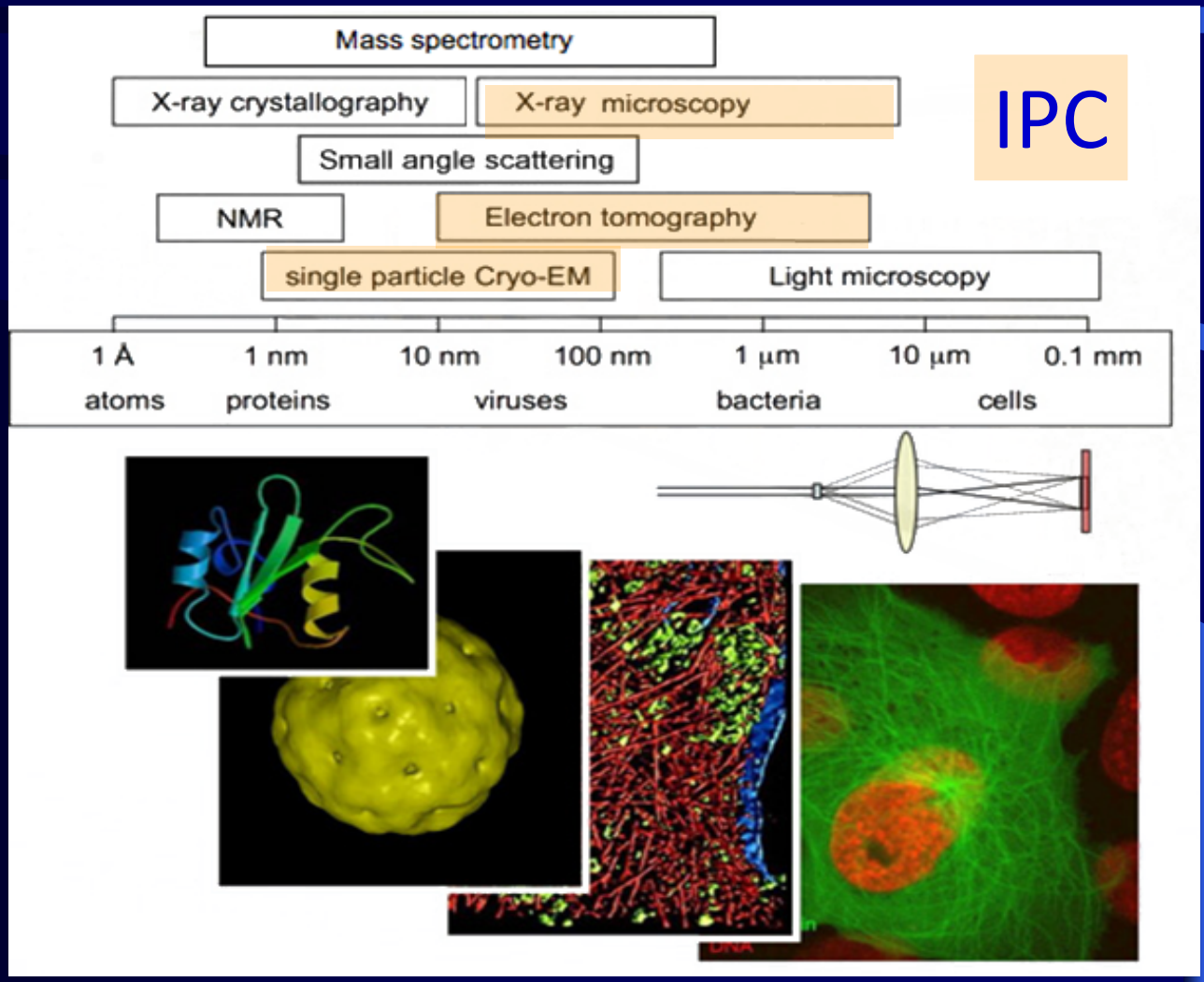
Macrophage and Bacterium 2,000,000X
David S. Goodsell, 2002

Biological problem

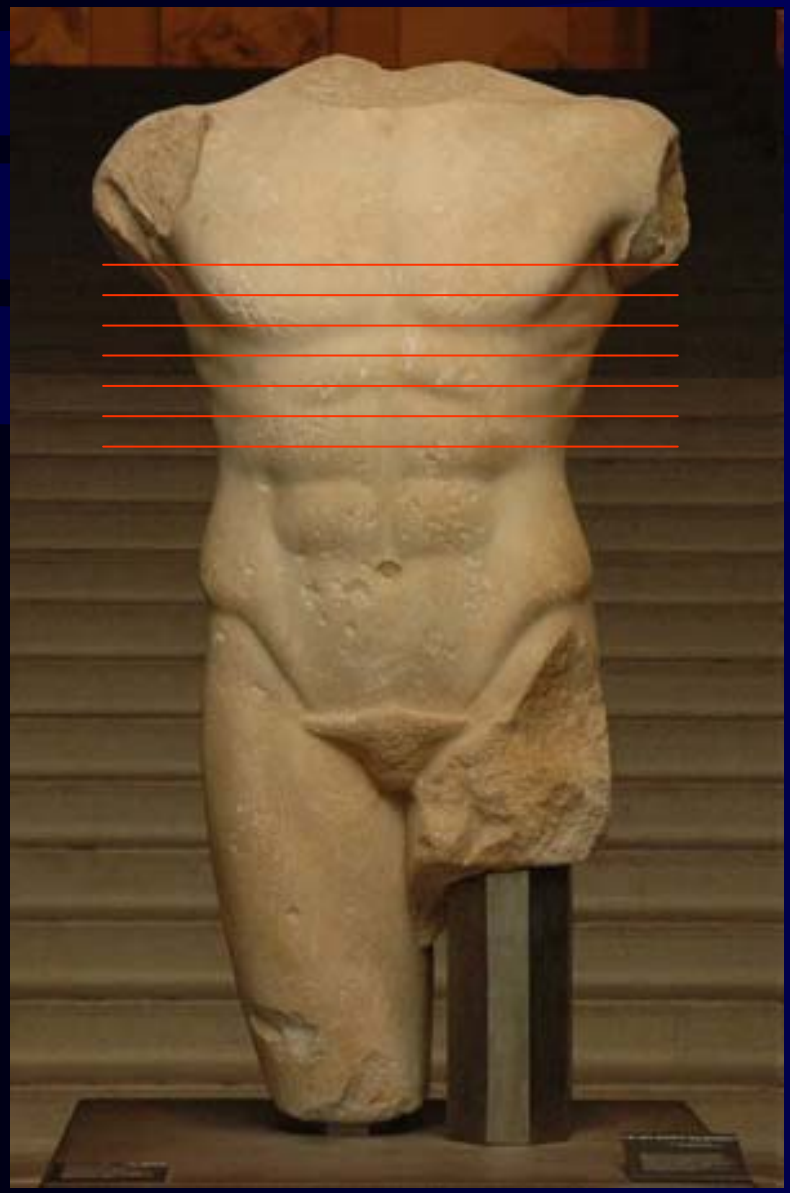
Thickness range

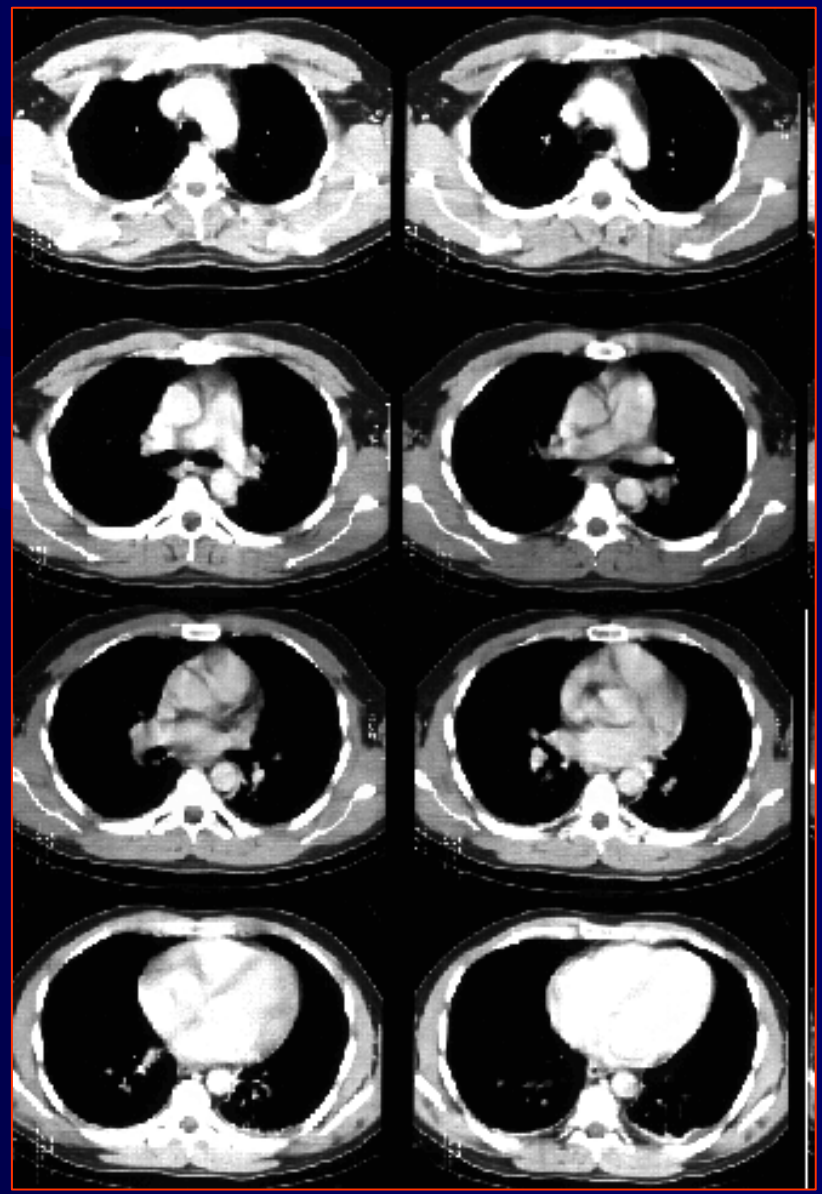


Resolution range

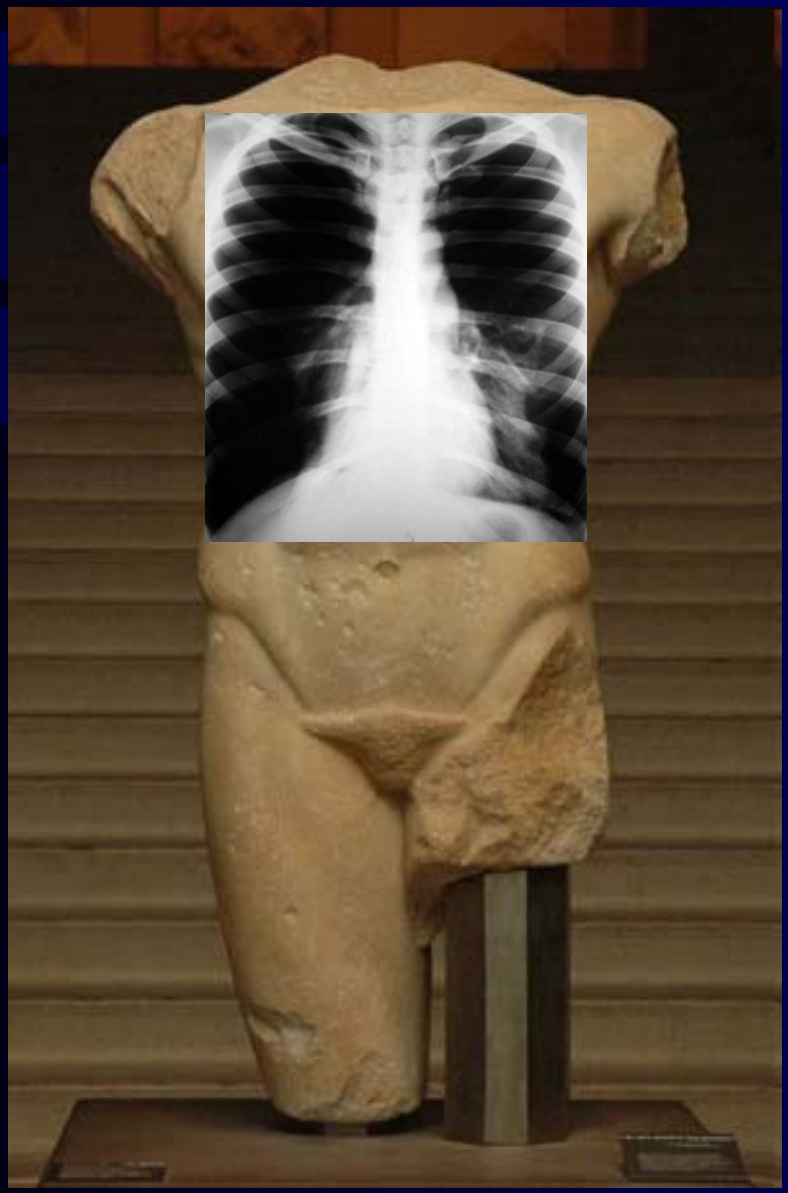


Why do we need image processing?





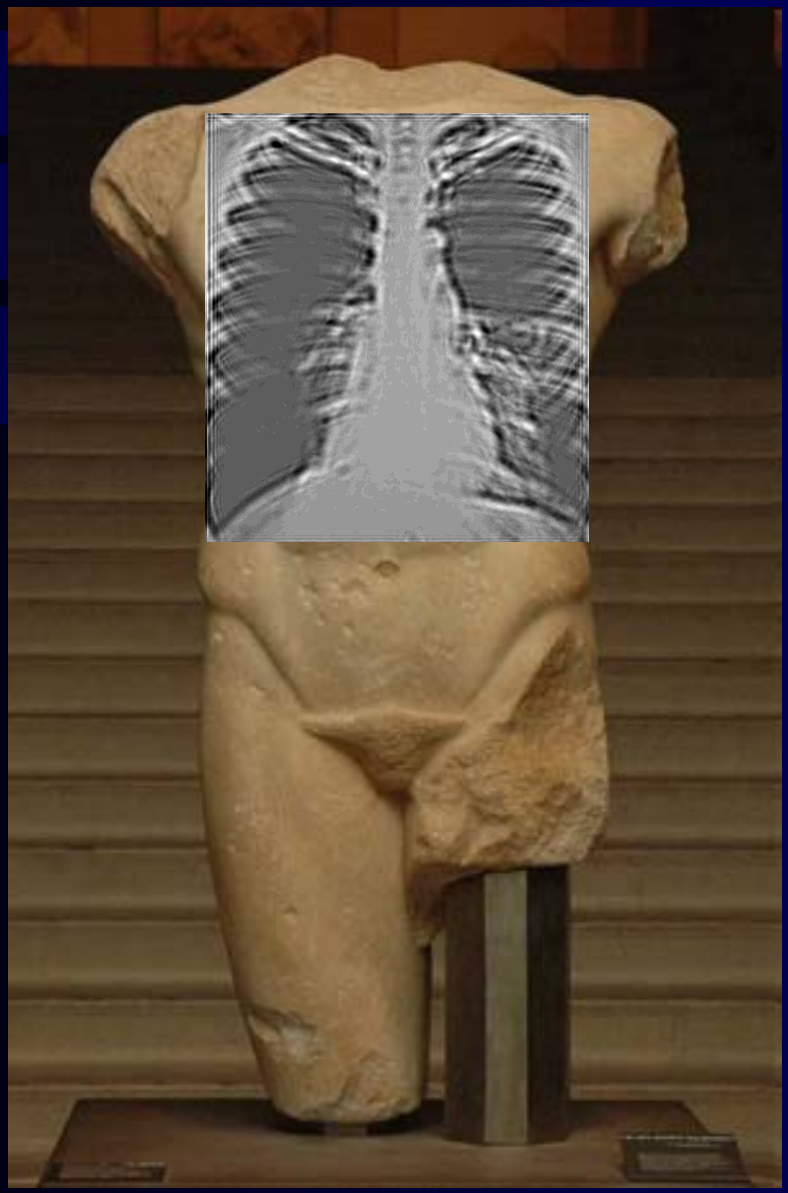
Why do we need image processing?



In the microscope:

3D information is collapsed

Why do we need image processing?



In the microscope:

- **3D information is collapsed**
- **2D information is “blurred”**

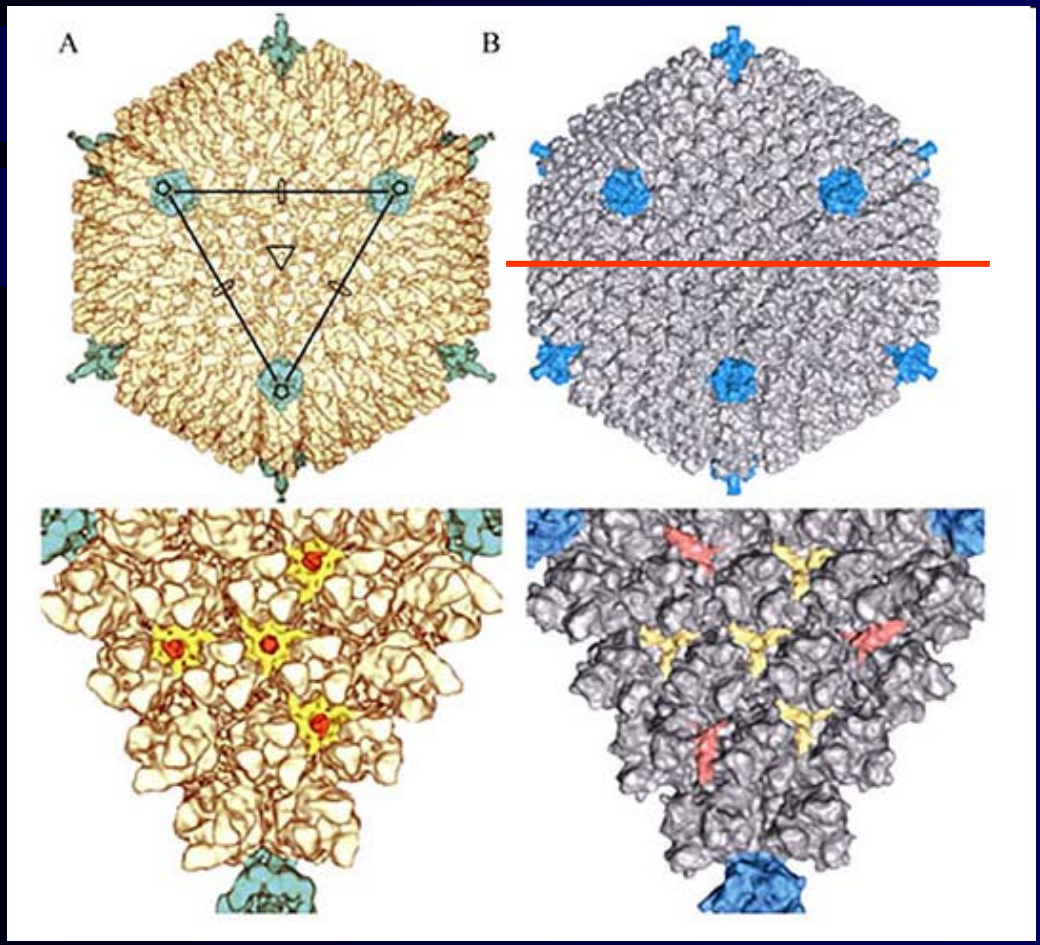
Why do we need image processing?

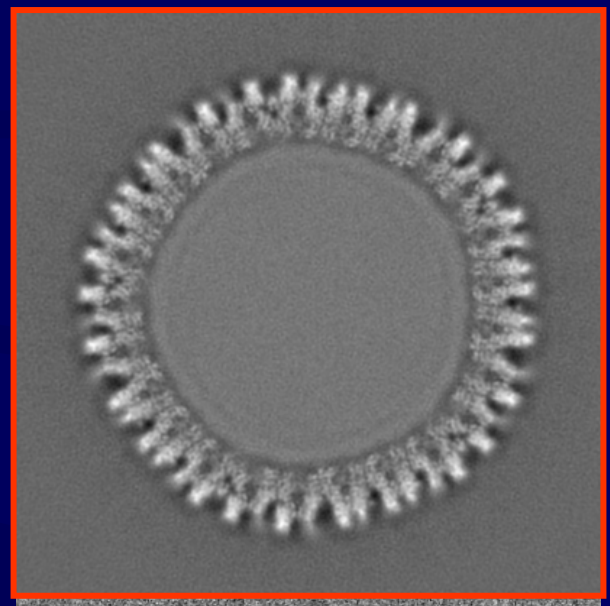
In the microscope:

- **3D information is collapsed**
- **2D information is “blurred”**
- **2D information is corrupted by noise**

Why do we need image processing?

Adenovirus





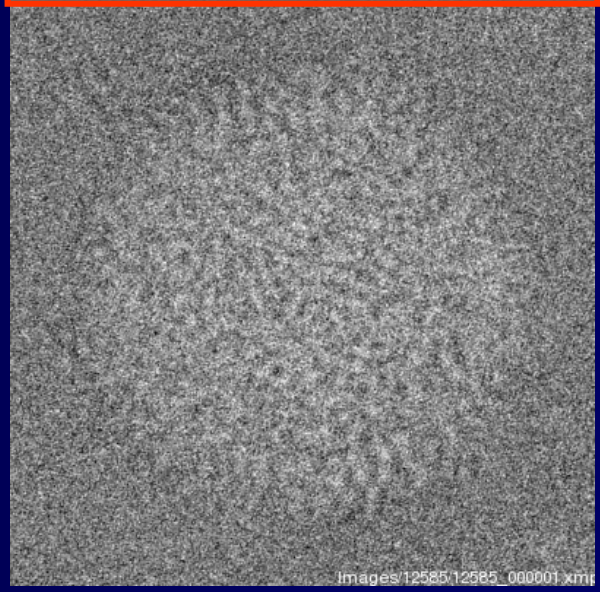
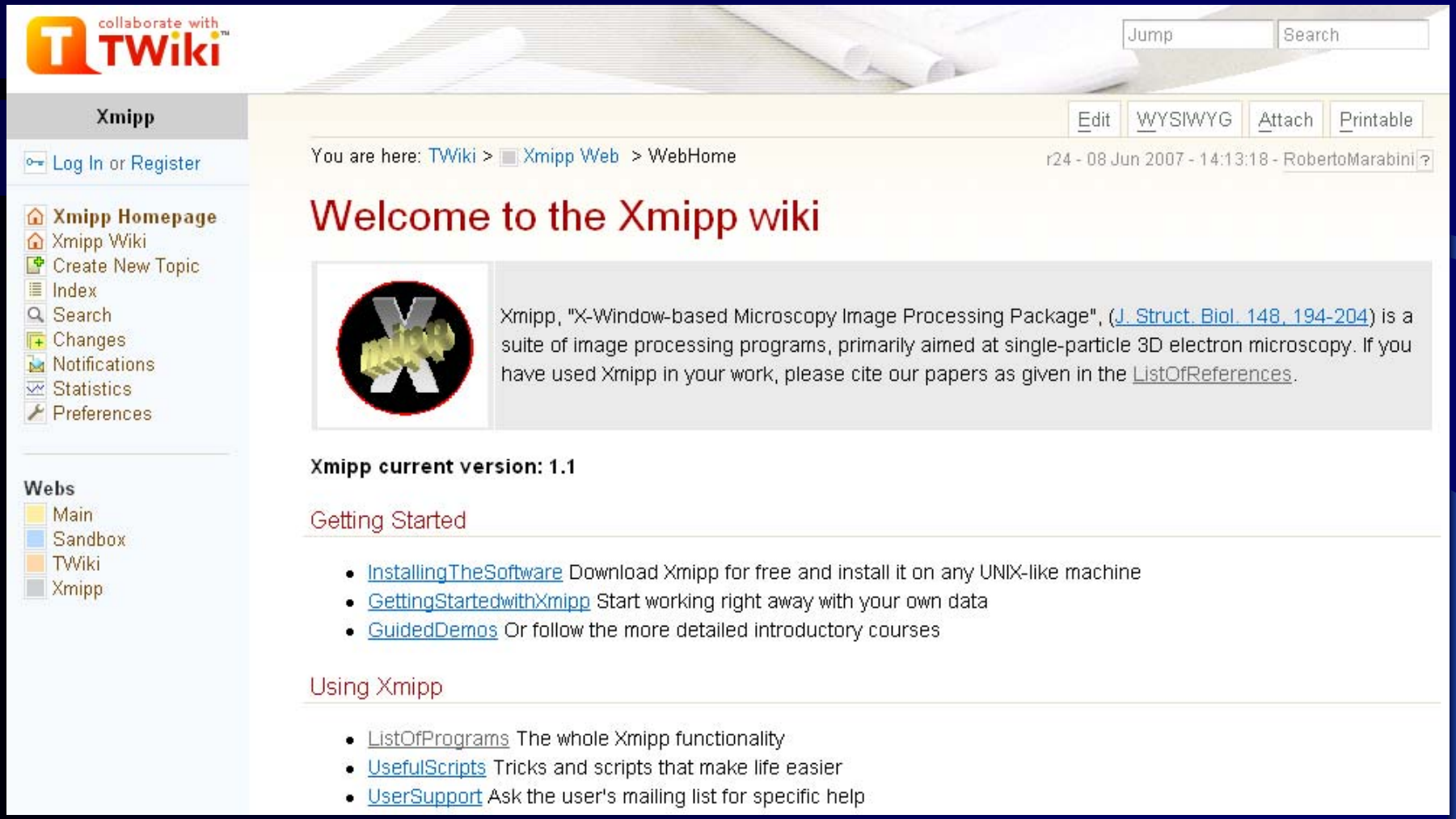


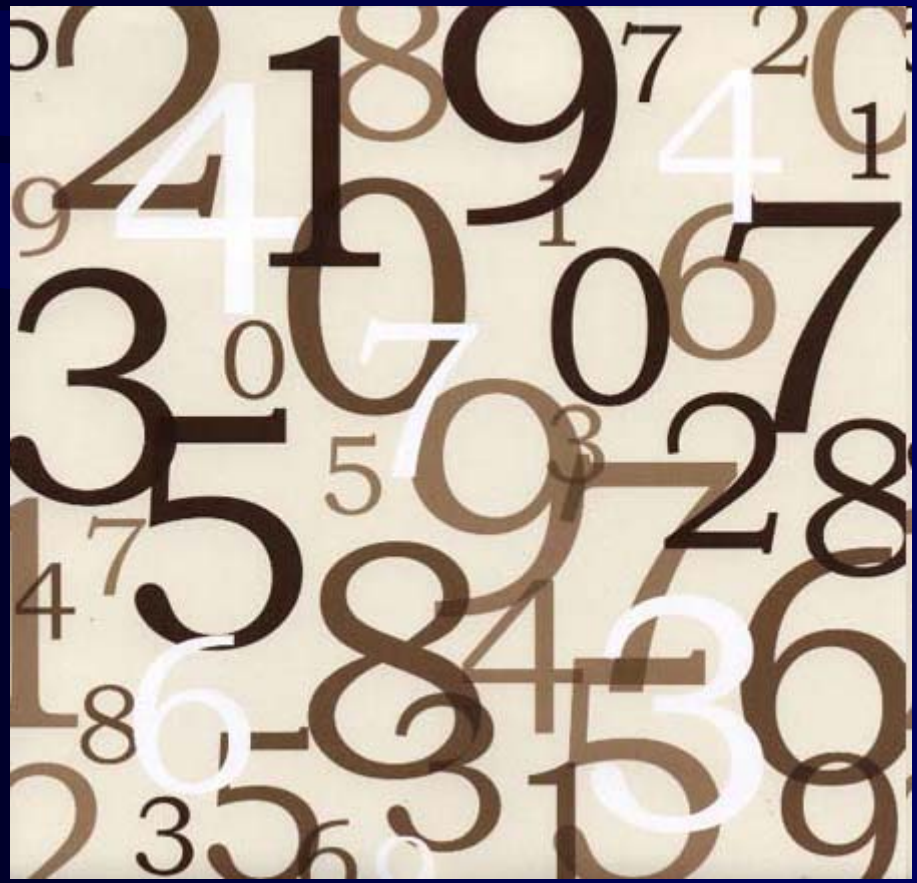
Image Processing Packages: Xmipp

Documentation: http://xmipp.cnb.csic.es

Email: xmipp@cnb.csic.es



Computational challenges: High Performance Computing



High Performance Computing

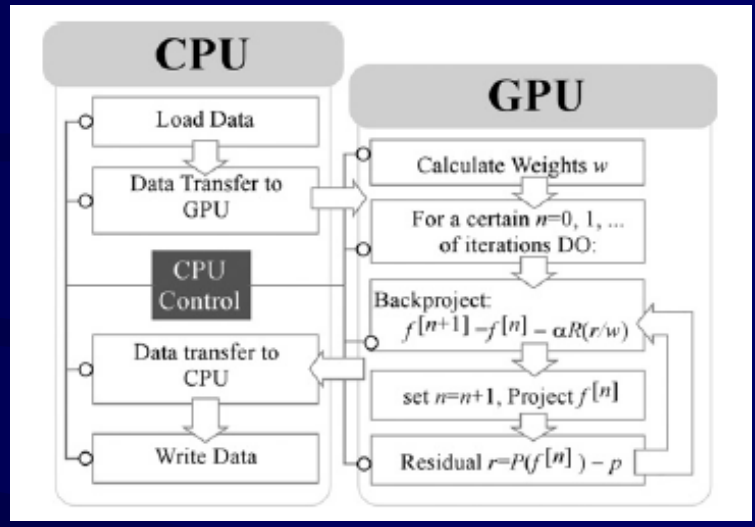
Electron Tomography:

- Image size: 2048x2048 (coming 4096x4096)
- Number of images: 140
- Data size: 140 x 2048x2048 x 8 = 4.37 GB
- Reconstruction size: 500 x 2048x2048 x 8 = 15.6 GB
- Acquisition rate: 6-12 tomograms/day

Computational needs:

- Alignment (6-12 hours)
- 3D reconstruction (12-24 hours)

High-Performance Computing: GPUs



Castaño-Díez, D. et al. *J. Structural Biology*, **2007**, *157*, 288-295

Table 1
Comparison of different running times of the SIRT algorithm

Iterations	Dimensions	NVIDIA GeForce 6600 GT	Intel Pentium 4, 2.4 GHz	Approx. speed up factor (1)	NVIDIA QuadroFX 4500	Intel Xeon 3.4 GHz	Approx. speed up factor (2)
1	200 × 200	0:03	0:15	4.9	0:01	0:12	7.1
10	200 × 200	0:06	1:44	17.3	0:02	1:28	29.7
50	200 × 200	0:19	8:23	26.5	0:07	7:03	49.7
100	200 × 200	0:35	16:42	28.6	0:14	14:05	57.3
1	512 × 512	0:06	1:15	12.5	0:02	1:34	36
10	512 × 512	0:23	13:44	35.8	0:09	11:24	73
50	512 × 512	1:40	65:00	39	0:39	55:00	84
100	512 × 512	3:15	135:00	41.5	1:16	110:00	85
1	1024 × 1024	0:17	6:35	23.2	0:06	5:34	49
10	1024 × 1024	1:23	45:00	32.5	0:32	38:00	71
50	1024 × 1024	6:17	n/p	n/p	2:26	186:00	76
100	1024 × 1024	12:25	n/p	n/p	4:48	n/p	n/p
1	2048 × 2048	1:01	28:00	27.7	0:23	26:00	67
10	2048 × 2048	5:21	186:00	34.8	2:03	154:00	75
50	2048 × 2048	24:35	n/p	n/p	9:27	n/p	n/p
100	2048 × 2048	48:38	n/p	n/p	18:44	n/p	n/p

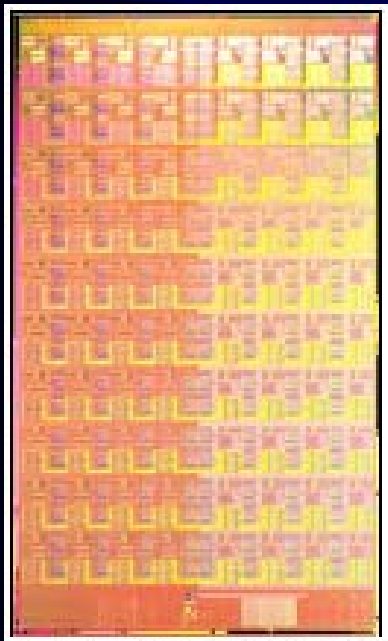
High-Performance Computing: Teraflop chips?

Overview Teraflops Research Chip

intel Leap ahead

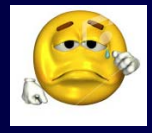
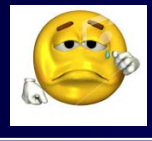
Intel's Teraflops Research Chip

Advancing multi-core technology into the tera-scale era.



Teraflops Research Chip

Future of HPC for Electron Tomography

	GPU	Multicore	Multicore + GPU
Memory requirements			
Deployment cost			
Development cost			
Execution time			

High Performance Computing

Single Particle Analysis

- Image size: 150x150
- Number of images: 100k (coming 1M)
- Data size: 100k x 150x150 x 8 = 16.7 GB
- Reconstruction size: 150x150x150 x 8 = 25MB
- Acquistion rate: 1 week (peak 250k/day)

Computational needs:

- 2D Classification: 2 months
- 3D Alignment and Reconstruction: 2 weeks
- 3D Classification: 8 months

High-Performance Computing: Multiprocessor Cores

	Machine8						Machine16					
	Reconstruction	Speed-up	Conversion	Speed-up	Total	Speed-up	Reconstruction	Speed-up	Conversion	Speed-up	Total	Speed-up
64 × 64 × 64												
Sequential	9.89	–	0.58	–	10.46	–	25.34	–	1.40	–	26.74	–
2 threads	4.71	**2.10**	0.34	**1.76**	5.05	**1.94**	12.97	**1.95**	0.81	**1.75**	13.78	**1.94**
4 threads	2.70	**3.66**	0.20	**2.95**	2.90	**3.37**	7.07	**3.58**	1.28	**1.11**	8.35	**3.21**
8 threads	1.73	**5.72**	0.14	**4.38**	1.86	**5.24**	6.12	**4.14**	0.40	**3.55**	6.52	**4.11**
16 threads	–	–	–	–	–	–	5.62	**4.51**	0.29	**4.88**	5.91	**4.53**
128 × 128 × 128												
Sequential	82.48	–	4.61	–	87.09	–	204.24	–	11.28	–	215.51	–
2 threads	40.13	**2.06**	2.63	**1.81**	42.76	**1.94**	103.74	**1.97**	5.98	**1.89**	109.71	**1.97**
4 threads	21.81	**3.78**	1.52	**3.13**	23.33	**3.56**	58.83	**3.47**	6.59	**1.71**	65.42	**3.30**
8 threads	11.68	**7.06**	0.92	**5.17**	12.60	**6.60**	44.30	**4.61**	2.73	**4.13**	47.03	**4.58**
16 threads	–	–	–	–	–	–	23.75	**8.60**	1.93	**5.86**	25.67	**8.40**
256 × 256 × 256												
Sequential	647.44	–	37.02	–	684.46	–	1633.42	–	90.90	–	1724.31	–
2 threads	314.84	**2.06**	20.45	**1.86**	335.29	**1.95**	832.00	**1.96**	50.36	**1.80**	882.35	**1.95**
4 threads	163.97	**3.95**	12.14	**3.13**	176.11	**3.72**	470.89	**3.47**	55.07	**1.65**	525.96	**3.28**
8 threads	91.71	**7.06**	7.12	**5.33**	98.83	**6.63**	354.79	**4.60**	23.94	**3.79**	378.73	**4.55**
16 threads	–	–	–	–	–	–	194.50	**8.40**	13.19	**6.88**	207.70	**8.30**
512 × 512 × 512												
Sequential	5260.07	–	302.24	–	5562.30	–	13050.93	–	731.49	–	13782.42	–
2 threads	2552.30	**2.06**	163.55	**1.87**	2715.85	**1.96**	6540.15	**1.97**	395.92	**1.88**	7036.07	**1.96**
4 threads	1281.50	**4.10**	96.04	**3.18**	1377.54	**3.87**	3415.78	**3.82**	357.52	**2.08**	3773.30	**3.65**
8 threads	706.61	**7.44**	57.25	**5.34**	763.86	**6.97**	2535.67	**5.15**	163.00	**4.56**	2698.66	**5.11**
16 threads	–	–	–	–	–	–	1495.56	**8.73**	99.47	**7.46**	1595.03	**8.64**

J.R. Bilbao-Castro, et al. J. Structural Biology, 165: 19-26 (2009)

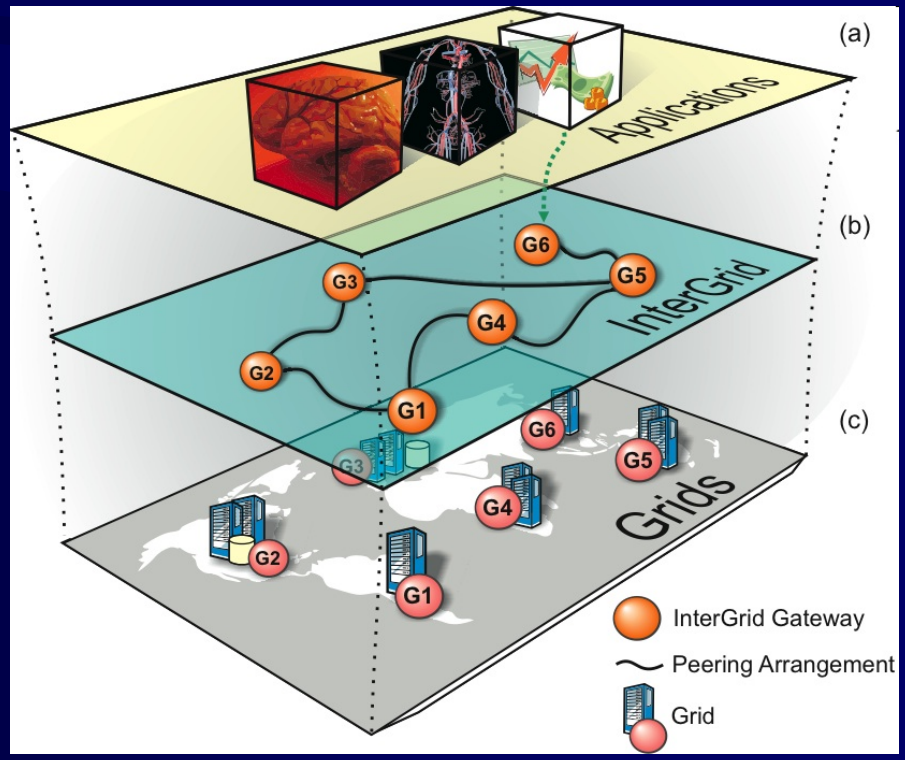
High-Performance Computing: Parallel computing

HPC capabilities in the most common packages in cryoEM

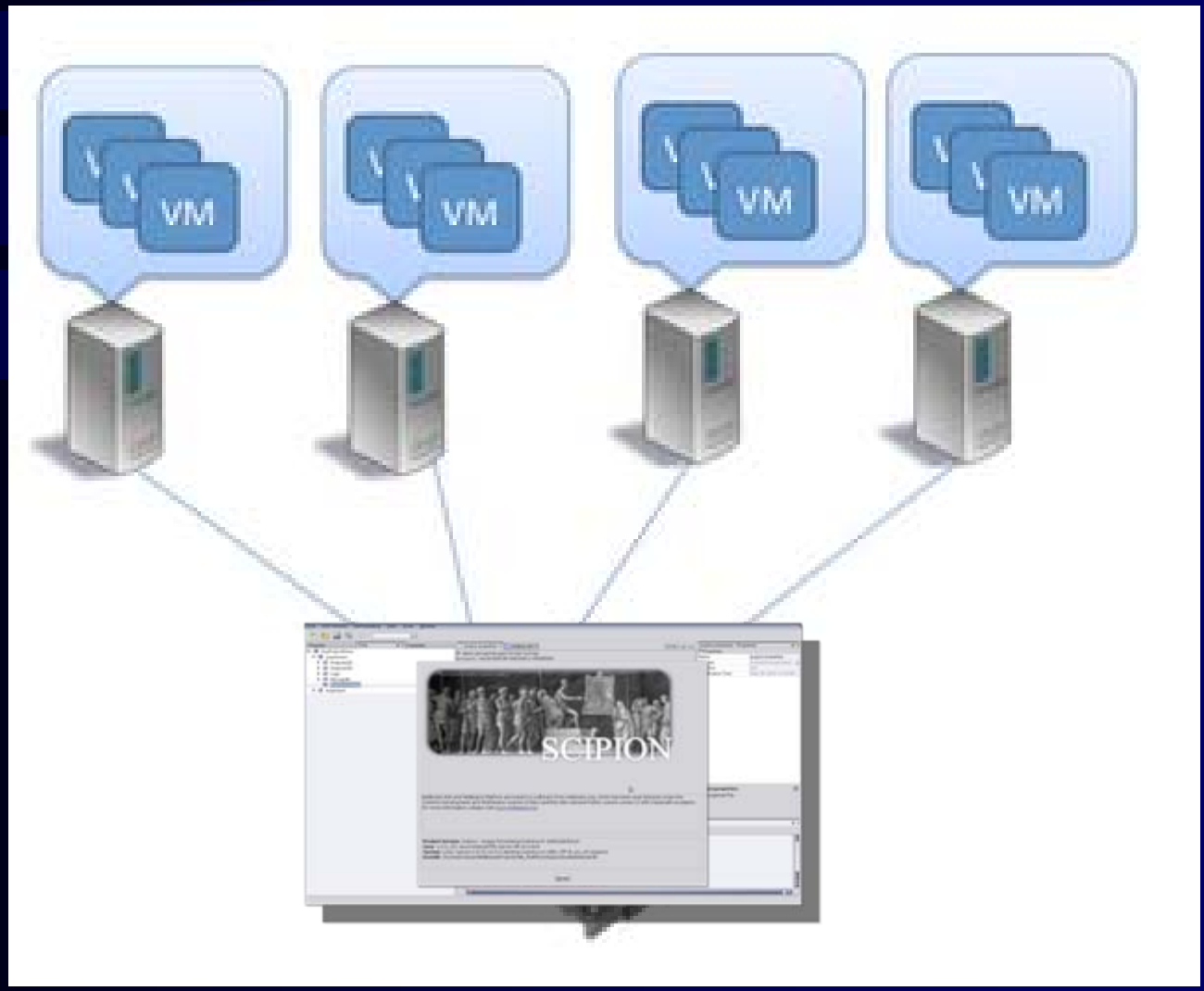
Package	Reference	Modality	Parallelized tasks	Implementation
AUTO3DEM	Yan et al. (2007)	Single particles'	Angular determination; reconstruction	MPI
BSOFT	Heymann and Belnap (2007)	Single particles	Reconstruction	Custom
BSOFT	Heymann et al. (2008)	Tomography	Reconstruction; denoising; resolution estimation	Custom
EMAN	Ludtke et al. (1999)	Single particles	Classification; angular determination; reconstruction; Helixhunter, foldhunter	Custom; MPI; OpenMP, pthreads
FREALIGN	Grigorieff (2007)	Single particles	Angular determination	Custom
IMAGIC	van Heel et al. (1996)	Single particles	Angular determination; reconstruction	MPI
IMIRS	Liang et al. (2002)	Single particles'	Angular determination; reconstruction	OpenMP, MPI
IMOD	Kremer et al. (1996)	Tomography	CTF correction; reconstruction; denoising; dual-axis tomogram combination	Custom
PRIISM/IVE	Chen et al. (1996)	Tomography	Reconstruction; alignment of two tilt series	Custom, GPUs
SPIDER	Frank et al. (1996)	Single particles	Angular determination; reconstruction; template matching (fitting)	OpenMP, MPI, custom
SPIDER	Frank et al. (1996)	Tomography	Reconstruction; template matching	OpenMP, custom
UCSF TOMOGRAPHY	Zheng et al. (2007)	Tomography	Reconstruction	MPI
XMIPP	Marabini et al. (1996)	Single particles	Classification & alignment via maximum-likelihood; angular determination; reconstruction	MPI, pthreads

J.J. Fernández, J. Structural Biology 164: 1-6 (2008)

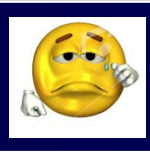
High-Performance Computing: Cloud computing



Location independency, resource allocation, data transfers



Future of HPC for Single Particles

	Local cluster	Cloud computing
Memory requirements		
Deployment cost		
Development cost		
Access cost		
Access time		

Summary

- **3D Electron Microscopy is a very intensive computational task demanding HPC**
- **Most important factor:**
  - Electron tomography: low execution time
  - Single particles: resource allocation
- **Different technologies have been explored**
- **Winning technology:**
  - Electron Tomography: Multicore+GPU
  - Single particles: Local clusters or cloud computing
- **More efficient development is needed**
- **Computational cost will have to be explicitly considered in Structural studies**